Search for new physics in $\sqrt{s} = 13$ TeV proton-proton collisions, using jet substructure techniques with the CMS detector at the LHC

Dominic Smith

A dissertation submitted to the University of Bristol and Vrije Universiteit Brussel in accordance with the requirements of the degree of PhD in the Faculty of Science, School of Physics, Bristol and Faculteit Wetenschappen en Bio-ingenieurswetenschappen, Vakgroep Fysica, Brussel.

Sep 2017

Published in: School of Physics, University of Bristol
Published in: Faculteit Wetenschappen & Bio-ingenieurswetenschappen,
Vrije Universiteit Brussel
Responsible Contact: Dominic Smith
Institute for High Energy Physics
Promoter: Prof. Dr. F. Blekman, Vrije Universiteit Brussel, BE
Promoter: Dr. H. Flaecher, University of Bristol, GB

ACKNOWLEDGEMENTS

The act of writing a thesis cannot be accomplished without of course the content. However, the most important aspect of my thesis, without which there would be none, is the support and encouragement I received throughout. Of the many names and faces I encountered during the last 4 years, each contributed in their own unique ways and for that I could not be more grateful.

I would like to begin by expressing my most profound gratitude to my supervisors, Prof. Dr Freya Blekman and Dr. Henning Flaecher. Naturally, I am grateful for the opportunity of doing a PhD under your supervision. However, I am most grateful for the patience and unwilling support I received from you both.

To my colleagues over the last 4 years, the great team on the L1 trigger and the members of the RA1 group, thank you for your company and support. I extend my thanks to my fellow PhD students, both in Brussels and Bristol. The journey of a PhD can often be a solitary experience, but having shared the sweet adventure with a great many persons is marvellous. To Doug and Tai, your presence and opinions throughout the years have been enormously influential - thank you for always being "there". A special mention must go to my dearest friend, Fionn, our companionship means more to me than you know.

I am also eternally grateful to my close friends outside the research world, those of whom accompanied me during the good times and more importantly, the bad times. You have always kept me grounded and provided a release when I needed one. Of course, I owe a great deal of thanks to Pauline. For your faith and enduring tolerance over the years, I cannot thank you enough.

Finally, I wish to express the greatest of gratitude to my family. There is no value to the ceaseless love, encouragement and faith you have given me. Through your energy and your virtue, you let me become who I am.

AUTHOR'S CONTRIBUTION

The research presented in this thesis was performed in collaboration with the Imperial College and the University of Bristol SUSY groups. Unless stated otherwise, the work presented in this thesis is that of the authors, or at least where the author was a major contributor. The specific contributions of the author concern the validation of the simulated samples, the statistical employment on signal models, the general cross-checking of the analysis, the studies into alternative angles for QCD multijet suppression and the development and implementation of jet substructure techniques. With regards to the investigation of jet substructure techniques, the author has been the main contributor to the running and restructuring of the methodology of all the analysis steps.

The authors contribution to the CMS experiment involve the full maintenance and commissioning of the Global Calorimeter Trigger, during the upgrade phase at the LHC. Furthermore, the author has contributed to roles of a wider responsibility in the full surveillance of all the Level 1 trigger systems, during a period which preceded and occurred during data taking at the CMS detector. Additional investigations into the upgrade of the Hadronic Forward Calorimeter have also been undertaken by the author.

ABSTRACT

An inclusive search for new physics with jets and missing transverse energy is presented, in *pp*-collisions at a centre-of-mass energy of $\sqrt{s} = 13$ TeV. The variables α_{T} and $\Delta\phi^{*}_{min}$ are employed to discriminate between events with genuine missing energy and events with a false source of missing energy, due to either a misreconstruction or a physics process. The use of jet substructure techniques as a means to tag jets from boosted particle decays, a phenomena associated to signal processes, is employed.

The search was performed on a dataset corresponding to an integrated luminosity of 12.9 fb^{-1}collected by the CMS detector at the LHC in 2016. The search categorises events according to the total jet multiplicity, multiplicity of jets arising from bottom quarks and the total hadronic energy. An additional event category, corresponding to the subjet multiplicity of top quark tagged or W boson tagged jets that followed the application of jet substructure techniques, is constructed to further categorise events. The addition of such a category allows a direct comparison to results obtained using the nominal event categories, and provided a direct gauge on the sensitivity that jet substructure techniques bring.

The observed event counts are found to be compatible with the expected contributions from Standard Model processes, and from which the results are interpreted in the context of squark and gluino pair produced simplified models, and a dark matter pseudoscalar mediated simplified model. The use of jet substructure techniques, to further classify events containing boosted top quarks or W bosons, shows an improvement in sensitivity in the exclusions limits set on the simplified models.

SAMENVATTING

Een zoektocht naar nieuwe fysica met jets en ontbrekende transversale energie wordt gepresenteerd. Hiervoor worden proton-proton botsingen met een massamiddelpuntsenergie van $\sqrt{s} = 13$ TeV gebruikt. De parameters α_T en $\Delta\phi^*_{min}$ worden gebruikt om het onderscheid te maken tussen events waar daadwerkelijk energie ontbreekt en events met een foutieve oorzaak voor de ontbrekende energie, ofwel door een foute reconstructie, ofwel door een fysisch proces. ‘Jet substructure’ technieken worden gebruikt om jets uit het verval van relativistische deeltjes, een fenomeen geassocieerd met het gezochte proces, te herkennen.

Er werd gezocht in een dataset met een grootte equivalent aan een geïntegreerde luminositeit van 12.9 fb^{-1} die werd verzameld door de CMS detector aan LHC in 2016. De zoektocht categoriseert events volgens ‘total jet multiplicity’, ‘jet multiplicity’ van bottom quarks en de totale hadronische energie. Een extra categorie, de ‘subjet multiplicity’ van jets met een top quark tag of W boson tag afkomstig van ‘jet substructure’ technieken, werd ook aangemaakt om events verder te kunnen onderverdelen. Het van deze categorie laat toe om een rechtstreekse vergelijking te maken met resultaten behaald via de normale categorieën, en geefteen schatting van de gevoeligheid die ‘jet substructure’ technieken bieden.

Het gemeten aantal events is consistent met de voorspellingen van Standaard Model processen. Hieruit worden de resultaten geïnterpreteerd in een vereenvoudigd model van squark en gluino paarproductie en een vereenvoudigd model van donkere materie met pseudoscalairen als mediator. Het gebruik van ‘jet substructure’ technieken om jets van relativistisch vervallende top quarks en W bosonen verder te categoriseren zorgt voor een verbetering van de limieten van de vereenvoudigde modellen.

TABLE OF CONTENTS

List of figures

LIST OF TABLES

CHAPTER 1

THE STUDY OF CHANGE

The study of change is an extremely ambiguous concept, involving a variety of seemingly diverse subjects that most originate from a common quality, our instinctive and inquisitive curiosity.

To better understand what is meant by this concept perhaps demands a revisit to our first evidence of "change", the birth of the universe. The theory, commonly known as the Big Bang, describes how the universe expanded from a state of high temperature and high density. Thereafter, a concatenation of events spanning billions of years has determined its evolution which is ongoing today. This journey has been a constant source of questioning for physicists particularly with regard to its origins, inevitably still speculative, as well as to the nature of its composition.

Understanding both the origin of the universe and how it is composed can be achieved by acknowledging other disciplines, notably History. The "study of change" associated to History is exemplified in the study of past events, achieved, in some way, by exploring the architecture of buildings, from the ingredients to the structural support within. In considering and examining these physical building blocks, the perspective of the overall understanding of the subject in question is widened.

This technique can be applied to our original consideration of the universe, in which the profoundest and most ancient architecture of our existence could be said to contain its own elementary and figurative building blocks. This prompted scientists to identify "atoms", from which they discovered dozens, each with variations in their properties and their ability to emit light. These atoms were later discovered to possess an interior architecture of their own, built from smaller particles, specifically electrons which orbited an atomic nuclei of a proton and were held crucially together by the strong force.

Such observations triggered the construction of theories, aimed to provide a framework needed to explain particle interactions and the known fundamental forces. With the introduction of the theory of relativity, relativistic quantum theories of particles were constructed and postulated the existence of antimatter.

The experimental and theoretical evolution of particle physics culminated in the construction of the Standard Model (SM) The SM of particle physics provides the deepest understanding of the structure of the universe. Of the fundamental forces of nature, the SM forms a quantum description of the electromagnetic interaction, the strong force and the weak force. Within the SM, interactions between the fundamental constituents of matter are mediated by such forces.

Despite its elegance and accuracy, the SM falls short at addressing some crucial points. The absent description of gravity and the inability to provide a viable Dark Matter (DM) candidate to explain the unknown mass content of the universe, are some examples of the SM shortcomings. These limitations, and in some cases discrepancies, in the theory have prompted the search for more comprehensive theories Beyond the Standard Model (BSM) that aim fill the gaps in knowledge left by the SM.

CHAPTER 2

THEORETICAL OVERVIEW

The field of High Energy Physics aims to examine several questions regarding our understanding of the universe. These questions can be reduced to two: firstly the composition of what we observe in our universe and secondly how its ingredients interact with one another.

The best explanations to such questions are encapsulated in the Standard Model (SM), a theory proposed in the 1960s and is the result of decades of research and extensive validation. This in turn provided the theoretical motivation necessary for the discovery of the Brout-Englert-Higgs boson in July 2012 at the European Organisation for Nuclear Research (CERN), Geneva [21, 22].

The sections that follow will provide a description of the SM and a motivation for an extension, in the form of new physics, which aims to satisfy some of the SMs weaknesses.

2.1 The Standard Model

The development of the SM has been facilitated by the cycle of theoretical breakthroughs inspiring experimental observations [23]. Each progressive refinement caused the cycle to repeat such that the world in which we live, at parochial energy scales accessible by experiment, is elegantly governed by the SM.

Of the four fundamental forces, the SM encapsulates three, namely the unified electromagnetic and the weak force (EW) and the strong force, with gravity yet to be included. The SM is a relativistic quantum field theory that expresses all matter as being composed of spin $\frac{1}{2}$ particles, called *fermions*.

The fermions of the SM are themselves classified into leptons and quarks, with each classification existing in three generations. The six lepton flavours of the SM are the electron e, the muon μ, and tau-lepton τ and their three corresponding neutrinos ν_e, ν_μ and ν_τ, as shown in Table 2.1.

Matter fermions: spin-$\frac{1}{2}$						
	Leptons			Quarks		
Generation	Particle	Mass (MeV)	Charge	Particle	Mass (MeV)	Charge
1	ν_e	<2.25×10^{-4}	0	d	4.8	$-\frac{1}{3}$
	e	0.511	-1	u	2.3	$+\frac{2}{3}$
2	ν_μ	<0.195	0	s	95	$-\frac{1}{3}$
	μ	106	-1	c	1270	$+\frac{2}{3}$
3	ν_τ	<18.2	0	b	4180	$-\frac{1}{3}$
	τ	1777	-1	t	173200	$+\frac{2}{3}$

Table 2.1 Summary of the particles of the SM of particle physics. Fermions, of spin-$\frac{1}{2}$ are shown, split into the three generations of leptons and quarks. Masses are taken from [19].

Interactions of the fermions are mediated by spin-1 *gauge bosons*. The massless photon (γ) mediates the electromagnetic interaction and the massive $W^\pm$ and Z bosons mediate the weak interaction. Furthermore, the strong force is mediated by 8 massless gluons (g). The properties of the gauge bosons of the SM, shown in Table 2.2, and their interactions are a result of their respective symmetry groups, discussed in more detail in Section 2.1.1.

Force carrying gauge bosons: spin-1				
Force	Particle	Symbol	Mass (GeV)	Charge
Electromagnetic	Photon	γ	0	0
	W boson	W^+	80.4	1
Weak	W boson	W^-	80.4	-1
	Z boson	Z	91.2	0
Strong	Gluons (8)	g	0	0
Higgs Boson: spin-0				
-	Higgs	H	125.09	0

Table 2.2 Summary of the gauge bosons of the SM. The force carrying bosons of spin-1 are shown, with the Higgs boson to complete the picture. Masses are taken from [19].

2.1.1 Symmetries and gauge invariance

In many theories in physics, the role of symmetries, structures of said theory for which physical observables remain unchanged under specific transformations, is highly significant.

Noether's Theorem relates a continuous symmetry, a transformation that leaves the Lagrangian unaffected, to a corresponding conserved current, that is associated with the symmetry [24].

Field theories in which the action remains invariant under a continuous symmetry that depends on spacetime is a *gauge theory.* In this context, the symmetry group associated with the field is a *gauge symmetry* and from which, introduces *gauge fields* to the theory which mediate a force.

To demonstrate the strength of this approach, consider a free theory of a fermion field, $\psi(x)$, with Lagrangian:

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

where γ^{μ} are the gamma matrices, ∂_{μ} is the partial derivative and m is the mass.

In this context, in shifting the phase of the fermion by $e^{i\theta}$, where θ is a constant, such that:

$$\psi(x) \rightarrow e^{i\theta}\psi(x), \tag{2.2}$$

the Lagrangian will remain invariant.

This type of transformation results in a global U(1) symmetry and forms a unitary *abelian*[1] group. Furthermore, as the Lagrangian is invariant under this transformation, there exists a conserved current associated with the U(1) global symmetry.

However, if the transformation were local such that θ can differ from space-time point to point, then the phase will change by $e^{i\theta(x)}$ and transform the field as:

$$\psi(x) \rightarrow e^{i\theta(x)}\psi(x). \tag{2.3}$$

Correspondingly, the Dirac Lagrangian will change according to:

$$\mathcal{L} \rightarrow \bar{\psi}e^{-i\theta(x)}(i\gamma^{\mu}\partial_{\mu} - m)e^{i\theta(x)}\psi = \mathcal{L} - \bar{\psi}\gamma^{\mu}(\partial_{\mu}\theta(x))\psi. \tag{2.4}$$

---

[1] In the context of QFT, an abelian gauge theory is one in which the generators of group commute.

Under this transformation, the Lagrangian is no longer invariant and as a result, a change is needed to retain the symmetry under the local U(1) transformation.

To restore the gauge invariance requires replacing the derivative ∂_μ in Equation 2.4, with the covariant derivative:

$$D_\mu \equiv \partial_\mu - iqA_\mu \tag{2.5}$$

where q is a coupling constant and A_μ is a new field.

It is clear that in order to regain gauge invariance and preserve the Lagrangian, D_μ, and as a result A_μ, need to transform as

$$D_\mu \to D'_\mu = e^{i\theta(x)} D_\mu e^{-i\theta(x)} \tag{2.6}$$

$$\partial_\mu - iqA'_\mu = e^{i\theta(x)}(\partial_\mu - iqA_\mu)e^{-i\theta(x)} = \partial_\mu - i\partial_\mu\theta(x) - iqA_\mu \tag{2.7}$$

which leads to a transformation of A_μ as

$$A_\mu \to A'_\mu = A_\mu - \frac{1}{q}\partial_\mu\theta(x) \tag{2.8}$$

The required U(1) local gauge invariance of the Lagrangian in Equation 2.1 is achieved by introducing the field A_μ, with associated gauge transformation properties. From which, the gauge invariant Lagrangian for a fermion expressed as:

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu\partial_\mu - q\gamma^\mu A_\mu - m)\psi. \tag{2.9}$$

introduces an interaction term of the fermion with the field A_μ.

To demonstrate the implication of such a transformation requires the introduction of one of the forces described by the SM.

The relativistic quantum field theory of electromagnetism is explained by a U(1) abelian gauge theory. In QED, the field A_μ which couples to electrons according to $-e\gamma^\mu A_\mu$, can be incorporated into the Lagrangian of Equation 2.9, such that the Lagrangian, under the application of a local phase transformation, alters as:

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi \to \bar{\psi}(i\gamma^\mu e^{-i\theta(x)} D'_\mu e^{i\theta(x)} - m)\psi. \tag{2.10}$$

The full QED Lagrangian can be written as:

$$\mathcal{L}_{QED} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \bar{\psi}(i\gamma^{\mu}\partial_{\mu} - m)\psi + e\bar{\psi}\gamma^{\mu}\psi A_{\mu} \tag{2.11}$$

where the vector field A_μ is coupled to a Dirac particle, in particular a photon to an electron of charge, -e, and $F^{\mu\nu}$ is the field strength tensor of the vector field, $\partial_\mu A_\nu - \partial_\nu A_\mu$.

As a result of this coupling, the theory of electromagnetism has formed into the theory of fermions, a resultant theory that is locally U(1) symmetric, following the introduction of a gauge field A_μ.

The principle of local gauge invariance can also be applied to the case where the transformation is *non-abelian*[2]. One of the essential differences between abelian and non-abelian gauge theories is the existence of self-interactions in the non-abelian case, a characteristic observed in the strong force through gluon self-interactions, discussed in more detail in Section 2.1.2.

The requirement that physics is invariant under local U(1) phase transformations, a fundamental symmetry of the universe, transforms a simple non-interacting theory of fermions into a theory of electromagnetism. Furthermore, coupling the fermion to the photon and allowing such a theory to remain locally U(1) symmetric, is satisfied only by requiring that the photon remain massless.

This restriction is not limited to the U(1) local gauge symmetry of QED, but to the gauge symmetries of the weak and strong interactions. While this is not a problem for the gauge bosons of the strong interaction, the gauge violating mass terms in the Lagrangian predict massless weak vector bosons and thus contradict experimental observation. In order to allow such massive particles, the symmetry must break in such a way that all successful symmetry predictions are still preserved. In addition, the weak vector bosons must acquire large masses whilst the massless gluons and photons remain massless.

Furthermore, the same feature applies to fermions whereby no mass terms are allowed while maintaining gauge invariance. A solution in this case is to introduce a new field which couples left handed particles to right handed particles to give an effective mass, as described in more detail in Section 2.1.3. This field is more commonly known as the *Brout-Englert-Higgs* field, the theory of which predicted the *Brout-Englert-Higgs Boson*.

---

[2] In the context of QFT, QCD is known as a non-abelian gauge theory as the generators of SU(3) do not commute.

2.1.2 Forces and Fermions

The strong force is described by the SU(3) gauge theory Quantum Chromodynamics (QCD), and introduces eight massless gluons to provide a description of the interaction between quark fields. The phase shift and the eight generators of the SU(3) group, have the form $e^{-i\theta^a T^a}$ and T^a respectively, where the latter are related to the Gell-Mann matrices by $T^a = \frac{1}{2}\lambda^a$.

The underlying symmetry of QCD, an invariance under SU(3) local phase transformations takes the form of Equation 2.12.

$$\psi(x) \rightarrow \psi^{'}(x) = (e^{-i\theta^a T^a})\psi(x) \tag{2.12}$$

and the Dirac Equation alters as such:

$$i\gamma^\mu(\partial_\mu + i(\partial_\mu\theta)\hat{T})\psi = m\psi \tag{2.13}$$

In a similar manner to QED, an infinitesimal transformation of ψ such that $\delta\psi_i = -i\theta^a(T^a)^i_j\psi_j$, yields a gauge-violating term as seen in Equation 2.14 and, more importantly, a Lagrangian that is no longer invariant.

$$\delta\mathcal{L} = \bar{\psi}^i(T^a)^i_j\gamma^\mu(\partial_\mu\theta^a(x))\psi_j. \tag{2.14}$$

In QCD, the necessary gauge invariance can be achieved by introducing, for each of the eight generators of the SU(3) group, a corresponding gauge field G_μ that interacts with some matter field.

The eight new fields G^a are the gluons of QCD, and it can be seen that the gauge invariant QCD Lagrangian is of the form:

$$\mathcal{L} = -\frac{1}{4}G^a_{\mu\nu}G_{a\mu\nu} + \bar{\psi}^i(i\gamma^\mu D_\mu - m)^i_j\psi. \tag{2.15}$$

where $G_{\mu\nu}$ represents the gluon field strength tensor, analogous to the QED field strength tensor seen in Equation 2.11, and D_μ is the covariant derivative used to express the introduction of a new field.

Through the construction of both *abelian* and *non-abelian* gauge theories, in which each of the Lagrangians remains invariant under local transformations of each group, the Quantum Field Theories of QED and QCD are complete.

The remaining weak force and the aforementioned electromagnetic force are responsible for interactions between particles possessing weak isospin, T^3, and electric charge, q, respectively. These forces are governed by vector bosons, of which the weak vector $W^{\pm}$ and Z^0 bosons mediate the weak force and the photon (γ) mediates the electromagnetic force, as seen in Equation 2.11.

While the two forces appear distinct, they are both married in the SM to form the unified electroweak theory, a theory that when spontaneously broken yields a massive scalar Brout-Englert-Higgs boson, discussed in more detail in Section 2.1.3.

In the weak sector, the charged-current weak interaction is associated with invariance under SU(2) local phase transformations. The observed form of the weak interaction couples only to left-handed particles states and right-handed antiparticle states, the implications of which mean that the gauge transformation can only effect particles of these states. To achieve this, right-handed particle states and left-handed antiparticle states are placed in weak isospin singlets. Furthermore, weak isospin doublets are composed of left-handed particle states and right-handed antiparticle states, thus motivating the symmetry group of the weak interaction to be referred to as $SU(2)_L$.

The special unitary group, $SU(2)_L$, is generated by the 2x2 matrices $T_i = \frac{1}{2}\sigma_i$, where σ_i are Pauli spin matrices.

The required SU(2) local gauge invariance can be satisfied by introducing three gauge fields, W_μ^k, thus corresponding to the three gauge bosons $W_\mu^{(1)}$, $W_\mu^{(2)}$ and $W_\mu^{(3)}$.

In the electroweak model of Glashow [25], Salam [26] and Weinberg [27] (GSW), the unitary group $U(1)_Y$ brings an additional generator, hypercharge Y, which couples to an additional gauge field B_μ.

The interaction term associated with the field B_μ, shown in Equation 2.16, has the same form as the interaction term from the U(1) symmetry group of QED.

$$g' \frac{Y}{2} \gamma^\mu B_\mu \psi \tag{2.16}$$

where, in contrast to the interaction term associated with QED, Qe is replaced by $g'\frac{Y}{2}$.

In processes which exchange a physical W boson, the four vector currents are written as:

$$j_\pm^\mu = \frac{g_W}{\sqrt{2}} \bar{\psi}_L \gamma^\mu \sigma_\pm \psi_L \tag{2.17}$$

where $\bar{\psi}_L$ contain the left-handed chiral adjoint spinors, and $\sigma_\pm$ is an associated Pauli ladder operator.

Of the four observed bosons of both QED and the weak interaction, the photon and Z boson, with the corresponding fields A_μ and Z_μ are neutral and, in the unified electroweak model, can be written as linear combinations of the B_μ and the neutral $W_\mu^{(3)}$ fields:

$$A_\mu = +B_\mu cos\theta_W + W_\mu^{(3)} sin\theta_W \tag{2.18}$$

$$Z_\mu = -B_\mu sin\theta_W + W_\mu^{(3)} cos\theta_W \tag{2.19}$$

where θ_W is the weak mixing angle, determined by the ratio of the electromagnetic coupling constant ($g^{'}$) and the weak coupling constant (g_W):

$$e = g_W sin\theta_W = g^{'} cos\theta_W \tag{2.20}$$

where e is the electron charge.

In the electroweak sector, the underlying gauge symmetry is the $U(1)_Y$ of weak hypercharge and the $SU(2)_L$ of the weak interaction, written more commonly as $U(1)_Y \otimes SU(2)_L$.

As the $SU(2)_L$ generators are the Pauli spin-matrices, of dimension 2, the fermion wavefunction is written in terms of two components and is thus termed a weak isospin doublet. Furthermore, as the observed form of the weak charged-current interaction couples only to left-handed particle chiral states and right-handed antiparticle chiral states, the weak isospin doublet takes the form of:

$$\psi_L(x) = \begin{pmatrix} \nu_L \\ e_L \end{pmatrix} \tag{2.21}$$

whereas the weak isospin singlet is simply $\psi_R(x) = e_R, \mu_R, \tau_R, \tau_R = e_R$. The wavefunctions for quarks follow a similar trend, resulting in five types of fermion wavefunction.

An important discrepancy currently observed in the universe is the excess of matter over antimatter, the source of which is potentially made possible by CP violation in the weak interaction. In particular, the combined effect of the charge conjugation operator $\hat{C}$ and the parity operator $\hat{P}$ is respected by the weak interaction[3]. However, as both QED and QCD interactions conserve C and P separately, and therefore CP, the only possible source in the SM where CP-violation can occur is in the weak interaction.

---

[3] Charge Parity (CP) symmetry.

In conclusion, the full $SU(2)_L \otimes U(1)_Y$ Lagrangian is simply written as:

$$\mathcal{L}_{EWK} = \bar{\psi} i \gamma^\mu D_\mu \psi - \frac{1}{4}(B_{\mu\nu}B^{\mu\nu} + W_{\mu\nu}W^{\mu\nu}) \tag{2.22}$$

It should be noted that as for QED, there are no boson mass terms in the electroweak Lagrangian. While this is acceptable for the photon field, the W and Z bosons are massive and so adding mass terms to the Lagrangian would break the gauge symmetry.

Alternatively, mass terms are generated using the Brout-Englert-Higgs mechanism [28–30]. Through introducing an extra scalar field, spontaneous symmetry breaking can occur and permit mass to the electroweak bosons, and SM fermions in both a gauge invariant way.

2.1.3 The Brout-Englert-Higgs Mechanism

Under the belief that $\mathcal{L}_{EWK}$ is an accurate description of the unified electroweak force, both the bosons and fermions must be massless. This requirement is enforced by the Lagrangian to retain gauge invariance. However, as observed in nature, the known vector bosons $W^{\pm}$ and Z^0, and indeed the fermions, are massive and thus demand further treatment to the Lagrangian.

In the SM, the electroweak gauge bosons acquire mass through their interactions with the Brout-Englert-Higgs field, the mechanism of which is labelled the Brout-Englert-Higgs Mechanism [28–30].

The simplest Brout-Englert-Higgs model, that contains the four necessary degrees of freedom for the gauge bosons, consists of two complex scalar fields, belonging to the symmetry group SU(2):

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{2.23}$$

The additional term in the Lagrangian appears as:

$$\mathcal{L}_H = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi) \tag{2.24}$$

where the covariant derivative, D_μ, and the potential of the field, V(ϕ), which takes the form $-\mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2$ are used.

It can be observed that if $\mu^2 < 0$ and $\lambda > 0$, the field ϕ will acquire a non-zero vacuum expectation value (VEV), $< \phi >= \sqrt{\frac{\mu^2}{2\lambda}}$, the distribution of which can be seen in Figure 2.1.

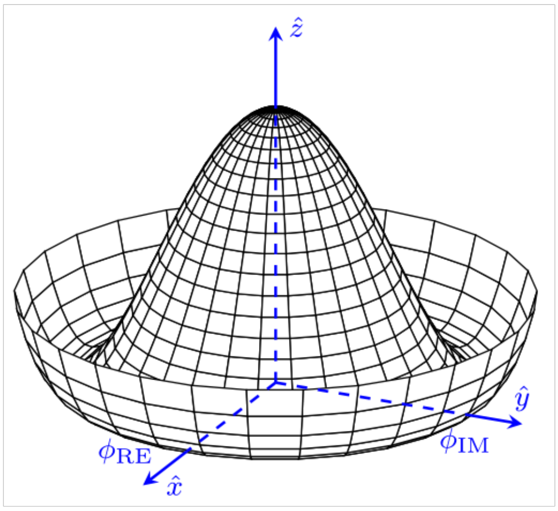


Fig. 2.1 Shape of the Brout-Englert-Higgs potential for $\lambda > 0$ as a function of the real and imaginary components of the complex field ϕ with $\mu^2 < 0$.

The act in determining where the field acquires a non-zero VEV spontaneously breaks the symmetry, and consequently results in the phase space of SU(2) to be no longer invariant.

After the spontaneous breaking of the symmetry, there will be a massive scalar and three massless Goldstone bosons, which ultimately give the longitudinal degrees of freedom of the $W\pm$ and Z bosons. The complex scalar field ϕ can be expanded about the physical vacuum, where for brevity, yields:

$$\phi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ \nu + h(x) \end{pmatrix} \tag{2.25}$$

where h(x) is the Brout-Englert-Higgs field, physical in the unitary gauge.

The masses of the gauge bosons can now be presented by considering the Lagrangian associated with $SU(2)_L \otimes U(1)_Y$. The terms that generates the masses of the gauge bosons is $(D_\mu\phi)^\dagger(D^\mu\phi)$:

$$(D_\mu\phi)^\dagger(D^\mu\phi) = \frac{1}{2}(\partial_\mu h)^2 + \frac{g_2^2\nu^2}{8}(W_\mu^+W^{+\mu} + W_\mu^-W^{-\mu}) + \frac{(g_1^2+g_2^2)\nu^2}{8}Z_\mu Z^\mu + A_\mu A^\mu \tag{2.26}$$

where the physical gauge bosons, $W_\mu^\pm$, Z_μ and A_μ for the fields W_μ^i and B_μ have been included.

In the Lagrangian, the mass terms will take the form $\frac{1}{2}m^2\chi_\mu\chi^\mu$ for the electroweak gauge bosons, while the photon will remain massless:

$$m_W = \frac{g_2 v}{2} \tag{2.27}$$

$$m_Z = \frac{\sqrt{g_1^2 + g_2^2}v}{2} = \frac{m_W}{\cos\theta_W} \tag{2.28}$$

$$m_\gamma = 0. \tag{2.29}$$

The Brout-Englert-Higgs mechanism, in particular the spontaneous symmetry breaking of the $SU(2)_L \otimes U(1)_Y$ gauge group, generates the masses of the W and Z bosons.

The fermion mass terms are acquired through Yukawa couplings with the Brout-Englert-Higgs field:

$$\mathcal{L}_{Yukawa} = -\frac{1}{\sqrt{2}}(\nu + h)(f_{mn}\bar{e}_{Lm}e_{Rn} + h_{mn}\bar{d}_{Lm}d_{Rn} + k_{mn}\bar{u}_{Lm}u_{Rn}) + \text{ hermitian conjugate} \tag{2.30}$$

where f_{mn}, h_{mn} and k_{mn} are the Yukawa coupling matrices between the different generations.

2.2 Beyond the Standard Model

Although the SM is an attractive and continuously successful theory [31], in describing the particles and their interactions of the universe, there is a global acknowledgement that it is a low-energy effective field theory valid below an energy scale Λ.

An immediate evidence of a weakness of the SM, is in its failure to incorporate the gravitational force into its description. Specifically, there is no interaction to explain the gravitational attraction between fundamental particles, the result of which means that any quantum description of gravity is yet to be found.

Furthermore, in the early universe, it is believed that matter and anti-matter were created in equal amounts through a process called "Baryogenesis". However, the observed dominance of matter over antimatter in the universe implies that some physical process created an asymmetry between the two types of matter. Such a process requires the violation of Charge-Parity (CP) symmetry, yet there are insufficient CP violating processes within the SM to account for this composition [32, 33].

The observation of neutrino oscillations, as a means to explain why fewer ν_e arrived from the Sun as what was theoretically predicted, implies that there is a mass difference between the different mass eigenstates of the neutrinos [34]. These differences of the mass eigenstates of neutrinos, which are a linear superposition of the weak flavour eigenstates, suggest that neutrinos do indeed have mass and from which show a clear contradiction to the massless, left-handed neutrinos predicted by the SM, in requiring local gauge invariance in the electroweak sector.

Other examples of limitations in physics that are not resolved by the SM are in certain cosmological observations. The form of velocity distributions of stars as they orbit the centre of mass of a galactic system, implies that there is a non-luminous component which contributes to the mass of the system [35]. This implication arose from the observation that for a spiral galaxy, such as the Milky Way, stars revolved around the centre of mass of the galaxy at equal or increasing speed, over a large range of distances. In contrast, the orbital velocities of moons and planets, that have most of their mass at the centre, decline with increasing distance. From this observation, it can be concluded that the mass of a galaxy has a significant contribution of dark matter (DM) This assumption is further indicated by studies of gravitational lensing [36] and the Cosmic Microwave Background (CMB) [37] [38], each indicating that the mass content on the universe is composed of DM.

Attempts to understand the characteristics of DM typically look for DM candidates colliding with matter, signatures from DM annihilation in space or the production of DM from SM collisions in a particle collider, the results of each currently show no direct observations.

The Hierarchy Problem

The experimental signs of imperfection of the SM are also accompanied by theoretical problems, in particular when calculating the radiative corrections to the Brout-Englert-Higgs boson mass.

By considering the Brout-Englert-Higgs boson mass, there are quantum corrections from fermion loops as shown in Figure 2.2.

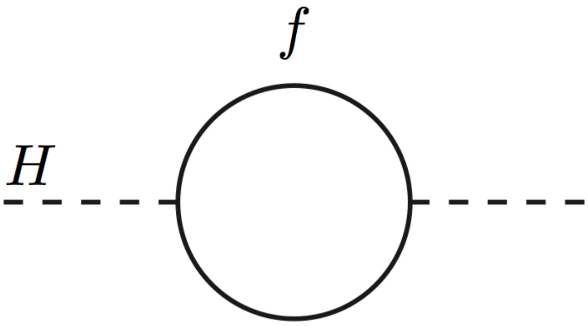


Fig. 2.2 Quantum corrections to the Brout-Englert-Higgs mass from a fermion loop.

Such quantum corrections ultimately yields a correction to the Brout-Englert-Higgs mass of the form:

$$\Delta m_H^2 \propto -|\lambda_f|^2\Lambda^2 \tag{2.31}$$

where Λ is an ultra-violet cut-off energy scale, typically the Planck scale ($\Lambda_{Planck} \sim 10^{19}$ GeV) beyond which the rules of SM are crucially no longer valid. To maintain the light Brout-Englert-Higgs boson, observed experimentally with a mass $\simeq$125 GeV [39], requires an accurate cancellation to the bare and corrected Brout-Englert-Higgs boson mass to a precision of 36 orders of magnitude.

This necessary “fine-tuning” is known as the *Hierarchy Problem* and despite being possible in the SM, is very unnatural and thus motivates the need for physics beyond the SM.

2.3 Supersymmetry

At its core, supersymmetry (SUSY) is the relation between fermions and bosons in the form of a continuous symmetry [40–45]. In SUSY, each SM particle has a superpartner "sparticle" which differs by a half unit of spin. As such, an operator of SUSY, $\hat{O}$, acting on a fermion (boson) state acts to change the spin by $\frac{1}{2}$, producing a superpartner boson (fermion) state, whilst maintaining electric charge, colour charge and weak isospin:

$$\hat{O}|f\rangle = |b\rangle; \qquad \hat{O}|b\rangle = |f\rangle \tag{2.32}$$

where $\hat{O}$ is the supersymmetric operator generating the transition. Through translations in superspace, as shown in Equation 2.32, particles can be transformed between fermionic and bosonic states.

However, no scalar particle with the same quantum numbers and mass of any fermion has been observed, and as such means that SUSY is a broken symmetry and that the mass scale of the SUSY particles is not known a priori. The implications of which imply that the masses of the superpartners exceed those of their SM counterparts, and thus cannot be predicted unless a SUSY breaking mechanism is first defined.

2.3.1 The Minimal Supersymmetric Standard Model

A generic model of SUSY is the *Minimal Supersymmetric Standard Model* (MSSM), in which there is a minimum number of new particle states, with one extra Brout-Englert-Higgs doublet [1]. Within the MSSM, for each of the SM fermions, f, there exists an associated left-handed, $\tilde{f}_L$, and right-handed, $\tilde{q}_R$, scalar superpartner[4]. Furthermore, the list of squarks is $\tilde{q}_L$, $\tilde{q}_L$ where $q = u, d, s, c, b, t$. The spin-1 SM gauge bosons have spin-1/2 gaugino partners, the gluino, winos and bino, $\tilde{g}$, $\tilde{W}^{\pm}$, $\tilde{W}^0$, $\tilde{B}^0$. The Higgs sector is extended to include two Higgs-doublets giving five mass-eigenstates, h^0, H^0, A^0 and $H^{\pm}$. Finally, the gauginos and the neutral higgsinos mix to give the neutralinos, $\tilde{\chi}^0_{i=1-4}$, and the winos and charged higgsinos mix to give the charginos, $\tilde{\chi}^{\pm}_{i=1-2}$. It should be noted that the neutralinos and charginos are linear combinations of superpartners to SM counterparts, and are thus not supersymmetric.

---

[4] The handedness does not refer to the helicity of the sfermions, but to that of their superpartners.

Names	Sparticle	Spin
Squarks (up-type)	$\tilde{u}_L$, $\tilde{u}_R$, $\tilde{c}_L$, $\tilde{c}_R$, $\tilde{t}_1$, $\tilde{t}_2$	0
Squarks (down-type)	$\tilde{d}_L$, $\tilde{d}_R$, $\tilde{s}_L$, $\tilde{s}_R$, $\tilde{b}_1$, $\tilde{b}_2$	0
Charged sleptons	$\tilde{e}_L$, $\tilde{e}_R$, $\tilde{\mu}_L$, $\tilde{\mu}_R$, $\tilde{\tau}_L$, $\tilde{\tau}_R$	0
Sneutrinos	$\tilde{\nu}_{eL}$, $\tilde{\nu}_{eR}$, $\tilde{\nu}_{\mu L}$, $\tilde{\nu}_{\mu R}$, $\tilde{\nu}_{\tau L}$, $\tilde{\nu}_{\tau R}$	0
Gluino	$\tilde{g}$	1/2
Neutralinos	$\tilde{\chi}^0_{i=1-4}$	1/2
Charginos	$\tilde{\chi}^{\pm}_{i=1-2}$	1/2
Brout-Englert-Higgs bosons	h^0, H^0, A^0, $H^{\pm}$	0

Table 2.3 Summary of the sparticles and their associated spin within the MSSM.

2.3.2 R-Parity

In the current representation of particle physics, a theory is constructed by specifying the particle content and any associated symmetries. All the terms subsequently allowed by the symmetries are included in the Lagrangian.

Under the representation of SUSY, there are terms that do not conserve lepton and baryon number, and as such consequently mediate proton decay as illustrated in Figure 2.3.

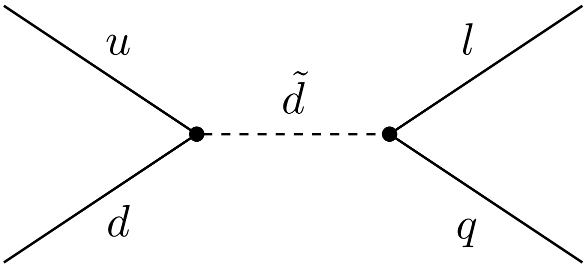


Fig. 2.3 Example of a lepton and baryon number interaction which would mediate proton decay.

Crucially, if such processes were permitted then the predicted lifetime of the proton would be very short, which is in stark contrast to established measurements, $\tau > 10^{33}$ years [46]. The result of which, indicates that interactions of the form above must be highly suppressed, if allowed at all.

A natural way to achieve this is by introducing a new quantum number, R-

Parity [47], which is defined as:

$$R_P = (-1)^{3B+L+2s} \left(= \begin{array}{c} +1 \text{ for SM particles} \\ -1 \text{ for SUSY particles} \end{array} \right), \tag{2.33}$$

where B is Baryon number, L is Lepton number and s is spin. For SM particles, $R_P = +1$, whereas SUSY particles have $R_P =$ -1.

R-Parity invariance is of great experimental relevance in that it implies all superpartner fields must be produced in pairs and must decay to states of which at least one is an RP odd particle. Furthermore, it also implies that the lightest supersymmetric particle (LSP) must be stable. Within the MSSM, the LSP is taken to be the lightest neutralino $\widetilde{\chi}^0$. This LSP is a weakly interacting stable particle and a possible candidate for the dark matter in the universe, as a WIMP [48].

2.3.3 Resolutions to the SM shortcomings

Unification of the Gauge Couplings

In many grand unified theories, the coupling constants of the SM, associated with U(1), SU(2) and SU(3) gauge symmetries, converge at some high energy unification scale, above which the three interactions unify.

While this is not possible in the SM, specifically there is no energy scale at which the coupling constants unite, the addition of SUSY finds that the coupling constants converge remarkably to a single value of $\alpha_{GUT} \approx \frac{1}{26}$ at 10^{16} GeV.

Including gravity

Of the four fundamental forces of nature, the SM fails to provide a mathematical description for gravity. A significant aim in developing a Theory of Everything (TOE) is to unify the classical theory of general relativity and the SM. One of the most active ways in which a theory of quantum gravity can be formed, is with string theory. In such a theory, SUSY is often a requirement, at some level, and as a result, any evidence of SUSY could lead towards a discovery of string theory.

Providing a DM candidate

In R-Parity conserving models, the LSP is electrically and colour-neutral from cosmological constants. In the MSSM, R-parity conservation requires that superpartners be

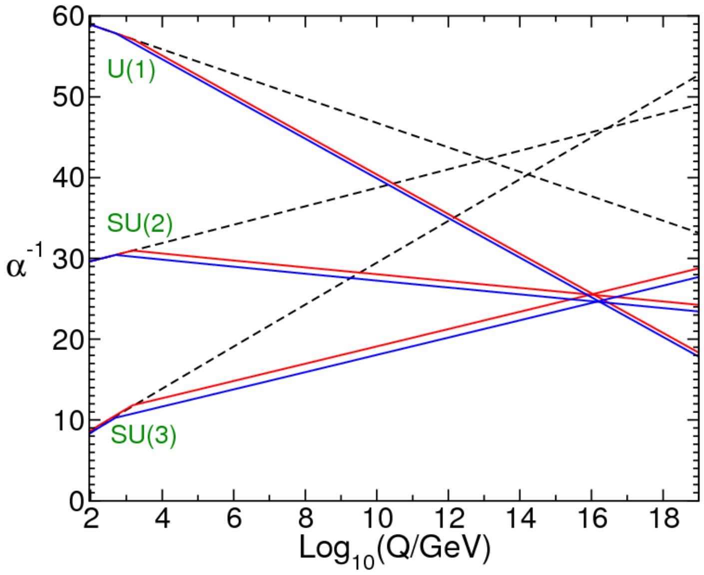


Fig. 2.4 The strengths of the gauge coupling parameters, α^{-1}_{1-3}, as a function of the energy scale Q. Dashed lines show the form of the couplings under a SM-only assumption, whereas the coloured lines show their form with the inclusion of the MSSM [1]

formed in pairs in collisions and that each superpartner decays to another superpartner in a chain that must end with the LSP. Therefore, the LSP is a natural and viable DM candidate. Within the MSSM, the possible DM candidates are the sneutrino with spin zero, the neutralino with a half integer spin, and the gravitino with spin 3/2. In addition, the LSP candidate that is most often considered is the lightest neutralino.

Hierarchy Problem

One of the prime motivations for SUSY is in the elegant solution it provides to the Hierarchy problem. As discussed in Section 2.2, quantum loops in the Brout-Englert-Higgs boson propagator contribute to the mass of the Brout-Englert-Higgs boson. While this in itself is not a problem, if the SM is to be part of a theory that is valid up to very high mass scales, such as that of a Grand Unified Theory, $\Lambda_{GUT} \approx 10^{16}$GeV, or the Planck scale $\Lambda_P \approx 10^{19}$GeV, these quantum corrections increase greatly and therefore make it difficult to keep the Brout-Englert-Higgs mass at the electroweak scale of 10^2 GeV.

SUSY can provide a natural solution to the aforementioned Hierarchy problem, as for every loop of particles there is a corresponding loop of sparticles, which provide a correction of opposite sign to the mass of the Brout-Englert-Higgs boson.

This can be seen by reconsidering the one loop corrections to the Brout-Englert-Higgs field, h, with massive fermions ψ and massive scalars ϕ, following a supersymmetric transition:

$$\begin{aligned} m_H^2 \sim m_{H0}^2 + \frac{\lambda_F^2}{4\pi^2}(\Lambda^2 + m_F^2) - \frac{\lambda_S^2}{4\pi^2}(\Lambda^2 + m_S^2) \\ + \text{ logarithmic divergences } + \text{ uninteresting terms } . \end{aligned} \tag{2.34}$$

As can be seen in Equation 2.34, there is a minus sign between the two quadratic terms in Λ. Further, if the couplings between the Brout-Englert-Higgs field and the fermion λ_F and scalar λ_S are equal then the quadratic contributions on Λ will cancel each other. Under the assumption that SUSY is an exact symmetry of nature, the sparticle masses would have the same masses as their respective particles, and hence the cancellation would be exact.

The largest contribution to the mass of the Brout-Englert-Higgs boson is from the top quark. Therefore, with regards to fine tuning the corrections to the Brout-Englert-Higgs mass, the supersymmetric partner to the top quark, the top squark, is of great interest. The correction to the Brout-Englert-Higgs mass due to top quark loop and the corresponding loop correction from the top squark is shown in Figure 2.5.

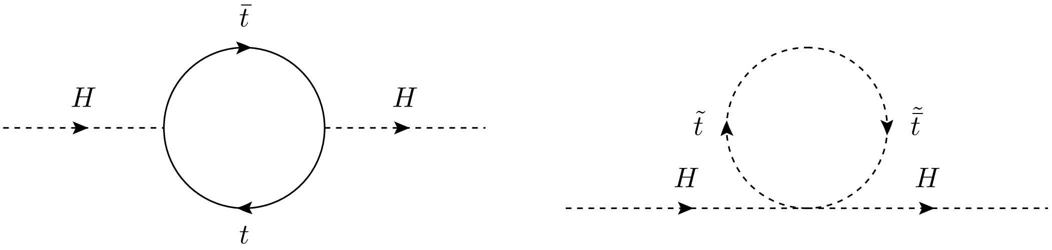


Fig. 2.5 The loop contributing to the Higgs mass due to the top quark, left, and the cancellation of the loop due to the top squark, right.

2.4 Experimental Stance

The motivations for SUSY, in particular the MSSM, as an extension to the SM aimed to resolve the aforementioned shortcomings, have driven experimentalists to focus on searching for signs of SUSY particles.

In the MSSM, in which R-Parity is conserved, sparticles are predicted to decay into final states containing the LSP, $\tilde{\chi}^0$, which being neutral escapes detection. As a result, the decay chain of a MSSM particle can have several stages, referred to as cascade decays. In such cases, a heavy sparticle, produced in a collision, decays down the SUSY spectrum where for each step of the decay, a SM particle is emitted until the remaining energy is small enough to emit only the LSP and a final SM particle. An example of a decay cascade for a gluino and a squark is illustrated in Figure 2.6.

The ensemble of possible decays, in addition to the energies involved, depend on the SUSY spectrum considered. Of the MSSM, the signatures of sparticle decays are typically multiple final state objects and a significant component of missing transverse momentum, p_{T}^{miss}, from the unobserved neutralinos, $\tilde{\chi}^0$.

2.4.1 Simplified Models Spectra

Historical searches for SUSY typically utilised models that constrained the large number of parameters associated to SUSY models. Despite their simplicity, such models produce complex mass hierarchies that can alter the topologies and signatures.

Therefore, to remain both as inclusive as possible and able to translate any results into consistent theories, signature-driven models known as Simplified Model Spectra (SMS) have been developed [49, 50]. In general, there are two free parameters associated with a SMS model: the mass of a produced mother particle m_{mother} and the mass of the LSP, m_{LSP}. Therefore, SMS consist of simplified decay scenarios which can later interpreted in the mass planes of a sparticle, typically in the mass of the $\tilde{g}$, $\tilde{t}$, $\tilde{b}$, $\tilde{q}$ and $\tilde{\chi}^0$ for example.

A consequence of the unknown masses of sparticles is the potential for different final kinematic properties assigned to the final states. In the case where the mass between the mother sparticle, m_{mother}, and the LSP, m_{LSP}, is small enough, the model is said to be “compressed”. This implies that the pair-produced mother sparticle has a mass that is close to that of the LSP, and as such poses a strong experimental challenge in identifying the softer decay products from events that could potentially also fall out of analysis acceptance. The reverse phenomenon, in which models have larger mass separations, is also theoretically possible. Such “uncompressed” regimes, illustrated in Figure 2.7, exhibit different characteristics with typically large energies detected.

In the situation where the mass of the pair-produced mother sparticle is much larger than that of the LSP, an excess of kinetic energy is formed, which can be transformed into a Lorentz boost on the decay products. The affect of such a Lorentz

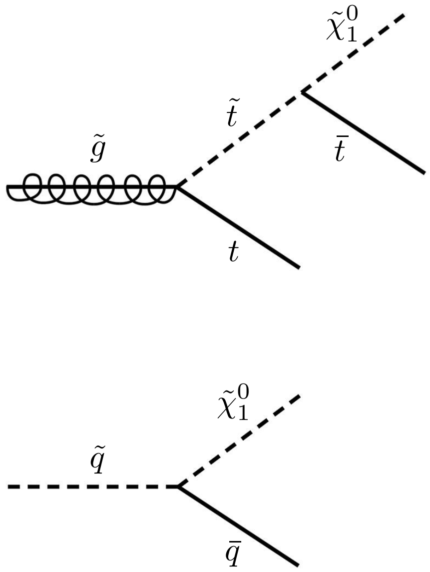


Fig. 2.6 Feynman diagram of the hadronic decay of a gluino (top) and squark (bottom) where $\widetilde{\chi}^0$ is the LSP.

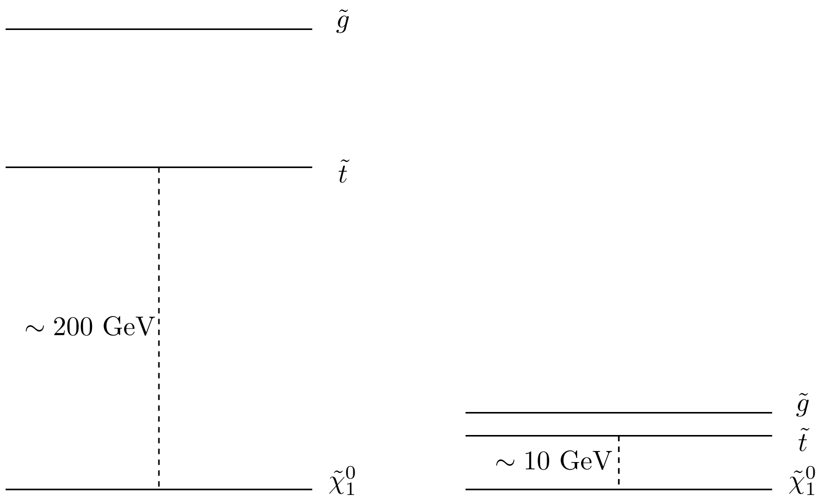


Fig. 2.7 An example SUSY spectrum in the bulk region, left, and for a compressed spectrum, right. The mass difference between the mother and the LSP, here the $\tilde{t}$ and $\tilde{\chi}^0$ receptively, is much reduced for the compressed scenario.

boost on particles reduces the angular separation between them and, in the context of decay products, alters the way in which they can be experimentally observed and the mother particle reconstructed. An example of which, illustrated in Figure 2.6, is when the mass of $\tilde{t}$ is sufficiently larger than $\tilde{\chi}^0$ such that the top quark emitted from the $\tilde{t}$ will receive a Lorentz boost. The consequence of which is that the decay products of the top quark would collimate and as such would appear as a different topology and reconstructed object in a detector capable of identifying and quantifying the effect of the hadronisation of such a quark.

The potential advantage of such a scenario is in the uniqueness to reconstruct a parent particle through the collimated decay products, a process likely mediated through the MSSM sparticles, and the boost they can provide, than through particles of the SM.

Direct gluino pair production

In the theory of the MSSM, in which R-Parity is conserved, sparticles are produced in pairs. Thereafter, each sparticle will decay in a chain to the LSP. The direct pair production of gluinos, $\tilde{g}$, with the advantage of a larger cross section with respect to stop quarks and as illustrated in Figure 2.8, is a key sparticle to be targeted. Such models in which gluinos are pair produced are referred to as *T1* models.

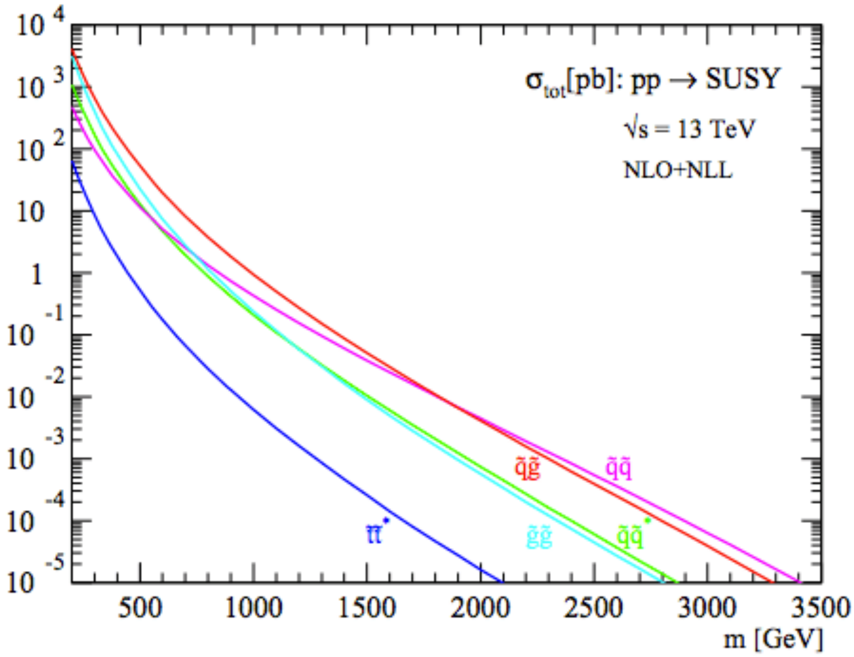


Fig. 2.8 NLO+NLL production cross sections for the case of equal degenerate squark and gluino masses as a function of mass at $\sqrt{s} = 13$ TeV [2].

In order to optimise any search for such a sparticle, the possible decay modes in different regions of parameter space need to be considered. For the pair production of $\tilde{g}$, a particular decay mode that can be considered is the three body decay to a light-flavour quark-antiquark pair and the LSP, $\tilde{g} \to q\bar{q}\tilde{\chi}^0$. By setting the branching ratio of a particular decay to unity, a $\tilde{g}$ can decay exclusively into bottom or top quark pairs. In the case where top quark pairs are the decay mode, a typical final state will consist of large hadronic activity from the quarks, and missing transverse momentum from the LSP. In a similar manner to the decay of the stop quark, the ability to measure the hadronic activity in a detector will be partially dependent on the mass splitting between the mother sparticle, $\tilde{g}$, and the LSP, $\tilde{\chi}^0$.

Direct stop pair production

As illustrated in Figure 2.6, the stop quark $\tilde{t}$ decay is a two body decay to a light-flavour quark and the LSP. Fundamentally, a typical final state expected from this decay mode would be hadronic activity and missing transverse momentum from the LSP. For the case when the mass difference between the stop quark and the LSP is large, typically greater than the mass of top quark, the decay mode is $\tilde{t} \rightarrow$ t $\tilde{\chi}^0$ [51], the schematic of which is illustrated in Figure 2.6.

However, when the mass difference is less than the top quark but above the mass of the W boson, $m_W < m_{\tilde{t}} - m_{\tilde{\chi}^0} < m_t$, the three body decay mode dominates $\tilde{t} \rightarrow$ b W $\tilde{\chi}^0$. Furthermore, when the mass difference is less than the W boson mass, both the flavour changing neutral current decay $\tilde{t} \rightarrow$ c $\tilde{\chi}^0$ and the four body decay $\tilde{t} \rightarrow$ b $\tilde{\chi}^0$ f $\bar{f}$ can occur, where in the latter case, f is a fermion [51].

2.5 Current Experimental Limits

There are many signatures of SUSY, each of which are typically categorised according to the final states. Hadronic searches, which are characterised by a high jet multiplicity, provide a fertile ground for searching for SUSY. Despite final states offering a more clean signature, such as leptons and photons, the larger branching fraction of W bosons to hadrons and the strong interaction of sparticles makes hadronic searches particularly sensitive.

Both the ATLAS [52] and CMS [5] collaborations have so far failed to identify an observation of events from a supersymmetric process, and as a result have placed limits for direct $\tilde{t}$ and $\tilde{g}$ production.

The current and previous limits on the masses of the stop quark and the gluino are summarised in Table 2.4.

Regarding the stop quark, results for both the ATLAS and CMS collaborations were interpreted for the decay of the stop quark into a quark and LSP. Furthermore, the ATLAS and CMS collaborations interpreted the decay of the gluino into several different decay modes. The decay of the gluino into a top anti-top quark pair and the LSP, provided upper limits on the gluino mass up to 1950 GeV for both the CMS and ATLAS collaborations.

Concerning the ATLAS collaboration, stop quark masses are currently excluded up 950 GeV on a dataset corresponding to an integrated luminosity of 36.1 fb^{-1}. On a dataset of the same magnitude, stop decay limits from the CMS collaboration have excluded stop quark masses up to 1.1 TeV. Both the limits of stop and gluino decay modes from the ATLAS and CMS collaboration are illustrated in Appendix B.

Limits on the mass of sparticles					
		ATLAS		**CMS**	
		Limit on mass (GeV)		Limit on mass (GeV)	
Sparticle	Decay mode	13.3 fb^{-1}	36.1 fb^{-1}	12.9 fb^{-1}	35.9 fb^{-1}
stop	$\tilde{t} \rightarrow \mathrm{t}\ \tilde{\chi}^0$	850	950	900	1150
gluino	$\tilde{g} \rightarrow \mathrm{t}\bar{\mathrm{t}}\ \tilde{\chi}^0$	1650 (18.2 fb^{-1})	1950	1700	1950

Table 2.4 Exclusion limits, per integrated luminosity, on the stop quark mass and the gluino mass for both the ATLAS [18] and CMS collaborations [17].

CHAPTER 3

THE LHC AND THE CMS DETECTOR

3.1 The Large Hadron Collider

Situated between the French-Swiss border, the Large Hadron Collider (LHC) is a 26.7 km circumference proton-proton synchrotron, located several dozen meters underground. The design purpose of the LHC is to collide beams of protons, and separately heavy ions, at a centre of mass energy $\sqrt{s} = 14$ TeV and with an instantaneous luminosity of $10^{34}\text{cm}^{-2}\text{s}^{-1}$.

In the context of proton-proton (pp) collisions, at the LHC there are two counter-rotating proton beams arranged in bunches that orbit at a rate of 400 MHz through radio frequency (RF) cavities, that are each able to focus the beam using multiple quadrupole and higher order magnets. The magnets are cooled to 1.9 K at a nominal current of 11.85 kA, such that the magnetic field required to maintain beam trajectory is achieved.

Prior to accelerating protons, the source of protons is obtained by placing hydrogen gas within an electric field to ionising the hydrogen molecule and yield protons and electrons.

The protons are then accelerated through the linear accelerator LINAC2, where the energy of the protons reaches 50 MeV. Thereafter, the protons are accelerated to 25 GeV by the Proton Synchrotron (PS) in which proton beams, composed of bunches of approximately $\mathcal{O}(10^{11})$ protons localised into less than 1 ns in the direction of motion, are formed. The final pre-accelerator in this chain, the Super Proton Synchrotron (SPS) raises the energy of the protons to 450 GeV prior to injection into the LHC, the chain of which is shown in Figure 3.1. Following additional energy adjustments, the proton beams are fed into the LHC via transfer lines TI2 and TI8, at which point the energy reaches 6.5 TeV.

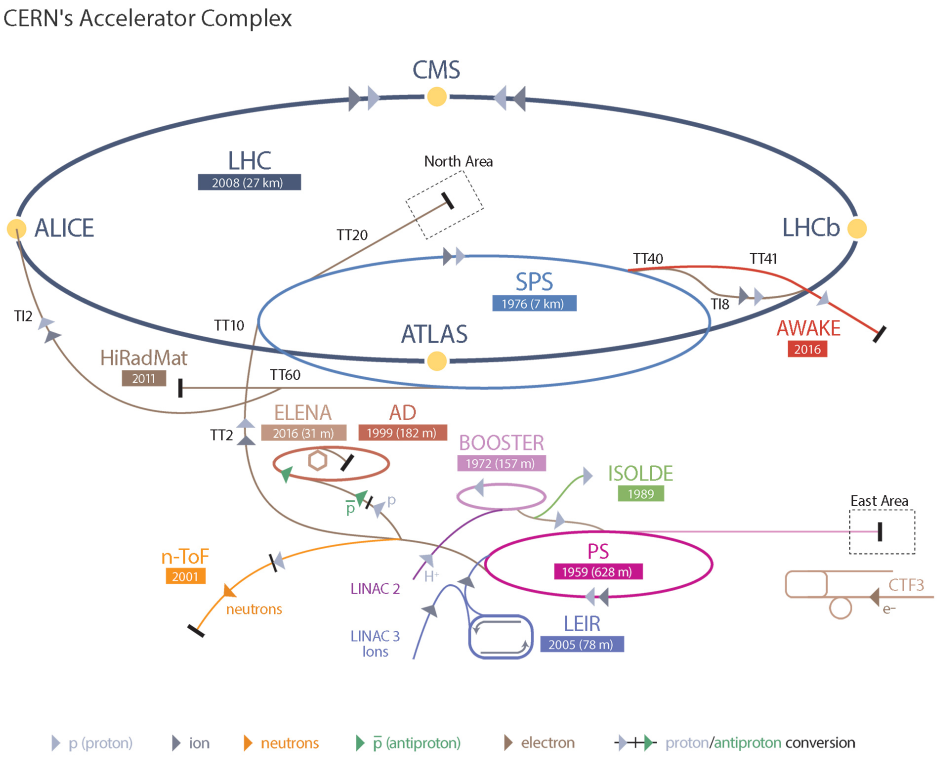


Fig. 3.1 The CERN accelerator complex [3].

As the beam of protons are injected into the LHC, they acquire a large transverse and longitudinal momentum which appears as low-rate collisions. To stabilise the momentum and phase of the oscillation, an oscillating electric field is employed in the direction of the beam. The structure of the beams consists of approximately 2076 bunches [53] of $\mathcal{O}(10^{11})$ protons, with a bunch spacing of 25 ns that in turn corresponds to a bunch crossing frequency of 40 MHz.

The rate of pp-collisions is dependent on the inelastic cross section and the instantaneous luminosity, given by the equation:

$$L = f\frac{n_1 n_2}{4\pi\sigma_x\sigma_y} \tag{3.1}$$

where f denotes the bunch collision frequency, the number of particles in the colliding bunches, n_1 and n_2, and the root-mean-square horizontal and vertical beam sizes σ_x and σ_y.

In high luminosity environments, simultaneous *pp* interactions while bunches cross, referred to as in-time pileup (PU), and overlapping particle decays from previous bunch crossings, referred to as out-of-time pileup (OOTPU), can cause a significant experimental challenge. These additional interactions pose problems for detector readout and event reconstruction, specifically when aiming to identify particular physics processes from selected events of a detector.

At the LHC, there are four main experiments; LHCb [54] (Large Hadron Collider Beauty experiment), ALICE [55] (A Large Ion Collider Experiment), and ATLAS [52], (A Toroidal LHC Apparatus) and the CMS [5] (Compact Muon Solenoid) detectors. While LHCb and ALICE study b-physics and heavy ion physics respectively, ATLAS and CMS are more general purpose detectors aimed at discovering new physics at the TeV energy scale.

3.2 The Compact Muon Solenoid Detector

The Compact Muon Solenoid (CMS) detector, illustrated in Figure 3.2, is a general all purpose detector situated at Point 5 on the LHC ring [5]. The detector is designed to perform high precision measurements of physics objects including muons, electrons, photons, taus and hadronic showers. The identification of physics objects is made possible through a 4π solid angle reconstruction with efficient particle identification and reconstruction.

The detector measures 21.6 m in length, 14.6 m in diameter and weighs 12500 tonnes [5]. It is composed of different subdetectors, built around a central 12.5 m long 3.8 T superconducting solenoid magnet, where each subdetector measures a certain characteristic or attribute of a particle. The CMS detector consists of a barrel region which encloses the aforementioned solenoid, and endcaps to extend for forward and backward coverage.

The common geometry associated to the CMS detector is a right-handed coordinate system with the z-axis as the longitudinal dimension along the direction of the beam, the x-axis to point perpendicularly towards the centre of the LHC ring, and the y-axis vertically upwards.

Furthermore, the position of a particle is described by the angular direction in the transverse plane with range $[-\pi, \pi]$ and an angle with respect to the z-axis, labelled ϕ and η respectively. The latter variable is defined “pseudo-rapidity” and represented as:

$$\eta = -\ln\ \tan\left(\frac{\theta}{2}\right), \tag{3.2}$$

where θ is the polar angle between the physics object and the z-axis in the y-z plane. The distance between particles is commonly described in terms of the variable $\Delta\mathrm{R} = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

Moving radially outwards, the detector is composed of a high-resolution pixel tracker, a granular strip tracker, an active lead tungstate electromagnetic calorimeter (ECAL), a sampling brass-plastic hadronic calorimeter (HCAL), an outer HCAL and three different muon detectors.

The solenoid measures the curved trajectory of charged particles through the magnetic field, to determine their trajectories and momenta. The calorimeters, located outside of the tracker, provide measurements of energy deposits from particles. An electron or photon passing through would cause electromagnetic showers in the crystals of the ECAL, which produce scintillation light that can be translated into the incident particle energy. Incident hadrons, when passing through the HCAL, shower and cause scintillation light to be produced and collected by photodetectors to determine the particle energy and position.

Due to the composite nature of the proton, inelastic collisions with large momentum transfer between quarks and gluons can occur. However, in a single bunch crossing, elastic scattering, where the momentum transfer is insufficient to force the proton to break up, can also occur. The consequence of both creates a densely populated environment in the detector and therefore poses some difficulty in probing any potential interaction of interest.

Furthermore, as a means to identify any new particle that is typically produced from a hard collision with significant energy, particle energy and momenta can be described in the transverse (x-y) plane. Through the conservation of energy and momentum transverse to the beam, any imbalance in momenta can be inferred as a particle passing through the detector with no trace. An example of which is a neutrino or a particle from a new physics process.

Hence, the aforementioned distinctive properties of the CMS detector are ideal to efficiently reconstruct particles and measure missing transverse energy, the observation of which could indicate signatures of new physics.

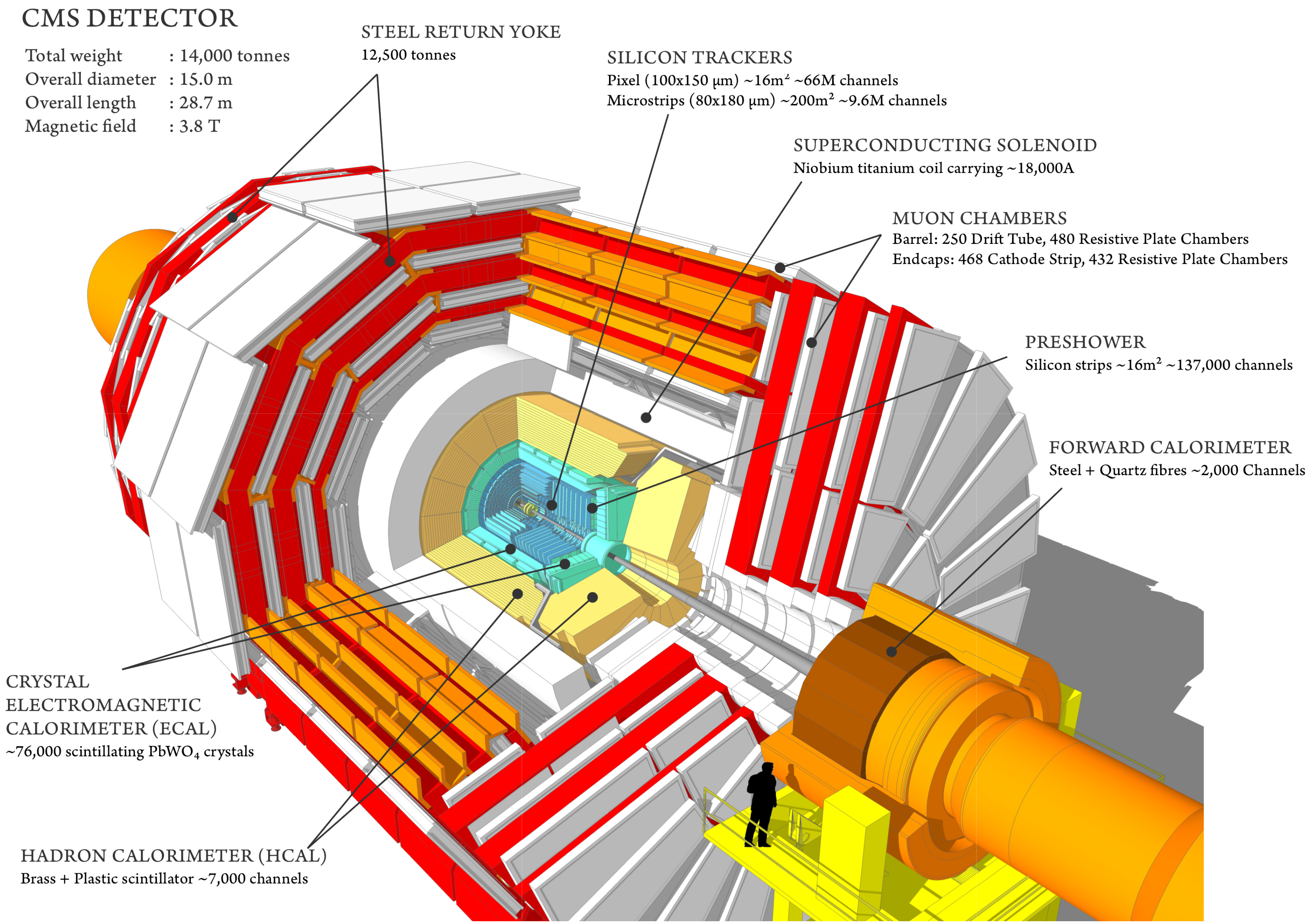


Fig. 3.2 Diagram of the Compact Muon Solenoid [4]

3.2.1 Tracker

The tracking system of CMS is designed to make precise and efficient measurements of charged particle trajectories as they emerge from the interaction point. Furthermore, secondary vertices are also able to be reconstructed, allowing the identification of heavy flavour decays such as bottom quark jets.

The tracking system is 5.8 m long with a diameter of 2.5 m and has a coverage of $|\eta| < 2.5$ [56]. To operate effectively in high instantaneous luminosity conditions and high particle fluxes, certain design criteria of the tracker are required. The granularity of the tracker should maintain a low detector occupancy and provide efficient and precise measurements of tracks, that will enable individual bunch crossing to be resolved. A consequence of such a feature is the need for the tracker to provide both a fast response and be radiation hard, to withstand the high particle fluxes. To meet these requirements, CMS utilises a silicon detector [56].

The silicon tracker system is composed of pixel-based inner and microstrip-based outer detectors. The pixel detector has three layers in the barrel, the closest with a radius of 4.4 cm. This is accompanied by end-caps with two pixel detector disks at distances $z = \pm 34.5$ and ± 46.5 cm. The radial region between 20 cm and 116 cm contains a silicon strip tracker, which is composed of three different subsystems: The Tracker Inner Barrel and Disks (TIB/TID), Tracker Outer Barrel (TOB) and Tracker Endcaps (TEC), the geometry of each is shown in Figure 3.3.

The CMS silicon strip tracker has a total of 9.3 million strips and an active area of silicon of 198 m^2. The sensors range in thickness and pixel pitch, depending on the tracker subsystem. The resultant momentum and spatial resolution of the tracker is excellent. For a 100 GeV muon, the transverse momentum resolution is 1-2 % up to $|\eta| = 1.6$ [5].

3.2.2 Electromagnetic Calorimeter

The Electromagnetic Calorimeter (ECAL) is designed to measure the energies of photons and electrons as they proceed in the detector, losing energy through radiation [56]. The ECAL covers a pseudorapidity range of $|\eta| < 3$, and is composed of over 75,000 lead tungstate crystals, which scintillate as particles deposit energy. Following which, this scintillated light is amplified by photodetectors.

Lead tungstate ($PbWO_4$) is chosen as the key material of which the ECAL is composed of due to its dense structure, short radiation length and small Molière

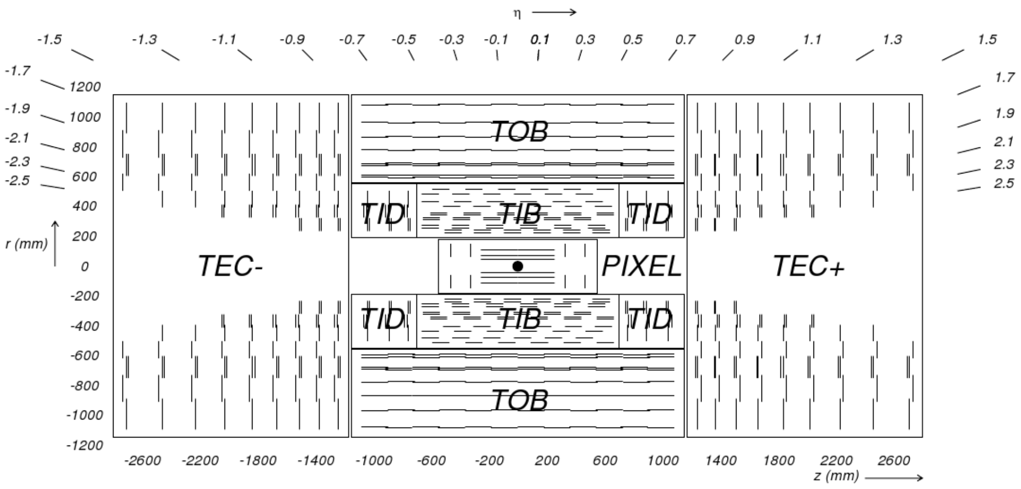


Fig. 3.3 The CMS tracker, shown in the r-z plane. The pixel detector is shown at the centre of the tracker, closest to the interaction region (shown by the black dot), and the strip detector surrounds it. The different subsystems of the strip detector are shown [5].

radius. This in turn, makes it well suited to be a compact and granular calorimeter, with the dimension of a crystal being 0.017 x 0.017 ($\Delta\eta \times \Delta\phi$).

The ECAL is composed of three subdetectors, the Electromagnetic Calorimeter Barrel (EB) which covers the range $0 < |\eta| < 1.479$, and the Electromagnetic Calorimeter Endcap (EE), covering the range $1.479 < |\eta| < 3$ and the Electromagnetic Preshower (ES) The crystals in the barrel have a front face of 22×22 mm^2 and a length of 23 cm, corresponding to approximately 25.8 radiation lengths (X_0). The crystals in the endcap have a front face of 28.6×28.6 mm^2 and length corresponding to 24.7 X_0. The energy of the incident electromagnetic particles is measured through the scintillation light produced in the crystals. The scintillation light yield from PbWO$_4$ crystals is relatively low (30 γ/MeV) and requires photoamplifiers capable to operate within a magnetic field and amplify the signal. Silicon avalanche photodiodes (APDs) are used as photodetectors in the barrel region while vacuum phototriodes (VPTs) are used in the endcaps as they are less sensitive to the high radiation conditions in the forward regions.

The energy resolution of the ECAL can be represented in Equation 3.3.

$$(\frac{\sigma_E{}^2}{E}) = (\frac{S}{\sqrt{E}})^2 \oplus (\frac{N}{E})^2 \oplus C^2 \tag{3.3}$$

where the parameter S is due to stochastic scattering, parameter N is due to noise and parameter C is an associated constant term. The parameters have been derived to be S = 2.8%, N = 12% and C = 0.3% [57]. For a 100 GeV particle the ECAL system has an energy resolution of $\sim 0.5\%$ [57].

3.2.3 Hadronic Calorimeter

The Hadronic Calorimeter (HCAL) is the second calorimeter in CMS, the aim of which is to measure energy depositions from hadronic showers [56]. The HCAL is a sampling brass calorimeter, built from alternating layers of brass absorber and plastic scintillator, of which there are 17 in the barrel (HB) and 19 in the endcaps (HE) The HB covers pseudorapidities up to $|\eta| < 1.3$, and the HE on each side covers a range of $1.3 < |\eta| < 3$. To achieve an almost hermetic calorimeter, there is a forward hadronic calorimeter (HF), covering an extended pseudorapidity range up to $|\eta| < 5.2$, and which compliments the other subsystems associated to the HCAL. Furthermore, to ensure adequate sampling depth for $|\eta| < 1.3$, the HCAL is extended outside of the vacuum tank of the magnet, supplementing the HB. The HO is used to identify late starting showers and to measure the shower energy deposited after HB, by using the solenoid coil as an absorber to promote interactions.

Hadron showers are created in the brass absorber plates, through nuclear interactions in the material. The plastic scintillator tiles receive the energy as scintillation light and reduces the frequency through wavelength shifting to then transfer the light down transparent fibres to hybrid photodetectors (HPD). After which, an electrical signal proportional to the incident hadron energy is produced.

Due to the harsher environment in the forward region, the HF uses a different technology to measure the energy depositions from hadronic particles. Steel absorber plates are embedded with scintillating quartz fibres, capable of detecting light emitted by charged particles in the shower.

The energy resolution of the HCAL is given by the expression in Equation 3.4 [58]

$$\frac{\sigma_H}{E} = \frac{94.3\%}{\sqrt{E}} \oplus 8.4\% \tag{3.4}$$

3.2.4 Solenoid

The core of the CMS detector houses a large-bore superconducting solenoid that generates a 3.8 T homogeneous magnetic field and provides sufficient influence to bend the trajectories of charged particles [56]. The solenoid is composed of five modules, each of inner diameter 6.3 m and 2.5 m in length, with a total weight of 220 tonnes. Each module has four winding layers of stabilised reinforced NbTi conductor, co-extruded with aluminium within a aluminium support.

The magnetic field is maintained by a 10,000 tone iron yoke that consists of five barrel wheels and six endcap disks [5]. To maintain the temperature to 4.5 K, a vacuum cryostat encloses the magnet and provides insulation and cooling with liquid helium.

3.2.5 Muon Systems

The muon system provides crucial identification and precise measurements of muon position and momenta [56]. Due to the clean experimental signature of muons, with respect to other particles, the muon system of CMS is a central design feature and demands high precision in the reconstruction of such muons.

The CMS muon system covers the entire kinematic range of the detector and employs three types of gaseous particle detectors to identify muons, as illustrated in Figure 3.4:

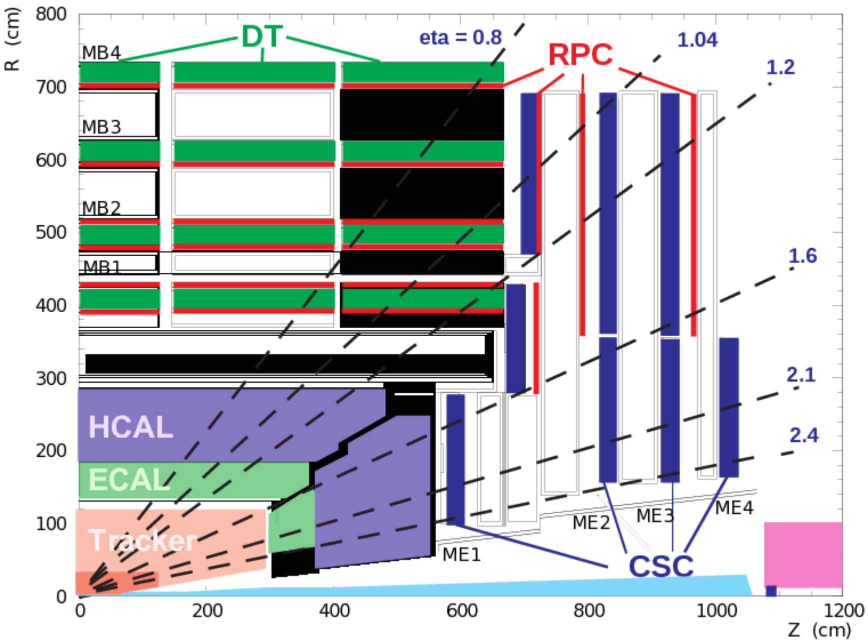


Fig. 3.4 Cross section of the CMS detector in the y-z plane, showing the position of the Drift Tube (DT), Resistive Plate Chamber (RPC) and Cathode Strip Chamber (CSC) subsystems, as well as other interior subdetector systems [6].

The central barrel region, $|\eta| < 1.2$, is equipped with Drift Tube (DT) detectors. The endcaps regions, $0.9 < |\eta| < 2.4$, where the muon rates and background levels are typically high, employs cathode strip chambers (CSC). An additional system of dual-layered Resistive Plate Chambers (RPCs) are utilised in both the barrel and endcaps, covering $|\eta|< 2.1$ and providing an independent signal.

With the tracker being able to supply information, the additional use of either the DTs or CSCs and RPCs allows CMS to have excellent muon reconstruction.

3.2.6 Trigger

As described in Section 3.1, bunches of protons at the LHC cross at a rate of 40 MHz, with an average of 20 collisions per crossing. This corresponds to approximately 1 MB of data per such crossing, which is equivalent to a data rate of $\mathcal{O}(10^{12})$ bytes per second.

While the inelastic cross section of proton-proton collisions at $\sqrt{s} = 13$ TeV is $\sigma \sim 71.3$ mb, the majority of the data being recorded by the detector contains relatively low energy soft scattering events, that are often of little interest. Furthermore, it is critical that the CMS detector swiftly retains possible signal events while satisfying the constraints on data bandwidth, storage and latency.

Due to the limitations on disk writing, specifically the speed at which the data can be saved to tape, and the bandwidth on the electronics of the Data Acquisition system (DAQ) employed at CMS, it is impossible to save all the data. To overcome this discrepancy between the production and storage rates, there must be a removal of some of the data, the acceptance or rejection of which is determined by trigger systems.

The core principle of a trigger system is to perform fast reconstruction and selection, to retain events of interest and discard those of lesser interest. In the CMS collaboration, events of interest are selected using a dedicated two-level trigger system, the Level-1 (L1) trigger executing its selections online, and the High-Level Trigger (HLT) executing its selections offline. A schematic diagram of the DAQ process, with the L1 and HLT trigger is shown in Figure 3.5.

Level-1 Trigger

The hardware-based L1 system is designed to reduce the event rate from 40 MHz to 100 kHz using information from the subsystems of the detector [59].

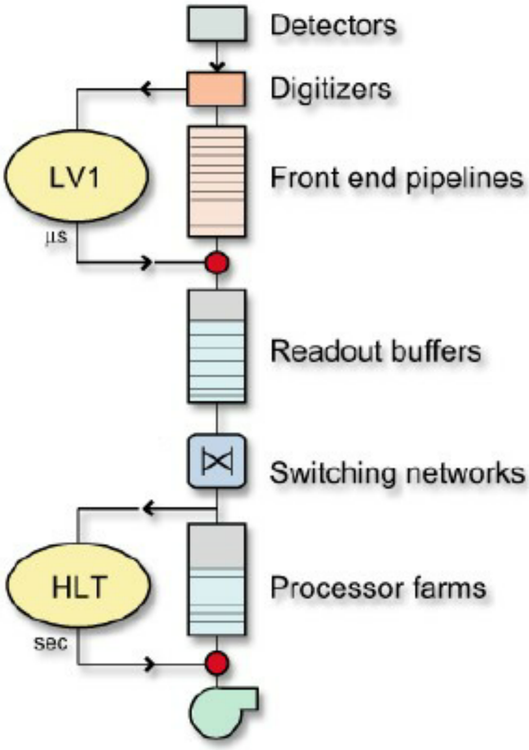


Fig. 3.5 Diagram of the CMS DAQ system showing the data flow from the detector through the L1 trigger and the HLT [7].

The subsystems used in the L1 trigger are both the electromagnetic and hadronic calorimeters and the muon chambers. The algorithms encoded in the L1 trigger execute decisions every 25 ns before either rejecting or passing the event to the HLT for further consideration. The trigger algorithms used to form the selections are implemented on a custom hardware platform composed of Field Programmable Gate Arrays (FPGAs) and Application Specific Integrated Circuits (ASICs)

During Run I and the commissioning period of Run II, the CMS L1 trigger was composed of a Regional Calorimeter Trigger (RCT), a Global Calorimeter Trigger (GCT), a Global Muon Trigger (GMT) and finally a Global Trigger (GT), an overview of which is shown in Figure 3.6.

However, due to the harsher environment expected from higher pileup and instantaneous luminosity of the CMS upgrade, following the long shutdown (LS1) in 2013, the L1 trigger was upgraded to maintain performance and keep the trigger rate below 100 kHz. The upgrades were divided into two stages, the former labelled Stage 1 and

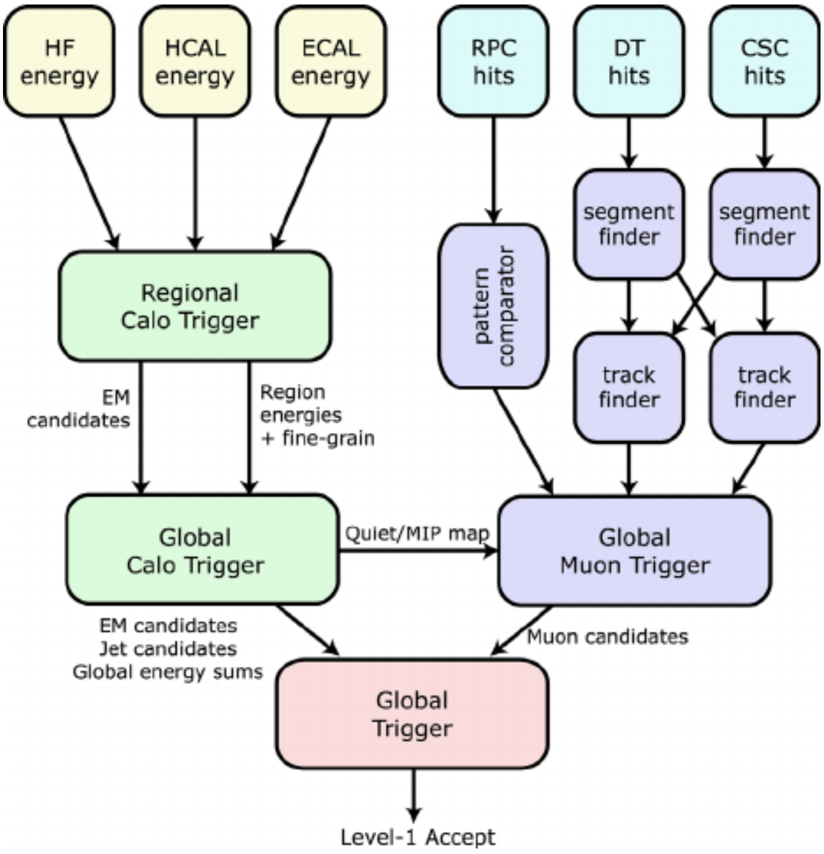


Fig. 3.6 An overview of the CMS L1 trigger system, showing the RCT, GCT, muon triggers and GT [8].

the latter labelled Stage 2.

The core upgrade of Stage 1 was to replace the GCT with the Layer-2 calorimeter trigger [9]. As part of a partial upgrade that commenced in 2015, the GCT was replaced with MP7 data processing cards capable of executing improved algorithms, such as a dedicated pileup subtraction on an event-by-event basis [60]. Such an upgrade demanded a detailed commissioning period, whereby both the GCT and the upgraded cards ran simultaneously to verify the logic of the algorithms and crucially, ensure that the data throughput and the computational power of the upgrade surpassed that of the GCT.

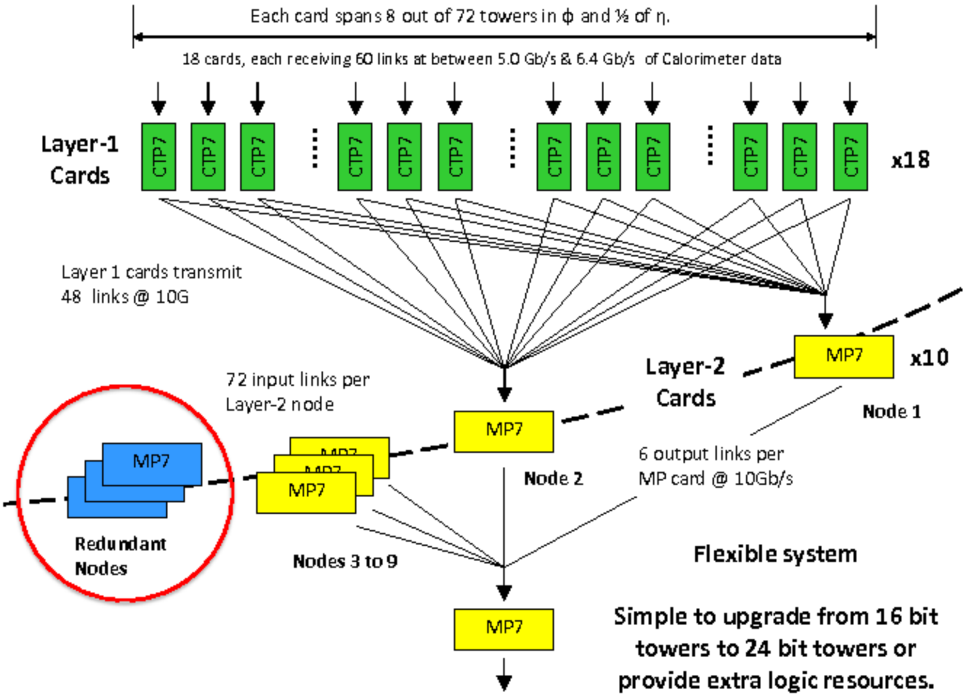


Fig. 3.7 An overview of the CMS Stage 2 architecture, showing the Layer-1 and Layer-2 calorimeter triggers [9].

The simultaneous running of Stage 1 and the GCT during the commissioning phase was achieved by duplicating the output of the RCT, as provided by the Optical Regional Summary Cards (oRSCs)

While the RCT layer was temporarily maintained, global algorithms were improved and refined pileup removal techniques were implemented. Furthermore, the copper links between the ECAL and the RCT were replaced with optical links.

During Stage 2, the Time Multiplexed Trigger (TMT) [61, 62] was employed and the Layer-1 calorimeter trigger, using CTP7 cards, replaced the RCT [63]. The full architecture of Stage 2, including the GCT upgrade to MP7 data process cards is illustrated in Figure 3.7.

As displayed in Figure 3.6, in the previous L1 trigger architecture, the RCT received data from the calorimeters whilst the GMT received its inputs elsewhere. From the outputs of the GCT and GMT, the GT makes a decision called the Level-1 Accept (L1A). This decision is then propagated to all detector subsystems so that the raw information may be read out for analysis at the HLT.

The current architecture of the L1 trigger issues a L1A following a decision based on the outputs of Layer-1 and Layer-2.

High-Level Trigger

The software based HLT is designed to reduce the event rate from the L1 trigger from 100 kHz to 100 Hz [64]. Due to the increase in processing time up to 50 ms, upon receiving a L1A signal, the HLT utilises the complete detector information, both at the full resolution and including the tracker, to improve on the reconstruction of the objects. At this stage, more detailed requirements on events can be constructed, which when applied select more events of specific interest and reject more background like events. The requirements can often be more sophisticated variables that may exploit characteristics of a particular physics process, and are therefore interpreted as trigger rules.

Events which satisfy the HLT selections are transferred to the CERN Tier-0 for permanent storage and reconstruction. A GRID computing infrastructure provides distributed data storage and event processing at dedicated computed sites [65].

3.3 Summary

The individual subdetectors of the CMS detector, and the means by which the protons produced at the LHC are brought to collision has been described in Section 3. The specific means by which particles and physics objects are formed at CMS is described in Section 4.

Chapter 4

Reconstruction and Objects

The data recorded at CMS requires a comprehensive description of the underlying aspects of the collision. Therefore, the precise reconstruction and identification of particles produced in the large multiplicity environment of highly energetic *pp*-collisions is necessary.

At CMS, particles produced from *pp*-collisions interact with the subdetectors and electrical signals are read out, as described in Section 3. To begin event reconstruction, the readouts from the various subsystems are combined and the energy deposits associated to each subsystem are measured. The readouts are then attributed to particles or are translated into physics objects, the precise definitions of which are provided by the Physics Object Groups (POGs) of the CMS collaboration.

Thereafter, selections are applied on any reconstructed object to reduce the probability of fake objects whilst aiming to maintain a high reconstruction efficiency of the object in question. Further, the control over the different levels of efficiency in fake object rejection and signal object reconstruction is provided by Working Points (WPs) For this purpose, different levels of WPs are defined, typically referred to as *loose*, *medium* or *tight*, and each of which offers differing levels of sensitivity to background rejection and signal efficiency.

4.1 Particle-Flow algorithm

Following the reconstruction of detector signals, the identification of individual particles and jets, which collectively are referred to as physics objects, is performed by the Particle-Flow (PF) algorithm [66]. The PF algorithm combines information from all the subdetectors, specifically the reconstructed tracks and calorimeter energy deposits, to provide particle-based event reconstruction. The process of utilising the information from all subdetectors significantly improves the measurements of particle properties, energies and trajectories than could be achieved by any one subdetector.

Of the hundreds of different particle species produced at the LHC, only a small fraction have sufficient lifetimes and interaction strengths to be directly measured by the detector subsystems. Amongst the stable species, the predominant in measurements of the CMS detector are: γ, $e^{\pm}$, $\mu^{\pm}$, $\pi^{\pm}$, $K^{\pm}$, $p^{\pm}$, K^{0} and n. These species are classified by the PF algorithm into five exclusive categories: photons, electrons, muons, neutral hadrons and charged hadrons [66].

The behaviour of the particle classification relies on subdetectors installed at CMS, including silicon tracker, the reconstructed tracks and calorimeter clusters[1]. These detector layers are referred to as elements and are passed to the PF algorithm. Any given particle is, in general, expected to return several PF elements in the various CMS subdetectors. Therefore, to fully identify a single particle requires a link algorithm to connect these elements. The link algorithm iteratively checks each pair of elements per event to produce blocks, where typically one, two or three elements arise from the same object. The reconstruction and identification of a set of particles from each block of elements is finally performed, the aim of which is to return both a set of individual particles, known as *PF-candidates*, and a global description of each event. With the blocks defined, the five classifications of the particle species can be identified, the details of which are discussed below:

Muon reconstruction

As a Minimally Ionising Particle (MIP), muons typically leave minimal energy deposits in the calorimeter systems and traverse the entire detector. The reconstruction of muons is performed using a combination of trajectories from the tracking detector and muon systems. At CMS, two algorithms are used to give complementary efficiency across the muon momentum spectrum, the global muon algorithm and the tracker muon algorithm [67].

The global muon algorithm matches hits in the muon chambers to tracks in the tracking system. Thereafter, a global muon track is fitted combining hits from the tracker track and standalone muon track using the Kalman filter technique [68]. The

---

[1] The calorimeter clusters are constructed from a calorimeter-cell seed in each calorimeter. The calorimeter-cell seed are chosen as cells with an energy deposit above a predefined threshold. Energy deposits above a certain requirement are added to the cluster until no more such cells are found.

tracker muon algorithm selects all tracks satisfying $p_{\mathrm{T}} > 0.5$ GeV and $p > 2.5$ GeV, where p is the total momentum. These are the extrapolated to the muon system, using the Kalman filter to account for energy loss in the calorimeters. Performing an extrapolation of tracks into the muon system, the tracker muon algorithm provides a higher reconstruction efficiency for low-momentum muons. Global muon reconstruction is more efficient for higher energy muons, which are likely to traverse the several muon stations [67].

The PF algorithm for muons is based on the global muon reconstruction [66, 67]. Muons are obtained by using reconstructed global muons, if the track momentum measurement of the global muon is in agreement with that of the combined muon track and within three standard deviations of the momentum resolution. Tracks associated with muons passing this selection are removed from the block and the expected energy deposition in the calorimeters along the muon trajectory are removed.

The PF muon identification is designed to identify prompt muons, produced by a hard process such as the decays of W and Z bosons or quarkonia states. Furthermore, the reconstruction of muons produced from the in-flight decays of hadrons or heavy quarks is also targeted by the PF muon identification.

Muons are required to be reconstructed by either of the PF, global or tracker muons algorithms. The latter two algorithms encode selections on various muon identification variables to target prompt muons and non-prompt muons, respectively. The *soft* muon selection requires the candidate to be a tracker muon, while the *tight* muon selection requires a global muon candidate [67].

Electron reconstruction

Electrons are reconstructed by matching energy deposits in the tracker and ECAL using two techniques, an ECAL driven reconstruction and a tracker driven reconstruction performed with the PF algorithm. The former technique relies on the clustering of electrons and their associated bremsstrahlung with rectangular ECAL windows, called superclusters. The reconstruction of an electron is determined from matching superclusters to tracks originating from the primary vertex, and is described in more detail elsewhere [69].

With the PF algorithm, electrons are obtained by identifying a combination of a number of tracking and calorimetric variables. As electrons produce a track in the inner tracking system and deposit energy in the ECAL, electron reconstruction is therefore seeded by a track (for low p_{T} electrons) or an ECAL supercluster (for isolated high p_{T} electrons).

Tracks consistent with supercluster hits are classified as electrons and the corresponding tracks and ECAL energy deposits are removed from the block.

Charged hadrons, photons and neutral hadrons

Charged hadrons are identified by using tracks that are compared to the surrounding subdetectors. If the measured track energy in a block is compatible with that of a linked calorimeter deposit, the tracks in the block as established as charged hadrons. The momentum and energy are calculated from the track momentum under a charged pion hypothesis.

In the situation where the measured calorimeter energy is larger than the measured track energy, the remaining ECAL energy is clustered to return photons. The remaining HCAL energy is then clustered as neutral hadrons. The PF algorithm is able to also resolve the particles arising from the decay of τ leptons. In considering the hadronic decay modes of the τ lepton, a PF jet is input to the hadrons-plus-strips (HPS) algorithm to identify PF hadronic τ leptons [70]. The full details of the reconstruction of the τ lepton physics object is detailed elsewhere [71].

4.1.1 Isolation

The level of hadronic activity around an identified lepton provides a gauge on the origin of the lepton. Prompt leptons, leptons produced in the primary *pp*-collisions, and non prompt leptons, produced in secondary decays of unstable particles, can be distinguished by measuring the level of isolation. The isolation is defined as the fraction of energy in a cone around the lepton, relative to the energy carried by the lepton itself.

The standard method of computing isolation used by CMS is PF relative isolation, I_{PF}^{rel}, defined within a cone of radius ΔR as:

$$I_{PF}^{rel} = \frac{1}{p_T^l}[\sum_{PF_{PV}} p_T^{CH} + \sum_{PF} p_T^{NH} + \sum_{PF} p_T^{\gamma} - \sum_{PU} p_T^{Neutral}] \tag{4.1}$$

where ${p_{\mathrm{T}}}^{l}$, ${p_{\mathrm{T}}}^{CH}$, ${p_{\mathrm{T}}}^{NH}$, ${p_{\mathrm{T}}}^{\gamma}$ are the momenta of the lepton, charged hadrons, neutral hadrons and photon, respectively. To reject particles from pileup, only charged hadrons that are associated with the primary vertex are utilised in the computation. A removal of the neutral contribution of pileup is performed, as neutral pileup cannot be associated with the primary vertex. The $\Delta\beta$ corrected and effective area (EA) correction are two methods used by CMS to estimate and correct for the contribution of neutral pileup.

To account for decays of massive particles, whereby jets produced in association with the lepton could be boosted in a collinear configuration, an alternative method for determining the isolation is used. The PF mini-isolation computation utilises a dynamic isolation cone size to account for the increasing hadronic activity within the isolation cone, and recover the efficiency for lepton selection in boosted decays. For a lepton with momentum $p_T^l < 50$ GeV, the radius of the cone is $\Delta R = 0.2$, and decreases with increasing lepton momentum to a minimum cone radius of $\Delta R = 0.05$ for a lepton with momentum $p_T^l > 200$ GeV.

4.1.2 Isolated tracks

Unreconstructed prompt leptons a large background for searches with hadronic final states. Such leptons, labelled as “lost leptons”, can be caused by both acceptance effects and the misreconstruction of processes, including e^-, μ^- decays and in hadronic decays of τ leptons. Due to the expected isolation of prompt leptons, events containing lost-lepton decays that occur in the tracker acceptance can be rejected by vetoing high energy isolated tracks. Isolated tracks are selected from isolated charged PF candidates that are associated with the primary vertex and which pass track quality selection.

4.2 Jets

As described by the SM, a quark or gluon cannot exist as a free particle and is confined by its colour. This behaviour forces the quark or gluon to hadronise and shower, the process of which produces a spray of hadrons following the direction of the initial quark or gluon, and is hence represented as a “jet”. Unlike electrons or muons, quarks and gluons can not be associated to one of the PF candidates, and thus require an alternative method of reconstruction.

The construction of a jet is performed by various algorithms, developed in order to cluster the spray of hadrons and assign them kinematic properties. To achieve this effectively, clustering algorithms are required to be insensitive to the addition of low p_{T} (soft) particles, *infrared safe*, and to collinear splitting of particles, *collinear safe.*

Recombination algorithms

The anti-k$_t$ clustering algorithm performs a sequential clustering of input objects, known as *protojets*, into jets [72]. The algorithm iterates through pairs of protojets and for each combination, two distance variables are constructed.

The first distance variable is a measure of the $(\eta - \phi)$ distance between pairs of pseudojets, d_{ij}, while the second variable measures the momentum space distance between a pseudojet and the beam, d_{iB}, both of which are defined as:

$$d_{ij} = min(k_{Ti}^{2p}, k_{Tj}^{2p}) \frac{\Delta R_{ij}^2}{R^2} \tag{4.2}$$

$$d_{iB} = k_{Ti}^{2p} \tag{4.3}$$

where i, j and k$_{Ti}$, k$_{Tj}$, are the respective indices and momenta of the protojets. The term ΔR_{ij}^2 is the separation between the pseudojets, i and j, and the term R is the radius parameter of the jet, which in turn determines the final size of the jet. The radius parameter is chosen as 0.4 and 0.8 in the PF event reconstruction at CMS, yielding two collections of jets. For each iteration of the algorithm, the minimum of the entire set of $\{d_{ij}, d_{iB}\}$ is found. In the situation where the minimum is d_{ij}, the four-momenta of the pseudojets i and j are combined to form a new pseudojet and the original pseudojets are removed. However, if the minimum is d_{iB}, the pseudojet i is defined as a jet and is then removed from the list of pseudojets.

The behaviour of the clustering algorithm, specifically the relative power on the energy versus geometrical scales, is dependent on the parameter p of Equation 4.2 and Equation 4.3. The value of p corresponding to the Cambridge-Aachen (CA) algorithm is 0 [73], which modifies Equations 4.2 and 4.3 to remove any dependency on the momentum. Due to the purely spatial character of the distance variables, the CA algorithm is often used in jet substructure studies. In particular, the sequential recombination in increasing angular distance is ideally suited to dealing with jets containing multiple or unknown angular scales.

For $p = 1$, the clustering algorithm favours clusterings that involve low p_{T} particles. This algorithm, referred to as the inclusive k$_t$ clustering algorithm, is often susceptible to the underlying event and PU, and can yield geometrically irregular jets [74]. Furthermore, a consequence of clustering soft particles first can lead to arbitrarily soft particles becoming jets in their own right.

Within CMS, jet clustering is predominantly performed with the anti-k$_t$ clustering algorithm, that utilises a negative exponent of p = -1. This enforces soft objects to be clustered with nearby hard objects. The resultant jets are, crucially, resilient to soft radiation and the jet area is unaffected, thus making the anti-k_t algorithm the preferential jet algorithm employed at CMS. The input objects to the anti-k_t algorithm are the particle candidates from the PF algorithm [66]. At CMS, as a means to construct a different type of jet, the choice of input object can be calorimeter energy deposits (CaloJets). However, as a much higher precision and energy resolution can be achieved with measurements from the tracking system than in the calorimeter, PF-reconstruction of jets provides a significant advantage over calorimeter reconstruction.

Jet energy scale corrections

Jet clustering algorithms, regardless of the distance parameter used, are developed such that the reconstructed jet can be compared to the generator level[2] parton energy. However, due to a contribution of additional interactions and the inefficiency in some regions of the detector, the raw energy measurements that form jets will not be registered and thus have to be corrected.

The corrections to the jet energy scale is performed using a factorised approach with dedicated jet energy scale (JES) calibrations [75]: At each stage, the jet four momenta is corrected and input to the next sequence of corrections. The applied sequential levels are:

L1 Pileup To remove the energy originating from PU events, simulated events are used to model the detector response. Thereafter, the effects from PU contributions can be mitigated when comparing reconstructed jets to generator level jets [76, 77]. The degree of PU energy contamination within the jet cone is correlated with the number of reconstructed vertices. The PU contributions are removed using charged hadron subtraction (CHS) where PF candidates not associated with the primary vertex are removed before clustering [76]. The CHS reduces PF-candidates associated with PU vertices prior to the clustering of jets, greatly reducing the energy contamination from charged PU vertices [78]. Thereafter, the neutral and a residual charge component of PU are corrected for using the jet area method [76]. The jet area method performs a jet-by-jet PU correction using the average PU energy density in the event, ρ, and the active

---

[2] In MC event generators, generator level refers to before the particle is reconstructed by a detector simulator.

area of each jet, A_j, by the relation:

$$p_T^{Corr} = p_T^{Raw} - \rho \cdot A_j \tag{4.4}$$

L2L3 Monte Carlo corrections Following the correction of PU effects, the transverse momentum of the L1 corrected jet is still different to that of the generator level jet. Therefore, the transverse momentum of the jet receives an additional correction as a function of p_{T} and η [75]. The correction to the jet p_{T} is given as:

$$p_{T,L2L3} = (\langle \frac{p_{T,L1}}{p_T^{gen}} \rangle [\eta, p_{T,L1}])^{-1} p_{T,L1} \tag{4.5}$$

where $\langle\rangle$ denotes the average in pre-defined categories of [η, $p_{T,L1}$].

L2L3 Residuals Final, additional corrections are applied to bring data and MC simulated events into agreement. The corrections are applied for jets observed in data only.

4.2.1 Identification of bottom quark jets

Jets originating from bottom quarks, labelled b-jets, can be identified through vertices that are displaced with respect to the primary interaction [79].

The bottom hadron, b-hadron, has a relatively large mass and lifetime with respect to gluons and light flavour quarks. This results in recognisable detector signatures that allows algorithms to exploit the unique characteristics of b-hadrons, and thus determine a probability that a jet originated from a bottom quark.

A unique effect of the large lifetime of the b-hadron is the ability to observe the charged particles track displacement from the primary vertex (PV). This displacement of a few millimetres, measured by displaced tracks, allows a secondary vertex to be reconstructed. Furthermore, the tracking system of CMS provides an impact parameter (IP), defined as the distance between the primary vertex and the track at their point of closest approach. An illustration of the procedure in identifying a b-jet is shown in Figure 4.1.

The Combined Secondary Vertex v2 (CSVv2) algorithm [11] is based on the CSV algorithm [80] and with a multivariate technique, combines the information of displaced tracks with the information of the reconstructed secondary vertices associated

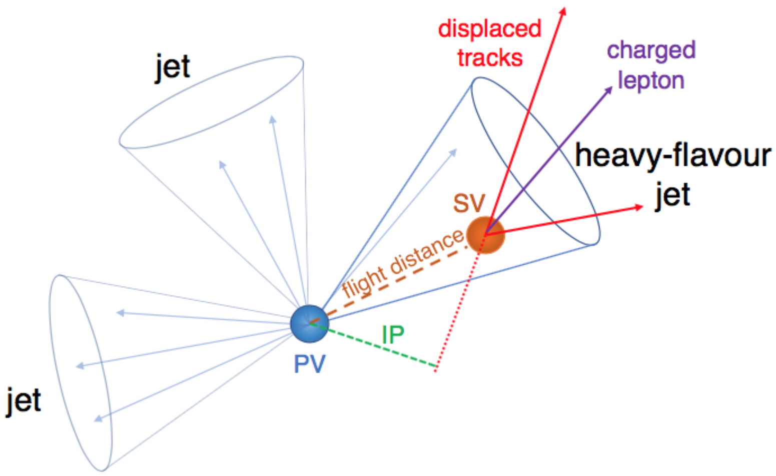


Fig. 4.1 Illustration of a b-jet with a secondary vertex (SV) from the decay of a b-hadron resulting in particle tracks that are displaced from the primary interaction vertex (PV) The impact parameter (IP) is shown as the green dotted line [10].

to the jet. The output of the CSVv2 is a discriminator value per jet, that can be used to discriminate between a jet originating from a bottom quark or from other sources. If the decay of a boosted particle contains a bottom quark, a phenomena expected in the decay of a boosted top quark, applying the CSVv2 algorithm on the constituents of the reconstructed jet of the top quark provides an additional discrimination power in identifying boosted top quarks. The constituents of the reconstructed jet that are input to the CSVv2 algorithm are hadronised partons, labelled subjets, and are described in more detail in Section 4.4.

The distribution of the resultant discriminator values of the CVSv2 algorithm on AK4 jets is shown in Figure 4.2. In this distribution, the mistag rates for b-jets and light-jets measured in $\sqrt{s} = 13$ TeV data and simulation are shown. Furthermore, the receiver operating characteristic (ROC) curves for b-jet identification algorithms is shown in Figure 4.3. From this distribution, the performance of the b-jet identification efficiency against the misidentification probability for jets in simulated $\mathrm{t\bar{t}}$ events is shown for a collection of b-jet identification algorithms. These are the Jet Probability (JP) algorithm, which computes the likelihood of the jet to originate from the PV using the associated tracks, and variations of the CSV algorithm [81].

Jets are hence identified as originating from bottom quarks by applying a criterion

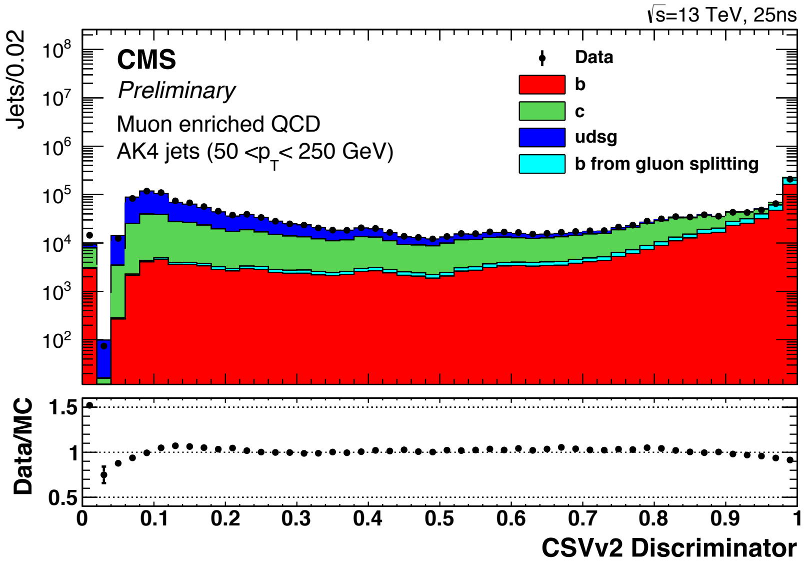


Fig. 4.2 Distribution of the CSVv2 Discriminator for jets defined using the anti-k$_T$ in a muon enriched jet sample. The markers correspond to data and the stacked histograms indicate the jet flavours. [11]

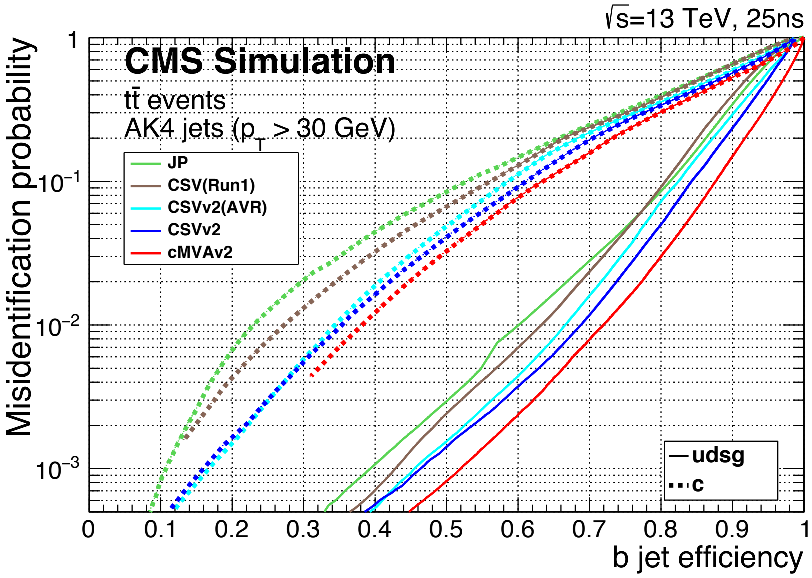


Fig. 4.3 Performance of b-jet identification efficiency algorithms against the probability for non b-jets to be misidentified as a b-jet [11].

on the discriminator values of the CSVv2 algorithm. A “medium” working point corresponds to a selection of > 0.80 on the CSVv2 algorithm discriminator variable. This results in a gluon/light-quark mistag rate of ∼1 % (where “light” means u, d and s quarks) and a tagging efficiency of $\sim$ 80 % [11].

4.3 Energy Sums

The total and missing energy in an event provides a useful gauge of the energy scale and can be used to discriminate between signal and background processes. Regarding BSM theories, in particular SUSY, where the decay of hypothesised massive particles yields a large energy transfer in both the production of visible and invisible particles, such variables allow the energy scale of the signal process to be probed.

Energy sums are computed from the scalar sum of particle energies. The total transverse energy E_{T} is defined as the scalar sum of all reconstructed PF-candidates. Furthermore, the total hadronic transverse energy, H_T, is computed from calibrated jets within a chosen acceptance and defined as $\sum_i^{\text{jets}} |\overrightarrow{p_{\mathrm{T}}}_i|$

$$\begin{aligned} \overrightarrow{\not{E}_T} &= -\sum_i \overrightarrow{p_{\mathrm{T}}}_i^{PF}, \quad & \not{E}_T &= \Big| -\sum_i \overrightarrow{p_{\mathrm{T}}}_i^{PF}| \\ \overrightarrow{\not{H}_T} &= -\sum_i^{\text{jets}} \overrightarrow{p_{\mathrm{T}}}_i, \quad & \not{H}_T &= \Big| -\sum_i^{\text{jets}} \overrightarrow{p_{\mathrm{T}}}_i\Big| \end{aligned} \tag{4.6}$$

A distinctive characteristic of neutrinos and hypothesised new particles of many BSM theories, including DM candidates, is their inability to be directly measured by the subdetectors of CMS. However, their presence can be inferred by the imbalance in transverse momentum. The missing transverse momentum vector $\overrightarrow{\not{E}_T}$, is defined as the negative vector sum of the momenta of PF-candidates in a given event, the magnitude of which is the labelled the total missing energy $\not{E}_T$.

Further, as a means to estimate the missing transverse energy, the hadronic missing energy, $\not{H}_T$, is computed from jets in a given event, while PF-$\overrightarrow{\not{E}_T}$ is constructed from PF-candidates and subject to type-I corrections [82]. The type-I correction is used to further improve the resolution of $\not{E}_T$ measurements by replacing the uncorrected transverse momentum of each PF jet with the corrected transverse momentum, as scaled by the jet energy corrections.

4.4 Jet Substructure

There are some cases whereby a jet clustering algorithm, with a distance parameter of $\Delta\mathrm{R} = 0.4$, as that employed at CMS, becomes inefficient in reconstruction. Many physics models beyond the SM predict final states containing highly energetic particles. The effect of such energy is typically translated into a Lorentz boost of the form $\gamma = \frac{E}{m}$, where E and m is the energy and mass of the particle respectively.

In the case of a top quark, whereby the decay mode is almost exclusively $t \rightarrow bW$, the separation between the W boson and b quark from the mother top quark, will be approximately $\Delta R = \frac{2m}{p_{\mathrm{T}}}$[3] [74]. If the top quark were to decay hadronically, a standard jet clustering algorithm would typically associate three separate jets. However, should the top quark receive a sufficient Lorentz boost, the resulting jets of the W boson and the b quark would be collimated to the point that standard reconstruction techniques begin to fail. Furthermore, when the separation between the decay products becomes smaller than the radius parameter of the jets, they often fail to be individually resolved by standard jet algorithms. A further impact on the collimated jets is due to the high-luminosity conditions at the LHC and the effect of PU interactions that can contaminate the jets.

Therefore, a jet clustering algorithm with a larger jet radius is important to allow the jet to capture enough of the hadronised particles and more accurately measure the jet mass and energy. However, as s larger jet radius is more susceptible to PU and the underling event (UE) dedicated techniques, described in more detail below, are required to estimate and remove their contribution. For such boosted topologies, jets at CMS are typically clustered using the anti-k$_t$ algorithm, with a larger opening angle of $\Delta\mathrm{R} = 0.8$, to construct another jet collection (AK8 jets) that aims to collect all the decay products into a single jet, labelled a *fatjet*.

The efficiency in using a jet clustering algorithm with a larger distance parameter to identify hadronically decaying boosted top quarks is illustrated in Figure 4.4. In the distribution, the efficiency is shown as a function of the top quark p_{T}, determined using simulated events.

In practice, identifying boosted objects demands techniques capable of distinguishing quarks from gluon jets and mitigating the effects of jet contamination within a fatjet. Different techniques are available to identify such fatjets, the core principle of

---

[3] For a quasi-collinear splitting into two objects i and j, the total mass $m^2 \simeq z(1-z)p_T^2\Delta R_{ij}^2$ where $z = p_{Ti}/p_T$ and $p_T = p_{Ti} + p_{Tj}$.

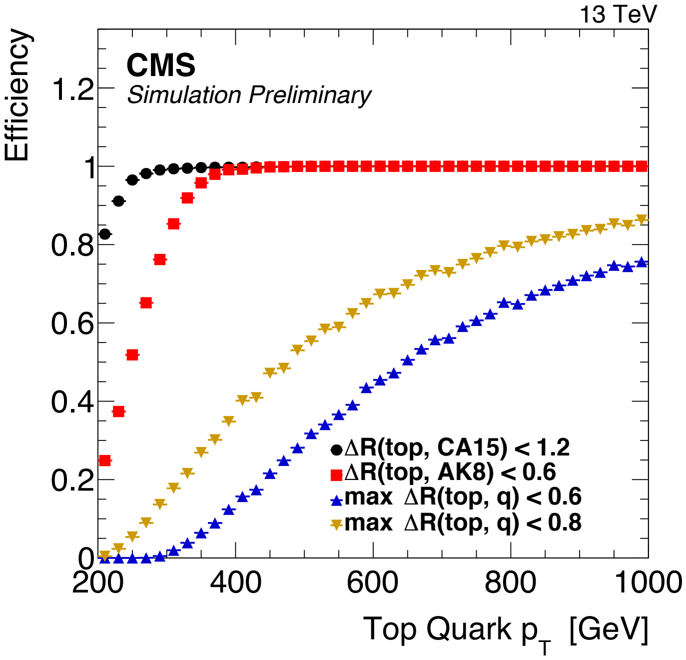


Fig. 4.4 Reconstruction efficiency, defined as the fraction of top quarks (with $p_{\mathrm{T}} > 200$ GeV) for which a reconstructed jet (with $p_{\mathrm{T}} > 200$ GeV) can be found within $\Delta R < 1.2$ ($\Delta R < 0.6$) for CA15 (AK8) jets (left)

each being the removal of soft and wide angle radiation produced by PU interactions, UE, or parton shower activity. A dedicated set of tools have been long developed and calibrated to attempt to distinguish the fatjets, formed from the decay of a highly boosted particle, from the large QCD multijet background.

Irrespective of a specific boosted particle reconstruction, eliminating uncorrelated UE or PU radiation from a fatjet is a crucial stage. Such methods, referred to as *grooming*, typically recluster fatjets following a dedicated sequential clustering algorithm, applying criterion at each stage of the reclustering to remove unwanted radiation. From which, highly energetic hadronic top quarks or W and Z bosons can be identified from fatjets. The ability to efficiently distinguish such boosted particles from QCD multijet processes could improve sensitivity to new physics.

Pruning and Softdrop

The recombination structure of the anti-k_t clustering algorithm defines jets that are assembled from a sequence of protojets, or subjets, is discussed in Section 4.2. Grooming tools typically aim to utilise the subjet structure within a fatjet to identify jets and thus events, likely to have originated from hadronic top quark or W boson decays. In practice, hadronic decays of very high p_{T} top quarks or electrowewak bosons could be collimated and collected under a single jet, whose invariant mass, following grooming, can provide a means to identify the origin of the jet.

Of the collection of grooming tools utilised, theoretically and empirically by CMS, the *pruning* [83] and *softdrop* [84] algorithms are predominantly used.

The pruning algorithm begins by defining variables that utilise the three-momenta and the masses of the fatjet and subjets. As jet clustering algorithms, including the anti-k_t algorithm, are formulated using variables that are invariant under longitudinal boosts and azimuthal rotations, the variables utilised in the pruning algorithm are also chosen to be invariant under the same transformations.

Under this assumption, the variables z and D, shown in Equation 4.7 and Equation 4.8, are constructed per fatjet, on which a criterion is applied to determine whether the fatjet represents the reconstruction of a particle of interest:

$$z \equiv \frac{min(p_{T1}, p_{T2})}{p_{Tp}} \tag{4.7}$$

$$D \equiv \frac{m_{jet}}{p_{Tp}} \tag{4.8}$$

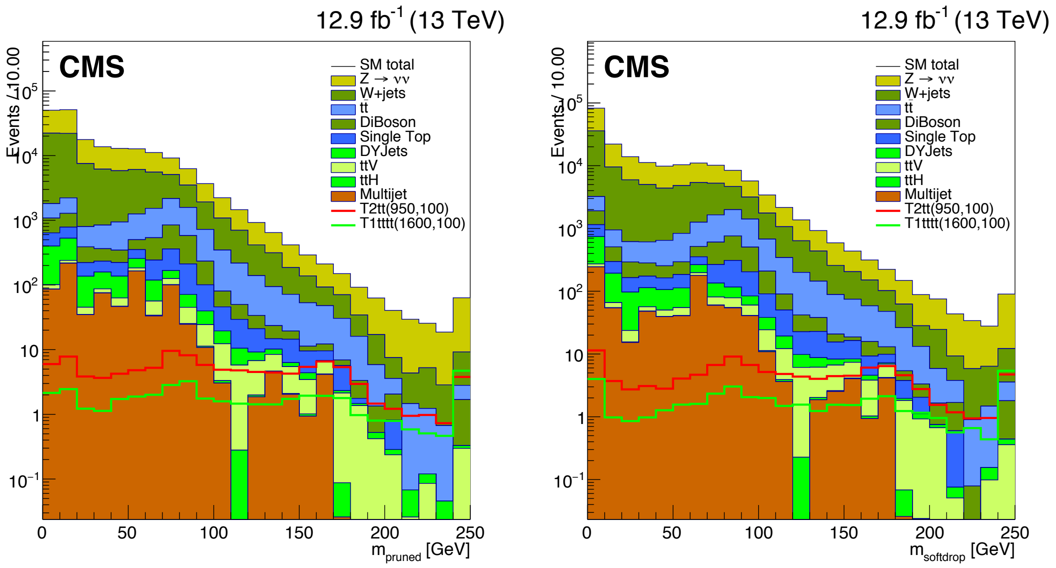
where p_{T1}, p_{T2} are the momenta of subjets of a fatjet, p_{Tp} is the momentum of the fatjet and m_{jet} is the mass of the fatjet.

The pruning algorithm begins by starting with the input objects to the AK8 jets, and reclustering them with the CA algorithm, requiring that each recombination satisfies the criterion on the aforementioned variables [83]:

$$\frac{min(p_{T1}, p_{T2})}{p_{Tp}} > z_{cut}, \quad \Delta R_{12} < D_{cut}\frac{m_{jet}}{p_{Tp}} \tag{4.9}$$

where ΔR_{12} is the angular separation between the two subjets. The two configurable parameters, z_{cut} and D_{cut}, are tuned to remove soft and wide angle particles respectively.

Following the full reclustering of a fatjet, with the application of the criteria of Equation 4.9, the resultant jet is referred to as the pruned jet from which additional variables can be constructed. The pruned mass, defined as the invariant mass of the subjets associated to a fatjet following the pruning algorithm, is used to discriminate between signal and background processes. Figure 4.5a shows the distribution of the pruned mass in a hadronic final state.



(a) Distribution of the pruned mass.

(b) Distribution of the softdrop mass.

Fig. 4.5 Distributions of the pruned and softdrop mass, for simulated SM events in a hadronic final state. The coloured histograms indicate simulated SM events and the solid lines represent two signal models, with final states containing top quarks.

In a similar method to the pruning algorithm, the softdrop declustering algorithm also aims to remove wide-angle soft radiation from a jet in order to mitigate any effects of contamination from initial state radiation (ISR), UE and PU.

The clustering of a jet with distance parameter R is reverted step by step, yielding at each iteration, two subjets j_1 and j_2 from the original jet, j. If the softdrop condition, shown in Equation 4.10, is satisfied then j is considered the final jet and the procedure ends. Otherwise, the subjet with the largest p_{T} is relabelled j and the procedure is repeated [84].

$$\frac{min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} > z_{cut}(\frac{\Delta R_{12}}{R})^{\beta} \tag{4.10}$$

where ΔR_{12} is the angular separation between the two subjets and z and β are tunable parameters.

The invariant mass of the subjets associated to a tagged jet, labelled the softdrop mass, can be used to identify heavy particles that decayed hadronically. This is illustrated in Figure 4.5b, where in the case of a hadronically decaying top quark, the softdrop algorithm successfully grooms the associated AK8 jet and assigns a softdrop mass criteria to identify top quarks. It should be noted that the pruning and softdrop algorithms can be used to identify either a boosted top quark or an electroweak boson. This can be achieved as the key principle of the groomers is to remove contamination from the UE and PU, and optimise the mass resolution of the jet before it can be subject to tagging criteria, as discussed in more detail in Section 4.4.1 and Section 4.4.2. This can be further illustrated in Figure 4.5a and Figure 4.5b, in which the distributions between the pruned and softdrop mass exhibit similar features.

A complete description of the different grooming techniques used at CMS can be found elsewhere [85, 86].

N-subjettiness

In addition to the algorithmic methods that are used to reduce the smearing effects of jet contamination from numerous sources, a study of the different energy flow in the decay pattern between signal and background jets can efficiently tag boosted objects.

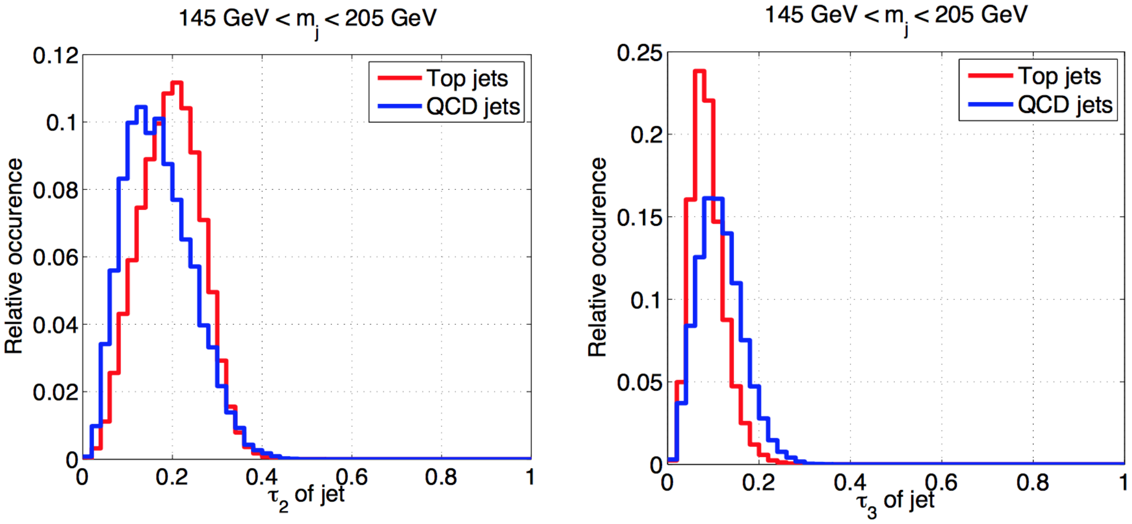
Such jet shape methods, used in conjunction with algorithmic techniques enhance the ability to identify heavy boosted objects in all-hadronic decay channels.

The *N-subjetittness* variable, τ_N, aims to exploit the energy distribution inside a jet by measuring the compatibility of a jet with the hypothesis that it is composed of i subjets [12]. It is defined as:

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} min\{\Delta R_{1,k}, \Delta R_{2,k}, ..., \Delta R_{N,k}\}. \tag{4.11}$$

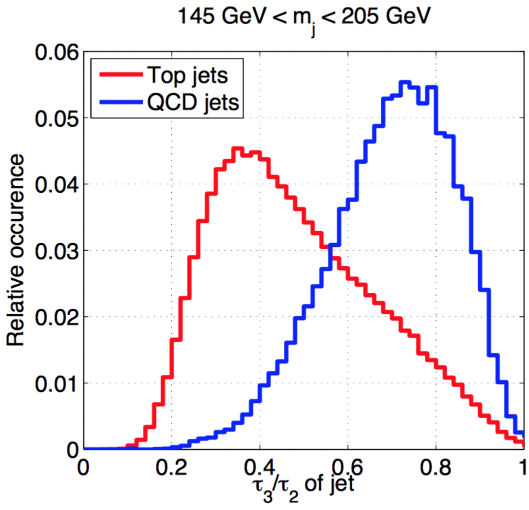
where k runs over the constituent particles in a given jet, $p_{T,k}$ are their transverse momenta, and $\Delta R_{J,k}$ is the distance in $(\eta - \phi)$ plane between a candidate subjet J and a constituent particle k. The normalisation factor d_0 is taken as $\sum_k p_{T,k} R_0$ where R_0 is the jet radius used in the original jet clustering algorithm. The subjets are reconstructed using the one-pass k_t axes approach [87].

Crucially, the τ_N observable has a small value if the jet is consistent with having N or fewer subjets, as almost every jet constituent will be close in ΔR to its own true subjet. In utilising the τ_N observable, a more discriminating method is to use the ratio τ_N/τ_{N-1}, illustrated in Figure 4.6, as a variable on which requirements can be made.



(a) Distribution of τ_2.

(b) Distribution of τ_3.



(c) Distributions of τ_3/τ_2 for boosted top and QCD jets.

Fig. 4.6 Distributions of τ_2, τ_3 and τ_3/τ_2 for boosted top and QCD jets. For the distributions shown, the jets are clustered with the anti-k$_t$ algorithm with a distance parameter of R = 0.8 and a kinematic requirement of $p_{\mathrm{T}} > 300$ GeV and $|\eta| < 1.3$. In addition, a requirement on the invariant mass of jets, m$_j$, is imposed such that 145 < m$_j$ < 205 GeV [12].

4.4.1 t-tagging

Hadronically decaying top candidates are identified with the softdrop algorithm that grooms the substructure of AK8 jets, according to the criterion of Equation 4.10.

The primary observable used to discriminate top quark jets in a background of light quark and gluon jets, is the large invariant mass of the jet itself. By reclustering the fatjet using the CA algorithm, the softdrop mass is calculated as the invariant mass of the sum of the constituent four vectors of the remaining subjets.

Furthermore, of the three expected subjets associated to a hadronically decaying top quark, one of which is expected to originate from a bottom quark. Therefore, to further distinguish a top quark tagged jet (t-tagged) from the fragmentation of light quarks or gluons, the softdrop subjets are subject to the CSVv2 algorithm [80].

In addition, for hadronically decaying top quarks, a topology with three subjets is expected and jets originating from light quarks or gluons will have fewer subjets. This, through the N-subjettiness variable, allows the ratio τ_3/τ_2 to favour three-prong boosted objects over two-prong boosted objects.

4.4.2 W-tagging

In a similar manner, boosted W bosons that decay hadronically will result in collimated decay products. On a collection of AK8 jets, W boson tagged jets (W-tagged) can be selected by reclustering the constituents with the CA algorithm, applying at each stage the pruning condition of Equation 4.9.

Furthermore, of the two expected subjets associated with a hadronically decaying W boson, the ratio τ_2/τ_1 provides further discrimination [88].

For both t-tagging and W-tagging, the combined use of the aforementioned discriminating variables provides a stronger efficiency to PU and a stronger background rejection efficiency, than if the variables were considered individually.

Following the processing of signals in the individual subdetectors, particles and physics objects are reconstructed through dedicated techniques, discussed in the aforementioned sections. Across the physics programme at CMS, a precise understanding of the particles and physics objects is necessary. This requires the reconstruction and identification of such objects to be obtained using specialist algorithms that suppress background contributions, mitigate PU and ensure high precision energy, position and temporal measurements. Thereafter, a more comprehensive understanding of the collision output is obtained and from which, certain SM processes can be identified and searches for new physics can be performed.

4.5 Summary

The methods and algorithms described in Section 4 are utilised in the reconstruction of particles and the construction of physics objects that are used across the CMS physics programme. The key components, by which the data collected at the CMS detector can be analysed, have been attained and the remaining aspects, required to confront data, are further discussed in Section 5.

CHAPTER 5

RECORDED DATA AND SIMULATED PROCESSES

A key concept in understanding the data collected at CMS is to use simulated events for both SM and signal processes. Such events are generated using software libraries that rely on Monte Carlo methods to perform random sampling and obtain numerical results, that can in turn be used to represent an event[1]. Collections of simulated physics processes, each containing events associated to a certain SM or signal process provides information to determine corrections and estimate systematic uncertainties, that can be used in an analysis.

5.1 Recorded data

During the course of 2016, the CMS collaboration collected *pp*-collision data at $\sqrt{s}$ = 13 TeV with a total recorded integrated luminosity of 35.9 fb^{-1} ± 2.5% [89]. A summary of the total integrated luminosity delivered by the LHC and recorded by CMS for *pp*-collisions in 2016 is shown in Figure 5.1. The data sample upon which the analysis is performed utilises the collection of certified data obtained by August 2016, which corresponds to an integrated luminosity of 12.9 fb^{-1}.

5.2 Simulated Standard Model processes

Simulated samples are necessary for background and signal prediction. In this description, backgrounds refer to the list of SM processes that resemble the signature of a signal event, either due to similar topologies or inefficiencies in object reconstruction.

---

[1] An event in this context is the result of an interaction following a *pp*-collision. An event contains detector information associated to the interaction and information on the kinematics of the particles, either reconstructed or generated.

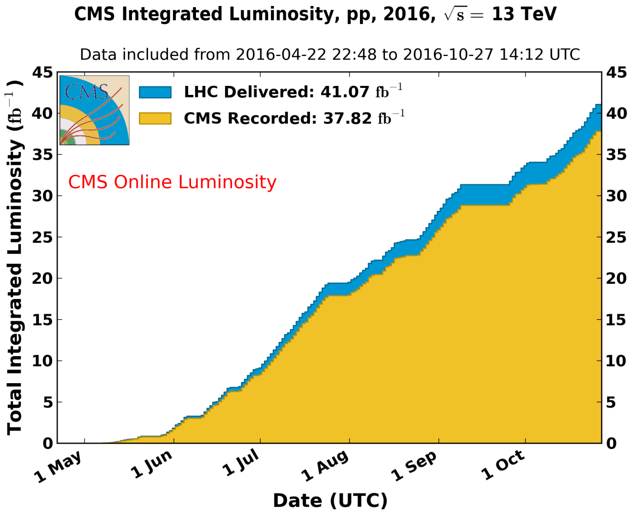


Fig. 5.1 The total integrated luminosity of data in pp-collisions at $\sqrt{s} = 13$ TeV during the 2016 campaign as a function of day delivered by the LHC (blue) and recorded by CMS (orange) during stable collisions [13] .

These physics processes can be simulated and then compared to recorded data, which allows a more comprehensive description of the fundamental aspects of the collision.

The simulation of CMS data attempts to model the hard scattering of the colliding partons. The fraction of the momentum carried by the parton, relative to the proton, is sampled from a parton distribution function (PDF). This matrix-element level hard-scatter is performed by MadGraph5 [90] with the hadronisation and showering of the hard scatter performed by Pythia8 [91]. A complete description of simulation techniques used at the LHC can be found elsewhere [92].

In the CMS collaboration, the detector simulator Geant [93] is used to input the generator level particles and produce digitised hits, analogous to those observed in the response of detector electronics. The individual components and their responses are reproduced to obtain a simulated event can be compared to one measured in data.

The details of simulated Monte Carlo (MC) samples, and their cross sections, used in this analysis during the run period of 2015 are listed in Table 5.1. These samples contain events that are simulated to represent SM processes that are associated as backgrounds in this analysis. Furthermore, the production of events simulating $t\bar{t}$, γ, W and Z boson production, each with associated jets is vital to the analysis, as these processes are expected to dominate the signal region.

Samples	Generator H_T^{parton} selection	Cross section (pb)
QCD	$100 < H_T^{parton} < 200$ GeV	2.785 x 10^7
	$200 < H_T^{parton} < 300$ GeV	1.717 x 10^6
	$300 < H_T^{parton} < 500$ GeV	3.514 x 10^5
	$500 < H_T^{parton} < 700$ GeV	3.163 x 10^4
	$700 < H_T^{parton} < 1000$ GeV	6.802 x 10^3
	$1000 < H_T^{parton} < 1500$ GeV	1.206 x 10^3
	$1500 < H_T^{parton} < 2000$ GeV	1.204 x 10^2
	$H_T^{parton} > 2000$ GeV	2.525 x 10^1
γ + jets	$100 < H_T^{parton} < 200$ GeV	9.226 x 10^3
	$200 < H_T^{parton} < 400$ GeV	2.300 x 10^3
	$400 < H_T^{parton} < 600$ GeV	2.774 x 10^2
	$H_T^{parton} > 600$ GeV	9.338 x 10^1
W(lν)+ jets	$100 < H_T^{parton} < 200$ GeV	1.627 x 10^3
	$200 < H_T^{parton} < 400$ GeV	4.352 x 10^2
	$400 < H_T^{parton} < 600$ GeV	5.918 x 10^1
	$600 < H_T^{parton} < 800$ GeV	1.458 x 10^1
	$800 < H_T^{parton} < 1200$ GeV	6.656 x 10^0
	$1200 < H_T^{parton} < 2500$ GeV	1.608 x 10^0
	$H_T^{parton} > 2500$ GeV	3.891 x 10^{-2}
Z($\nu\nu$)+ jets	$100 < H_T^{parton} < 200$ GeV	3.450 x 10^2
	$200 < H_T^{parton} < 400$ GeV	9.368 x 10^1
	$400 < H_T^{parton} < 600$ GeV	1.346 x 10^1
	$> H_T^{parton}$ 600 GeV	5.170 x 10^0
Z/γ^*(ll)+ jets	$100 < H_T^{parton} < 200$ GeV	1.715 x 10^2
	$200 < H_T^{parton} < 400$ GeV	5.258 x 10^1
	$400 < H_T^{parton} < 600$ GeV	6.761 x 10^0
	$H_T^{parton} > 600$	2.718 x 10^0
$t\bar{t}$	Inclusive	8.318 x 10^2
t	Inclusive	3.560 x 10^1
WW	Inclusive	1.246 x 10^1
WZ	Inclusive	6.610 x 10^1
ZZ	Inclusive	3.180 x 10^1

Table 5.1 Summary of the simulated background samples utilised in the analysis, with their generated H_T^{parton} requirement and respective cross section at $\sqrt{s} = 13$ TeV.

5.3 Simulated signal processes

The simulation of signal processes is performed at the parton level with MadGraph5 [90]. Thereafter, the simulation of hadronisation and showering is performed using Pythia8 [91]. The production cross sections relevant to SUSY signal models are calculated to NLO[2] and Next-to-leading logarithm precision [2]. The detector simulator is the FastSim detector simulation software [94]. In comparison to the FullSim detector simulator used for all the simulated SM processes, the FastSim detector simulator targets the processing time required to generate samples. In employing detector response parametrisations tuned using FullSim samples, the FastSim detector simulator avoids the CPU-intensive demands of the Geant framework. This enables the production of simulated events for a wider range of possible signal scenarios and provides a larger region of the parameter space that can be probed.

5.4 Corrections to simulations

Events from simulation are prone to mismodelling from numerous sources and thus require corrections to better describe the measurements observed in data.

Pileup reweighting

During *pp*-collisions, in an environment where the bunch crossing interval is short and the instantaneous luminosity is high, several pileup interactions are anticipated and the effect of which needs to be accounted for in simulation. Pileup interactions are simulated by generating low p_{T} *pp*-interactions and processing the associated detector signals with Geant [93]. It should be noted that number of pileup interactions is chosen such that it closely resembles the distribution of the expected number of additional low p_{T} *pp*-collisions. Such an elementary profile is constructed in simulation due to the fact that the number of actual pileup interactions is dependent on the instantaneous luminosity of each bunch crossing[3]. Upon the recording of data, the instantaneous luminosity is measured and multiplied by the total inelastic cross section of *pp*-collisions at $\sqrt{s}$ = 13 TeV to form a distribution of the number of pileup interactions. The

---

[2] The hard interaction of a process can be simulated at different theoretical orders, such as Leading order (LO) 2→1 or 2→2 processes. Some processes can be simulated to higher orders or Next-to-leading order (NLO)

[3] The bunch crossing refers to the spacing between the bunches of protons injected at the LHC.

total inelastic cross section is measured to be 69.2 mb ± 4.6% [95]. Thereafter, the pileup profile formed in simulation is reweighted to match that observed in data. A distribution of the number of pileup interactions determined in data, simulated events and reweighted simulated events is given in Figure 5.2.

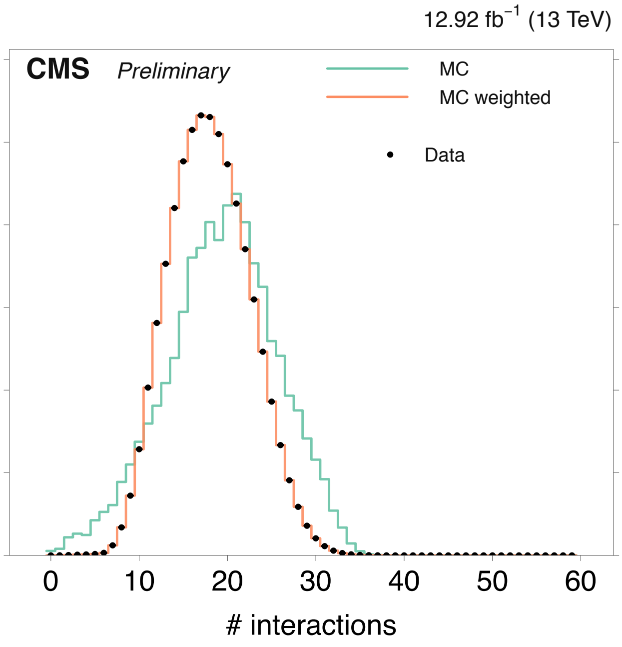


Fig. 5.2 The distribution of the average numbers of the inelastic interactions per colliding bunch pair per lumi-section in the data, corresponding distribution in the simulated events, and that of the reweighted simulated events [14].

The effects of reweighting simulated events according to their pileup profile shows a satisfactory agreement to that of data. In particular, the distribution of the number of reconstructed vertices in simulation, prior to reweighting, peaks at a higher pileup multiplicity than that observed in data which, after reweighting, is then brought into agreement.

Lepton and photon scale factors

When using physics objects, the modelling of the efficiency in which objects are selected is often different between simulation and data. Furthermore, certain parameters which are input to simulating physics processes are not precisely known nor calculable. Therefore, to account for such differences in efficiency, a scale factor correction is determined. The efficiency in simulation, $\varepsilon_{simulation}$, is compared to that in determined in data, ε_{data}, with the scale factor (SF) constructed as the ratio, $\varepsilon_{simulation}/\varepsilon_{data}$.

An example of which is in the reconstruction of muons, whereby the ability to accurately reconstruct a muon in simulation is not guaranteed to mirror that in data, due to an inefficiency in the muon identification, isolation or trigger selection.

Scale factors are typically determined as a function of p_T and η of the object in question, and using dedicated control samples. For muon reconstruction efficiency, muon scale factors are measured using the "tag-and-probe" method in a control region in which a leptonically decaying Z boson is reconstructed [67]. Thereafter, events in simulation are reweighted to improve the reconstruction efficiency to that observed in data.

B tagging scale factor correction

The scale factors for b-tagging are computed using the ratio of the efficiency in data to the efficiency in simulation, for identifying a jet as originating from a bottom quark. The resultant scale factors are dependent on the simulated jet flavour, jet p_T and jet η. The simulated samples are then corrected by determining an event weight, computed by calculating the probabilities of a particular configuration of b-tagged jets in data and simulation.

For each event, the probability in simulation and data for a particular configuration of b-tagged jets is given as:

$$P(\text{simulation}) = \sum_{\text{i=tagged}} \varepsilon_i^b \sum_{\text{j=not tagged}} (1 - \varepsilon_j^b) \tag{5.1}$$

$$P(\text{data}) = \sum_{\text{i=tagged}} SF_i \varepsilon_i^b \sum_{\text{j=not tagged}} (1 - SF_j \varepsilon_j^b) \tag{5.2}$$

where SF is the scale factor. With the aforementioned probabilities, the b-tag scale factor weight for each event is determined by the ratio:

$$w = \frac{P(\text{data})}{P(\text{simulation})} \tag{5.3}$$

In the use of jet substructure techniques, the method for determining the event weight from the scale factors in tagging jets as top quarks or W bosons is computed in the same manner. The details of how the scale factors are computed and how their associated uncertainties impact the analysis is described in more detail in Section 7.4.1.

Trigger efficiencies

Inefficiencies in the trigger requirements when obtaining events in data requires the simulated event yields to be corrected. Simulated samples in a region are subject to the trigger criterion associated to that region, with the difference between the measured efficiency in data and simulation corrected by a scale factor.

5.4.1 Correcting SM sample cross sections

H_T^{parton} binned samples

A fundamental requirement on the simulated samples is for it to contain a sufficient number of events to reduce any potential statistical error and provide an accurate representation of data. This can be achieved by simulating events within a restricted kinematic range, before utilising the full phase space of the given range. In practise, dedicated simulated samples that are categorised according to parton-level H_T, H_T^{parton}, defined as the scalar sum of the transverse energies of the quarks and gluons involved in the hard scatter process, are produced. The categorised samples correspond to the processes W+jets, Z($\nu\nu$), QCD, $t\bar{t}$ + jets, γ + jets and DY + jets. Each sample is produced using cross-sections calculated at leading-order (LO), with some being additionally translated into next-to-next-to-leading-order (NNLO) via scaling factors (k-factors) derived from their corresponding inclusive sample cross-sections.

The H_T^{parton} categorised samples are ensured to be consistent with both their inclusive counterpart and their neighbouring samples, by applying certain criteria. For every final state quark or gluon in the hard-scatter, the H_T of the event is determined. After which, the cross-section of the sample and the number of generated events are used to compute a weight per process and sample. Thereafter, the collection of samples are stitched together for a given process, with the predetermined weight applied. The modelling of the cross-sections, per H_T^{parton} range, is determined by measuring the rate of change across the stitched samples.

A particular amendment to the above mentioned method is applied for $t\bar{t}$ samples. The value of H_T^{parton} that is close to the true generator level value, is obtained by removing the contributions from the decay product of the top quark. The refined value of H_T^{parton}, labelled $H_T^{parton}(\text{no } t)$, is the scalar sum of the transverse energies of the gluons and quarks in the hard scatter which itself or none of whose ancestor is a top or anti-top quark.

Chapter 6

Analysis strategy

This chapter and the next describe a search for events with jets and missing transverse momentum, with a particular emphasis on the potential of incorporating jet substructure techniques to enhance sensitivity to signal processes. The methodology of the analysis is constructed to target the signature of BSM physics, through hadronically decaying SM particles and the trace of a BSM candidate.

Motivated by the complicated topologies expected from BSM particle decays, in particular the phenomena of boosted hadronically decaying particles yielding collimated decay products, the use of jet substructure tools are explored as a means to provide additional sensitivity by identifying events that contain such processes.

6.1 Challenging topologies for a BSM search with jets

In the context of SUSY, a number of free parameters can manifest itself into a variety of complicated decay topologies. A particular challenge in searching for signatures of such a BSM theory, is the ability to restrict and understand the backgrounds originating from SM processes.

In previous campaigns of the analysis, the rejection of the main background, QCD multijet production, is driven by targeted variables, while any remaining electroweak (EWK) backgrounds with genuine $\not{E}_T$, and any residual QCD, are predicted through dedicated techniques. However, the increase in centre-of-mass energy to $\sqrt{s} = 13$ TeV in Run II provides a significant enhancement in sensitivity to the production of BSM particles. Therefore, a good understanding of the backgrounds and their characteristics are critical in estimating and suppressing their effect on the performance of the analysis.

6.1.1 The need for substructure

As the masses of many hypothesised particles of BSM theories are unknown, any interpretation of a BSM model requires several potential masses to be considered. In the context of SUSY, the masses of the sparticles and LSPs have a direct connection to the topology of the final state.

An example of which is when the mass difference between the parent particle and the LSP decreases, the amount of available kinetic energy for the decay products reduces and typically results in decay products with low p_{T}. The correlation between the mass splitting, between the parent sparticle and the LSP, and the p_{T} of the decay products also has an opposite effect whereby the decay products have increasingly large p_{T} when the mass difference increases.

The effect of such a p_{T} on a hadronically decaying particle, will result in decay products that collimate under a single jet, and thus demand jet substructure techniques to probe their interior structure. As previously described in Section 4.4, this phenomena can be related to a hadronically decaying top quark or W boson, whereby large mass splittings between sparticles provide the boost required to force collimation.

Thereby, the ability to reconstruct hadronically decaying boosted particles, has the potential to provide a strong discrimination power with final states containing such particles and SM backgrounds.

In the following chapters, an overview of the α_{T} analysis is presented, with an additional focus on the extension to incorporate jet substructure techniques.

6.2 The α_{T} analysis

The α_{T} analysis is a hadronic SUSY search which utilises the dimensionless kinematic variable α_{T}, to effectively suppress the QCD multijet background. In addition, the α_{T} variable maintains a strong acceptance for a range of signal models. During Run I of the LHC, the α_{T} analysis was employed to search for SUSY, on datasets of 4.98 fb^{-1}at $\sqrt{s} = 7$ TeV and 18.5 fb^{-1}at $\sqrt{s} = 8$ TeV [96–100].

A key feature of the analysis is the effective suppression of the QCD multijet background, whose large cross section dominates the SM processes in *pp*-collisions at the LHC. To effectively reduce the contribution of events originating from QCD multijet processes, the analysis utilises three sensitive variables: α_{T}, $\Delta\phi^{*}_{min}$ and $\not{H}_T/\not{E}_T$, which will be discussed in the following sections.

Furthermore, due to detector and reconstruction effects, mismeasurements of jets can allow the QCD multijet background to acquire a false source of missing energy (fake-$\not{E}_T$) and thus satisfy requirements on the aforementioned variables, sensitive to QCD multijet processes. The sources of this fake-$\not{E}_T$ can be classified into the following categories: *detector inefficiency*, *detector effects*, *misreconstruction* and *acceptance* effects:

- Detector inefficiencies can cause a significant proportion of the energy of any physics object to be lost, due to regions that are uninstrumented or with a reduced response. Such under-measurements can lead to an artificial momentum imbalance whereby the $\not{E}_T$ vector is approximately in the azimuthal direction of the problematic region.

- Fake additional energy can added to the event due to effects such as “hot cells” that constantly record energy depositions in the ECAL and HCAL. Furthermore, spontaneous discharges of the calorimeter photodetectors can additional result in an artificial energy imbalance.

- Secondary particles that are produced outside the detector from beam interactions, known as “beam halo”, can interact with muon chambers, causing fake muons to be reconstructed, or may deposit energy in the calorimeters as they traverse the detector [101]. Muons originating from the atmosphere, known as “cosmic muons”, are identified when their trajectories are more than 1 cm away from the beam axis. Furthermore, in the reconstruction algorithm, described in Section 4.1, such muons are removed from the list of reconstructed particles if the measured $\not{H}_T$ is consequently reduced.

- Misreconstruction effects, such as the under or over correcting of jet energies and tracking failures can introduce missing energy. This can apply to one or more physics object in the event.

- Physics objects that fall outside of momentum and pseudorapidity acceptances are not included in energy sums, occasionally resulting in an artificial energy imbalance. Such momentum and pseudorapidity thresholds are typically required because of imperfect detector coverage.

It is also possible for QCD multijet processes to posses a genuine $\not{E}_T$ signature, through the leptonic decay of heavy flavour quarks. In such cases, the lepton fails the isolation requirements that are applied to ensure the hadronic search channel of the analysis.

Following the reduction of the QCD multijet background to a sub-dominant level, with respect to the other backgrounds, the analysis utilises four control regions to estimate the remaining backgrounds. A hadronic control region is formed to estimate the residual QCD multijet background, and a photon control region (γ+jets), and two muon control regions (μ+jets and $\mu\mu$+jets), are used to predict alternative SM backgrounds that contain EWK processes. This background is composed of several different processes, the dominant being top quark pair production and vector boson production in association with jets. The ability to estimate the contributions of these processes is important as each process is capable to produce jets with a genuine $\not{E}_T$ source from neutrinos in leptonic decays.

6.2.1 The α_{T} variable

The dimensionless variable α_{T}, was originally proposed as the α variable utilising the full jet three-momentum [102], before the transverse variant was introduced. The motivation behind the variable aims to exploit the balanced configuration of a jet pair, with no significant missing momentum $\not{E}_T$ or $\not{H}_T$. Further, the α_{T} variable is intrinsically robust against jet energy mismeasurements in multijet systems.

For a dijet system, α_{T} is defined as:

$$\alpha_{\mathrm{T}} = \frac{E_T^{j_2}}{M_T} \tag{6.1}$$

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

For a perfectly formed dijet event, with no associated $\not{E}_T$, the jets are produced with balanced transverse momenta in a back-to-back configuration and the value of α_{T} associated to this scenario is 0.5.

Furthermore, if the jets remain back-to-back but at least one jet is mismeasured, the corresponding value of α_{T} is less than 0.5. However, if the two jets are produced with genuine $\not{E}_T$ and as such recoil against their source, they will not be back-to-back, and more crucially, α_{T} can obtain values above 0.5.

The α_{T} variable can be generalised to any higher jet multiplicity by reducing the system into two *pseudo-jets*. Every possible combination of pseudo-jets is considered, with the system choosing the one that minimises ΔE_T, defined as 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.

The resultant, more generalised, form of α_{T} is:

$$\alpha_T = \frac{H_T - \Delta H_T}{2\sqrt{H_T^2 - \not{H}_T^2}}. \tag{6.3}$$

The behaviour of α_{T} measured in data for backgrounds with genuine and fake missing energy is illustrated in Figure 6.1. The distribution of the QCD multijet background can be seen to strongly peak for values of α_{T} below 0.5 before sharply falling with increasing α_{T}. The origin of the tail of the QCD distribution, that appears above $\alpha_{\mathrm{T}} = 0.55$, is considered residual contamination from events with fake-$\not{E}_T$. Conversely, the EWK backgrounds, possessing a source of genuine $\not{E}_T$, demonstrate a broader distribution of α_{T}, such that for regions with $\alpha_{\mathrm{T}} > 0.55$, the dominant background becomes events originating from EWK processes.

It should be noted that in rare cases of significant jet energy mismeasurements, QCD multijet processes can be reconstructed with larger values of α_{T}, resulting in residual QCD multijet contamination that requires further consideration.

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

As a further means to suppress the QCD multijet background, after the α_{T} selection criteria, the topological $\Delta\phi^*_{min}$ variable aims to reduce residual contamination from QCD multijet processes through heavy flavour decays and fake-$\not{E}_T$ arising from jet mismeasurements.

The variable $\Delta\phi^*_{min}$ is computed as the minimum azimuthal separation between the transverse momentum vector of each jet, $\vec{p}_T^{\,j_i}$, and the corresponding $\overrightarrow{\not{H}_T}$ when the jet is removed, $\vec{\not{H}}_T^{j_i} = \vec{\not{H}}_T + \vec{p}_T^{\,j_i}$. This leads to the definition in Equation 6.4, with the

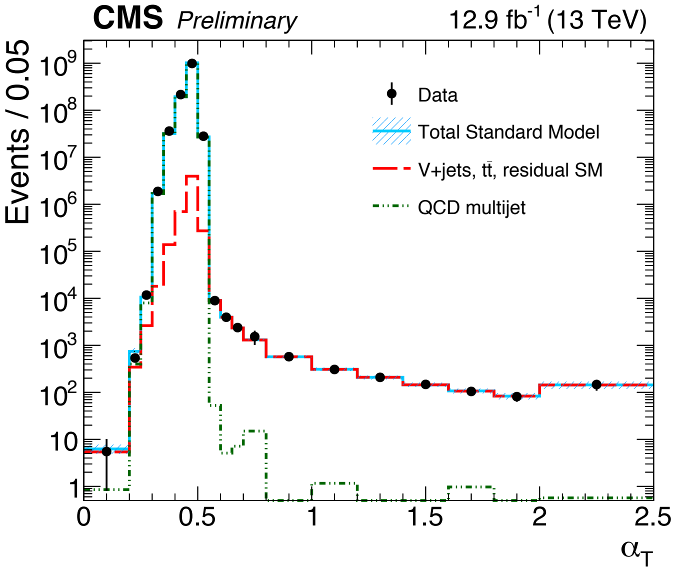


Fig. 6.1 Distribution of $\alpha_{\rm T}$ for data and simulated events at $\sqrt{s} = 13$ TeV. The electroweak (red solid line) and QCD (green dashed line) are determined from simulation. The data and simulated events are subject to a criteria $H_T > 300$ GeV and $p_T^{j2} > 100$ GeV, where $j2$ is the sub-leading jet, for $\alpha_{\rm T} < 0.55$ and the full signal region selection, discussed in Section 6.7, for $\alpha_{\rm T} > 0.55$ [14].

schematic shown in Figure 6.2.

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

Alternatively, the minimum azimuthal separation between the jet vector and $\overrightarrow{\not{H}_T}$, $\Delta\phi$, are commonly used in suppressing sources of mismeasurements. However, due to their construction, they typically include the highest energy jet in the computation and are therefore susceptible to severe jet mismeasurement.

The improvement of $\Delta\phi^*_{min}$ with respect to the traditional $\Delta\phi$ variable, constructed using a jet $p_{\rm T}$ vector and $\overrightarrow{\not{H}_T}$, is the removal of the probe jet in the computation, that acts to improve robustness against such instances of mismeasurement. In the case where a jet is mismeasured due to detector inefficiencies, $\overrightarrow{\not{H}_T}$ is typically in the same direction as the mismeasured jet and as such returns small values of $\Delta\phi^*_{min}$, as illustrated in Figure 6.2. Furthermore, for QCD multijet events with genuine-$\not{E}_T$, where the neutrino is emitted collinearly to the jet direction, $\overrightarrow{\not{H}_T}$ is typically aligned in the same direction as the jet, thus also yielding small values of $\Delta\phi^*_{min}$.

The behaviour of $\Delta\phi^*_{min}$ is illustrated in Figure 6.3 for both data and simulation. In

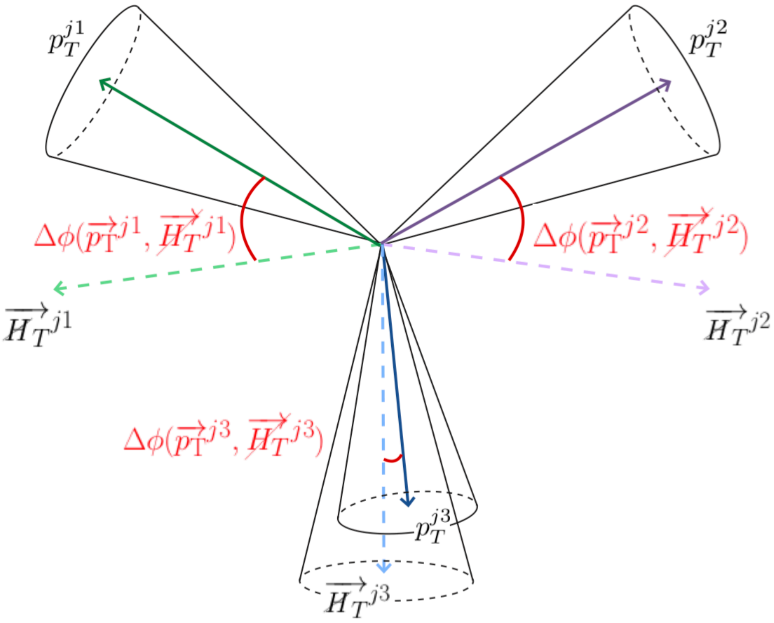


Fig. 6.2 Figurative representation of the $\Delta\phi^*_{min}$ variable in a mismeasured three-jet event configuration. The momenta of each jet is represented as the solid arrows, and $\not{H}_T$ vector of each jet is shown by the dashed line. The relation between each jet and $\not{H}_T$ vector is represented as the different shades of the same colour. The angle between each jet and the associated $\not{H}_T$ vector is shown by the red arc. The jet that minimises the angle is considered the mismeasured jet, the angle of which is $\Delta\phi^*_{min}$.

the distribution the low value region of $\Delta\phi^*_{min}$ is abundant in QCD multijet processes, whereby the higher values of $\Delta\phi^*_{min}$ are occupied by the EWK backgrounds as the jets recoil against the invisible system.

6.2.3 The missing energy ratio $\not{H}_T/\not{E}_T$

The final control variable used in the suppression of QCD multijet events is the ratio $\not{H}_T/\not{E}_T$, which aims to target events with fake-$\not{E}_T$ arising from severe mismeasurement and threshold effects.

If several jets fall just below the threshold, they can provide a source of fake-$\not{H}_T$. Consequently, the threshold on the particles from which $\not{E}_T$ is calculated is lower than that of $\not{H}_T$. Therefore, the ratio of these two variables can effectively remove the background from physics objects that have fallen below threshold.

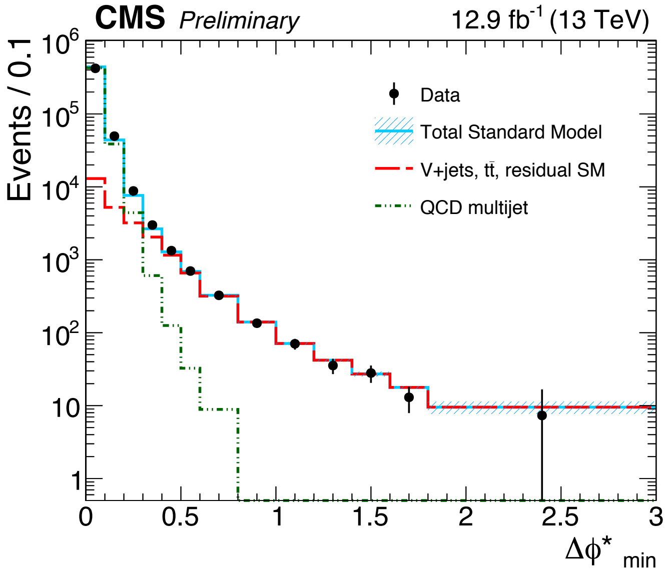


Fig. 6.3 Distribution of $\Delta\phi^*_{min}$ for data and MC simulation at $\sqrt{s} = 13$ TeV. The electroweak (red solid line) and QCD (green dashed line) are determined from simulation. The data and simulated events are subject to a criteria $H_T > 800$ GeV and $p_T^{j2} > 100$ GeV, where $j2$ is the sub-leading jet [14].

6.3 Object selection

The $\alpha_{\rm T}$ analysis utilises the collection of physics objects described in Section 4 to select and reconstruct events. Despite the hadronic search region of the analysis, a large range of physics objects are used.

Leptons and photons are used to construct the control regions, used in the prediction of the backgrounds. In addition, the veto of events containing such particles in the signal region ensures a hadronic composition.

To characterise the hadronic activity within an event, two jet collections are formed. The former aims to quantify the hadronic activity within the event, while the latter is used in the study of reconstructing boosted particles that decayed hadronically, and is discussed in Section 6.4.

6.3.1 Jets

In the analysis, jets are reconstructed with the anti-k$_t$ clustering algorithm, with a radius parameter of ΔR = 0.4 (AK4 jets) The input to the clustering algorithm are the

PF-candidates, as discussed in Section 4.1. Further, additional CHS and jet area corrections are applied to suppress PU energy contamination, before relative and absolute JECs are applied, as described in Section 4.2.

Such reconstructed jets, are required to satisfy the *loose* PFJet ID WP, defined by the selections listed in Table 6.1.

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

Table 6.1 The *loose* jet ID requirements.

Jets that originate from bottom quarks are identified using the CSVv2 algorithm, through a *medium* WP [11]. This corresponds to a requirement of > 0.80 on the discriminator value of the aforementioned algorithm, with a misidentification probability for light-parton jets $\sim 1\%$, and a charm-quark mis-tag rate $\sim 10\%$ and a tagging efficiency $\sim 80\%$ [79].

To ensure good reconstruction and suppress the QCD multijet background whilst maintaining a strong signal efficiency, kinematic selections are applied on the reconstructed jets. Jets are required to be centrally produced, $|\eta| < 3$, with $p_{\text{T}} > 40$ GeV. On the jet with the largest transverse momentum, the pseudorapidity requirement is tightened to $|\eta| < 2.5$ and $p_{\text{T}} > 100$ GeV.

6.3.2 Muons

In constructing either the signal region or the control regions, two types of selections are made on muons. To ensure good reconstruction and efficiency with respect to trigger selections, muons in the control region are subject to a *tight* ID WP with a selection of $p_{\text{T}} > 30$ GeV and $|\eta| < 2.1$. This provides a $\sim 95\%$ selection efficiency. In addition, muons in the control region have a PF relative isolation requirement, defined in Section 4.1.1, of $I_{PF}^{rel} < 0.15$ computed with a cone size of $\Delta R = 0.4$.

For the purpose of vetoing muons in the signal region, a *loose* ID WP is used, with a selection of $p_{\mathrm{T}} > 10$ GeV and $|\eta| < 2.1$. Further, the mini-isolation algorithm is utilised with effective-area PU correct $I_{PF}^{mini} < 0.2$ [103].

6.3.3 Photons

Photons are selected using the cut-based *tight* ID WP, which provides a $\sim 71\%$ selection efficiency and $\sim 88\%$ background rejection. Isolation requirements, applied to provide an effective control of fake-prompt photons is shown in Table 6.2.

Isolation type	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	$1.40 + 0.0014 \times p_{\mathrm{T}}{}^{\gamma}$	$1.40 + 0.0091 \times p_{\mathrm{T}}{}^{\gamma}$

Table 6.2 Photon isolation selection requirements on the 2016 analysis (*tight* working point).

The main source of misidentified photons is of the hadronic nature, and as such, the charged hadron isolation is employed and is defined as the scalar sum of energy of all charged hadronic PF-candidates reconstructed inside a cone of Δ R < 0.4. In addition, a neutral hadron and a photon isolation requirement are implemented, as recommended by the *tight* working point. For photons satisfying the requirements in Table 6.2, a kinematic selection of $p_{\mathrm{T}} > 25$ GeV and $|\eta| < 2.5$ is applied on veto photons as part of the construction of the signal region. In forming the γ+jets control region, a tighter requirement of $p_{\mathrm{T}} > 200$ GeV and $|\eta| < 1.45$ is applied.

6.3.4 Electrons

Electrons are subject to a veto in the signal region, following a selection using the *loose* ID WP, which corresponds to a $\sim 90\%$ selection efficiency. Electrons are required to be isolated with a PF relative mini isolation requirement of $I_{PF}^{mini} < 0.1$, with the effective-area corrected using a maximum cone size of $\Delta R = 0.2$.

Furthermore, to ensure good reconstruction, reconstructed electrons are required to satisfy the kinematic selections of $p_{\mathrm{T}} > 10$ GeV and $|\eta| < 2.5$.

6.3.5 Isolated tracks

Isolated tracks are used to identify single prong decays of the τ lepton and misreconstructed leptons, as discussed in Section 4.1.2. They are selected by requiring $p_{\mathrm{T}} > 10$ GeV and must be associated with the primary vertex (PV), $\Delta z(\text{track, PV}) < 0.05$ cm. An isolation requirement on the tracks is employed with a PF relative isolation requirement of $I_{PF}^{rel} < 0.1$, computed from charged PF-candidates associated with the PV.

A summary of the physics objects and their kinematic requirements, used in the α_{T} analysis, is presented in Table 6.3.

Object		Kinematic Selection
Jet (AK4)	Central jets	$p_{\mathrm{T}} > 40$ GeV, $\lvert\eta\rvert < 2.4$
	Leading central jets	$p_{\mathrm{T}} > 100$ GeV, $\lvert\eta\rvert < 2.4$
	Forward jet (veto)	$p_{\mathrm{T}} > 40$ GeV, $\lvert\eta\rvert > 2.4$
Jet (AK8)	Central jets	$p_{\mathrm{T}} > 200$ GeV, $\lvert\eta\rvert < 3.0$
Photon	γ+jets control region (relative isolation)	$p_{\mathrm{T}} > 200$ GeV, $\lvert\eta\rvert < 1.45$
	Veto (relative isolation)	$p_{\mathrm{T}} > 25$ GeV, $\lvert\eta\rvert < 2.5$
Muon	μ+jets and $\mu\mu$+jets control regions (relative isolation)	$p_{\mathrm{T}} > 30$ GeV, $\lvert\eta\rvert < 2.1$
	Veto (mini isolation)	$p_{\mathrm{T}} > 10$ GeV, $\lvert\eta\rvert < 2.1$
Electron	Veto (mini isolation)	$p_{\mathrm{T}} > 10$ GeV, $\lvert\eta\rvert < 2.5$
Isolated track	Veto	$p_{\mathrm{T}} > 10$ GeV, $\lvert\eta\rvert < 2.5$

Table 6.3 Kinematic selections on the physics objects used in the α_{T} analysis.

6.3.6 Energy sums

The total missing energy, $\not{E}_T$, is computed as the vector sum of the momentum of all PF-candidates within $|\eta| < 5$, with a type-I correction applied as discussed in Section 4.3. The $\not{E}_T$ variable is utilised in the definition of the transverse mass, M_T, which forms part of the selection criteria that defines the muon control sample. In the control regions, $\not{E}_T$, is computed without the momentum of the corresponding lepton or photon.

The total hadronic energy, H_T, and the hadronic missing energy, $\not{H}_T$, are constructed with the jets outlined in Section 6.3.1, and are thus subject the kinematic requirements of $p_{\mathrm{T}} > 40$ GeV and $|\eta| < 3$.

6.4 Jet substructure

Another collection of jets are reconstructed with the anti-k$_t$ clustering algorithm, using an enlarged distance parameter $\Delta R = 0.8$ (AK8 jets), and with CHS and jet area corrections applied to compensate any PU energy contamination. Relative and absolute JECs are also applied.

On the collection of reconstructed AK8 jets, the N-subjettiness algorithm is employed to measure the consistency of a jet containing N or fewer subjets. As displayed in Equation 4.11, the value of τ_N quantifies to what degree a particular jet is composed of N subjets. In addition, jets with $\tau_N \approx 0$ have all of their energy aligned with the candidate subjet directions and therefore have N or fewer subjets. Furthermore, jets with $\tau_N >> 0$ have a large fraction of their energy distributed away from the candidate subjet directions and therefore have at least N+1 subjets. Therefore, an effective approach to discriminate between jets containing N subjets and jets originating from QCD multijet processes, is to take the ratio of jet shapes: τ_N/τ_{N-1} [12].

Following the application of the softdrop and pruning algorithms, the invariant mass of the subjets belonging to a tagged jet can be used to form m$_{softdrop}$ and m$_{pruned}$, respectively. As the invariant mass of the jet is a key observable to identify top quarks and W bosons, targeted selections can be applied to discriminate against a background of light quark and gluon jets.

In the analysis, there is no minimum requirement on the number of AK8 jets per event. However, to ensure that any AK8 jet is centrally produced and has a sufficient transverse momentum to force their decay products into a single jet, a selection criteria of $p_{\mathrm{T}} > 200$ GeV and $|\eta| < 3$ is applied on each AK8 jet.

6.4.1 t-tagging

The *softdrop* algorithm [84] is applied on each reconstructed AK8 jet to remove soft and wide-angle radiation produced by PU interactions, underlying events (UE), or parton shower activity.

An additional selection of $p_{\mathrm{T}} > 400$ GeV is applied on each AK8 jet to ensure that the decay products of a hadronically decaying top quark candidate be contained within the AK8 jet. The identification of boosted top quarks from AK8 jets (t-tagging) is performed with a combination of both the N-subjettiness and the variables determined from the softdrop algorithm. Furthermore, as one of the expected subjets of a top quark jet is expected to originate from a bottom quark, the subjets are thus subject to the CSVv2 algorithm to provide additional sensitivity in rejecting jets originating from both QCD multijet processes and EWK backgrounds.

t-tagging	
Variable	**Selection**
p_{T} (AK8 jet) (GeV)	≥ 400
$m_{softdrop}$ (GeV)	$\{105{,}220\}$
τ_3/τ_2	< 0.5
subjet b-tag discriminator value	> 0.46

Table 6.4 Summary of t-tagging requirements used in the analysis. [20].

The WP used in identifying boosted top quarks, that provided a background efficiency of $\sim 0.1\%$, was the softdrop working point 2 (SD WP2) [104], the conditions of which are shown in Table 6.4.

6.4.2 W-tagging

Of the collection of reconstructed AK8 jets, the *pruning* algorithm is also applied to remove soft and wide-angle radiation and identify subjets that correspond to the decay products of hadronically decaying W bosons (W-tagging) [83]. The configurable parameters of the pruning condition, displayed in Equation 4.9, are chosen to be $z = 0.1$ and $R = 0.5$ [20]. The pruned mass (m_{pruned}) is calculated from the sum of the four-vectors of the jet constituents, following the pruning requirements.

Jets are tagged as W-tagged jets following the *medium* ID WP, the conditions of which are shown in Table 6.5.

W-tagging	
Variable	**Selection**
p_{T} (AK8 jet) (GeV)	≥ 200
m_{pruned} (GeV)	$\{65{,}105\}$
τ_2/τ_1	< 0.45

Table 6.5 Summary of W-tagging requirements of the analysis. [20].

As both the pruning and softdrop algorithms are applied on each AK8 jet, in the case of a single jet being identified as both a top quark jet and a W boson jet, the jet is simply treated as a top quark candidate and the W boson is discarded.

6.5 Trigger strategy

A core principle of the analysis is to retain as much phase space as possible, and thus retain sensitivity to the many signatures that SUSY and DM exhibit.

To achieve this, dedicated signal triggers that utilise the α_{T} variable are employed. Furthermore, events with significant hadronic activity and missing energy are retained using triggers that rely on H_T and $\not{H}_T$ variables that are constructed within the HLT.

To efficiently collect events with low values of H_T, the α_{T} variable is employed within the HLT. Requirements on a combination of α_{T} and H_T for a variety of different H_T thresholds are made to limit the trigger rate and maintain acceptance of events with low H_T.

In addition, monojet signatures are selected with a $\not{H}_T$-$\not{E}_T$ cross-trigger, that has thresholds of $\not{H}_T > 90$ GeV and $\not{E}_T > 90$ GeV and a jet ID selection. The full details of the triggers used in the analysis are shown in Table 6.6.

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

Table 6.6 Trigger thresholds of the Level-1 hardware trigger and HLT for the hadronic signal region and the leptonic control regions.

6.6 Pre-selection

A basic experimental method for searching for signatures of BSM physics, is to select events that exhibit characteristics expected from signal, thus forming a signal region. The sensitivity of such a method is dependent on the ability to remove events originating from background processes.

In addition, targeted filters and vetoes are designed to control events containing fake sources of $\not{E}_T$, acquired through the means discussed in Section 6.2.

6.6.1 Event filters

Certain requirements that evolved during the commissioning period of the detector in Run II are applied. These filters are designed to remove source of fake-$\not{E}_T$ and remain insensitive to real physics events. These include:

- Reject events with CSC detector hits and energy deposits in the calorimeters that are caused by beam interactions exterior to the detector. These *beam halo* effects are removed by the CSC beam halo filter.

- Events with anomalous signals in the HCAL and ECAL, are removed by the HBHE noise and isolation filter, the EE bad supercluster noise filter and the ECAL Trigger Primitive (TP) filter.

- The *bad track filter* removes events with failures in the track reconstruction, by targeting events with misidentified straight tracks and a large p_{T}.

6.6.2 Event vetoes

In addition, events with characteristics that are not controlled by the event filters are subject to further criteria. Such requirements veto or retain an event based on the following conditions:

- Events where the lead jet has a PF Charged Hadron Energy Fraction (CHEF) $< 10\%$ are vetoed. This requirement targets residual beam halo effects that can form fake jets through the beam halo calorimeter energy deposits.

- Isolated tracks are used to identify single prong decays of the τ lepton and misreconstructed leptons. As discussed in Section 4.1.2, lost lepton events can arise from a range of sources including misreconstructed e^{-}, μ^{-} decays and in hadronic decays of τ leptons.

  Such events are suppressed with an isolated track veto, which removes events with an isolated track with $p_{\mathrm{T}} > 10$ GeV and $|\eta| < 2.5$.Further, for the leptonic control regions, a minimum $\Delta R > 0.02$ requirement is applied between the reconstructed lepton and the isolated track to ensure that the isolated track is not associated to the reconstructed lepton.

- The requirement of $|\eta| < 3$ on jets ensures that no significant energy is deposited in the region of the detector that is not covered by the tracker. For any event containing jets that are outside this range, *forward jets*, the event is vetoed.
- The $\not{H}_T/\not{E}_T < 1.25$ requirement, described in Section 6.2.3, rejects events containing $\not{H}_T$ due to energy present in the event that is not clustered into jets.

6.6.3 Kinematic selections

Kinematic selections are applied on events, with a requirement of at least one jet with $p_{\mathrm{T}} > 100$ GeV, with a lowered requirement of $p_{\mathrm{T}} > 40$ GeV on all other jets. An additional requirement of $|\eta| < 3$ on the jets ensures that no significant energy is deposited outside the acceptance of the detector. Furthermore, the QCD multijet background consists of many low p_{T} jets that are prone to deposit a significant amount of energy in the forward region of the detector. Therefore, the requirement has sensitivity in suppressing this background with little effect to the acceptance of signal models.

Additional kinematic selections are applied with requirements on the total hadronic energy and hadronic missing energy of $H_T > 200$ GeV and $\not{H}_T > 130$ GeV, respectively. This ensures that the events retained are energetic with a missing energy signature suitable for consideration.

A summary of the pre-selection requirements, is shown in Table 6.7.

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

Table 6.7 Summary of the pre-selection criteria.

6.7 The signal region

After applying a pre-selection criteria, a collection of targeted selections are made to construct a signal region. The signal region is subject to selections that favour final states composed of jets and $\not{E}_T$, and are chosen to remain sensitive to potential BSM signatures. In addition, events with mismeasurements and contributions from SM processes with genuine $\not{E}_T$ aim to be minimised.

Events containing identified leptons, as described in Section 4, are removed to enforce a purely hadronic topology in the signal region. Furthermore, any event that contains a reconstructed photon is also removed.

Due to the hadronic nature of the signal region, a dominant expected background are QCD multijet processes. The contribution of which is significantly reduced by applying H_T dependent α_{T} requirements. Events satisfying this condition are then subject to a $\Delta\phi^*_{min}$ requirement to ensure any residual background is removed. Thereafter, events with significant energy deposits that have not been clustered as jets are removed by rejecting events with $\not{H}_T/\not{E}_T > 1.25$. The requirements specific to the signal region are shown in Table 6.8:

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 6.8 The H_T- α_{T} requirements applied in the signal region. For all H_T bins satisfying $H_T > 800$ GeV, no α_{T} cut is applied, with only the $\not{H}_T > 130$ GeV requirement applied.

The scaling of the α_{T} thresholds is driven by the trigger selections in the signal region and the evolution of the α_{T} behaviour of the QCD multijet background with H_T. For events with $H_T > 800$ GeV, no α_{T} requirement is made due to the decreased efficiency of the α_{T} selections with high jet multiplicities.

Incorporating jet substructure

In the event categories in which t-tagging and W-tagging is performed, as detailed in Table 6.11, a minimum requirement of $p_{\mathrm{T}} > 400$ GeV and $p_{\mathrm{T}} > 200$ GeV respectively is applied on any AK8 jet.

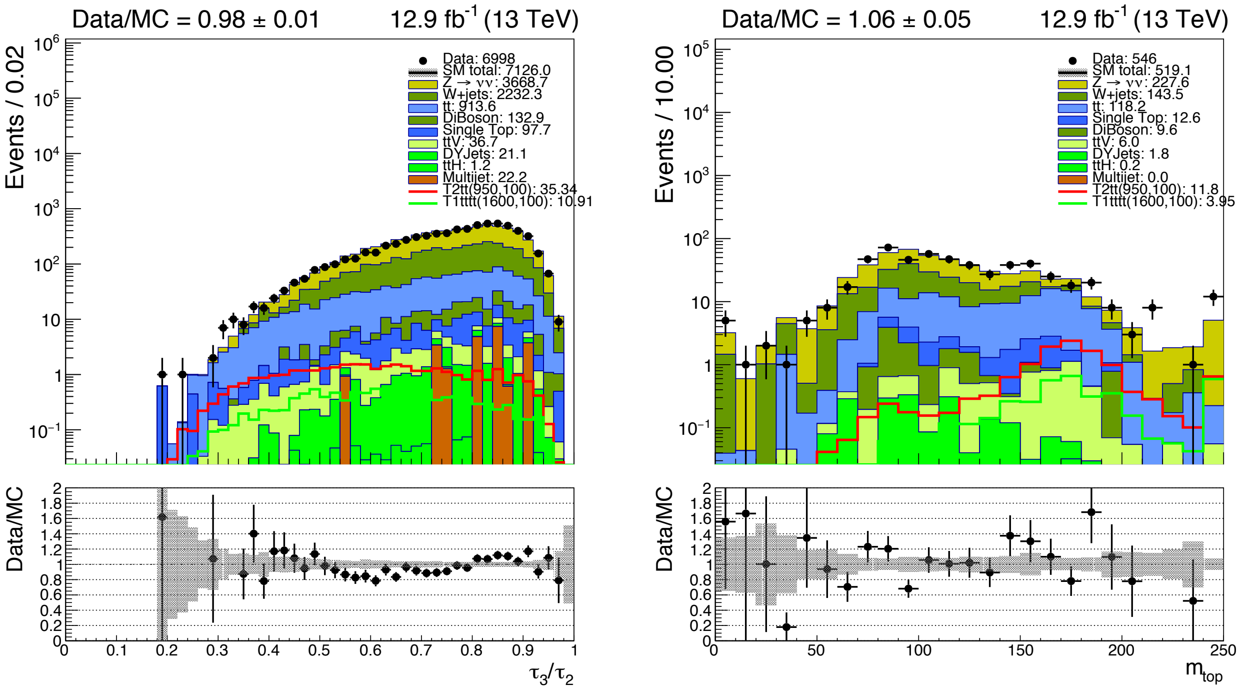
The distributions of the primary observables used in t-tagging and W-tagging, and respectively listed in Table 6.4 and Table 6.5, are shown in Figure 6.4 and Figure 6.5 for both data and simulated events. Furthermore, two benchmark SUSY signal models for a compressed scenario, in which the mass splitting between the parent sparticle and LSP is large, are overlayed. The signal model shown in red, and labelled T2tt, is the pair production of squarks whereby each squark subsequently decays into a top quark and a LSP. In addition, the signal model shown in green, and labelled T1tttt, is the pair production of gluinos whereby each gluino decays into a $\mathrm{t\bar{t}}$ pair and a LSP[1] The SUSY models, described in more detail in Section 8.4, are chosen due to the presence of top quark pairs in their final state, that under the compressed regime, could acquire a boost and thus be sensitive to jet substructure techniques.

In each distribution, the requirements of the signal region are applied and the variable associated to the distribution is subject to only the conditions of its counterpart. As an example, in the case of t-tagging, the distribution for $\mathrm{m}_{softdrop}$ shown in Figure 6.4b is subject to only the requirements on τ_3/τ_2, as listed in Table 6.4. In a similar method, the distribution of τ_3/τ_2, shown in Figure 6.4a is subject to the requirements of $\mathrm{m}_{softdrop}$, as listed in Table 6.4. This method of visualising a variable is often referred to as an "N-1" distribution, where a requirement on a variable, denoted N, is investigated upon removing a requirement associated to the one in question. This provides an insight into the sensitivities across different requirements on variables, and in the case of t-tagging and W-tagging demonstrates the individual strength each requirement has on its counterpart.

In the distributions of Figure 6.4 and Figure 6.5, there is shown to be a satisfactory modelling, in data and simulated events, of the key variables used in the application of jet substructure techniques. Furthermore, the distributions of each mass variable, $\mathrm{m}_{softdrop}$ and m_{pruned}, upon the application of the counterpart variable, τ_3/τ_2 and τ_2/τ_1, show that the majority of events are situated near the mass of the particle that is aiming to be reconstructed. This phenomena is further highlighted for the signal models, where the majority of events peak near the mass that each algorithm aims to target. It should be noted that although the agreement between data and simulated events is satisfactory, any disagreement is not critical in the analysis. Furthermore, the estimation of the backgrounds in the signal region is obtained through control regions that, together with the simulated events satisfying the signal region requirements, provide a more accurate estimation of the EWK contribution in the signal region. In addition, the final background estimation is obtained in a way such that any observed disagreement between data and simulated events is accounted for in a fitting procedure, described formally by a likelihood model and discussed in more detail in Section 8.1.

---

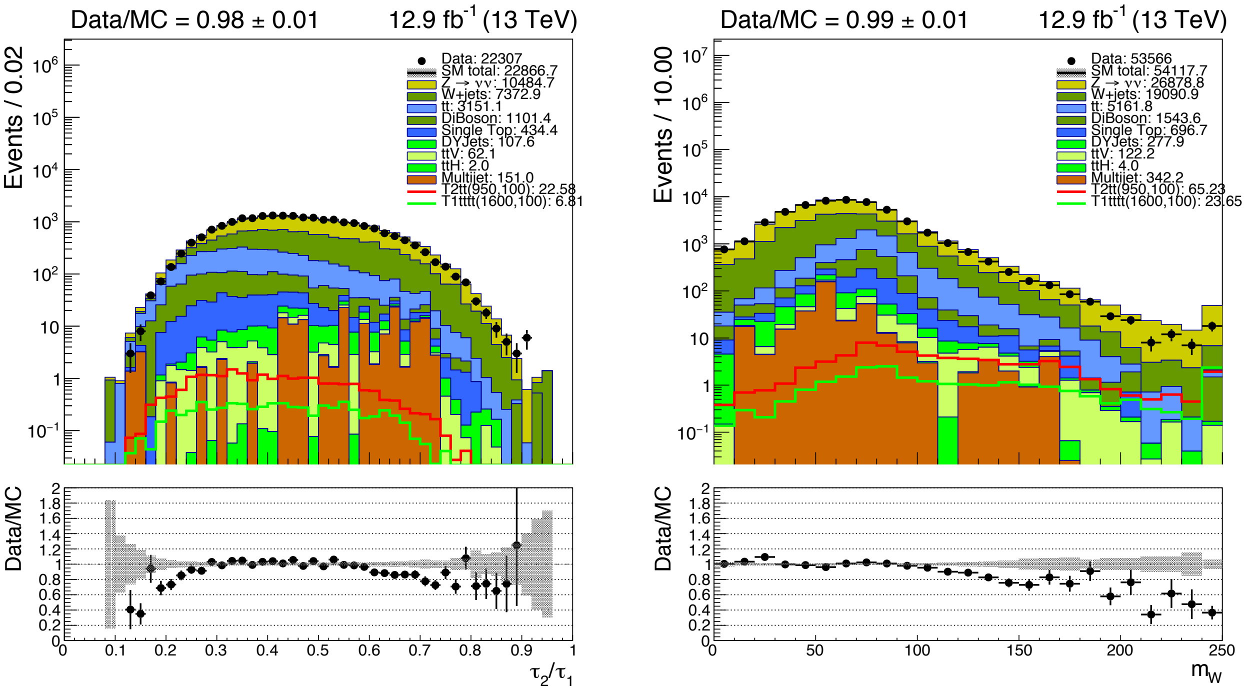
[1] An example of the gluino and stop decay, relevant to T2tt and T1tttt, is shown in Figure 8.7.



(a) Distribution of the ratio, τ_3/τ_2 , in the signal region.

(b) Distribution of the $m_{softdrop}$ variable, in the signal region.

Fig. 6.4 “N-1” distributions of the key variables used in t-tagging, in the signal region. In each case, there is no requirement on the subjet b-tag CSVv2 discriminator value. Each distribution is subject to the requirements of the signal region. The values in the legend denote the number of events satisfying the requirements of the signal region.



(a) Distribution of the ratio, τ_2/τ_1 , in the signal region.

(b) Distribution of the m$_{pruned}$ variable, in the signal region.

Fig. 6.5 “N-1” distributions of the key variables used in W-tagging, in the signal region. Each distribution is subject to the requirements of the signal region. The values in the legend denote the number of events satisfying the requirements of the signal region

6.8 The control regions

The estimation of the SM background in the signal region is performed using data-driven techniques in dedicated control regions. For this purpose, $\mu\mu$+jets and γ+jets control regions are used in the estimation of the irreducible $\mathrm{Z} \to \nu\bar{\nu}$ background, and a μ+jets control region is used in the $\mathrm{W}+\mathrm{jets}$ and $\mathrm{t\bar{t}}$ backgrounds.

The presence of QCD multijet processes in either control region is suppressed by the core requirement of isolated muons or a photon. As a result, the $\Delta\phi^{*}_{min}$ requirement is not applied in the control regions and the α_{T} requirements are not applied in the muon control regions. This statistically increases the datasets of each control region and thus reduces the statistical uncertainties on the background estimations.

Distributions of the key variables in the μ+jets, $\mu\mu$+jets and γ+jets control regions are shown in Section 7.2.

6.8.1 The μ+jets control region

The μ+jets control region is constructed to select events that contain W bosons that are produced in association with jets. Events are required to contain a muon that satisfies both the tight WP ID and the kinematic requirement of $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.1$. In addition, the muon is required to be well separated from other jets by requiring $\Delta R(\mu,\mathrm{jet}) > 0.5$. Furthermore, to select W bosons and ensure a purer control region, a requirement on the transverse mass between the $\not{E}_T$ and the muon is applied, $30 < M_T(\mu,\not{E}_T) < 125$ GeV. This effectively reduces any residual events that contain QCD multijet processes, and therefore does not demand a $\Delta\phi^{*}_{min}$ and α_{T} requirement.

Events in the μ+jets control region are also subject to t-tagging and W-tagging, which constructs a phase space similar to the signal region. The SM background prediction for such events of the signal region, using the μ+jets region, is discussed in more detail in Section 7.

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

As a means to select events containing Z bosons, the $\mu\mu$+jets control region is used. To ensure the Z bosons decay to a muon pair, and are produced in association with jets, exactly two muons are required to pass the tight WP ID with the additional kinematic selection of $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.1$. Furthermore, the invariant mass of the muon pair is required to be close to the Z boson mass, by requiring $66.2 < \mathrm{M}_{\mu_1\mu_2}$ < 116.2 GeV.

As the events of the μ+jets and $\mu\mu$+jets control region are dominated by the decays of massive W and Z bosons respectively, no α_{T} and $\Delta\phi^{*}_{min}$ selections are required.

6.8.3 The γ+jets control region

To form the γ+jets control region, a single tight photon is required to be reconstructed with $H_T > 400$ GeV. This ensures that the photon is produced significant hadronic energy. The requirement on the photon is such that $p_{\mathrm{T}} > 200$ GeV and $|\eta| < 1.45$, which ensures a high trigger efficiency. A requirement on the separation between the photon and each jet such that, $\Delta R(\gamma, \text{jet}) > 1$ ensures that the photon is well isolated and did not originate from a jet. Photon samples are more easily contaminated with QCD multijet events than in the muon control regions. To mitigate the QCD multijet contamination in the γ+jets control region, the α_{T} requirements, shown in Table 6.8, are applied.

6.9 Event categorisation

To better separate signal and background topologies, the search region of the analysis is divided into categories of discriminating variables. The variables are constructed to separate events with characteristics common to signal models from those originating from background processes.

Events in the α_{T} analysis are categorised according to the multiplicity of AK4 jets: n_{jet}, their total hadronic energy H_T, the number of jets identified as originating from bottom quarks, n_{b}, and the magnitude of the missing hadronic energy $\not{H}_T$.

As a means to enhance sensitivity, events in this analysis are further categorised according to the total number of subjets associated to either a t-tagged or W-tagged jet per event, the details of which are discussed in more detail in Section 6.9.3.

6.9.1 Jet categories

As the expected topology of SUSY processes is rich in hadronic activity, the jet multiplicity, n_{jet}, is a strong discriminator between events containing background processes and those containing signal processes. Furthermore, categories with increasing n_{jet}

will contain far fewer events originating from SM processes, while for different signal models, the jet multiplicity can vary. In SUSY processes, squark pair production can result in final states with typically fewer jet multiplicities compared to the pair production of gluinos, where final states with high n_{jet} can be expected.

In addition, signal models with smaller mass separations, known as compressed spectra, exhibit less kinetic energy in their final states and as such, produce jets with low momenta that are challenging to distinguish from SM backgrounds. To achieve sensitivity to such topologies, events are also categorised according to their jet topology: *monojet*, *asymmetric* and *symmetric*, which are defined by requirements on the transverse momentum on the second leading jet.

Regarding the signature of pair produced sparticles, where at least two high p_{T} jets are expected, the symmetric selection is defined by requiring the second leading jet to have $p_{\mathrm{T}} > 100$ GeV. This requirement aims to retain acceptance to such models, while suppressing the SM backgrounds.

The compressed spectrum case is targeted by the asymmetric and monojet categories, requiring the second leading jet to have $40 < p_{\mathrm{T}} < 100$ GeV and $p_{\mathrm{T}} < 40$ GeV, respectively. Sensitivity to such scenarios is achieved by targeting events with significant initial state radiation (ISR) recoiling against the decay system that results in a signature of a single hard jet and a soft or absent second jet. In addition to providing a greater acceptance to compressed spectra, the categorisation of events according to their jet topologies is also sensitive to dark matter models.

While the compressed spectra of signal models, that typically result in soft final state particles, is accessible via the different jet topologies, the uncompressed scenario will be typically restricted to one. Such topologies, whereby the mass splitting between the parent sparticle and the LSP is large, will result in typically harder jets that will then populate the symmetric category.

Thus, in addition to the separate jet categories, events are categorised according to their jet and b-jet multiplicities, where $n_{\mathrm{jet}} = 1,2,3,4$ and ≥ 5. Similarly, the b-jet categories cover $n_{\mathrm{b}} = 0,1,2$ and ≥ 3, where the maximum value of n_{b} cannot exceed that of n_{jet}.

The n_{jet} and n_{b} categorisation is illustrated in Table 6.9.

6.9.2 Hadronic energy categories

To further separate different signal and background processes, events in a given jet category shown in Table 6.9 are categorised according to H_T. This provides a measure

Selection	Categorisation														
n_{jet}	2			3				4				$\geq$5			
n_{b}	0	1	2	0	1	2	3	0	1	2	3	0	1	2	3

Table 6.9 n_{jet} and n_{b} event categorisation in the α_{T} analysis .

of the mass scale of the interaction through the activity of the jets within an event.

The α_{T} analysis has a minimum H_T requirement of > 200 GeV, following which a collection of H_T categories in varying width allow events to distribute. The H_T categories are chosen to be: H_T = 200-250, 250-300, 300-350, 350-400, 400-500, 500-600, 600-800, and > 800 GeV. To ensure that events contain sufficient missing energy, a minimum $\not{H}_T$ requirement of > 130 GeV is made.

A complete summary of the event categorisation for the n_{b}, n_{jet} and H_T variables is shown in Table 6.10.

Selection	Categorisation
n_{jet} multiplicity	$n_{\text{jet}} = 1$ (Monojet)
	$n_{\text{jet}} = 2, 3, 4, \geq 5$ (Symmetric/ Asymmetric)
b-jet multiplicity	0, 1, 2, ≥ 3 ($n_{\text{b}} \leq n_{\text{jet}}$)
H_T	200–250, 250–300, 300–350, 350–400, 400–500, 500–600, 600–800, ≥ 800
$\not{H}_T$	> 130 GeV

Table 6.10 Nominal event categories for the analysis.

6.9.3 Incorporating substructure

As previously discussed, different signal models can affect the kinematics and composition of final states. In compressed scenarios, relaxing the p_{T} requirements on the sub-leading jet to form additional jet categories, such as the asymmetric and monojet categories, allows events of this nature to be retained. In more uncompressed regimes, whereby the kinetic energy available is often large, the typical p_{T} of particles is expected to increase.

Under this assumption, jet substructure techniques are used to identify jets that have originated from highly boosted, hadronically decaying top quarks or W bosons, as described in Section 6.4.

Event categorisation

As a means to provide sensitivity in incorporating jet substructure techniques, events associated to both t-tagged and W-tagged jets are further categorised. A key distinction between a t-tagged jet and a W-tagged jet is the subjet structure associated to each, where the expected number of subjets associated to a boosted top quark and W boson is three and two respectively.

Therefore, to classify events containing both t-tagged and W-tagged jets, the variable n_{subjets} is used. The method to simultaneously consider t-tagged and W-tagged jets is achieved by choosing the value of n_{subjets} to be the sum of the union of the subjets of the set of t-tagged and W-tagged jets in an event.

If a jet is both t-tagged and W-tagged, then the t-tagged jet is retained and the W-tagged is discarded.

Mathematically, n_{subjets} is written as:

$$n_{\text{subjets}} = \sum N^{top}_{subjets} \cup N^{W}_{subjets} \tag{6.5}$$

where $N^{top}_{subjets} \equiv N_{tops} \times 3$, and is defined as the total number of subjets associated to all t-tagged jets in the event. As an example, if an event contains one t-tagged and one W-tagged jet, the corresponding value of n_{subjets} is five.

As the ability to efficiently tag boosted top quarks or W bosons is dependent on the p_{T} of such a particle, the application of t/W-tagging is subject to a H_T and n_{jet} requirement. In this case, events in the final H_T region of the analysis, > 800 GeV, are subject to further categorisation according to n_{subjets}.

Furthermore, as different signal models result in different jet multiplicities, two scenarios are considered. For certain signal models where the jet multiplicity is expected to be large, events in the final symmetric n_{jet} region, ≥ 5, are subject to t/W-tagging and thus categorised according to n_{subjets}. Similarly, for signal models that yield typically low jet multiplicities, events in the $n_{\text{jet}} = 3$, 4 and ≥ 5 symmetric region are subject to t/W-tagging and then categorised according to n_{subjets}.

The range of n_{subjets} is set as $\{0, 2, 3, 4, 5, \geq 6\}$, whereby an event in a tagging region with no t-tagged or W-tagged jets would yield $n_{\text{subjets}} = 0$. In an non-tagging region, in which no t/W-tagging is performed, events are not categorised according to n_{subjets}.

Therefore, the nominal event categorisation of Table 6.10 is expanded to form:

n_{jet} selection	n_{b} selection	Final H_T threshold	n_{subjets} selection	
			High n_{jet} multiplicity	Low n_{jet} multiplicity
1	0, 1	600 GeV	≥ 0	≥ 0
2	0, 1, 2	800 GeV	≥ 0	≥ 0
3	0, 1, 2, ≥ 3	800 GeV	≥ 0	0, 2, 3, 4, 5, ≥ 6
4	0, 1, 2, ≥ 3	800 GeV	≥ 0	0, 2, 3, 4, 5, ≥ 6
≥ 5	0, 1, 2, ≥ 3	800 GeV	0, 2, 3, 4, 5, ≥ 6	0, 2, 3, 4, 5, ≥ 6

Table 6.11 Extended event categorisation according to n_{subjets} in the symmetric region for two scenarios, one with an expected high jet multiplicity and the second with an lower expected jet multiplicity.

6.10 Summary

The selections described in this section are chosen to strongly suppress the dominant QCD multijet background as well as significantly reducing the electroweak backgrounds. This is achieved in the construction of the signal region, aimed to provide inclusive sensitivity to a wide range of BSM models. Furthermore, the use of jet substructure techniques to tag jets as originating from top quark decays or a W boson decay provides a means to target the characteristic boosted topologies of signal models. The events that remain in the signal region are finely categorised to better separate signal and background topologies. Thereafter, to extract a potential signal contribution, the residual backgrounds in the signal event categories must be accurately predicted, as described in Section 7.

Chapter 7

Background Estimation

The ability to precisely measure the SM backgrounds is a key requirement in any search for new physics. Following the formation of the signal region, to maximise the signal sensitivity the analysis utilises dedicated methods to estimate the number of events originating from SM processes. Due to the difference in topology, these backgrounds are classified under two categories, the electroweak and the QCD multijet backgrounds. Thereafter, dedicated background estimation methods are employed, using transfer factors formed with the analysis control regions and targeted QCD multijet background estimation methods to estimate the electroweak and QCD multijet background respectively.

7.1 Estimation of SM processes with genuine $\not{E}_T$

An accurate modelling of the EWK backgrounds is achieved using both simulated MC samples and data-driven corrections, that provide predictions of the SM background contribution in the signal region. These samples are generated with the methods discussed in Section 5, with the corresponding cross sections for each sample tabled in Table 5.1. The simulations are subject to corrections discussed in Section 5.4, aimed to improve the agreement between data and simulation.

7.1.1 The transfer factor method

The control samples are constructed in a way to obtain a collection of events in data that are in a phase space similar to that of the signal region. From which, each event category in a control region can be extrapolated to the signal region through a *transfer factor*. Although the EWK backgrounds are generally well described by simulation, any mismodelling is mitigated through the use of control regions in data that, when used in conjunction with the correction from the transfer factor, provides a more accurate prediction of the EWK contribution in the signal region.

The transfer factor method relies on simulated samples to transform the observed yield in a given H_T, jet (n_{jet}), b-tag (n_{b}) multiplicity category of a control sample, $N^{\text{control}}_{\text{obs}}(n_{\text{jet}}, n_{\text{b}}, H_T)$, into a predicted yield for the corresponding category of the hadronic signal region, $N^{\text{signal}}_{\text{pred}}(n_{\text{jet}}, n_{\text{b}}, H_T)$. The choice of event categorisation in the control samples is identical to that for the signal region, as defined in Section 6.

Each transfer factor is simply a ratio of the yields obtained from MC simulation for the same category of that in signal region and a given control sample.

In the nominal event categories whereby no t-tagging and no W-tagging is applied, the transfer factor is of the form:

$$\text{TF} = \frac{N^{\text{signal}}_{\text{MC}}(n_{\text{jet}}, n_{\text{b}}, H_T)}{N^{\text{control}}_{\text{MC}}(n_{\text{jet}}, n_{\text{b}}, H_T)} \tag{7.1}$$

where $N^{\text{signal}}_{\text{MC}}$ is the total simulated event yield for all background processes in the signal region, and $N^{\text{control}}_{\text{MC}}$ is the total event yield in a control region, taken from simulation.

In this way, predictions of the total EWK contribution in the signal region, $N^{\text{signal}}_{\text{pred}}$, can be made based on the following relation:

$$N^{\text{signal}}_{\text{pred}}(n_{\text{jet}}, n_{\text{b}}, H_T) = \frac{N^{\text{signal}}_{\text{MC}}(n_{\text{jet}}, n_{\text{b}}, H_T)}{N^{\text{control}}_{\text{MC}}(n_{\text{jet}}, n_{\text{b}}, H_T)} \times N^{\text{control}}_{\text{obs}}(n_{\text{jet}}, n_{\text{b}}, H_T) \tag{7.2}$$

where $N^{\text{control}}_{\text{obs}}$ is the total observed event yield in the control region, taken from data.

When constructing the transfer factors, the MC expectations for the following SM processes are considered: W + jets (N_{W}), $\text{t}\bar{\text{t}}$ + jets ($N_{\text{t}\bar{\text{t}}}$), $Z \to \nu\bar{\nu}$ + jets ($N_{Z\to\nu\bar{\nu}}$), DY + jets (N_{DY}), γ+jets (N_{γ}), single top + jets production via the s, t, and tW-channels (N_{top}), WW+ jets, WZ + jets, and ZZ + jets ($N_{\text{di-boson}}$), and $\text{t}\bar{\text{t}}$V or $\text{t}\bar{\text{t}}$H ($N_{\text{t}\bar{\text{t}}\text{X}}$). The transfer factor procedure uses different control regions to predict two categories of background processes, the irreducible $Z \to \nu\bar{\nu}$ process and all other remaining processes. The prediction of the $Z \to \nu\bar{\nu}$ background in a signal region category is performed with each control sample, including the μ+jets control region. Furthermore, the other remaining background processes are predicted with only the μ+jets control region.

The selection criteria for the data control samples closely resemble that of the signal region, differing mainly through the use of a lepton or photon object *tag* (that is ignored in the calculation of jet-based kinematic variables such as H_T, $\not{H}_T$, α_{T}) and minimal additional kinematic requirements to obtain W, Z, and $\mathrm{t\bar{t}}$-enriched event samples. The same selection criteria are designed to suppress signal contamination in the control samples so that unbiased data-driven estimates for the SM backgrounds in the signal region can be made.

A core principle of the transfer factor method is to account for differences in cross sections and branching ratios, acceptance and reconstruction efficiencies, or kinematic requirements between the signal and control regions. Any dependence on an event category is largely attributable to differences in acceptance due to the presence or otherwise of α_{T} or $\not{H}_T$ requirements.

Many systematic effects are expected to cancel largely in the transfer factors. However, a systematic uncertainty is assigned to each transfer factor to account for theoretical uncertainties and effects such as the mismodelling of kinematics and instrumental effects, as described in more detail in Section 7.4.

Furthermore, a fitting procedure that provides the final result is defined formally by the likelihood model described in Sec 8.1. The observation in each bin of the signal sample is modelled as Poisson-distributed about the sum of a SM expectation (and a potential signal contribution). The components of this SM expectation are related to the expected yields in the control samples via transfer factors derived from simulation. 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.

Details on the MC samples used are given in Section 5. All MC samples are normalised to the integrated luminosity of the appropriate data sample.

Incorporating substructure

As the control samples aim to provide a collection of events that are in a phase space similar to the signal region, for events of the signal region that are further categorised according to n_{subjets}, the transfer factor approach is adapted.

Regarding the two jet multiplicity scenarios, detailed in Table 6.11, in which jet substructure techniques are performed, event categories that are further classified according to n_{subjets} provide an additional term in the transfer factor. In this situation, the construction of transfer factors acquires an additional n_{subjets} dimension, as shown in Equation 7.3

$$\text{TF} = \frac{N_{\text{MC}}^{\text{signal}}(n_{\text{jet}}, n_{\text{b}}, n_{\text{subjets}}, H_T)}{N_{\text{MC}}^{\text{control}}(n_{\text{jet}}, n_{\text{b}}, n_{\text{subjets}}, H_T)} \tag{7.3}$$

As the $\mu\mu$+jets and γ+jets control regions are constructed specifically to predict the $Z \to \nu\bar{\nu}$ background, the topologies in the control regions are not sensitive to t/W tagging. As a result, only the μ+jets control region is used in the prediction across n_{subjets}. Any event category in the signal region which is not subject to further event categorisation according to n_{subjets}, utilises the μ+jets, $\mu\mu$+jets and γ+jets control regions.

For signal models predicting low jet multiplicities, the event categories in the μ+jets control region used to predict the SM backgrounds mirror those in the signal region, as illustrated in Table 7.1.

Signal region category	Predicted with		
	μ+jets	$\mu\mu$+jets	γ+jets
n_{jet}, H_T ($n_{\text{jet}} \leq 2$)	n_{jet}, H_T ($n_{\text{jet}} \leq 2$)	n_{jet}, H_T ($n_{\text{jet}} \leq 2$)	n_{jet}, H_T ($n_{\text{jet}} \leq 2$)
$n_{\text{jet}} = 3$, $H_T > 800$ GeV, n_{subjets}	$n_{\text{jet}} = 3$, $H_T > 800$ GeV, n_{subjets}	-	-
$n_{\text{jet}} = 4$, $H_T > 800$ GeV, n_{subjets}	$n_{\text{jet}} = 4$, $H_T > 800$ GeV, n_{subjets}	-	-
$n_{\text{jet}} \geq 5$, $H_T > 800$ GeV, n_{subjets}	$n_{\text{jet}} \geq 5$, $H_T > 800$ GeV, n_{subjets}	-	-

Table 7.1 The control region event categories used to predict the n_{subjets} categories ($n_{\text{jet}} \geq 3$ and $H_T > 800$ GeV) for low expected n_{jet} multiplicities and the remaining event categories, across all n_{b} multiplicities.

For signal models predicting larger jet multiplicities, the event categories $n_{\text{jet}} \geq 5$ and $H_T > 800$ GeV are further categorised according to n_{subjets}, as shown in Table 7.2. In making predictions of the EWK background in the signal region for such categories, the alternative $n_{\text{jet}} \geq 3$ and $H_T > 800$ GeV of only the μ+jets control region is used when constructing the transfer factor, as shown in Table 7.2. This constructs a region, both kinematically suited in SM background prediction for events with n_{subjets}, and statistically abundant, as illustrated in Figures 7.4 and 7.5.

7.2 Key distributions of the control regions

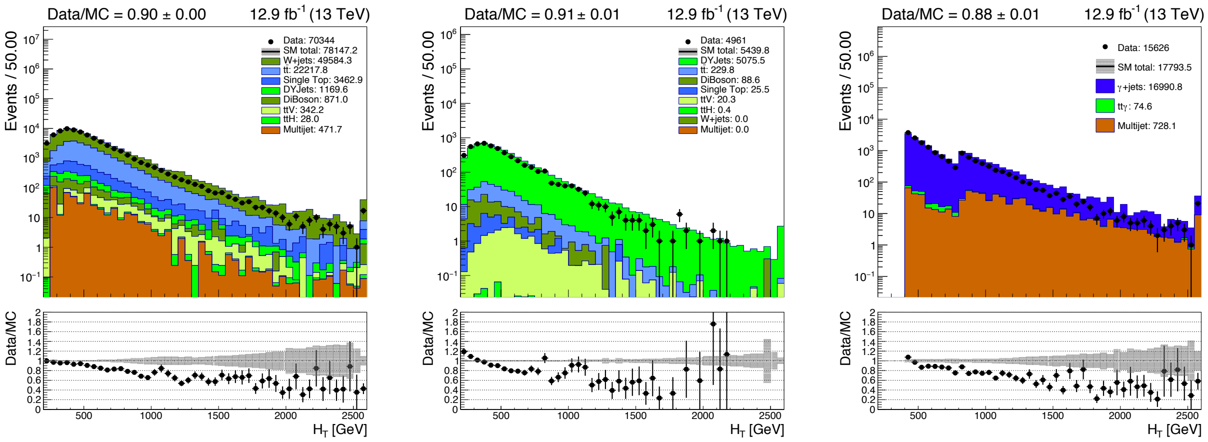
The modelling of the H_T, n_{jet} and n_{b} variables in the control regions are displayed in Figures 7.1, 7.2 and 7.3. Across each control region, a satisfactory agreement in the modelling of the key variables across the control regions is observed. The level of

Signal region category	Predicted with		
	μ+jets	$\mu\mu$+jets	γ+jets
n_{jet}, H_T ($n_{\text{jet}} < 5$)	n_{jet}, H_T ($n_{\text{jet}} < 5$)	n_{jet}, H_T ($n_{\text{jet}} < 5$)	n_{jet}, H_T ($n_{\text{jet}} < 5$)
$n_{\text{jet}} \geq 5$, $H_T > 800$ GeV, n_{subjets}	$n_{\text{jet}} \geq 3$, $H_T > 800$ GeV, n_{subjets}	-	-

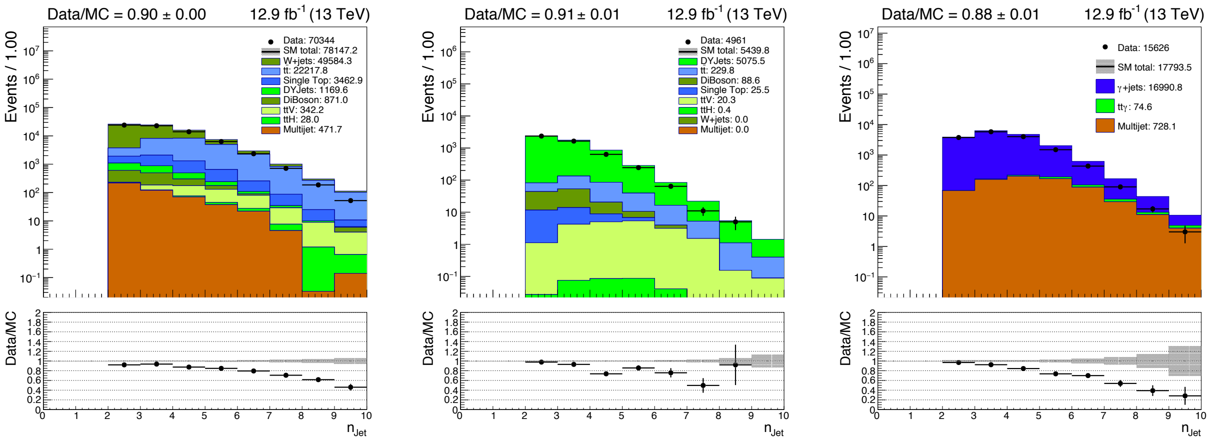
Table 7.2 The control region event categories used to predict the n_{subjets} categories ($n_{\text{jet}} \geq 5$ and $H_T > 800$ GeV) for large expected n_{jet} multiplicities and the remaining event categories, across all n_{b} multiplicities.

agreement is not a critical requirement in the analysis as any observed mismodelling is corrected through the transfer factors, with any potential discrepancy observed in the control region used in the correction of the modelling in the signal region.

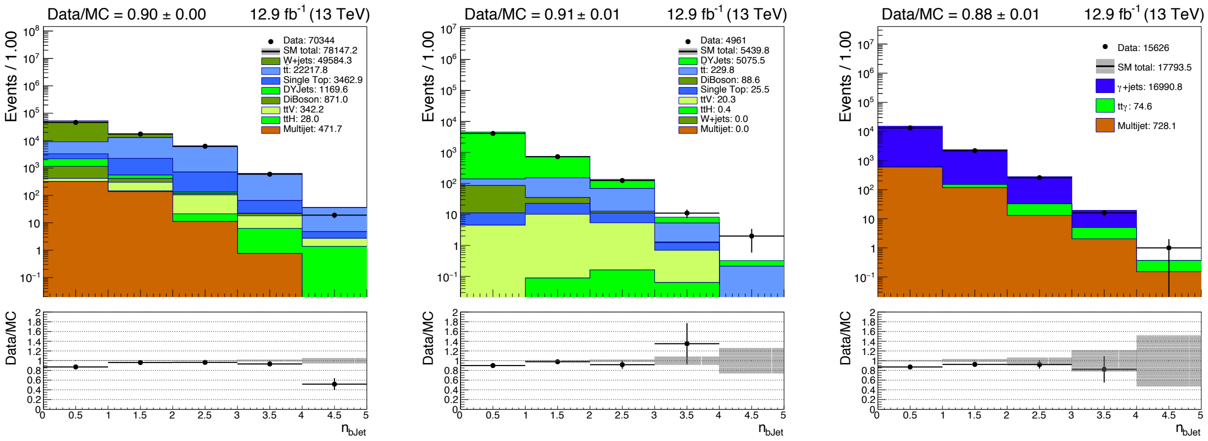
The purpose of the control regions are to predict the EWK backgrounds in the signal region by connecting each category of a control region to the corresponding category of the signal region by means of a transfer factor. The final background estimation, from which the results are interpreted are determined with a fitting procedure, described formally as the likelihood model, and discussed in more detail in Section 8.1.



(a) μ+jets, H_T (b) $\mu\mu$+jets, H_T (c) γ+jets, H_T

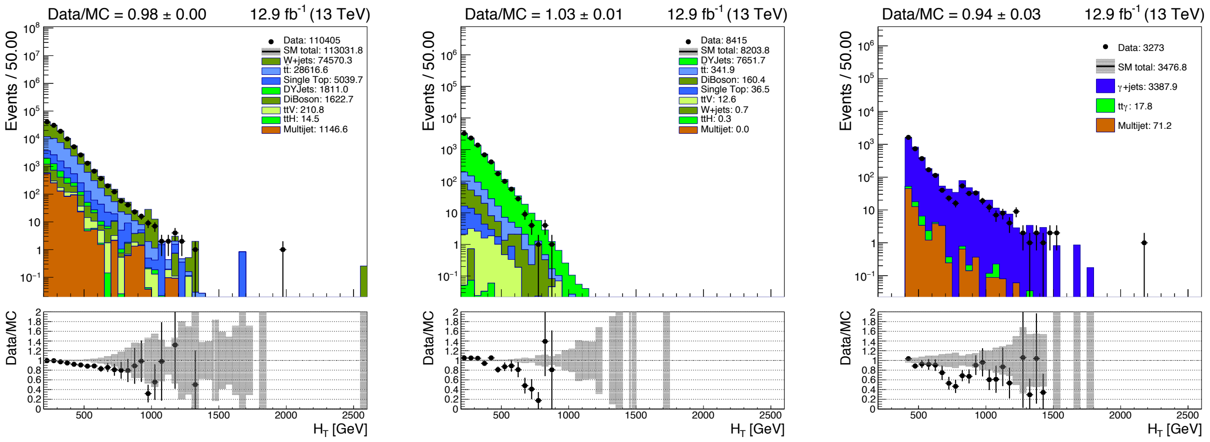


(d) μ+jets, n_{jet} (e) $\mu\mu$+jets, n_{jet} (f) γ+jets, n_{jet}

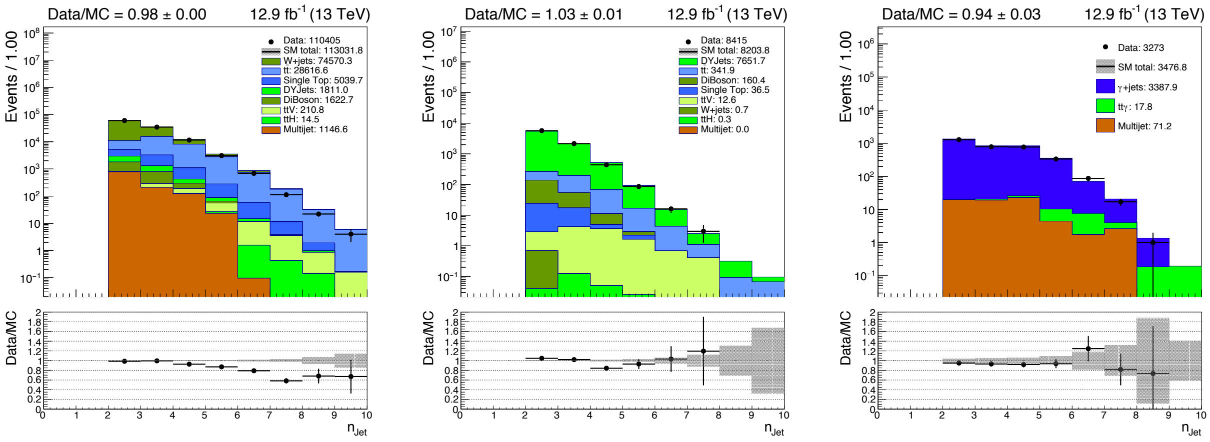


(g) μ+jets, n_{b} (h) $\mu\mu$+jets, n_{b} (i) γ+jets, n_{b}

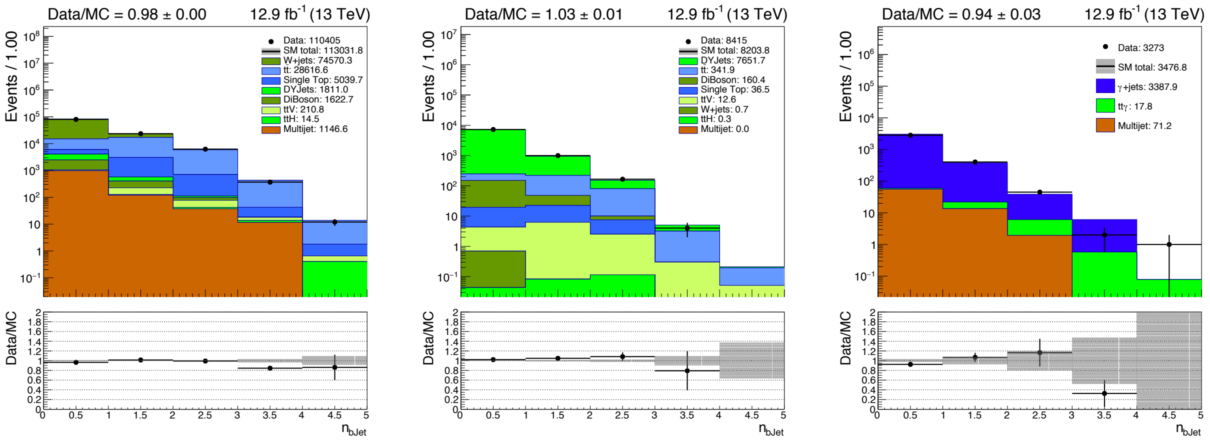
Fig. 7.1 Distributions of the key analysis variables, H_T, n_{jet} and n_{b}, measured in data (markers) and simulation (coloured histograms), for the symmetric categories in the μ+jets, $\mu\mu$+jets and γ+jets control regions. The observed slopes in the H_T distribution for the γ+jets control region is a result of the α_{T} requirements that are applied, as discussed in Section 6.8.3.



(a) μ+jets, H_T (b) $\mu\mu$+jets, H_T (c) γ+jets, H_T

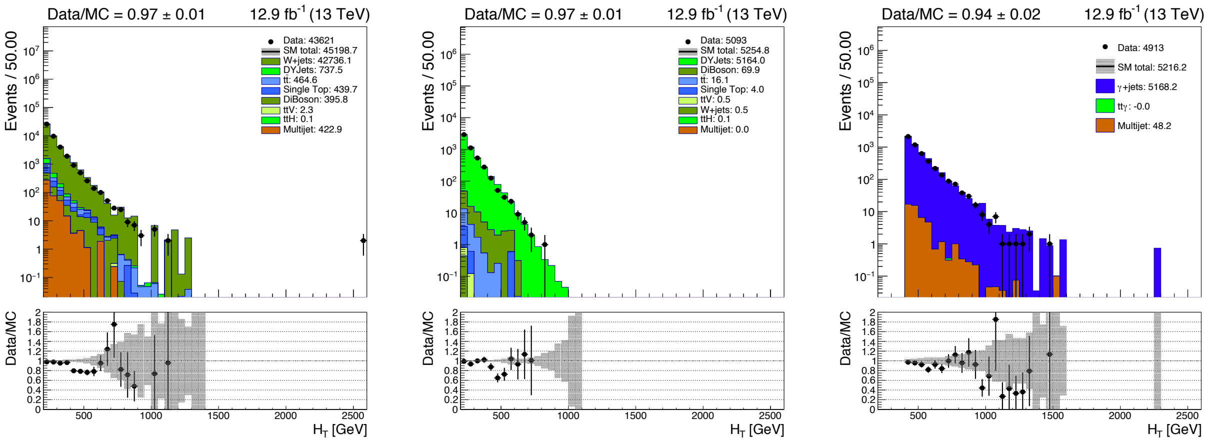


(d) μ+jets, n_{jet} (e) $\mu\mu$+jets, n_{jet} (f) γ+jets, n_{jet}



(g) μ+jets, n_{b} (h) $\mu\mu$+jets, n_{b} (i) γ+jets, n_{b}

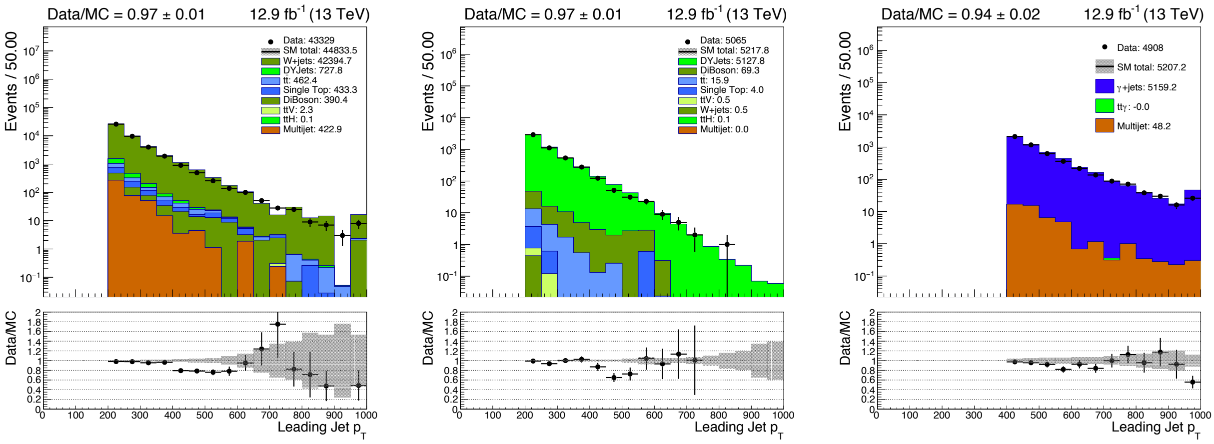
Fig. 7.2 Distributions of the key analysis variables, H_T, n_{jet} and n_{b}, measured in data (markers) and simulation (coloured histograms), for the asymmetric categories in the μ+jets, $\mu\mu$+jets and γ+jets control regions.



(a) μ+jets, H_T

(b) $\mu\mu$+jets, H_T

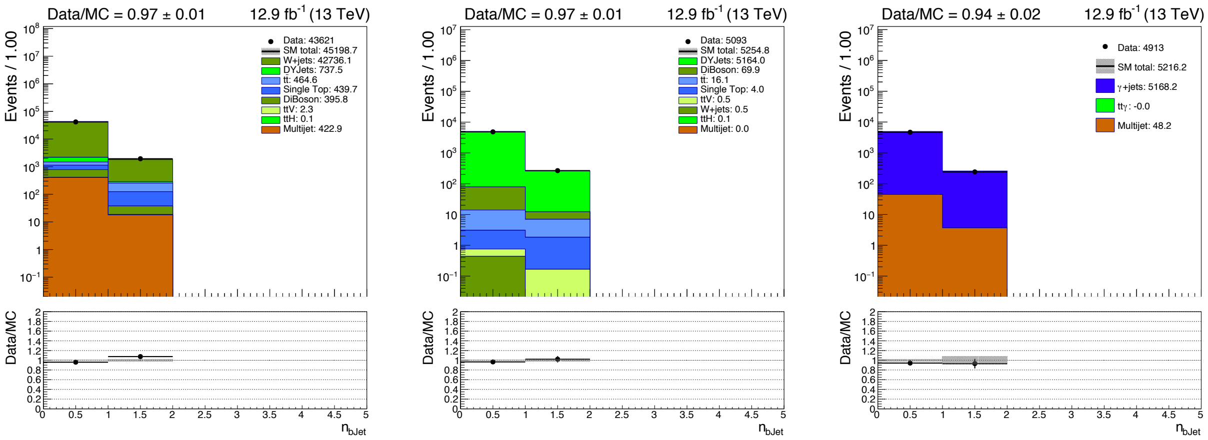
(c) γ+jets, H_T



(d) μ+jets, Leading jet p_{T}

(e) $\mu\mu$+jets, Leading jet p_{T}

(f) γ+jets, Leading jet p_{T}



(g) μ+jets, n_{b}

(h) $\mu\mu$+jets, n_{b}

(i) γ+jets, n_{b}

Fig. 7.3 Distributions of the key analysis variables, H_T, leading jet p_{T} and n_{b}, measured in data (markers) and simulation (coloured histograms), for the monojet categories in the μ+jets, $\mu\mu$+jets and γ+jets control regions.

7.2.1 Key distributions of substructure variables

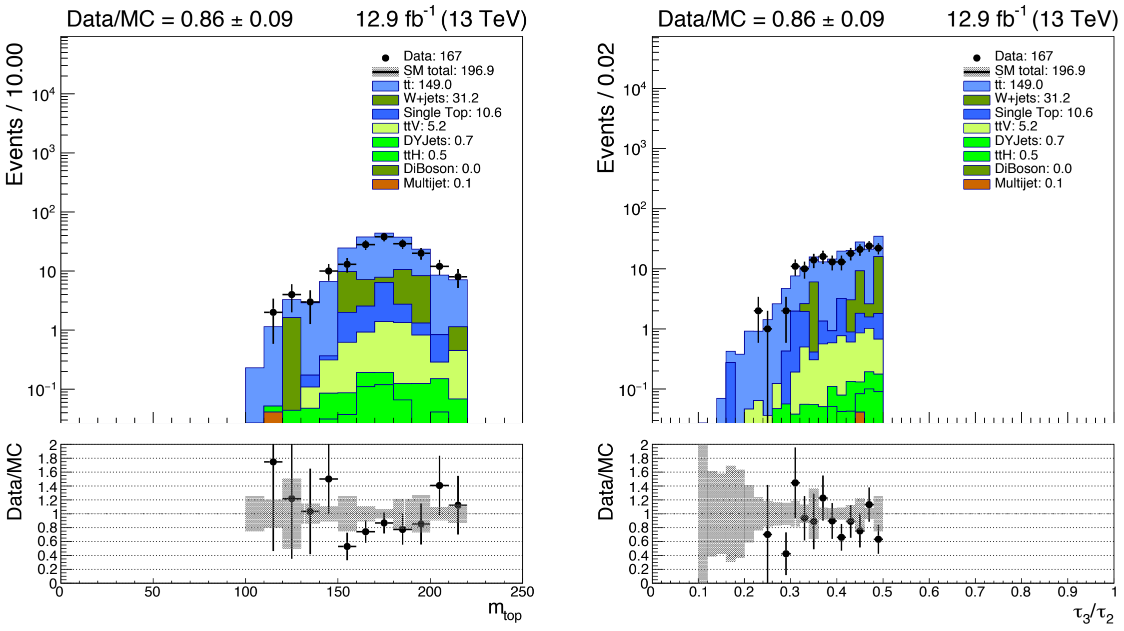
Specific to t-tagging, the softdrop mass, the N-subjettiness ratio τ_3/τ_2 and the subjet b-tag discriminate are used, and the distributions of which in the μ+jets region are shown in Figure 7.4.

Regarding W-tagging, whereby the pruning algorithm and the N-subjettiness ratio τ_2/τ_1 is applied, the distributions and modelling between data and simulation are shown in Figure 7.5.

The ratio distribution on the lower half of each distribution shows the agreement between data and simulation.

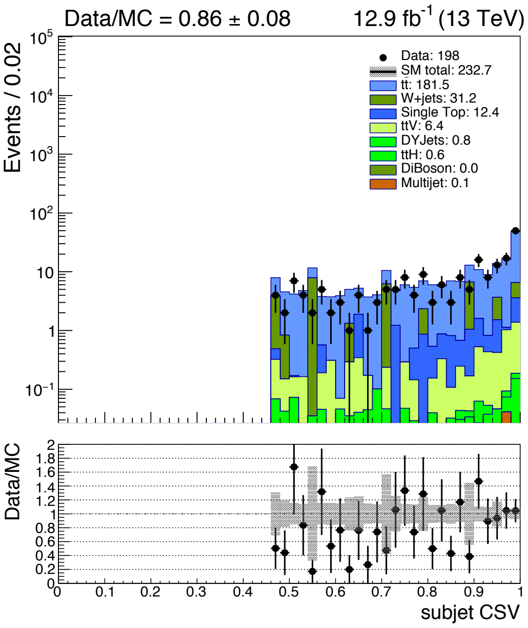
In each distribution, a requirement of $n_{\text{jet}} \geq 3$ and $H_T > 800$ GeV is applied, analogous to the construction of the μ+jets control region used in predicting the n_{subjets} event categories of the signal region for signal models with a large expected n_{jet} multiplicity, as detailed in Table 6.11.

Any disagreement between data and simulation in the control region is not a critical weakness to the analysis, as any observed discrepancies are used in the correction in the modelling in the signal region. Specifically, the final background predictions from which the results are interpreted are determined via a likelihood fit, discussed in more detail in Section 8.1. In addition, in using the transfer factors to construct ratios in simulation, any mismodelling in simulation is expected to cancel for all control and signal regions.



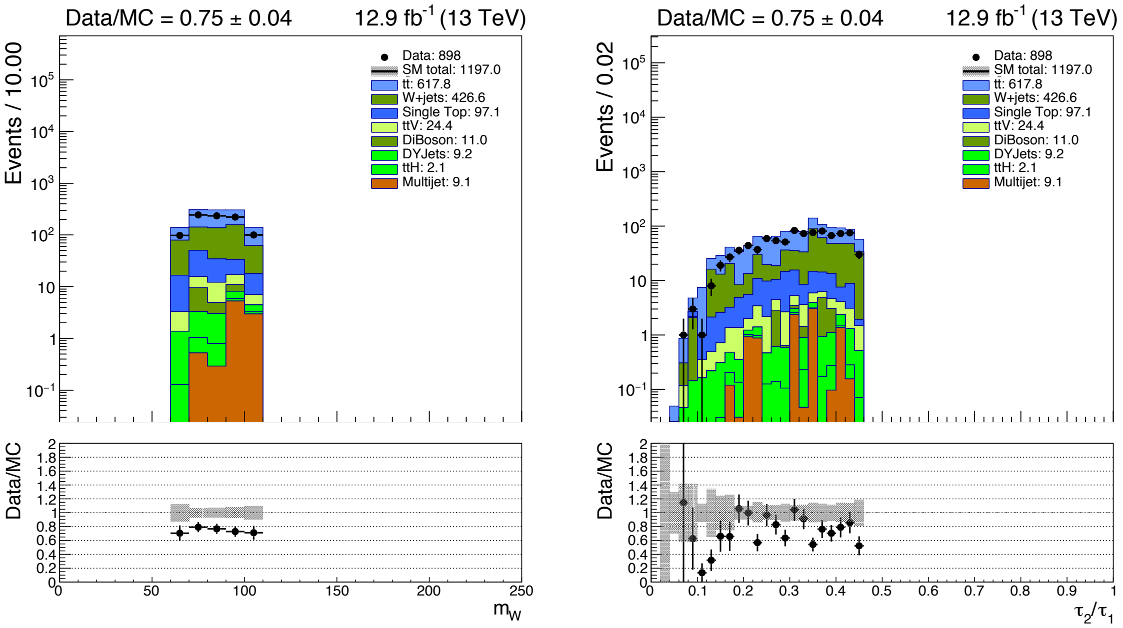
(a) μ+jets, $\mathrm{m}_{softdrop}$

(b) μ+jets, τ_3/τ_2



(c) μ+jets, subjet b-tag

Fig. 7.4 Distributions of the key variables used in t-tagging, in the μ+jets control region, with the selection $n_{\mathrm{jet}} \geq 3$ and $H_T > 800$ GeV.



(a) μ+jets, m$_{pruned}$

(b) μ+jets, τ_2/τ_1

Fig. 7.5 Distributions of the key variables used in W-tagging, in the μ+jets control region, with the selection $n_{\text{jet}} \geq 3$ and $H_T > 800$ GeV.

7.3 Estimating the QCD multijet background

One of the core strategies of the analysis is in the suppression and control of the dominant QCD multijet background. Through the collection of dedicated variables, discussed in Section 6, designed to reduce the QCD multijet background and events with a fake source of $\not{E}_T$, the levels of QCD multijet events are at a negligible level.

The precise estimation of the QCD multijet background is a particularly challenging procedure, with effects from theoretical uncertainties and statistical limitations in simulated samples. Furthermore, the ability to accurately model and quantify detector and misreconstruction effects, a means by which QCD multijet events enter the signal region, is difficult. As a result, any use of simulated events to predict the level of QCD multijet background is susceptible to large uncertainties.

Therefore, to better control the accuracy and uncertainty of the QCD multijet prediction, a dedicated data-driven background estimation method is employed. The method constructs three sidebands to the signal region that are formed with data.

7.3.1 QCD-enriched sidebands

The procedure begins by determining in simulation, the pass and fail ratio of the $\not{H}_T/\not{E}_T$ sideband, R, as shown in Equation 7.4.

$$R_{QCD}(H_T, n_{\text{jet}}) = \frac{N^{QCD}_{Pass}(H_T, n_{\text{jet}})}{N^{QCD}_{Fail}(H_T, n_{\text{jet}})} \tag{7.4}$$

where N^{QCD}_{Pass} and N^{QCD}_{Fail} represent the number of QCD multijet events that pass and fail the $\not{H}_T/\not{E}_T < 1.25$ selection respectively. The determination of the variable R is computed per H_T and n_{jet} category, inclusive in the n_{b} and $\not{H}_T$ dimensions, and is equivalent to a transfer factor.

Each ratio, R, is then used as a multiplier on the predicted QCD multijet event counts per H_T and n_{jet} category, in a $\not{H}_T/\not{E}_T$ data sideband. A data-driven estimate of the $\not{H}_T/\not{E}_T$ sideband, denoted as B in Table 7.3, is obtained with a maximum likelihood fit. In the likelihood fit, the EWK backgrounds are determined using the transfer factor method of Section 7.1.1, with the same selection as the control regions. However, the $\not{H}_T/\not{E}_T$ requirement is inverted.

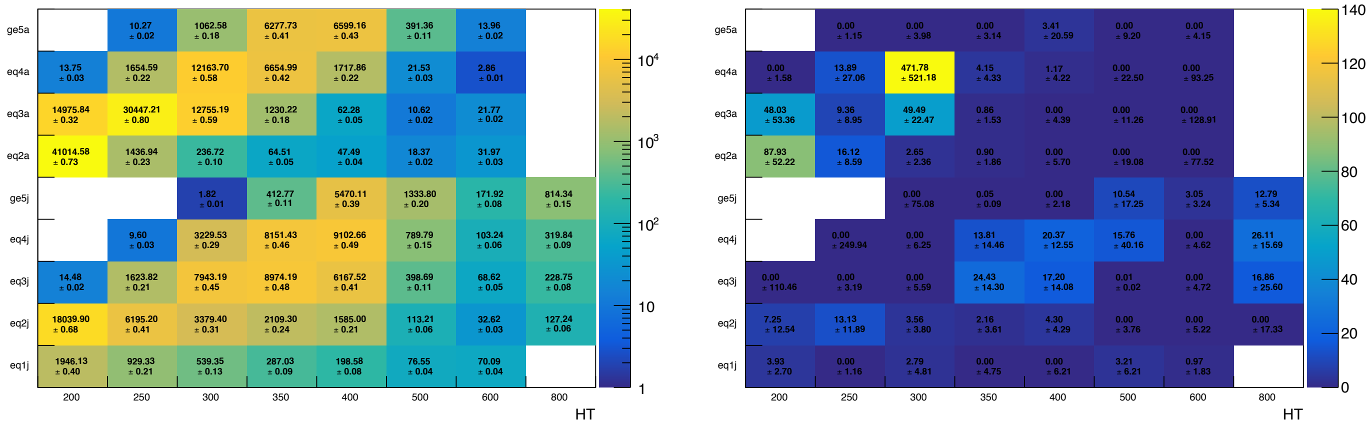
$$N^{QCD}_{Pass}(H_T, n_{\text{jet}}) = R_{QCD}(H_T, n_{\text{jet}}) \cdot N^{QCD}_{Fail}(H_T, n_{\text{jet}}) \tag{7.5}$$

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

Table 7.3 Definition of sidebands used in the determination of the QCD background contributions in the signal region.

After the contribution of the electroweak background in the $\not{H}_T/\not{E}_T$ sideband is estimated, the remaining data counts are attributed to QCD multijet events. Therefore, the product of the data-driven estimate of QCD multijet counts in the $\not{H}_T/\not{E}_T$ sideband, B, and the ratio R provides an estimate of the level of QCD multijet events in each H_T and n_{jet} category of the signal region, as by the relation in Equation 7.5. The results of this prediction are shown in Figure 7.6b, with the ratio between the multijet and non-multijet background predictions, predicted with the procedure defined in Section 7.1.1, shown in Figure 7.6c.

The predictions summarised in Figure 7.6c show that the α_{T} and $\Delta\phi^*_{min}$ requirements in the signal region suppress the QCD multijet contamination in all event cat-

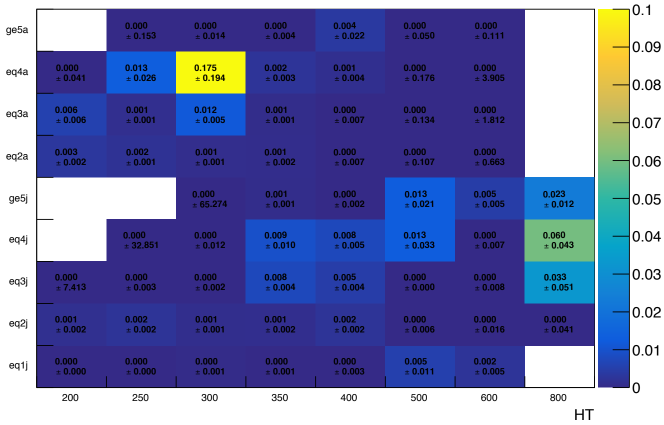


	200	250	300	350	400	500	600	800
ge5a		10.27 ± 0.02	1062.58 ± 0.18	6277.73 ± 0.41	6599.16 ± 0.43	391.36 ± 0.11	13.96 ± 0.02	
eq4a	13.75 ± 0.03	1654.59 ± 0.22	12163.70 ± 0.58	6654.99 ± 0.42	1717.86 ± 0.22	21.53 ± 0.03	2.86 ± 0.01	
eq3a	14975.84 ± 0.32	30447.21 ± 0.80	12755.19 ± 0.59	1230.22 ± 0.18	62.28 ± 0.05	10.62 ± 0.02	21.77 ± 0.02	
eq2a	41014.58 ± 0.73	1436.94 ± 0.23	236.72 ± 0.10	64.51 ± 0.05	47.49 ± 0.04	18.37 ± 0.02	31.97 ± 0.03	
ge5j			1.82 ± 0.01	412.77 ± 0.11	5470.11 ± 0.39	1333.80 ± 0.20	171.92 ± 0.08	814.34 ± 0.15
eq4j		9.60 ± 0.03	3229.53 ± 0.29	8151.43 ± 0.46	9102.66 ± 0.49	789.79 ± 0.15	103.24 ± 0.06	319.84 ± 0.09
eq3j	14.48 ± 0.02	1623.82 ± 0.21	7943.19 ± 0.45	8974.19 ± 0.48	6167.52 ± 0.41	398.69 ± 0.11	68.62 ± 0.05	228.75 ± 0.08
eq2j	18039.90 ± 0.68	6195.20 ± 0.41	3379.40 ± 0.31	2109.30 ± 0.24	1585.00 ± 0.21	113.21 ± 0.06	32.62 ± 0.03	127.24 ± 0.06
eq1j	1946.13 ± 0.40	929.33 ± 0.21	539.35 ± 0.13	287.03 ± 0.09	198.58 ± 0.08	76.55 ± 0.04	70.09 ± 0.04	

	200	250	300	350	400	500	600	800
ge5a		0.00 ± 1.15	0.00 ± 3.98	0.00 ± 3.14	3.41 ± 20.59	0.00 ± 9.20	0.00 ± 4.15	
eq4a	0.00 ± 1.58	13.89 ± 27.06	471.78 ± 521.18	4.15 ± 4.33	1.17 ± 4.22	0.00 ± 22.50	0.00 ± 93.25	
eq3a	48.03 ± 53.36	9.36 ± 8.95	49.49 ± 22.47	0.86 ± 1.53	0.00 ± 4.39	0.00 ± 11.26	0.00 ± 128.91	
eq2a	87.93 ± 52.22	16.12 ± 8.59	2.65 ± 2.36	0.90 ± 1.86	0.00 ± 5.70	0.00 ± 19.08	0.00 ± 77.52	
ge5j			0.00 ± 75.08	0.05 ± 0.09	0.00 ± 2.18	10.54 ± 17.25	3.05 ± 3.24	12.79 ± 5.34
eq4j		0.00 ± 249.94	0.00 ± 6.25	13.81 ± 14.46	20.37 ± 12.55	15.76 ± 40.16	0.00 ± 4.62	26.11 ± 15.69
eq3j	0.00 ± 110.46	0.00 ± 3.19	0.00 ± 5.59	24.43 ± 14.30	17.20 ± 14.08	0.01 ± 0.02	0.00 ± 4.72	16.86 ± 25.60
eq2j	7.25 ± 12.54	13.13 ± 11.89	3.56 ± 3.80	2.16 ± 3.61	4.30 ± 4.29	0.00 ± 3.76	0.00 ± 5.22	0.00 ± 17.33
eq1j	3.93 ± 2.70	0.00 ± 1.16	2.79 ± 4.81	0.00 ± 4.75	0.00 ± 6.21	3.21 ± 6.21	0.97 ± 1.83	

(a) Predicted QCD counts in $\not{H}_T/\not{E}_T$ sideband, B

(b) QCD multijet predictions in the signal region



	200	250	300	350	400	500	600	800
ge5a		0.000 ± 0.153	0.000 ± 0.014	0.000 ± 0.004	0.004 ± 0.022	0.000 ± 0.050	0.000 ± 0.111	
eq4a	0.000 ± 0.041	0.013 ± 0.026	0.175 ± 0.194	0.002 ± 0.003	0.001 ± 0.004	0.000 ± 0.176	0.000 ± 3.905	
eq3a	0.006 ± 0.006	0.001 ± 0.001	0.012 ± 0.005	0.001 ± 0.001	0.000 ± 0.007	0.000 ± 0.134	0.000 ± 1.812	
eq2a	0.003 ± 0.002	0.002 ± 0.001	0.001 ± 0.001	0.001 ± 0.002	0.000 ± 0.007	0.000 ± 0.107	0.000 ± 0.663	
ge5j			0.000 ± 65.274	0.001 ± 0.001	0.000 ± 0.002	0.013 ± 0.021	0.005 ± 0.005	0.023 ± 0.012
eq4j		0.000 ± 32.851	0.000 ± 0.012	0.009 ± 0.010	0.008 ± 0.005	0.013 ± 0.033	0.000 ± 0.007	0.060 ± 0.043
eq3j	0.000 ± 7.413	0.000 ± 0.003	0.000 ± 0.002	0.008 ± 0.004	0.005 ± 0.004	0.000 ± 0.000	0.000 ± 0.008	0.033 ± 0.051
eq2j	0.001 ± 0.002	0.002 ± 0.002	0.001 ± 0.001	0.001 ± 0.002	0.002 ± 0.002	0.000 ± 0.006	0.000 ± 0.016	0.000 ± 0.041
eq1j	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.001	0.000 ± 0.001	0.000 ± 0.003	0.005 ± 0.011	0.002 ± 0.005	

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

Fig. 7.6 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 Appendix A. The result of multiplying the observed QCD multijet events predicted in (a) by the translation factor from the sideband to the signal region determined with simulation is shown in (b). The ratio of expected multijet background events in the signal region divided by non-multijet backgrounds is shown in (c).

egories to percent-level or smaller, with respect to the non-multijet background. The resultant prediction of the QCD multijet background is included as a background contribution to the likelihood model described in Section 8.1.

Due to limitations in simulated events for QCD multijet production, the prediction of the QCD multijet events in the signal region using the full collection of event categories is not possible. Therefore, the calculation is performed using only the n_{jet} and H_T event categories in which an adequate amount of simulated events allow a valid prediction. Furthermore, the QCD shapes for the n_{b}, $\not{H}_T$ and n_{subjets} event categories are taken from the EWK shapes in simulation and normalised to the QCD multijet counts.

7.3.2 Validation of the method

Although the method of predicting the QCD multijet background in the signal region is predominantly data-driven, the ratio R is constructed with simulation. To validate the ratio, a QCD sideband in data is constructed by requiring $\Delta\phi^*_{min} < 0.5$. In this sideband a data-driven estimation of the QCD counts is carried out in two regions, those with $\not{H}_T/\not{E}_T < 1.25$ and those with values $\not{H}_T/\not{E}_T > 1.25$.

From which, a maximum likelihood fit, analogous to that described in Section 7.3 is performed. The fit utilises the control regions to estimate the EWK backgrounds, and then removes their contribution from the sideband to assume that the remaining data counts in the sideband originate from QCD multijet processes. Under this construction, it is possible to construct a data-driven ratio of QCD multijet counts with $\not{H}_T/\not{E}_T < 1.25$ and those with $\not{H}_T/\not{E}_T > 1.25$, denoted $R^{data}_{\Delta\phi^*_{min}<0.5}$. Furthermore, the same ratio can be constructed using simulation, $R^{simulation}_{\Delta\phi^*_{min}<0.5}$, and compared to $R^{data}_{\Delta\phi^*_{min}<0.5}$.

The modelling of the ratio, R, is validated by comparing the agreement between the ratio computed from electroweak-corrected measurements in data, $R^{data}_{\Delta\phi^*_{min}<0.5}$ to the ratio derived from simulation, $R^{simulation}_{\Delta\phi^*_{min}<0.5}$. The distribution of the double ratio, $R^{data}_{\Delta\phi^*_{min}<0.5}$ / $R^{simulation}_{\Delta\phi^*_{min}<0.5}$, is shown in Figure 7.7, in which categories with statistics in data or simulation to make the calculation are omitted. To account for any disagreement between data and simulation, a fully correlated systematic uncertainty of 100% is taken.

The distribution in Figure 7.7 shows the modelling of the two ratios, in which any bin with insufficient statistics is omitted. Any disagreement between simulation and data is accounted by a fully correlated systematic uncertainty of 100%, taken on the

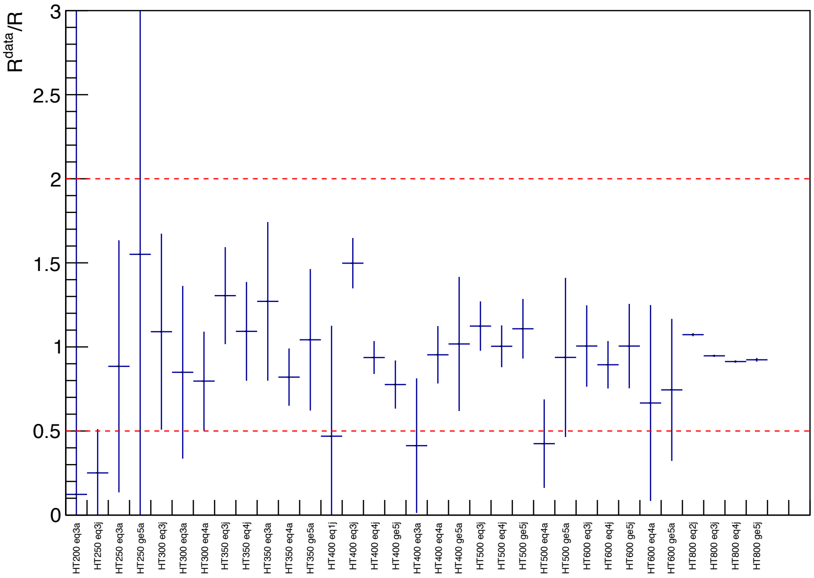


Fig. 7.7 The ratio of the measurement of R, the pass and fail ratio for the $\not{H}_T/\not{E}_T$ selection, from data and simulation in the $\Delta\phi^*_{min} < 0.5$ sideband. The ratio and distribution is made per H_T and n_{jet} event category. For visualisation purposes, the dotted red lines indicate a 100% up and 50% down uncertainty to demonstrate the control of any disagreement.

predicted QCD multijet contamination in the signal region.

7.4 Systematic uncertainties

Following the construction of the data-driven techniques used for estimating the SM backgrounds in the signal region, different sources of systematic uncertainties need to be considered when predicting event yields. Such uncertainties arise in the theoretical prediction and modelling of the physics processes that are being predicted.

The magnitudes of these systematic uncertainties are determined with two methods, a simulation-driven method and a data-driven *closure test* method. Together, these methods provide an estimate of the systematic uncertainty on the transfer factors by probing the levels at which the transfer factors are sensitive to relevant uncertainties.

The simulation driven approach has the advantage of independently probing the effect of known and simulated sources of systematic uncertainty. However, to account for unknown sources of systematics the closure test method is more advantageous, with no reliance on simulation and any associated potential mismodelling.

A summary of all the systematic uncertainties on the transfer factors considered in the analysis, with representative ranges and correlation model is shown in Table 7.6. Furthermore, additional sources of systematic uncertainties for the prediction and correction of the yield from signal models is incorporated in the likelihood model, and is discussed in more detail in Section 8.1.1.

7.4.1 Simulation driven uncertainties

To account for potential differences in the theoretical prediction and modelling of physics processes, several corrections are applied in the analysis, each of which presents an associated uncertainty.

The simulation-driven method estimates the systematic uncertainty in transfer factors by measuring the relative change in the event yields, by varying a correction upwards and downwards within its 1σ uncertainty. The determined uncertainties are then propagated to each of the transfer factors.

For events that are further classified according to n_{subjets}, the same method is applied and the relative change of their respective transfer factors under variations is examined. The effect of the variations on the event categories that are not subject to t/W tagging are shown in Appendix C.

t/W tagging efficiency

The scale factors for top quark and W boson tagging, that are independently determined for the working point of Tables 6.4 and 6.5, have associated uncertainties. For t-tagging, the scale factors are first determined from a pure sample of boosted top quark jets in data. This is achieved by selecting semi-leptonic $\mathrm{t\bar{t}}$ events with requirements on a single muon and AK4 jet in the same hemisphere. Following which, an AK8 jet with $p_{\mathrm{T}} > 400$ GeV and $|\eta| < 2.4$ is required in the opposite hemisphere to the muon [104]. This jet, represented as a probe jet, is subject to the conditions of t-tagging and the efficiency in t-tagging in data and in simulation is then determined.

The scale factors, determined as the ratio of the efficiency in data and the efficiency in simulation, are evaluated both inclusively and as a function of the p_{T} of the AK8 jet. It should be noted that the scale factors associated to the working point of t-tagging are determined with the application of subjet b-tagging. Therefore, the individual scale factors determined in subjet b-tagging is not applied.

In a similar approach, the scale factors for W-tagging are obtained from a sample requiring one energetic lepton and an AK8 jet in the final state, where the AK8 is aimed at containing the two quarks from a boosted W boson. Further, to create a $\mathrm{t\bar{t}}$ enriched region, the presence of at least one b-tagged AK4 jet with $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.4$ is required [20]. As both events in the signal and μ+jets control region are categorised according to n_{subjets}, the effect of the uncertainties of the t/W tagging scale factors is expected to be low. However, as the scale factors aim to account for the differences in t/W tagging efficiency between data and simulation, any uncertainty associated to the scale factors must be taken into account.

To determine the effect of the scale factor uncertainties, the correction applied to the event weight, that is determined through a similar manner as described in Section 5.4, is recomputed. The correction is varied to $\pm\ 1\sigma$ from its central value and the prediction from simulation is re-evaluated.

For low expected n_{jet} multiplicities in the signal region, the effect on the transfer factors through variations of the t/W tagging scale factors by their uncertainty, in the μ+jets control regions, is shown in Table 7.4. The tables show the relative change in the transfer factors for the μ+jets control region when used to predict the $\mathrm{t\bar{t}} + \mathrm{W}$ backgrounds in the signal region.

For signal models with a higher expected n_{jet} multiplicity, where the construction

n_b	n_jet	n_subjets	Variation	
			Up	Down
0	3	0	0.98	1.01
	4	0	0.98	1.00
	≥ 5	0	0.99	1.02
		2	0.99	1.00
1	3	0	0.0	0.0
	4	0	0.98	1.02
		3	1.18	1.19
	≥ 5	0	0.99	0.98
		2	1.03	1.03
		3	1.03	1.04
2	3	0	0.0	0.0
	4	0	1.17	1.19
	≥ 5	0	0.99	0.99
		2	1.07	1.05
		3	1.08	1.07
		4	1.16	1.14
≥ 3	3	0	0.0	0.0
	≥ 5	0	1.08	1.07

n_b	n_jet	n_subjets	Variation	
			Up	Down
0	3	0	1.05	0.95
	4	0	1.04	0.95
	≥ 5	0	1.05	0.96
		2	1.04	0.95
1	3	0	0.0	0.0
	4	0	1.05	0.96
		3	1.23	1.13
	≥ 5	0	1.03	0.95
		2	1.09	0.98
		3	1.08	0.99
2	3	0	0.0	0.0
	4	0	1.23	1.13
	≥ 5	0	1.04	0.95
		2	1.09	1.03
		3	1.12	1.03
		4	1.17	1.13
≥ 3	3	0	0.0	0.0
	≥ 5	0	1.12	1.04

Table 7.4 The relative change in the μ+jets $\rightarrow \mathrm{t\bar{t}} + \mathrm{W}$ transfer factors, in populated n_subjets event categories, when varying the t-tagging scale factors (left) and W-tagging scale factors (right) within its uncertainties, in $H_T > 800$ GeV category. Variations according to the $+\sigma$ (left) and $-\sigma$ (right) are shown.

of the transfer factor method is altered to use a loosened $n_\mathrm{jet} \geq 3$ category, the μ+jets region is investigated under variations in both the t and W-tagging efficiencies, as shown in Table 7.5.

n_b	n_jet	n_subjets	Variation	
			Up	Down
0	≥ 5	0	1.01	1.01
		2	1.00	1.01
1	≥ 5	0	0.99	0.97
		2	1.03	1.03
		3	1.04	1.04
2	≥ 5	0	0.99	0.98
		2	1.06	1.05
		3	1.07	1.06
		4	1.14	1.13
≥ 3	≥ 5	0	1.07	1.07

n_b	n_jet	n_subjets	Variation	
			Up	Down
0	≥ 5	0	1.05	0.99
		2	1.04	1.03
1	≥ 5	0	1.04	1.01
		2	1.03	1.05
		3	1.08	1.04
2	≥ 5	0	1.03	1.01
		2	1.13	1.08
		3	1.14	1.03
		4	1.16	1.13
≥ 3	≥ 5	0	1.07	1.03

Table 7.5 The relative change in the μ+jets $\rightarrow \mathrm{t\bar{t}} + \mathrm{W}$ transfer factors, in populated n_subjets event categories, when varying the t-tagging scale factors (left) and W-tagging scale factors (right) within its uncertainties, in $H_T > 800$ GeV category. Variations according to the $+\sigma$ (left) and $-\sigma$ (right) are shown.

The effect of varying the t/W scale factors on the transfer factors, that are constructed with the μ+jets control region to predict ttW in a signal region with large n_{jet} multiplicities for signal models, are in the range 1 - 20 %. Similarly, for signal models with an expected low n_{jet} multiplicity, the uncertainty on the transfer factor is measured to be typically 1 - 20 %.

Jet energy scale

The uncertainty associated to the jet energy corrections, detailed in Section 4, is used to vary the corrections and probe any effect on the transfer factors. However, any effect due to the jet energy corrections is expected to be mitigated as the construction of the transfer factor, between regions with similar binning, will result in the effects to cancelled.

Furthermore, due to threshold effects, some events have the potential to migrate between different event categories, and thus alter the selection efficiency. These effects are measured in the μ+jets, $\mu\mu$+jets and γ+jets control regions for each event category, including those that are further categorised according to n_{subjets}, providing an estimate of a 1 - 15 % systematic uncertainty in the magnitude of the transfer factor.

b-tagging efficiency

As described in Section 5.4 scale factors are applied on simulated events to correct for any differences in the efficiency, between simulated events and data, in tagging jets arising from bottom-quarks. To determine the effect the uncertainty of the scale factors has on the transfer factors, and thus any prediction in the signal region, the scale factors are varied within their estimated uncertainty. Thereafter, the relative change in the transfer factors is presented as a function of H_T and jet category. The uncertainty on the transfer factor through these variations is found to be in the range 1 - 5%. It should be noted that scale factors associated with the identification of b and c jets are varied together, while the scale factors associated with light jets are varied separately.

Photon trigger uncertainty

The effect of varying the photon trigger uncertainties is investigated on the transfer factors. To make a conservative estimate, the magnitude of the systematic uncertainty is taken as the size of the trigger inefficiency. The relative change in the γ+jets transfer factor is in the range 0 - 3%.

Signal trigger uncertainty

To correct for potential inefficiencies in the trigger selection, the efficiency of the signal triggers are determined using electron and muon reference triggers. Event categories that are relevant to the trigger inefficiencies are the low H_T and $\not{H}_T$ event categories of the analysis, in which the triggers are not fully efficient. To determine the systematic uncertainty associated to the trigger inefficiencies, the difference in the measured efficiency between the electron and muon reference triggers is computed. Upon varying the systematic uncertainty, the relative change on the transfer factor is measured to be in the range 0 - 5%.

Top p_{T} reweighting

The p_{T} spectrum of top quarks exhibits a different behaviour between simulated events and $\mathrm{t\bar{t}}$ enriched regions in data. To account for such differences, a scale factor is derived as a function of the p_{T} of the top quark at generator level of the simulation [105] [106], and a correction factor is determined. To account for the uncertainty on the top p_{T} reweighting, a systematic uncertainty on this correction is taken as the size of the correction itself. Thereafter, the relative change in the transfer factors upon varying the uncertainty is measured in the range 1 - 30%.

Lepton and photon efficiencies

The process in which events from $\mathrm{t\bar{t}}$ and W processes can enter the signal region is through their leptonic decays, with the lepton failing identification, reconstruction or falling out of acceptance. As a result, there are associated uncertainties with these acceptance and reconstruction effects that can impact the determination of the transfer factors for the $\mathrm{t\bar{t}}$ and W background.

The variation in the lost lepton background is measured by using a simulated sample with no lepton veto. With the sample, the variation in the probability to reconstruct a lepton in the signal region, and the variation for reconstructing a muon for the μ+jets control region is determined with appropriate correlation. The systematic uncertainties on the trigger, lepton identification and isolation efficiencies are varied separately by their ±1σ correction. 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.

Furthermore, uncertainties in the scale factors associated to leptons can effect the event yields in the muon control regions. The change in the transfer factors upon varying the scale factors of the muons, used to define the μ+jets and $\mu\mu$+jets control regions are in the range of 0 - 3%.

PU reweighting

The reweighting of the simulated PU distribution to match that observed in data is dependant on the minimum bias cross section, a parameter that has an associated uncertainty of $\sim$ 5%. This uncertainty is used to calculate the scale factors for PU reweighting, which are varied to determine a relative change in the transfer factors of the order of 1 - 5%.

The other known sources of systematic uncertainty are subject to the same procedure across the analysis event categories, with their relative uncertainty on the transfer factors summarised in Table 7.6. In addition, when interpreting signal models, the simulated events for the signal contribution are corrected to account for systematic uncertainties, the details of which are discussed in Section 8.4.1.

7.4.2 Data driven uncertainties

To account for sources of systematic uncertainty that are not derived from the variations of scale factors in simulation, an additional method of estimating systematic uncertainties is applied, the core principle of which relies on closure tests.

A closure test determines an experimental uncertainty by measuring the agreement between a predicted yield determined by a transfer factor, N_{pred}, and an observed yield in a data control region N_{obs}.

The method relies on constructing a transfer factor to extrapolate from one sub-region of a particular control region into another control region sub-region. Mathematically, it is equivalent to Equation 7.2, whereby a prediction is made from one control region to another, and from which any potential sources of bias in the transfer factors can be determined.

The compatibility of a given test is expressed as a ratio, with $(N_{obs} - N_{pred})/N_{pred}$, where N_{pred} are the number of predicted events in the sample and N_{obs} are the number of observed events. The closure of a test is defined as the statistical significance of a deviation from zero, from which any displacement from zero is used to approximate a systematic uncertainty.

Each test is performed separately for each n_{jet} category inclusive in n_{b} and the corresponding systematic uncertainty is determined per H_T category. To determine the systematic uncertainty in each H_T category, the n_{jet} categories are merged into their symmetric and asymmetric categories and the ratio, (N$_{obs}$ - N$_{pred}$)/N$_{pred}$, is added in quadrature with its statistical error.

Modelling of α_{T} and $\Delta\phi^{*}_{min}$

In the construction of the control regions, no $\Delta\phi^{*}_{min}$ requirement is made and only the γ+jets control region is subject to an α_{T} requirement. To determine the compatibility of using the muon control regions, without an α_{T} or $\Delta\phi^{*}_{min}$ requirement, to make background predictions in the signal region, events in the μ+jets control region in data with an α_{T} or $\Delta\phi^{*}_{min}$ requirement are used to predict the number of events with that requirement inverted.

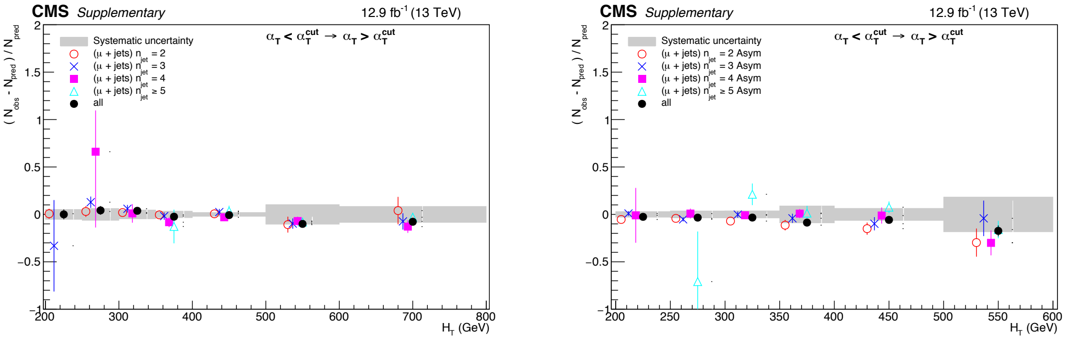
The inversion of the α_{T} and $\Delta\phi^{*}_{min}$ variables is tested independently in the μ+jets region. The contribution of the systematic uncertainty is taken only from the α_{T} closure tests for the H_T categories $<$ 800 GeV, and the $\Delta\phi^{*}_{min}$ closure tests with H_T $>$ 800 GeV.

The results of the α_{T} and $\Delta\phi^{*}_{min}$ closure tests are shown in Figure 7.8 as a function of H_T and n_{jet}, inclusive of n_{b} and n_{subjets}. In each distribution, the shaded band represents the systematic uncertainty and the closure for each jet multiplicity and the quadrature sum is shown by the coloured and solid black lines respectively. The systematic uncertainty in the modelling of the variables is derived to be in the range 3 - 30 %.

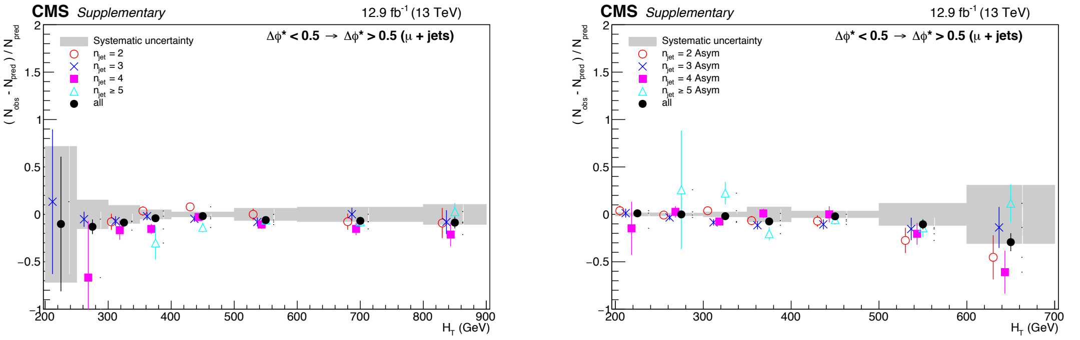
Modelling of the W/Z ratio

Of the control regions described in Section 6.8, each is used in the transfer factor method to predict the $Z \rightarrow \nu\bar{\nu}$ background in a signal region event category. While the $\mu\mu$+jets and γ+jets control regions are constructed to target this prediction with accuracy, the $W \rightarrow \ell\nu$ and $\mathrm{t\bar{t}}$ enriched μ+jets control region is susceptible to mismodelling.

In order to validate the use of the μ+jets control region to predict the $Z \rightarrow \nu\bar{\nu}$ background, closure tests are performed in data using the μ+jets and $\mu\mu$+jets control



(a) α_{T} modelling, symmetric category (b) α_{T} modelling, asymmetric category



(c) $\Delta\phi^*_{min}$ modelling, symmetric category (d) $\Delta\phi^*_{min}$ modelling, asymmetric category

Fig. 7.8 The measures of the closure, as a function of H_T, in the modelling of the α_{T} variable (top) and $\Delta\phi^*_{min}$ variable (bottom) for the symmetric (left) and asymmetric (right) jet categories. In each test, the closure per jet multiplicity is represented by the coloured points, and the their quadrature sum is represented as the solid black points. The shaded band represents estimated total systematic uncertainty [14].

regions. Event yields in the μ+jets control region are used to predict event yields in the $\mu\mu$+jets control region, using transfer factors derived from simulation. The results of the closure test target the modelling of the W/Z ratio in simulation, and the systematic uncertainty is measured to be in the range 4 - 15 %.

Modelling of W/Z acceptance due to polarisation effects

In *pp*-collisions, high p_{T} W bosons are predominantly produced in a left-handed state [107]. The effect of the polarisation manifests itself in the decay spectra of the daughter particles when the W boson decays. In the leptonic decay of left-handed W^{+} bosons of a certain p_{T}, neutrinos are produced with a higher p_{T} than the charged lepton, with the converse being true in the leptonic decay of left-handed W$-$ bosons. Therefore, the leptonic decay of the expected left-handed W$\pm$ bosons produces an asymmetry in the p_{T} distributions between the leptons, with leptons originating from W+ bosons having lower p_{T} than those produced in W$-$ decays.

The result of such an asymmetry can be viewed in the topologies of the μ+jets control region and the signal region, where the former relies on lepton p_{T} for acceptance, while the latter relies on the neutrino p_{T} for acceptance. To validate the modelling of this asymmetry, event counts in simulation in a W+ boson enriched sample are used to predict event counts in a W^{-} enriched sample. From which, a closure test is performed in the μ+jets control region resulting in an estimated systematic uncertainty in the range 2 - 10%.

Modelling of the Z/γ ratio

In a similar manner to the closure test that probes the W/Z ratio in simulation, driven by the uncertainty in using the μ+jets control region to predict $Z \to \nu\bar{\nu}$ processes, the use of the γ+jets control region to also predict the $Z \to \nu\bar{\nu}$ background is investigated.

In order to ascertain the accuracy of this modelling, event yields in the γ+jets control region are used to predict events in the $\mu\mu$+jets control region, using transfer factors constructed from simulation. The results of the test return systematic uncertainties in the range 6 - 11 %.

Modelling of the boosted top/W admixture

For event categories subject to t-tagging and W-tagging, and thus further classified according to n_{subjets}, the prediction of the $\mathrm{t\bar{t}}$ and W + jets background is performed by the μ+jets control region.

In order to validate the modelling of the relative composition of boosted W bosons and top quarks, event yields with a W-tagged jet, where $n_{\text{subjets}} = 2$, are used to predict event yields with a t-tagged jet, where $n_{\text{subjets}} = 3$. The closure test for this validation is performed in the μ+jets control region as a function of n_{jet} and inclusive in n_{b} and H_T, and is shown in Figure 7.9. The systematic uncertainty derived from these tests is in the range of 3 - 70 %.

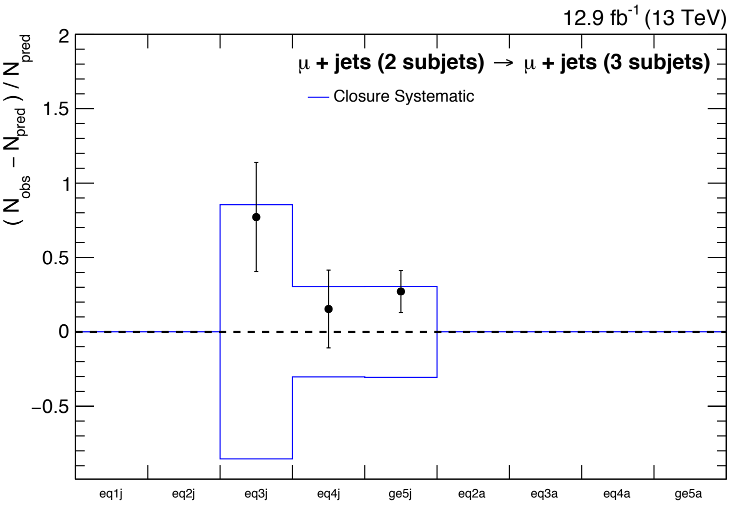


Fig. 7.9 The measure of the closure, as a function of n_{jet}, in the modelling of n_{subjets}. The binning label is detailed in Appendix A .

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

In all event categories, irrespective of those classified according to n_{subjets}, the prediction of the $\mathrm{t\bar{t}}$ background is performed by the $W \to \ell\nu$ enriched μ+jets control region. To probe the modelling of the relative composition of events from W + jets and $\mathrm{t\bar{t}}$ processes, predictions from a light flavour enriched region to a heavy flavour enriched, of the μ+jets control region, are made. The construction of a light flavour enriched region is achieved by requiring $n_{\text{b}} = 0$. By means of a transfer factor constructed in simulation, event yields in this region are then propagated to a heavy flavour region, constructed by requiring $n_{\text{b}} = 1$.

The closure test, which predicts event yields in data where $n_{\mathrm{b}} = 0$ to event yields in data with $n_{\mathrm{b}} = 1$ in the μ+jets control region, probes the sensitivity of the transfer factors to the relative admixture of events from the W + jets and $\mathrm{t\bar{t}}$ processes. The systematic uncertainty derived from these tests is in the range of 4 - 30 %.

7.4.3 Uncertainties in the $\not{H}_T$ dimension

In addition to the systematic uncertainties in the normalisation of the yields in the event categories of the analysis, the uncertainty in the modelling of the $\not{H}_T$ distribution is considered. The aforementioned systematic uncertainties, described in Section 7.4.1, are altered according to their $\pm 1\sigma$ variations and treated as an uncertainty that varies the $\not{H}_T$ shape.

Furthermore, as the $\not{H}_T$ distribution is constructed from simulation, an additional uncertainty is determined. To achieve this, a data-driven closure test method is constructed and the level of closure of such tests is used to create alternative $\not{H}_T$ shapes to cover the measured systematic differences. The procedure begins by investigating, each control region, the agreement between data and simulated events of the $\not{H}_T$ shape for each event category. The agreement between the two distributions is parametrised with a linear orthogonal polynomial, to allow a linear fit with one parameter, as in Equation 7.6.

$$f_n(x) = \sum_{k=0}^{n} (p_k) \cdot (\bar{x} - x)^k \tag{7.6}$$

where $\bar{x}$ is the weighted mean of the distribution and p_k are constants. To determine the uncertainty for each of the background sources, a combined linear fit of the predictions in all control regions is performed with a requirement of $\not{H}_T > 130$ GeV applied to ensure that the control regions are kinematically similar to the signal region. Thereafter, the uncertainty is taken as the quadrature sum of the value of the best fit parameter, p, with 1σ deviation. This allows alternative $\not{H}_T$ shapes to be derived based on the $\pm 1\sigma$ variation of the uncertainty.

Systematic	Method	Relative uncertainty on transfer factor				Correlation model
		$\mu + jets \rightarrow$ $Z \rightarrow \nu\bar{\nu}$	$\mu\mu + jets \rightarrow$ $Z \rightarrow \nu\bar{\nu}$	$\gamma + jets \rightarrow$ $Z \rightarrow \nu\bar{\nu}$	$\mu + jets \rightarrow$ $\mathrm{t\bar{t}} + \mathrm{W}$	
$\alpha_{\mathrm{T}}/\Delta\phi^{*}_{min}$ 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/$\mathrm{t\bar{t}}$ 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.
t-tag/W-tag admixture	data-driven tests	-	-	-	$4 - 25\%$	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
t-tagging efficiency	MC variations	$1 - 3\%$	-	-	$1 - 3\%$	fully correlated
W-tagging efficiency	MC variations	$1 - 3\%$	-	-	$1 - 3\%$	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

Table 7.6 Summary of the systematics on the transfer factors considered in the analysis, with representatives ranges of uncertainties and the correlation assumed, for the predictions of the $\mathrm{t\bar{t}}$, W and $Z \rightarrow \nu\bar{\nu}$ background components.

7.5 Alternative angular variables for QCD multijet suppression

While the α_{T} variable is constructed to exploit the topology of a dijet QCD process, the $\Delta\phi^*_{min}$ variable relies on the azimuthal separation between the transverse momentum vector of each jet, $\overrightarrow{p_{\mathrm{T}i}}$ and the missing transverse energy vector, $\overrightarrow{\not{H}_T}$.

Mathematically, $\Delta\phi^*_{min}$ is written as Equation 6.4, in which the metrics typically include the highest energy jet in the computation and as such, are therefore susceptible to jet mismeasurement. The sensitivity in rejecting events with fake $\not{H}_T$, through a jet mismeasurement or a high p_{T} neutrino from a heavy flavour decay, arises from the expectation that the magnitude of the jet $\overrightarrow{p_{\mathrm{T}}}$ changes, due to such mismeasurement. In addition, as the direction of the jet $\overrightarrow{p_{\mathrm{T}}}$ is not expected to change, the fake $\not{H}_T$ is created in a direction close to that of the jet $\overrightarrow{p_{\mathrm{T}}}$, and is thus targeted by the $\Delta\phi^*_{min}$ requirement.

The mathematical structure of $\Delta\phi^*_{min}$ can be expanded further, by the law of cosine as shown in Equation 7.7.

$$\cos\Delta\phi^*_i = \frac{\overrightarrow{p_{\mathrm{T}i}} \cdot (\overrightarrow{\not{H}_T} + \overrightarrow{p_{\mathrm{T}i}})}{|\overrightarrow{p_{\mathrm{T}i}}||\overrightarrow{\not{H}_T} + \overrightarrow{p_{\mathrm{T}i}}|} \tag{7.7}$$

where $\Delta\phi^*_i$ is the angle between $\overrightarrow{p_{\mathrm{T}i}}$, denoted as the p_{T} vector of a jet, and the $\overrightarrow{\not{H}_T}$ recalculated without the jet i [15].

The geometric relations between $\Delta\phi^*_i$, $\overrightarrow{p_{\mathrm{T}i}}$ and $\overrightarrow{\not{H}_T}$ is shown in Figure 7.10a, in which $\overrightarrow{\not{H}_T}$ is written as $\overrightarrow{H}^{miss}_{\mathrm{T}}$. In this figure, the jet $\overrightarrow{p_{\mathrm{T}}}$ and $\overrightarrow{\not{H}_T}$ are drawn on the transverse momentum plane that is rotated and scaled such that the vector $\overrightarrow{\not{H}_T}$ points along the x axis with unit length.

An important metric in discriminating between events containing multijet processes and those originating from a potential signal process, are the variables $\overrightarrow{p_{\mathrm{T}i}}$ and $\overrightarrow{\not{H}_T}$. The two variables, when taken as the ratio in Equation 7.8, can provide an additional discrimination strength, and thus yield an alternative representation of $\Delta\phi^*_{min}$ [15], as shown in Equation 7.9:

$$f_i \equiv \frac{|\overrightarrow{p_{\mathrm{T}i}}|}{|\overrightarrow{\not{H}_T}|} \tag{7.8}$$

$$\cos \Delta\phi_i^* = \frac{f_i + \cos \Delta\phi_i}{\sqrt{1 + f_i^2 + 2 f_i \cos \Delta\phi_i}} \tag{7.9}$$

The sensitivity of f_i in discriminating between events, can be considered when a jet $\overrightarrow{p_\mathrm{T}}$ and $\overrightarrow{\not{H}_T}$ are back-to-back. In this situation, whereby the value of $\Delta\phi_i$ is large, if the value of f_i is larger than one, thus implying that the jet p_T is greater than $\not{H}_T$, the corresponding value of $\Delta\phi^*_{min}$ is small and as a result, the event will be rejected. Furthermore, if the value of f_i is less than one, and no other jet in the event yields a small value of $\Delta\phi^*_{min}$, the event will be retained.

The discrimination strength lies in the behaviour of QCD multijet processes to typically yield low values of f_i, due to either a jet p_T over-measurement or a neutrino from a semileptonic heavy flavour decay. In contrast, events containing a genuine source of $\not{H}_T$, such as signal events containing invisible particles, have a smaller chance to contain jets with p_T greater than $\not{H}_T$. In this context, jets produced in signal events will share the recoil of the total p_T associated to the invisible particles.

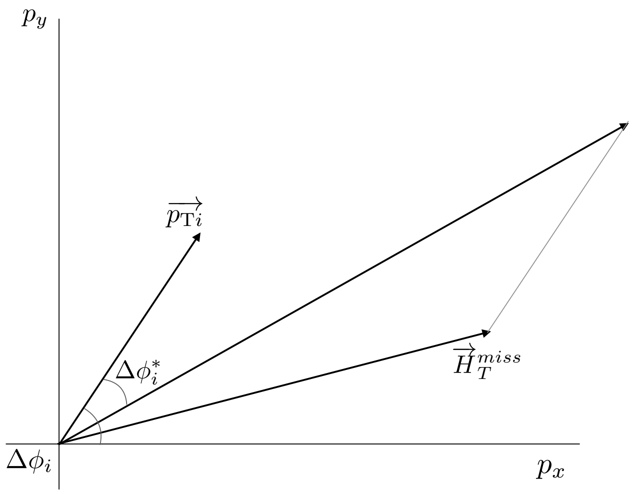
A figurative representation of the variable f_i, and the relation between $\overrightarrow{p_{\mathrm{T}i}}$ and $\overrightarrow{\not{H}_T}$ is shown in Figure 7.10b. The angle $\Delta\phi^*_{min}$, the jet $\overrightarrow{p_\mathrm{T}}$ and the variable f_i is drawn in transverse momentum plane, with the vector $\overrightarrow{\not{H}_T}$ pointing along the x axis with unit length.

7.5.1 Minimising $\not{H}_T$

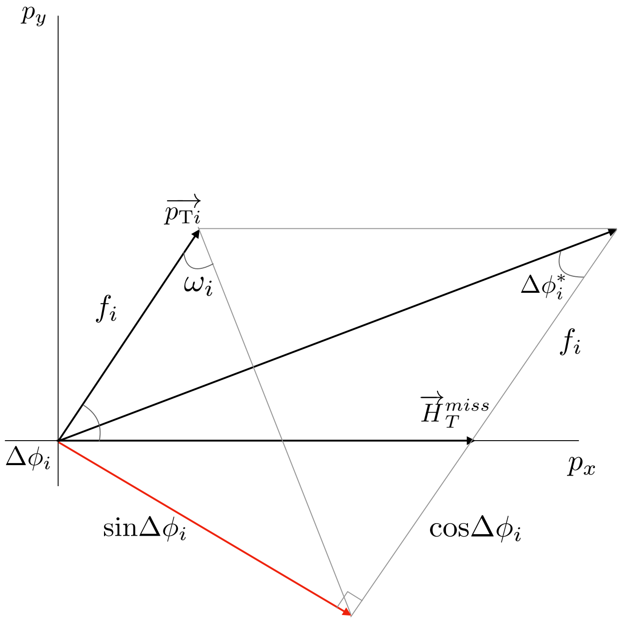
A crucial distinguishing feature between the variable $\Delta\phi_i$ and $\Delta\phi^*_{min}$ is the omission of the jet i. In the former case, the angle is performed between the jet i and the $\not{H}_T$ vector, while in the latter scenario the probe jet i is removed in the calculation of $\overrightarrow{\not{H}_T}$, and the angle between the two is computed as in Equation 6.4.

The method of removing the probe jet in the computation of $\Delta\phi^*_{min}$ can be considered alternatively as the determination of the azimuthal angle between $\overrightarrow{p_{\mathrm{T}i}}$ and $\overrightarrow{\not{H}_T}$ when the magnitude of $\overrightarrow{p_{\mathrm{T}i}}$ is varied to zero. Under this construction, the angle $\Delta\phi^*_{min}$ is the minimum angle between $\overrightarrow{p_{\mathrm{T}i}}$ and $\overrightarrow{\not{H}_T}$ in the variation of $p_{\mathrm{T}i}$.

In QCD multijet processes, a jet mismeasurement or neutrinos from a semi-leptonic heavy flavour decay can alter the magnitude of $\overrightarrow{p_{\mathrm{T}i}}$, but not the direction. If $p_{\mathrm{T}i}$ can be varied, the magnitude of the mismeasured jet can become close to its true value. Thus, in the case of a mismeasured jet, the large $\overrightarrow{\not{H}_T}$ will be minimised by varying the p_T of one of the jets in the event. Furthermore, if there is a genuine source of $\overrightarrow{\not{H}_T}$, from a potential signal process, the act of varying the p_T of any jet in the event will



(a) The relations between $\overrightarrow{p_{\mathrm{T}i}}$, $\Delta\phi^*_{min}$ and $\overrightarrow{\not{H}_T}$ in the transverse momentum plane.



(b) The relations between f_i, $\sin\Delta\phi_i$ and $\cos\Delta\phi_i$.

Fig. 7.10 Figurative representations of $\Delta\phi^*_{min}$ and f_i in the transverse momentum plane.

have little effect on minimising $\overrightarrow{\not{H}_T}$.

As can be inferred from Figure 7.10b, the tangent of $\Delta\phi_i^*$ can be expressed as:

$$\tan\Delta\phi_i^* = \frac{\sin\Delta\phi_i}{f_i + \cos\Delta\phi_i} \tag{7.10}$$

where sin$\Delta\phi_i$ is the minimised $\overrightarrow{\not{H}_T}$ when varying $p_{\mathrm{T}i}$, and cos$\Delta\phi_i$ is the variation to $p_{\mathrm{T}i}$. The physical interpretation of Equation 7.10 is the ratio of the minimum value of $\overrightarrow{\not{H}_T}$ when the magnitude of the jet p_{T} is varied, and the magnitude of the jet p_{T} that yields the minimum $\not{H}_T$, given the condition $f_i + \cos\Delta\phi_i \geq 1$.

As seen in Figure 7.10b, the tangent of the angle ω_i, defined as $\frac{\sin\Delta\phi_i}{f_i}$ can be interpreted as the ratio of the minimised $\overrightarrow{\not{H}_T}$ and the original jet p_{T}. An important observation of ω_i is when $\Delta\phi_i$ is an acute angle, which represents the situation whereby the jet p_{T} and $\overrightarrow{\not{H}_T}$ are on the same side. In a situation where a QCD multijet process has a fake source of $\not{H}_T$, due to a jet p_{T} mismeasurement, the value of $\Delta\phi_i$ is less than ω_i, $\Delta\phi_i^* < \omega_i$.

Conversely, when the jet p_{T} and $\overrightarrow{\not{H}_T}$ are on opposite planes with an angle $\Delta\phi_i > \frac{\pi}{2}$, the fake source of $\not{H}_T$ is due to a jet p_{T} over-measurement, and thus $\Delta\phi_i^* > \omega_i$.

7.5.2 min(χ)

The observation of the different characteristics of ω and $\Delta\phi_i^*$ for different jet p_{T} and $\overrightarrow{\not{H}_T}$ topologies, leads to the construction of an alternative angle χ_i, as defined in Equation 7.11.

$$\chi_i = \begin{cases} \omega_i \text{ if } \Delta\phi_i \leq \pi/2 \\ \Delta\phi_i^* \text{ if } \Delta\phi_i^* \leq \pi/2 \\ \pi/2 \text{ otherwise} \end{cases} \tag{7.11}$$

where i corresponds to a jet in the event. For every jet considered in an event, the minimum value of χ_i, defined as min(χ) is used to select events.

The alternative use of min(χ), as a replacement to $\Delta\phi_{min}^*$ has the ability to reject more QCD multijet processes with fake sources of $\not{H}_T$, than applying requirements on $\Delta\phi_{min}^*$. Furthermore, the signal acceptances from incorporating min(χ) remain at the same level to those observed when using $\Delta\phi_{min}^*$, and in some cases improve after using min(χ). This is illustrated in Figure 7.11, whereby the selection efficiency of QCD multijet events is shown against the selection efficiency for two benchmark

MSSM models. Each distribution shows the selection efficiency for the α_{T}, $\Delta\phi^*_{min}$ and $\min(\chi)$ requirements, at different values.

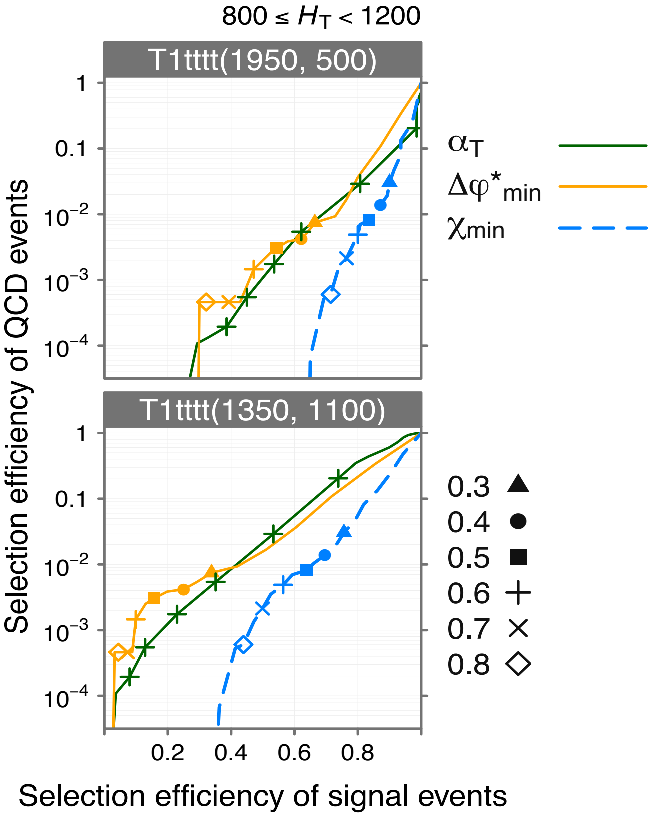


Fig. 7.11 Receiver operatic characteristic (ROC) curves of the α_{T}, $\Delta\phi^*_{min}$ and $\min(\chi)$variables for QCD multijet background events and two production modes of the signal model T1tttt, in an H_T range [15]. The markers indicate the values of the variables. The definition of the signal model is described in more detail in Section 8.4.

From the distributions shown in Figure 7.11, the $\min(\chi)$ variable demonstrates a weaker selection efficiency for QCD multijet events at a higher signal selection efficiency, with respect to the α_{T} and $\Delta\phi^*_{min}$ requirements.

As a means to further quantify the sensitivity of $\min(\chi)$, the selection efficiency for the SM backgrounds and two alternative benchmark models is compared against $\Delta\phi^*_{min}$, as seen in Table 7.7. The two benchmark models of the MSSM are described in Section 8.4.

In Table 7.7, events are required to satisfy a requirement of $\min(\chi) > 0.6$, while the requirement of $\Delta\phi^*_{min}$ is > 0.5.

Process	Event selection				
	Before	$\Delta\phi^*_{min}$		$\min(\chi)$	
		Yield	Selection efficiency (%)	Yield	Selection efficiency (%)
QCD multijet	3828731	2464	0.064	2392	0.062
Electroweak.	286,059	223,937	78.3	140,970	49.3
T1tttt(1600,100)	60	14.	23.3	27.5	45.8
T2tt(950,100).	98.	59.	60.2	73.	74.5

Table 7.7 A comparison between the alternative angle $\min(\chi_i)$ and $\Delta\phi^*_{min}$ for QCD multijet processes and two benchmark signal models. The total number of event counts after the full analysis selection, omitting the $\Delta\phi^*_{min}$ requirement is shown, in the signal region is shown. The selection efficiency is defined as the ratio of event counts before and after a requirement on $\Delta\phi^*_{min}$ or $\min(\chi)$.

The results of fully incorporating $\min(\chi)$, with the requirement $\min(\chi) > 0.6$, as a replacement to $\Delta\phi^*_{min}$ is shown in Section 8.4.

Chapter 8

Results and interpretations

The ability to claim the presence or absence of a new physics signal demands a well founded statistical treatment. On the data sample corresponding to 12.9 fb^{-1}of pp-collisions, events are analysed and categorised, with predictions of the contribution of SM processes made on each category.

8.1 Likelihood model

To measure the compatibility of the results with the SM backgrounds, likelihood models are constructed.

Events in both the signal region and control regions are categorised according to H_T, n_{jet}, n_{b} and n_{subjets}, labelled H_T^{cat}, whereby an additional index i is used in the signal region to denote the corresponding $\not{H}_T$ bin.

For a given number of observed events, $n^{H_T^{cat}}_{\text{had},i}$, a predicted number of background events, $b^{H_T^{cat}}_{\text{had},i}$, and an expected number of signal events $s^{H_T^{cat}}_{\text{had},i}$, the likelihood function in the signal region, can be written as:

$$\mathcal{L}^{H_T^{cat}}_{\text{had}} = \prod_i \text{Poisson}(n^{H_T^{cat}}_{\text{had},i} | \, b^{H_T^{cat}}_{\text{had},i} + s^{H_T^{cat}}_{\text{had},i}). \tag{8.1}$$

where the product runs over each $\not{H}_T$ bin.

In order to test the compatibility with the Standard Model only hypothesis, a likelihood function is constructed for each control region, j. Given the observed number of events, $n^{H_T^{cat}}_{\text{CR},j}$, the predicted number of background events, $b^{H_T^{cat}}_{\text{CR},j}$, and the predicted number of signal events, $s^{H_T^{cat}}_{\text{CR},j}$, the likelihood function takes the form:

$$\mathcal{L}^{H_T^{cat}}_{\text{CR,j}} = \text{Poisson}(n^{H_T^{cat}}_{\text{CR,j}} | \, b^{H_T^{cat}}_{\text{CR,j}} + s^{H_T^{cat}}_{\text{CR,j}}). \tag{8.2}$$

where the predicted number of signal events, $s_{\text{CR},j}^{H_T^{cat}}$, accounts for signal contamination in the control region and is estimated with simulated events.

The prediction of the background events in the signal region, $b_{\text{had},i}^{H_T^{cat}}$, and in the corresponding control region, $b_{\text{CR},j}^{H_T^{cat}}$, are connected by a transfer factor, TF, as discussed in Section 7.1.1. In the fitting procedure, this connection is treated as a floating parameter, which binds the background events in the signal and control regions to float together. The systematic uncertainties affecting the transfer factors, as discussed in Section 7.4, are incorporated as nuisance parameters. The treatment of systematic uncertainties in the likelihood model is discussed in Section 8.1.1.

The total likelihood is then constructed as the product of the individual likelihoods, across all the H_T^{cat} categories, and control regions:

$$\mathcal{L} = \prod_{H_T^{cat}} \mathcal{L}_{\text{had}}^{H_T^{cat}} \times \prod_{j} \mathcal{L}_{\text{CR,j}}^{H_T^{cat}} \tag{8.3}$$

The procedure by which the likelihood fit is performed is discussed in more detail in Section 8.1.2.

8.1.1 Incorporation of systematic uncertainties

As mentioned in Section 7, the prediction of the background events in the signal and control regions, are connected by transfer factors. To incorporate this into the likelihood model, the yields of the $Z \rightarrow \nu\bar{\nu}$ and combined $\mathrm{t\bar{t}}/W$ backgrounds between the signal and control region are correlated by a floating parameter. Such a parameter allows the background prediction in the signal region based on the yields of the control region yields to float together.

The data-driven systematic uncertainties that affect the transfer factors, described in more detail in Section 7.4.2, are integrated into the likelihood model through log-normal distributed nuisance parameters that act on the floating parameter, and are taken as uncorrelated between each of the H_T^{cat} categories.

Furthermore, additional systematic effects determined from simulation, as discussed in Section 7.4.1, are incorporated through shape uncertainties on the transfer factors that are fully correlated across all of the H_T^{cat} categories.

Uncertainties on the product of the experimental acceptance times efficiency ($A \times \varepsilon$), are also treated as shape uncertainties correlated across all the H_T^{cat} categories. The systematic uncertainties derived from simulation can also cause events to migrate across different $\not{H}_T$ bins. To account for this, alternative $\not{H}_T$ templates for the $\pm\sigma$ variation of each source of uncertainty are introduced. The uncertainty due to the limited statistics in the simulated events that are used to populate the template histograms is incorporated as additional nuisance parameters.

8.1.2 Fitting

In the analysis, the fit is performed in two stages. In the first stage, the background predictions in the signal region, $b^{H_T^{cat}}_{\mathrm{had},i}$ are extracted from the control region only likelihood function of Equation 8.2. In this method, the signal contribution is not included and the signal contamination term, $s^{H_T^{cat}}_{\mathrm{CR},j}$, is set to zero. This fit provides a direct gauge on the compatibility with the Standard Model only hypothesis, in which the contribution from signal processes is zero. In the second step of the fitting procedure, the full likelihood function, of Equation 8.3, is employed with all the correlations between the control regions and the signal region taken into consideration. It should be noted that in the full likelihood fit, the signal contribution in the signal region is not included.

In the interpretation of signal models, the likelihood is profiled against all nuisance parameters and includes a signal contribution term, $s^{H_T^{cat}}_{\mathrm{had},i}$, used to derive expected exclusion limits and sensitivities on various signal models, as shown in Section 8. The formalism for setting limits is discussed in more detail in Section 8.3.1.

8.2 Results

Using the likelihood fit discussed in Section 8.1, the SM backgrounds across the event categories of the signal region are predicted, and compared to the observations in data. In the absence of an excess of observed events, with respect to the expected background yields, limits are set on the production cross section for a range on new physics models, using the CL$_s$ method [108] described in more detail in Section 8.3.1. This result has been released publicly at [14]. The results of incorporating jet substructure techniques, as described in this thesis, are shown in the following sections.

The total predicted event yields for SM processes across the n_{jet}, n_{b}, H_T and n_{subjets} event categories are shown in Figures 8.1, 8.2 and 8.3, for the monojet, asymmetric and symmetric jet categories, respectively. For the event categories that are subject to further classification according to n_{subjets}, the total predicted event yields for the

SM processes is shown in Figure 8.4.

In each case, the event yields observed in data are shown as black circles with their error bars. The histograms correspond to the event yields of SM processes, predicted with the TF methods described in Section 7. In the bottom panel of each distribution, the *pull* is defined as the difference between the number of observed events and the number of predicted events, divided by the 1σ uncertainty. The value of the pull is determined from each of the two likelihood fits, as described in Section 8.1.2, where control region only fit is labelled as "pre-fit", and the full likelihood fit is labelled as "post-fit".

Across the analysis event categories, the observed yields in data are compatible with the SM predictions, whereby any pull $\sim 3\sigma$ is significantly reduced when employing the full likelihood fit. The pulls of the observation compared to the prediction from the control region only fit, is shown in Appendix E for each jet topology.

The distribution of the prediction of the SM backgrounds as a function of $\not{H}_T$ is shown in Figure 8.5 for a collection of n_{jet}, n_{b} and H_T categories that are not further considered with respect to n_{subjets}. As no $\not{H}_T$ shape information is used in the monojet topology, the full collection of categories is shown in Figure 8.1.

For signal models with a large expected n_{jet} multiplicity, examples of $\not{H}_T$ distributions for the SM background and a benchmark signal model, T1tttt, in event categories further classified by n_{subjets}, is shown in Figures 8.6. Regarding signal models whereby a lower n_{jet} multiplicity is expected, the $\not{H}_T$ distributions for the SM background and a benchmark signal model, T2tt, is shown in Appendix D for a collection of event categories. For the T1tttt and T2tt models, a collection of benchmark mass points are chosen to represent various scenarios. In each distribution, the data yields and statistical errors are represented as black markers.

8.3 Signal model interpretation

The observed event yields are interpreted in the context of signal models, of a specific theory, which provide a range of production and decay modes for different mass spectra. Further, as a means to explore the sensitivity gain in incorporating jet substructure techniques, both a supersymmetric simplified model and a dark matter model with expected top quarks and W bosons are interpreted, the details of which are mentioned in Section 8.4 and Section 8.5.

To further explore the reach of the analysis and compare in detail the sensitivity between the default event categories and those further classified according to n_{subjets}, several benchmark model mass points that yield differing kinematics are considered. As the topology of the final state is strongly dependent on the mass points of the sparticles, a compressed and uncompressed scenario is considered. For the model T1tttt, the compressed and uncompressed benchmark mass points are $\mathrm{m}_{\tilde{g}}$ = 900 GeV and m_{LSP} = 550 GeV, and $\mathrm{m}_{\tilde{g}}$ = 1600 GeV and m_{LSP} = 100 GeV, respectively. Regarding the T2tt model, the compressed and uncompressed benchmark mass points are $\mathrm{m}_{\tilde{t}}$ = 300 GeV and m_{LSP} = 200 GeV, and $\mathrm{m}_{\tilde{t}}$ = 950 GeV and m_{LSP} = 100 GeV, respectively. The interpretation of the results across the full mass plane of the T2tt and T1tttt modes are discussed in Section 8.4.

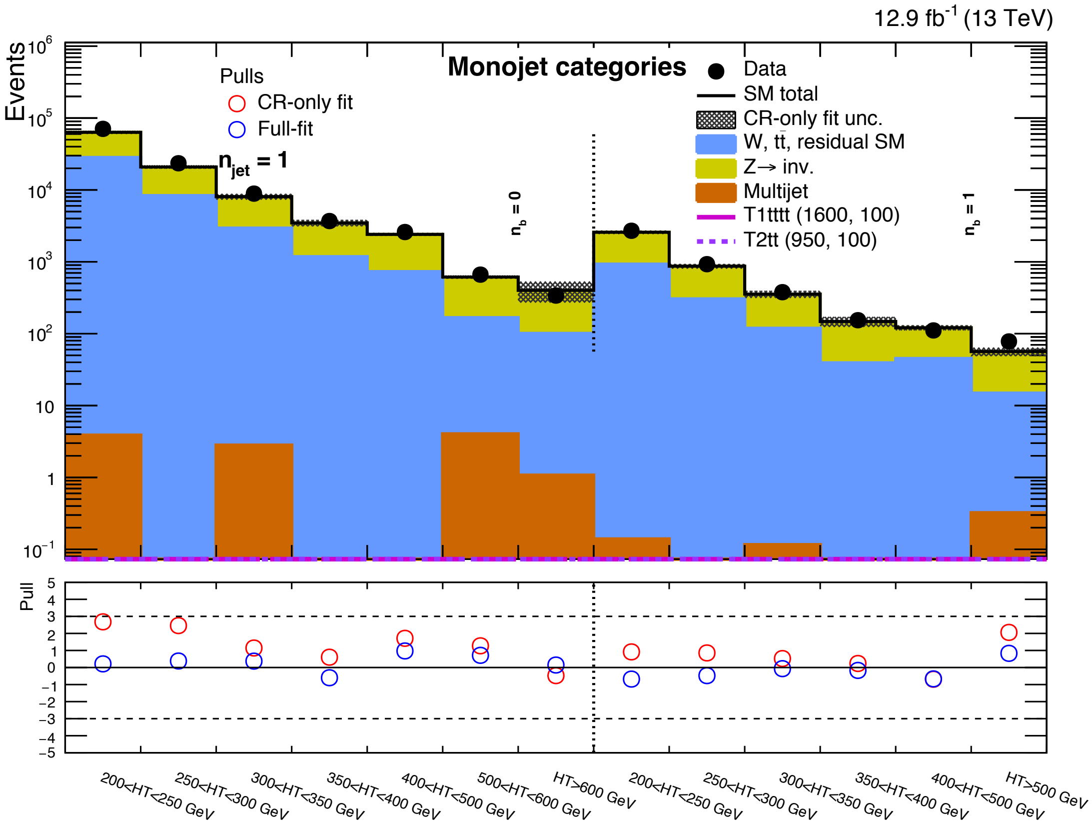


Fig. 8.1 The total event yields in data (solid black circles) and the SM expectations with their associated uncertainties (black histogram with shaded band) as a function of n_b and H_T for the **monojet** topology ($n_\mathrm{jet} = 1$) in the hadronic signal region. The bottom panel shows the significance 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).

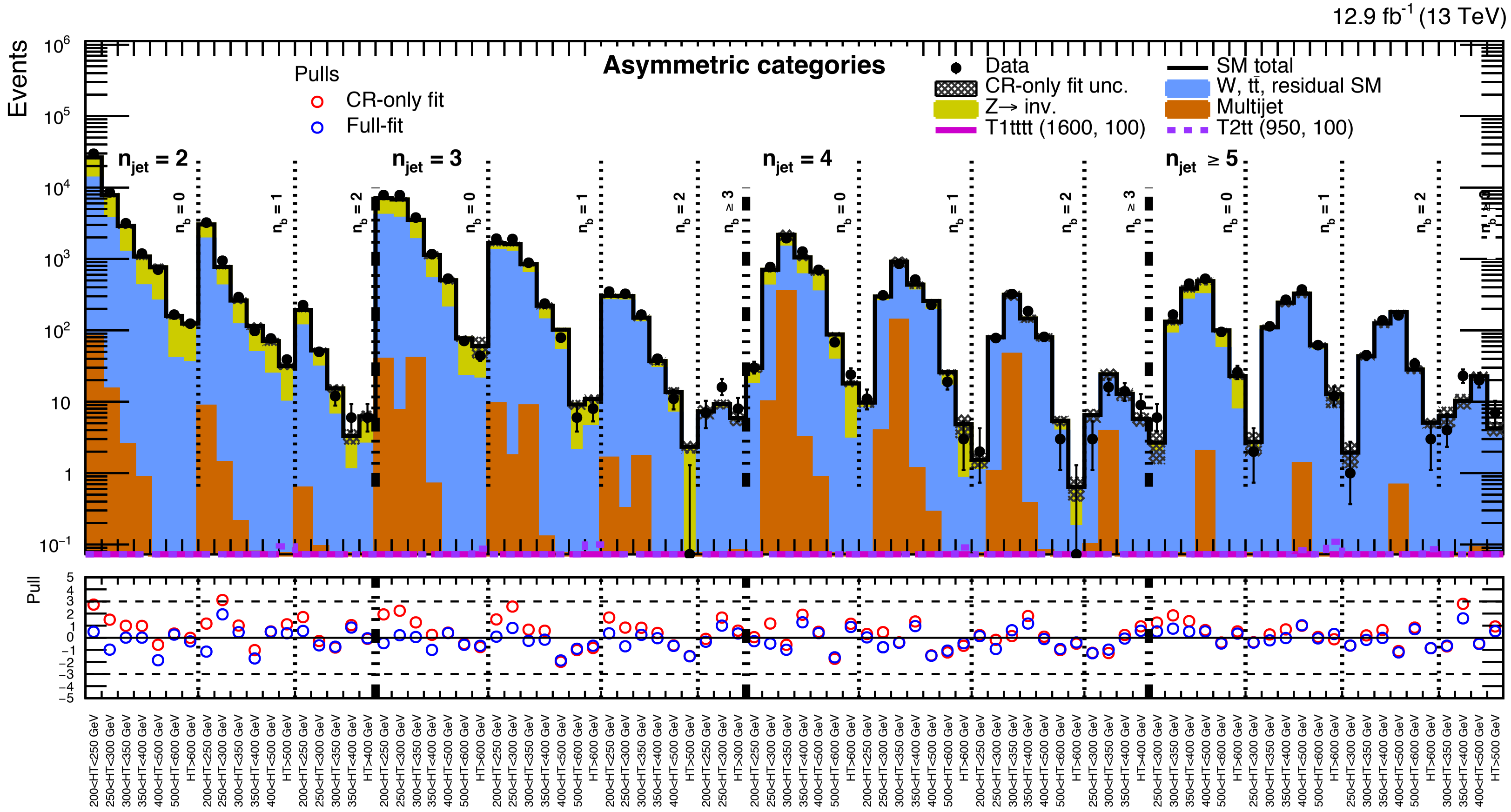


Fig. 8.2 The total event yields in data (solid black circles) and the SM expectations with their associated uncertainties (black histogram with shaded band) integrated over $\not{H}_T$ as a function of n_{jet}, n_{b} and H_T for the **asymmetric** topology in the hadronic signal region. The bottom panel shows 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 hadronic signal region (blue circles).

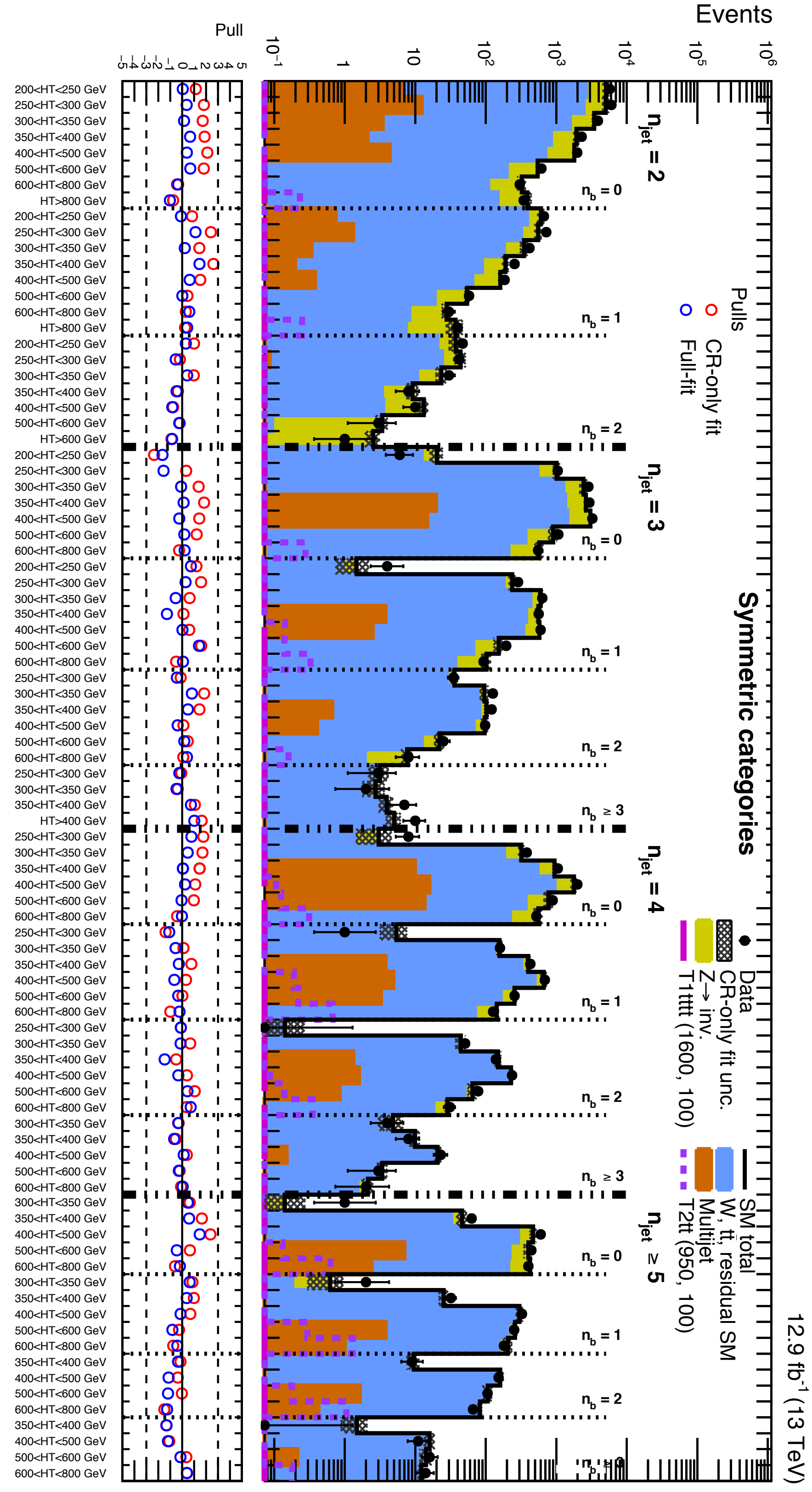


Fig. 8.3 The total event yields in data (solid black circles) and the SM expectations with their associated uncertainties (black histogram with shaded band) integrated over H_T as a function of n_{jet}, n_{b}, and second last H_T category for the **symmetric** topology in the hadronic signal region. The bottom panel shows 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 hadronic signal region (blue circles).

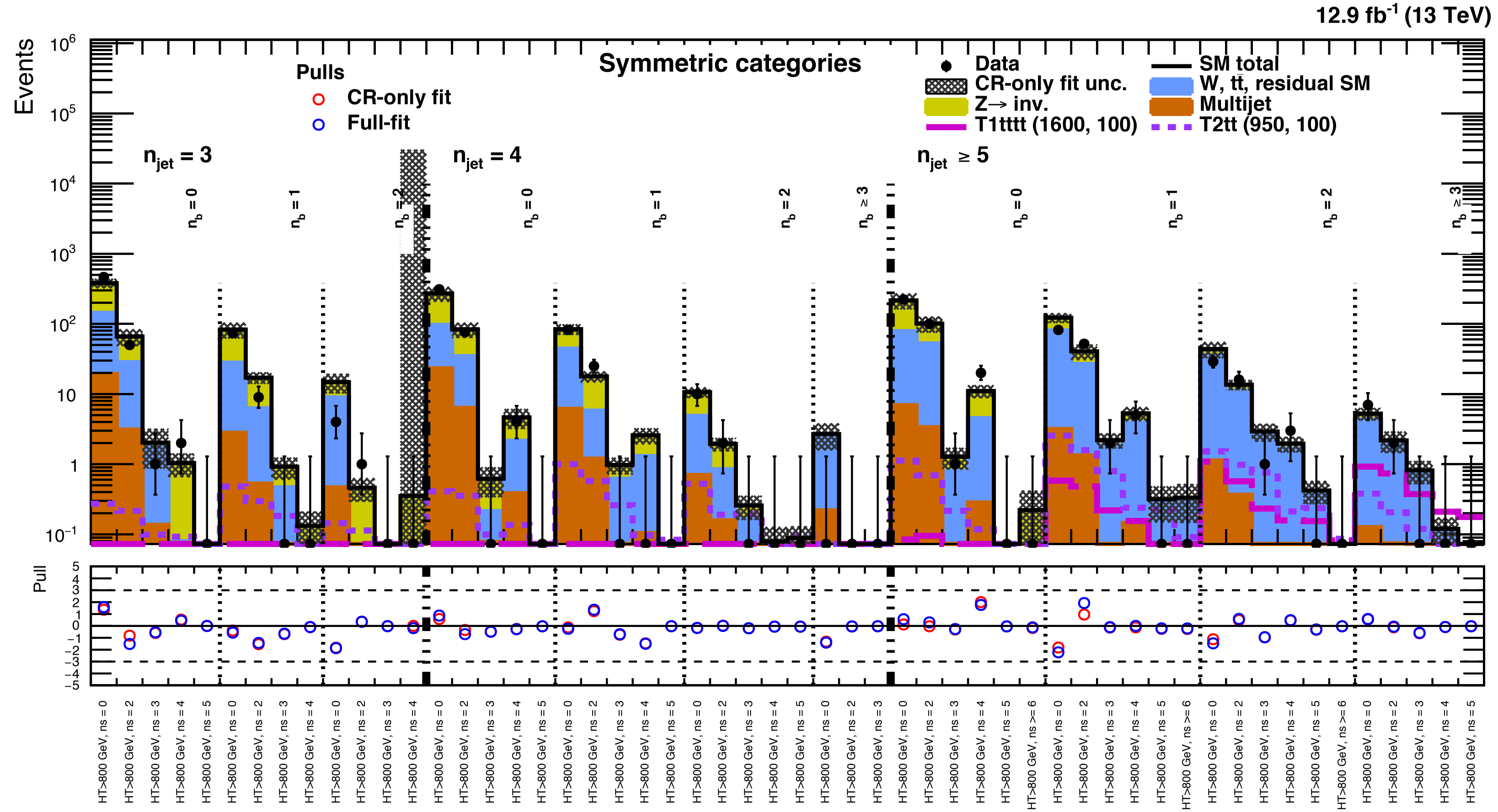
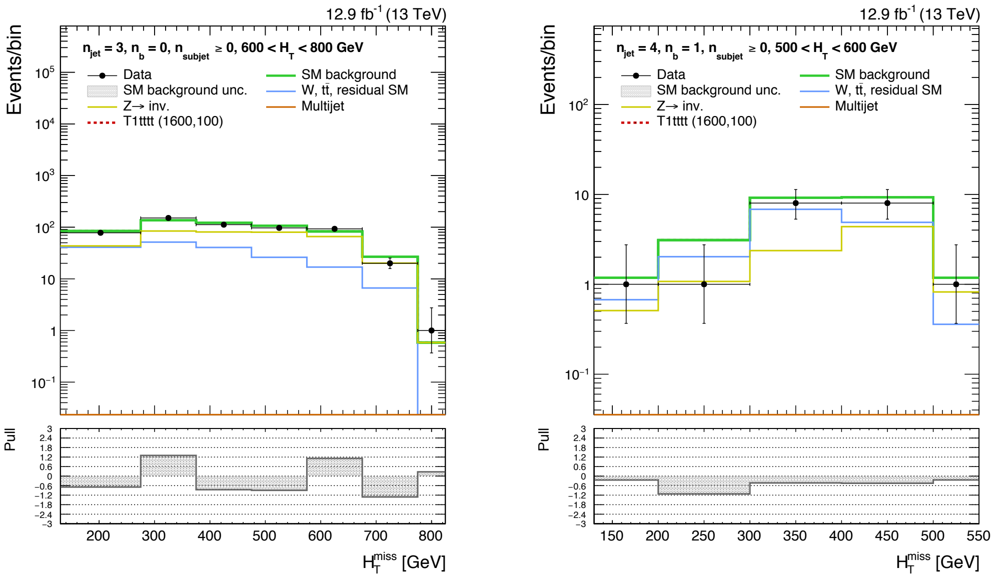
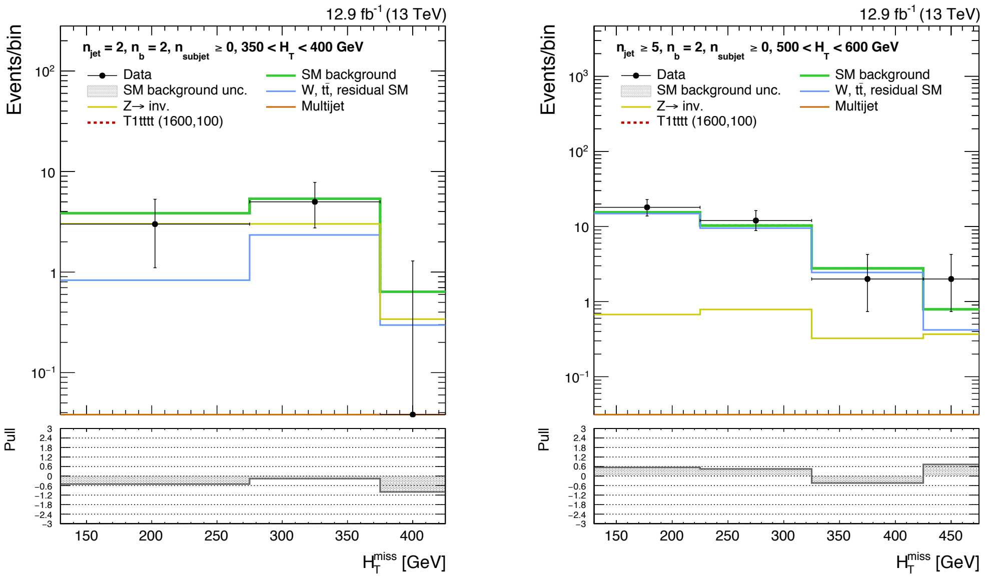


Fig. 8.4 The total event yields in data (solid black circles) and the SM expectations with their associated uncertainties (black histogram with shaded band) integrated over $\not{H}_T$ as a function of $n_{\text{jet}} = 3$, $4 \geq 5$, n_{b}, and the final H_T event category that corresponds to **further event categorisation by** $n_{\textbf{subjets}}$. The jet category is the **symmetric** topology in the hadronic signal region. The bottom panel shows 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 hadronic signal region (blue circles).



(a) $\not{H}_T$ shape in the symmetric topology for $n_b = 0$, $n_{jet} = 3$, $n_{subjets} \geq 0$ and 600 $< H_T <$ 800 GeV.

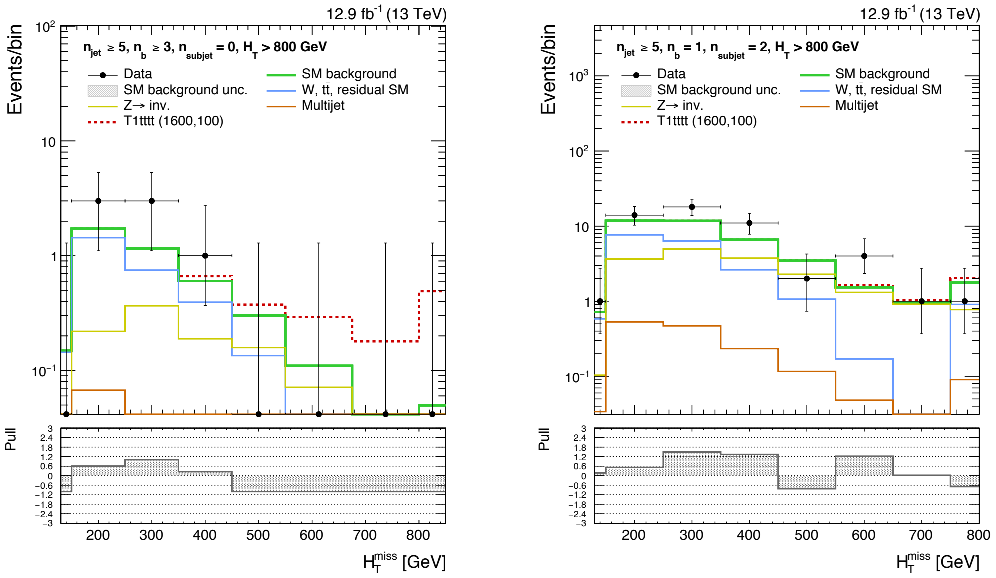
(b) $\not{H}_T$ shape in the asymmetric topology for $n_b = 1$, $n_{jet} = 4$, $n_{subjets} \geq 0$ and 500 $< H_T <$ 600 GeV.



(c) $\not{H}_T$ shape in the symmetric topology for $n_b = 2$, $n_{jet} = 2$, $n_{subjets} \geq 0$ and 350 $< H_T <$ 400 GeV.

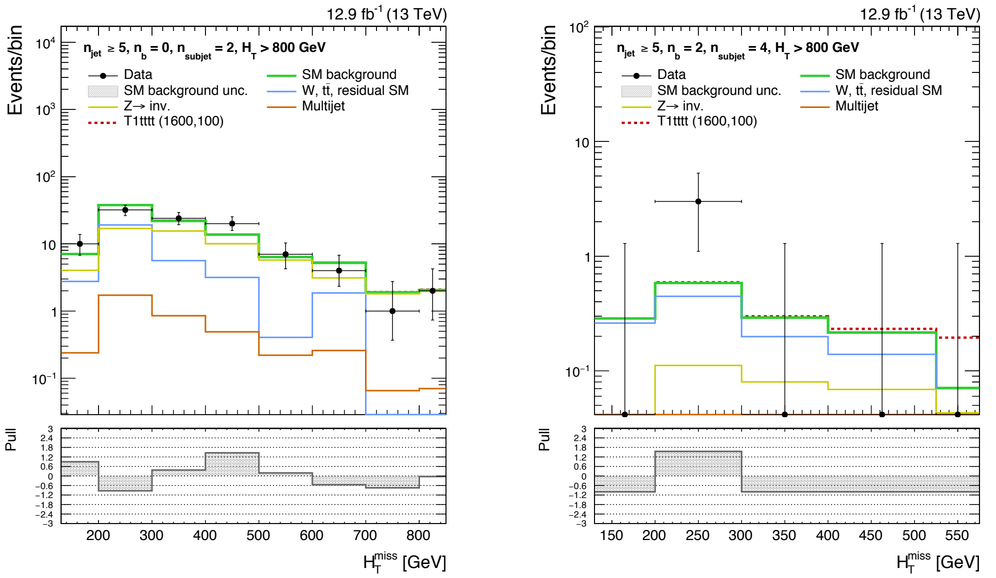
(d) $\not{H}_T$ shape in the asymmetric topology for $n_b = 2$, $n_{jet} \geq 5$, $n_{subjets} \geq 0$ and 500 $< H_T <$ 600 GeV.

Fig. 8.5 The total event yields in data and the SM expectation with their associated uncertainties as a function of $\not{H}_T$ for events in the signal region. The distributions below represent the significance of deviations, denoted as a pull, observed in data with respect to the SM expectations.



(a) $\not{H}_T$ shape in the symmetric topology for $n_\mathrm{b} \geq 3$, $n_\mathrm{jet} \geq 5$, $n_\mathrm{subjets} = 0$ and H_T $>$ 800 GeV.

(b) $\not{H}_T$ shape in the symmetric topology for $n_\mathrm{b} = 1$, $n_\mathrm{jet} \geq 5$, $n_\mathrm{subjets} = 2$ and H_T $>$ 800 GeV.



(c) $\not{H}_T$ shape in the symmetric topology for $n_\mathrm{b} = 0$, $n_\mathrm{jet} \geq 5$, $n_\mathrm{subjets} = 2$ and H_T $>$ 800 GeV.

(d) $\not{H}_T$ shape in the symmetric topology for $n_\mathrm{b} = 2$, $n_\mathrm{jet} \geq 5$, $n_\mathrm{subjets} = 4$ and H_T $>$ 800 GeV.

Fig. 8.6 The total event yields in data and the SM expectation with their associated uncertainties as a function of $\not{H}_T$ for events in the signal region further categorised according to n_subjets. The distributions below represent the significance of deviations, denoted as a pull, observed in data with respect to the SM expectations.

8.3.1 The CL$_s$ method

Limit setting

The procedure of placing limits on new physics processes is performed with respect to the hypothesis that no new physics processes is present in the data. Such an assessment, in terms of the statistical procedure, aims to compare the signal plus background hypothesis, in which new physics processes are assumed, to a background only hypothesis. Thereafter, an observable, x, is determined, for which each hypothesis predicts a different outcome. For each outcome of the hypotheses' predictions of x, a probability model is derived and is used to construct a test statistic. The distribution of a test statistic is often constructed through the generation of pseudo-experiments to obtain distributions for each hypothesis. However, when the expected number of events is sufficiently large, an asymptotic approximation of the pseudo-experiments can be made, whereby the pseudo-experiments are replaced by a single representative data set[1] [109]. The compatibility of a signal model with observations in data is determined by the CL$_s$ method to determine upper limits at a confidence level for the signal plus background hypothesis [108].

The test statistic, q_μ, based upon which upper limits are derived is constructed as:

$$q_\mu = \begin{cases} -2\ln\lambda(\mu), & \text{if } \mu \geq \hat{\mu} \\ 0 & \text{otherwise} \end{cases}, \tag{8.4}$$

where:

$$\lambda(\mu) = \frac{L(\mu, \theta_\mu)}{L(\hat{\mu}, \hat{\theta_\mu})}, \tag{8.5}$$

represents the profile likelihood ratio, in which μ is the signal strength[2], θ_μ is a collection of nuisance parameters for a given μ, $\hat{\mu}$ is the Maximum Likelihood (ML) value of μ, and $\hat{\theta}$ is the corresponding ML set of nuisance parameters. The signal strength parameter, μ, represents the fraction of the theoretical signal cross section that the expected signal yield is being calculated, such that a value of $\mu = 0$ corresponds to a background only hypothesis, and $\mu = 1$ represents the signal plus background hypothesis with the signal yield at a nominal theoretical cross section.

---

[1] Such a data set is referred to as the Asimov set.

[2] When constructing a histogram of the measured variable x, the expectation value for the number of events n in the i-th bin of the histogram can be written as $E[n_i] = \mu\mathrm{s}_i{+}\mathrm{b}_i$ where s_i and b_i represent the number of predicted signal and background events, respectively.

The CL$_s$ procedure

For a given observation of a test statistic, as measured in data, $q_{\mu,obs}$, the probability of finding a data set with equal or greater incompatibility to the signal plus background hypothesis can be constructed as:

$$p_{S+B} = \int_{q_{\mu,obs}}^{\infty} f_{S+B}(q_\mu|\mu)dq_\mu \tag{8.6}$$

where $f_{S+B}(q_\mu|\mu)$ represents the probability density function describing the q_μ distribution, of the signal plus background hypothesis. In a similar manner, the background-only hypothesis can be dealt with as follows:

$$1 - p_B = \int_{-\infty}^{q_{\mu,obs}} f_B(q_\mu|0)dq_\mu \tag{8.7}$$

The value of the integral, for a certain hypothesis is referred to as the p-value for that hypothesis. The compatibility of a signal model with observations in data is then determined by the parameter CL$_s$:

$$CL_s = \frac{p_{S+B}}{1 - p_B} \tag{8.8}$$

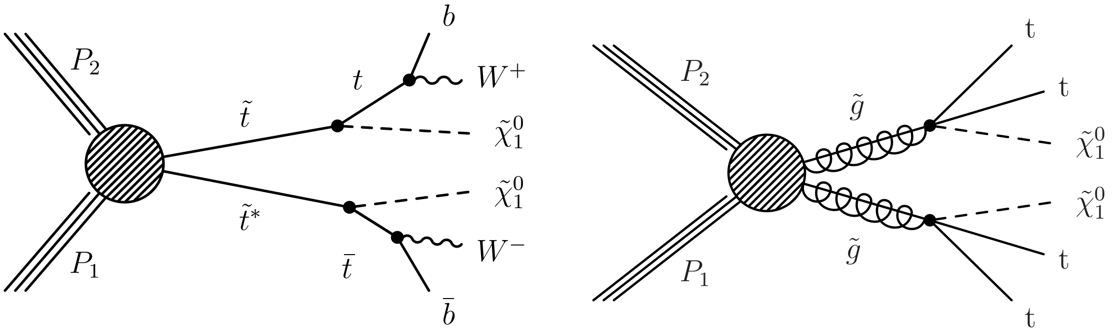
A model is considered to be excluded at 95% confidence level when CL$_s \geq 0.05$ [108].

8.4 Interpretations in Simplified SUSY Models

In the absence of any significant deviation of data from the SM backgrounds, upper limits are set on the parameters of certain SUSY models. The simplified models interpreted include the pair production of gluinos, referred to as *T1* models, and the pair production of squarks, *T2* models, the Feynman diagrams of two specific models are illustrated in Figures 8.7.

Per model, a full scan of the m$_{SUSY}$ - m$_{LSP}$ mass plane is performed. To gauge which mass ranges of a model can be excluded by the results in the event categories, a large variety of different mass points are produced in simulation. In the context of interpreting *T1* models, a large number of simulated events are produced for varying gluino and LSP masses, m$_{\tilde{\mathrm{g}}}$ and m_χ. Similarly, in interpreting the *T2* models, vast events are simulated for varying squark and LSP masses, m$_{\tilde{\mathrm{q}}}$ and m_χ.

For each mass point, a 95% CL upper limit is set on the production cross section to pair produce sparticles. The limit is produced under a background plus signal hypothesis with a modified frequentist approach. The CL_s method [108], is used to compute the limits on the range of excluded parameter space for signal models, with the one-sided profile likelihood ratio as the test statistic.



(a) T2tt (b) T1tttt

Fig. 8.7 Feynman diagrams of the signal models interpreted. The left diagram, denoted T2tt, shows top squark pair production with each top squark decaying into a top quark and the LSP. The right diagram shows gluino pair production with each gluino decaying into a pair of top quarks and the LSP.

8.4.1 Uncertainties on signal models

The signal models interpreted in the analysis are subject to systematic uncertainties on the experimental acceptance and efficiency, $A \times \varepsilon$. These uncertainties are considered across all (H_T, n_{jet}, n_{b}) and the n_{subjets} categories, when used.

In a similar method to determining the systematic uncertainties on the individual SM backgrounds, as discussed in 7.4, further systematic uncertainties are considered on each simplified model. The following systematic uncertainties for the prediction and correction of the signal yield are utilised:

Luminosity

The overall normalisation contains an uncertainty from the luminosity. A systematic uncertainty of 6.2% on the measurement of the total integrated luminosity is applied.

Trigger efficiency

A systematic uncertainty measured using the difference between the electron and muon reference triggers, as described in Section 7.4.1, is applied. The systematic uncertainty is found to be in range of 0 - 4%.

Pileup

A 5% systematic uncertainty on the minimum bias cross section, as discussed in Section 7.4.1, is applied.

ISR

In the analysis, the selection of signal events, and therefore signal acceptance, is dependent on the number of reconstructed jets. As a result, the modelling of ISR must be reliable. To account for differences observed in the ISR modelling between data and simulation, a correction is made.

In both data and simulation, the p_{T} spectrum of the gluino-gluino system which recoils against an ISR jet is measured. A systematic uncertainty of half the correction factor applied to the number of ISR jets in simulation is taken and found to be in the range 1 - 20%. The choice of PDF set, or variations therein, predominantly affects $A \times \varepsilon$ through the changes in the p_{T} spectrum of the system recoil, which is covered by the ISR uncertainty.

Object efficiencies

In a similar manner to the background processes, the simulated events for the signal contribution are corrected for uncertainties in the lepton reconstruction efficiency, b-tagging efficiency and JEC. The systematic uncertainties are propagated and treated as correlated across the event categories.

A summary of the uncertainties is shown in Table 8.1.

Systematic source	Type	Correlated	Typical magnitude (%)
Luminosity	Normalisation	Yes	6.2
Monte Carlo statistics	Norm. + shape	No	1–50
Jet energy scale (AK4)	Norm. + shape	Yes	3–10
Jet energy scale (AK8)	Norm. + shape	Yes	3–10
b-tag efficiency scale factors	Norm. + shape	Yes	5–40
t-tag efficiency scale factors	Norm. + shape	Yes	5-10
W-tag efficiency scale factors	Norm. + shape	Yes	5-10
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

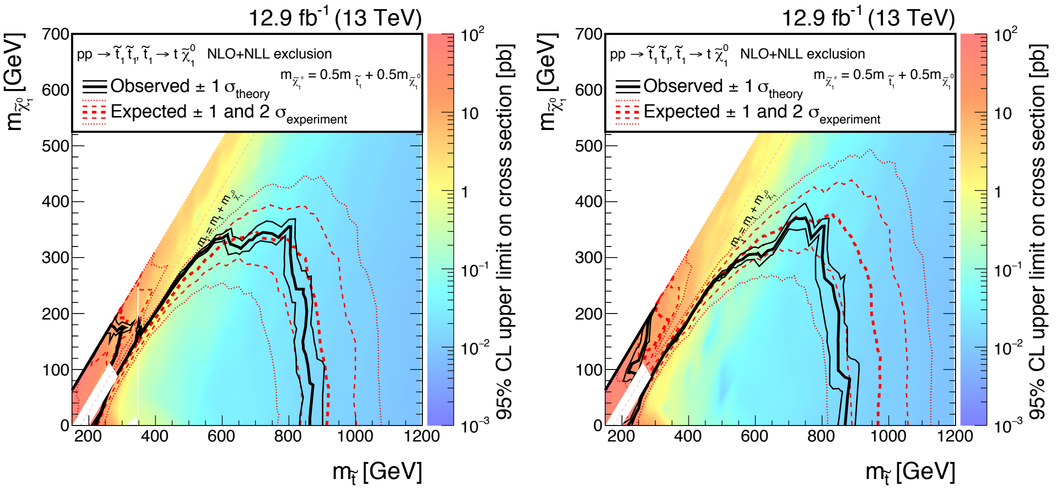
Table 8.1 The magnitude of uncertainties in the signal model yields in the case of a T2bb simplified model.

8.4.2 Exclusion Limits

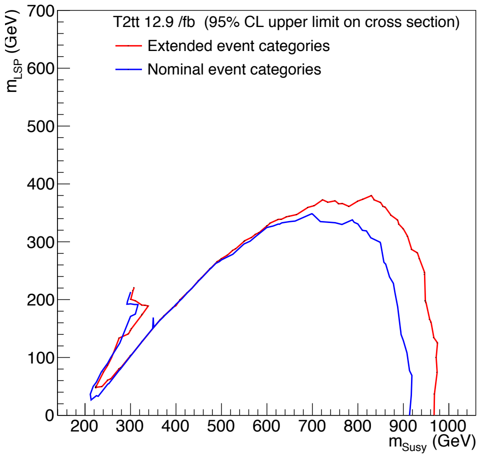
For each mass point of the simplified models mentioned in Section 8.4, an upper limit at a 95 % confidence level (CL) on the production cross section is determined, using the CL$_s$ method [108] with the asymptotic formulae used to model the test statistic distribution [109].

The results for a full scan of the gluino and squark pair produced models are summarised in Figure 8.8 and Figure 8.9. The colour scale of these distributions shows the observed 95% CL upper limit on the production cross section for each pair of m$_{SUSY}$ and m$_{LSP}$ masses. For each model, two results are presented with the former being the results obtained using the nominal event categorisation and the latter shows the results obtained when events are further categorised according to n_{subjets}. In addition, a collection of the most sensitive n_{jet}, n_{b}, H_T and n_{subjets} categories are shown for the collection of benchmark mass points.

In using the nominal event categories, stop quark masses are excluded up to 920 GeV and an upper limit on the production cross section is set $\sigma(m_{\tilde{t}=920\ GeV}) =$ 11.1 fb [110]. An improvement is observed for events that are further categorised according to n_{subjets}, and stop quark masses up to 970 GeV are excluded. This corresponds to an upper limit on the production cross section of $\sigma(m_{\tilde{t}=970\ GeV}) =$ 7.3 fb [110], and thus an approximate 30% improvement, with respect to the result obtained using the nominal event categories.



(a) Nominal event categories. (b) Nominal and n_{subjets} event categories.



(c) Comparison of the expected limit on the cross section, between the nominal event categories and event categories extended with n_{subjets}.

Fig. 8.8 The 95% observed upper limit on the cross section (histogram) for the T2tt model, with the expected (dotted red line) and observed (black line) exclusion contours, on the dataset corresponding to 12.9 fb^{-1}, on both the nominal event categories (left) and the extended categories (right)

Table 8.2 and Table 8.3 show most sensitive n_{jet}, n_{b} and H_T categories of the benchmark mass points of the SUSY model, T2tt, for the default event categories and those extended with n_{subjets} respectively. In each table, the most sensitive category, determined by the expected r$_{value}$, and the number of events per corresponding category is shown for a collection of benchmark mass points. In this context, r_{value} represents the limit on the signal strength[3] , determined when considering all the event categories and the $\not{H}_T$ dimension. Therefore, it should be noted that the events in the event categories in Table 8.2 and Table 8.3 are further classified according to their $\not{H}_T$ shapes. In addition, the signal yield, the background prediction and the observation in each category is shown.

When interpreting the results a combined fit over all the event categories is performed. The result of the combined fit, denoted by a single r_{value} is what determines the result observed in the mass planes of Figure 8.8 and Figure 8.9.

Signal Model	Rank (by expected limit)	Category	Signal	Prediction	Observation	Expected r_{value}	Observed r_{value}
T2tt (300,200)	1	$\geq$ 5j, 1b, 800 $-$ inf	12.85	165.93 ± 36.81	141	2.32	2.28
T2tt (300,200)	2	$\geq$ 5j, 0b, 600 $-$ 800	28.50	392.23 ± 18.98	402	2.63	2.38
T2tt (300,200)	3	$\geq$ 5j, 0b, 500 $-$ 600	26.76	389.63 ± 58.76	443	3.33	2.27
T2tt (300,200)	4	$\geq$ 5a, 0b, 400 $-$ 500	28.33	466.31 ± 48.62	528	3.52	6.33
T2tt (300,200)	5	$\geq$ 5a, 0b, 500 $-$ 600	7.73	95.72 ± 7.20	95	3.52	4.40
T2tt (300,200)	6	$\geq$ 5j, 1b, 600 $-$ 800	14.58	181.35 ± 8.54	181	3.80	4.30
T2tt (950,100)	1	$\geq$ 5j, 2b, 800 $-$ inf	3.79	62.43 ± 13.69	49	1.99	3.80
T2tt (950,100)	2	$\geq$ 5j, 1b, 800 $-$ inf	5.47	165.93 ± 36.81	141	2.57	2.68
T2tt (950,100)	3	$\geq$ 5j, $\geq$ 3b, 800 $-$ inf	0.82	8.64 ± 1.09	9	4.39	3.84
T2tt (950,100)	4	4j, 2b, 800 $-$ inf	1.05	13.53 ± 1.58	12	6.53	4.43
T2tt (950,100)	5	$\geq$ 5j, 2b, 600 $-$ 800	1.05	72.57 ± 9.08	66	8.16	7.17
T2tt (950,100)	6	4j, 1b, 800 $-$ inf	2.03	110.16 ± 9.73	107	9.16	8.82

Table 8.2 Expected and observed limits for most sensitive categories of all analysis event categories of the T2tt model, on two benchmark models. For each event category, the number of signal, background predicted and observed events are shown.

Signal Model	Rank (by expected limit)	Category	Signal	Prediction	Observation	Expected r-value	Observed r-value
T2tt (300,200)	1	$\geq$ 5j, 1b, $\geq$ 0s, 600 $-$ 800	14.58	206.23 ± 30.45	181	3.48	3.83
T2tt (300,200)	2	$\geq$ 5a, 0b, $\geq$ 0s, 500 $-$ 600	7.73	99.79 ± 7.07	95	3.48	4.17
T2tt (300,200)	3	$\geq$ 5j, 1b, 2s, 800 $-$ inf	4.62	40.58 ± 10.08	52	3.64	6.30
T2tt (300,200)	4	$\geq$ 5j, 0b, 0s, 600 $-$ 800	28.50	423.32 ± 29.26	402	3.98	4.16
T2tt (300,200)	5	$\geq$ 5a, 0b, $\geq$ 0s, 400 $-$ 500	28.33	487.75 ± 59.48	528	4.17	6.23
T2tt (300,200)	6	$\geq$ 5j, 0b, 0s, 500 $-$ 600	26.76	402.21 ± 59.81	443	4.30	2.82
T2tt (950,100)	1	$\geq$ 5j, 2b, 3s, 800 $-$ inf	0.76	2.92 ± 0.83	1	3.95	5.31
T2tt (950,100)	2	$\geq$ 5j, 1b, 3s, 800 $-$ inf	0.80	2.20 ± 0.53	2	4.27	5.06
T2tt (950,100)	3	$\geq$ 5j, 2b, 0s, 800 $-$ inf	1.52	43.76 ± 11.38	29	4.36	6.58
T2tt (950,100)	4	$\geq$ 5j, 1b, 0s, 800 $-$ inf	2.55	122.23 ± 19.18	82	4.45	4.63
T2tt (950,100)	5	$\geq$ 5j, 2b, 2s, 800 $-$ inf	0.98	13.59 ± 2.28	16	4.80	7.74
T2tt (950,100)	6	$\geq$ 5j, 1b, 2s, 800 $-$ inf	1.58	40.58 ± 10.08	52	4.80	6.12

Table 8.3 Expected and observed limits for most sensitive categories of all analysis event categories of the T2tt model, on two benchmark models, when incorporating the n_{subjets} event category. For each event category, the number of signal, background predicted and observed events are shown.

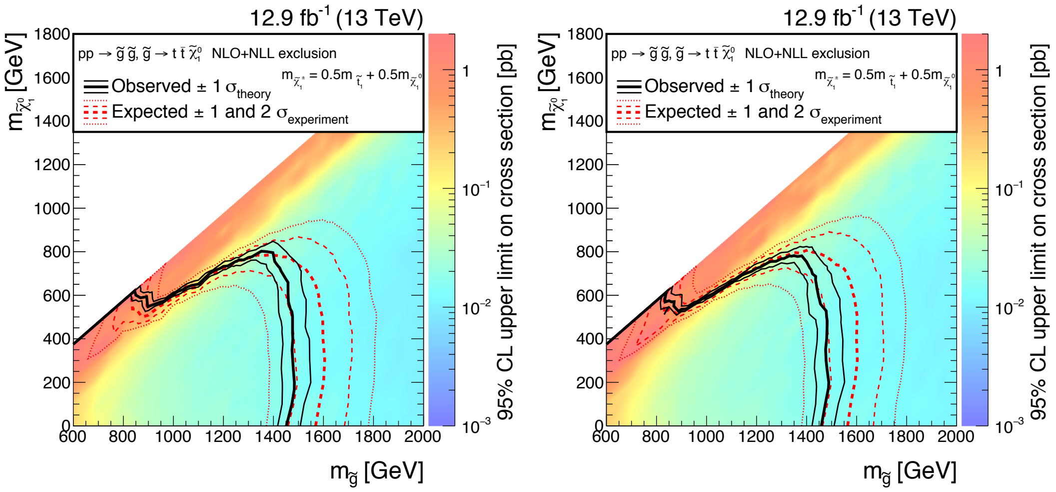
The most sensitive category for the compressed scenario for model T2tt, is one in which no jet substructure techniques are applied. A reason behind this observation is

---

[3] This variable is equivalent to the variable μ introduced in Section 8.3.1.

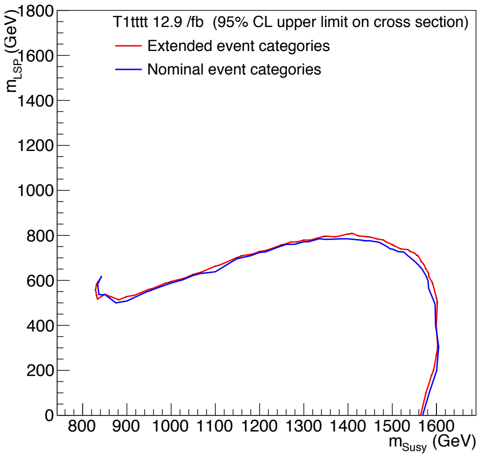
due to the soft p_T spectra of the final state particles, that results in a large separation between the decay products of hadronically decaying top quarks and W bosons. Therefore, the degree of separation between the decay products is inadequate to fully cluster them into a single jet that can be subject to jet substructure techniques. Furthermore, in the uncompressed scenario, where a larger amount of kinetic energy is available due to the large mass splitting between the stop quark and the LSP, the most sensitive categories are those containing a top quark tagged jet. The sensitivity in this event category is drawn from the discrimination strength in both tagging a boosted hadronically decaying top quark, amongst the SM backgrounds, and in utilising the $\not{H}_T$ dimension. This is discussed in more detail in Section 8.4.3.

The results for a full scan of the gluino produced model, T1tttt, are summarised in Figure 8.9. The 95 % CL upper limit on the cross section is shown across the m$_{SUSY}$ - m$_{LSP}$ mass plane. The colour scale of these distributions shows the 95% CL upper limit on the production cross section for each pair of m$_{SUSY}$ and m$_{LSP}$ masses. For gluino-mediated models, gluino masses are excluded up to 1600 GeV. This corresponds to an upper limit on the production cross section of $\sigma(m_{\tilde{g}=1600\ GeV}) = 8.1$ fb [110]. For the event categories that are further categorised according to n_subjets, there is no observed gain in the excluded gluino mass. Thereby, gluino masses are also excluded up to 1600 GeV, and the results are discussed in more detail in Section 8.4.3.



(a) Nominal event categories.

(b) Nominal and n_{subjets} event categories.



(c) Comparison of the expected limit on the cross section, between the nominal event categories and event categories extended with n_{subjets}.

Fig. 8.9 The 95% observed upper limit on the cross section (histogram) for the T1tttt model, with the expected (dotted red line) and observed (black line) exclusion contours, on the dataset corresponding to 12.9 fb^{-1}, on both the nominal event categories (left) and the extended categories (right) The results of interpreting T1tttt are discussed in more detail in Section 8.4.3.

Table 8.4 and Table 8.5 show most sensitive n_{jet}, n_{b}, H_T and n_{subjets} categories of the benchmark mass points for the model T1tttt. The sensitivity is determined by the expected upper limit on the signal strength, r_{value}.

Signal Model	Rank (by expected limit)	Category	Signal	Prediction	Observation	Expected r-value	Observed r-value
T1tttt (1600,100)	1	$\geq$ 5j, $\geq$ 3b, 800 − inf	2.53	8.64 ± 1.09	9	1.31	1.11
T1tttt (1600,100)	2	$\geq$ 5j, 2b, 800 − inf	2.26	62.43 ± 13.69	49	2.62	5.63
T1tttt (1600,100)	3	$\geq$ 5j, 1b, 800 − inf	1.50	165.93 ± 36.81	141	7.91	9.39
T1tttt (1600,100)	4	$\geq$ 5j, 0b, 800 − inf	0.26	324.98 ± 20.67	344	83.25	37.71
T1tttt (1600,100)	5	4j, 2b, 800 − inf	0.05	13.53 ± 1.58	12	206.50	172.99
T1tttt (1600,100)	6	4j, $\geq$ 3b, 800 − inf	0.01	2.61 ± 1.09	0	416.88	233.54
T1tttt (1450,200)	1	$\geq$ 5j, $\geq$ 3b, 800 − inf	5.83	8.64 ± 1.09	9	0.63	0.55
T1tttt (1450,200)	2	$\geq$ 5j, 2b, 800 − inf	5.58	62.43 ± 13.69	49	1.16	2.32
T1tttt (1450,200)	3	$\geq$ 5j, 1b, 800 − inf	3.37	165.93 ± 36.81	141	4.70	4.94
T1tttt (1450,200)	4	$\geq$ 5j, 0b, 800 − inf	0.59	324.98 ± 20.67	344	63.25	29.18
T1tttt (1450,200)	5	4j, 2b, 800 − inf	0.05	13.53 ± 1.58	12	199.00	165.51
T1tttt (1450,200)	6	$\geq$ 5j, $\geq$ 3b, 600 − 800	0.04	11.14 ± 1.71	14	224.25	283.89
T1tttt (900,550)	1	$\geq$ 5j, $\geq$ 3b, 800 − inf	2.80	8.64 ± 1.09	9	3.09	3.15
T1tttt (900,550)	2	$\geq$ 5j, 2b, 800 − inf	4.50	62.43 ± 13.69	49	5.89	5.23
T1tttt (900,550)	3	$\geq$ 5a, $\geq$ 3b, 500 − inf	1.51	3.98 ± 0.92	7	6.45	12.33
T1tttt (900,550)	4	$\geq$ 5j, 1b, 800 − inf	3.25	165.93 ± 36.81	141	11.72	8.29
T1tttt (900,550)	5	$\geq$ 5j, $\geq$ 3b, 600 − 800	0.85	11.14 ± 1.71	14	15.06	13.59
T1tttt (900,550)	6	$\geq$ 5a, 2b, 600 − inf	0.54	4.62 ± 0.63	3	15.31	13.17

Table 8.4 Expected and observed limits for most sensitive categories of all analysis event categories of the T1tttt model, on three benchmark models of the nominal analysis.

Signal Model	Rank (by expected limit)	Category	Signal	Prediction	Observation	Expected r-value	Observed r-value
T1tttt (1600,100)	1	$\geq$ 5j, $\geq$ 3b, 0s, 800 − inf	0.92	5.26 ± 1.06	7	3.10	2.63
T1tttt (1600,100)	2	$\geq$ 5j, $\geq$ 3b, 2s, 800 − inf	0.74	2.18 ± 0.74	2	3.67	4.74
T1tttt (1600,100)	3	$\geq$ 5j, 2b, 0s, 800 − inf	1.09	43.76 ± 11.38	29	4.30	7.88
T1tttt (1600,100)	4	$\geq$ 5j, 2b, 2s, 800 − inf	0.57	13.59 ± 2.28	16	6.95	11.05
T1tttt (1600,100)	5	$\geq$ 5j, $\geq$ 3b, 3s, 800 − inf	0.37	0.81 ± 0.30	0	8.09	5.81
T1tttt (1600,100)	6	$\geq$ 5j, $\geq$ 3b, 4s, 800 − inf	0.21	0.12 ± 0.05	0	10.16	9.62
T1tttt (1450,200)	1	$\geq$ 5j, $\geq$ 3b, 0s, 800 − inf	2.29	5.26 ± 1.06	7	1.39	1.15
T1tttt (1450,200)	2	$\geq$ 5j, $\geq$ 3b, 2s, 800 − inf	1.64	2.18 ± 0.74	2	1.63	2.08
T1tttt (1450,200)	3	$\geq$ 5j, 2b, 2s, 800 − inf	1.96	13.59 ± 2.28	16	2.05	3.37
T1tttt (1450,200)	4	$\geq$ 5j, 2b, 0s, 800 − inf	2.25	43.76 ± 11.38	29	2.51	4.14
T1tttt (1450,200)	5	$\geq$ 5j, $\geq$ 3b, 4s, 800 − inf	0.60	0.12 ± 0.05	0	3.64	3.40
T1tttt (1450,200)	6	$\geq$ 5j, $\geq$ 3b, 3s, 800 − inf	0.68	0.81 ± 0.30	0	4.02	3.25
T1tttt (900,550)	1	$\geq$ 5j, $\geq$ 3b, 0s, 800 − inf	1.75	5.26 ± 1.06	7	4.02	4.51
T1tttt (900,550)	2	$\geq$ 5a, $\geq$ 3b, $\geq$ 0s, 500 − inf	1.51	5.26 ± 1.06	7	5.59	8.19
T1tttt (900,550)	3	$\geq$ 5j, 2b, 0s, 800 − inf	2.77	43.76 ± 11.38	29	5.92	5.49
T1tttt (900,550)	4	$\geq$ 5j, $\geq$ 3b, 4s, 800 − inf	0.37	0.12 ± 0.05	0	6.34	5.71
T1tttt (900,550)	5	$\geq$ 5j, 2b, 2s, 800 − inf	1.54	13.59 ± 2.28	16	6.53	8.59
T1tttt (900,550)	6	$\geq$ 5j, $\geq$ 3b, 2s, 800 − inf	0.68	2.18 ± 0.74	2	7.41	8.20

Table 8.5 Expected and observed limits for most sensitive categories of all analysis event categories of the T1tttt model, on three benchmark models, when extending the event categorisation by n_{subjets}.

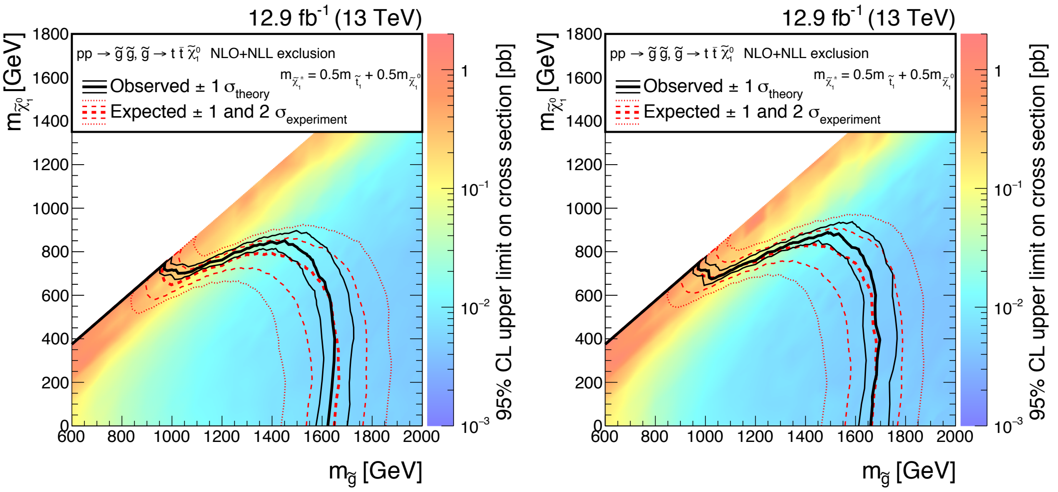
8.4.3 Discussion

As seen in Figure 8.9c, the act of further categorising events according to n_{subjets} does not improve the mass reach of gluinos with respect to the default event categories. A reason behind this observation is due to the strong sensitivity already present in the event categories and the $\not{H}_T$ dimension. Of the sensitive categories for the T1tttt model, shown in Table 8.5, the most sensitive event category contains no n_{subjets} and high n_{b}. For these event categories, it is observed that the background shape across the $\not{H}_T$ dimension is similar to the corresponding default event categories, without categorisation according to the n_{subjets} variable. As a result, no additional discrimination is gained when further categorising events according to n_{subjets}. In contrast, the most sensitive category for the T2tt model, shown in Table 8.3, contains a top quark tagged jet and $n_{\text{subjets}} = 3$. The act of tagging a top quark jet, for the model T2tt, eliminates the dominant $Z \to \nu\bar{\nu}$ background and results in a steeper shape of the $\not{H}_T$ distribution for the SM backgrounds. Therefore, a gain in sensitivity is observed for the T2tt model between the default event categories and those further classified according to n_{subjets}.

The limits achieved by the analysis are comparable to exclusions obtained by other CMS SUSY searches using the 12.9 fb^{-1}dataset. A particular comparison can be drawn to a search utilising a customised top tagger, in which collections of four jets, clustered using the default distance parameter of $R = 0.4$, are divided into all possible sets of three sets [111]. A reconstructed top quark is then one of three trijet combinations, that must carefully satisfy customised requirements. While this analysis obtains a stronger limit for gluino mediated models, in particular T1tttt, the SUSY model T2tt is excluded up to stop quark masses of 950 GeV, a limit matched when further categorising events according to n_{subjets}, as observed in Figure 8.8b.

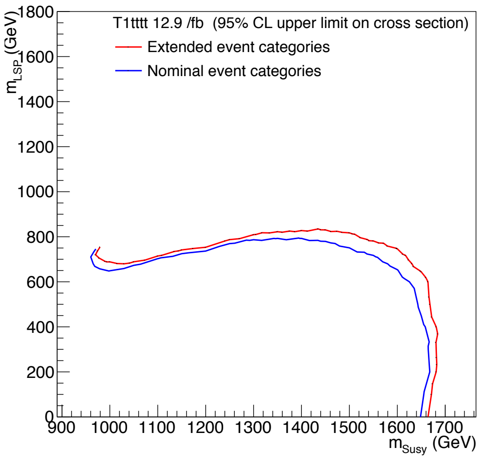
8.4.4 Utilising alternative angular variables

As discussed in Section 7.5, alternative angular variables can be employed as a replacement to $\Delta\phi^*_{min}$, to improve signal acceptance and retain the same level of QCD multijet rejection. To investigate the potential improvement of such a replacement, the full analysis workflow is repeated with the min(χ) variable, whereby events are required to satisfy a value of min(χ)$>$ 0.6. Thereafter, the data-driven QCD multijet prediction which utilises the sideband method, as described in Section 7.3, is repeated. To quantify the gain in signal acceptance, the SMS model, T1tttt is interpreted with both the nominal event categories and those extended with n_{subjets}.



(a) Nominal event categories, when utilising min(χ) instead of $\Delta\phi^*_{min}$.

(b) Nominal and n_{subjets} event categories, when utilising min(χ) instead of $\Delta\phi^*_{min}$.



(c) Comparison of the expected limit on the cross section, between the nominal event categories and n_{subjets} event categories, when utilising min(χ) instead of $\Delta\phi^*_{min}$.

Fig. 8.10 The 95% observed upper limit on the cross section (histogram) for the T1tttt model, with the expected (dotted red line) and observed (black line) exclusion contours, on the dataset corresponding to 12.9 fb^{-1}, on both the nominal event categories (left) and the extended categories (right) In each interpretation, the alternative min(χ)variable was used instead of $\Delta\phi^*_{min}$.

As seen in Figure 8.10, by replacing the variable $\Delta\phi^*_{min}$ with the alternative variable min(χ), as discussed in Section 7.5, the gluino mass reach has improved beyond the

1600 GeV gluino mass reach obtained with $\Delta\phi^{*}_{min}$. The limit on the gluino mass corresponds to an upper limit on the production cross section. Therefore, an upper limit on the gluino mass of 1675 GeV is an upper limit on the production cross section $\sigma(m_{\tilde{g}=1675\ GeV}) = 5.39$ fb [110]. Furthermore, by further categorising events according to n_{subjets}, a small gain in sensitivity is observed, which corresponds to approximately a 15% improvement on the upper limit on the production cross section.

8.5 Interpretations in Dark Matter Models

The generic signature of DM pair production in colliders is $\not{E}_T$ from the dark matter particle with recoiling energetic visible particles that satisfy the kinematic requirements of the analysis. Therefore, the phase space covered by the event categories is sensitive to DM production in association with both light and heavy flavoured jets. In the latter scenario, should the recoil energy be large enough, heavy flavoured jets could acquire a boost and thus force their decay products to collimate. To probe topologies of this nature, events further categorised according to n_{subjets} are interpreted with a simplified DM model. The event categories that are subject to jet substructure techniques correspond to the loose n_{jet} scenario, as described in Section 6.9. Signal models with a low expected n_{jet} multiplicity are subject to the t-tagging and W-tagging conditions, as detailed in Table 6.11.

8.5.1 Simplified Dark Matter models

The heavy flavour simplified models are similar to light flavour models, the difference being the production of the mediator is associated with a $\mathrm{t\bar{t}}$ or $\mathrm{b\bar{b}}$ pair. The couplings of the mediator with the standard model and dark matter particles are given by g_{SM} and g_{DM}, respectively. The recommendations on the choice of couplings is $g_{\text{SM}} = 1$, $g_{\text{DM}} = 1$ for (pseudo)scalar models, and $g_{\text{SM}} = 0.25$, $g_{\text{DM}} = 1$ for (axial-)vector models [16].

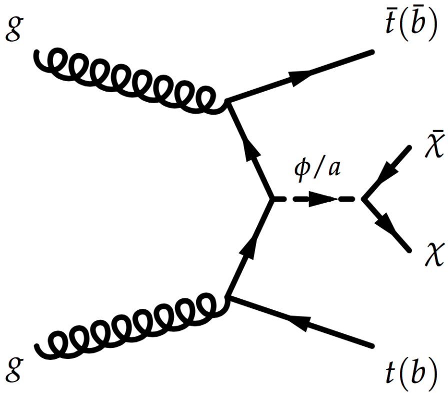


Fig. 8.11 Feynman diagram of the pair production of Dark Matter particles in association with $\mathrm{t\bar{t}}$ or $\mathrm{b\bar{b}}$.[16]

The results presented in Section 8.2 are interpreted in the context of a simplified model of dark matter as upper limits at a 95 % CL on the signal strength and production cross section, as a function of the mediator and DM mass.

The simplified model interpreted, labelled DM+$\mathrm{t\bar{t}}$, has final states with relatively large jet multiplicities and top quarks, an exemplary Feynman diagram of which is shown in Figure 8.11. The associated production cross sections for a range of DM and mediator masses, for pseudoscalar mediated DM+$\mathrm{t\bar{t}}$ models is shown in Table 8.6.

Sample	M_Φ	m_χ	$\mathrm{N}_{generated}$	σ (pb)
DM+$\mathrm{t\bar{t}}$ PS	10	1	252054	4.577e-01
DM+$\mathrm{t\bar{t}}$ PS	20	1	249253	4.117e-01
DM+$\mathrm{t\bar{t}}$ PS	50	10	231909	3.110e-01
DM+$\mathrm{t\bar{t}}$ PS	50	1	255516	3.114e-01
DM+$\mathrm{t\bar{t}}$ PS	100	10	246115	1.946e-01
DM+$\mathrm{t\bar{t}}$ PS	100	1	249971	1.945e-01
DM+$\mathrm{t\bar{t}}$ PS	200	50	232399	8.546e-02
DM+$\mathrm{t\bar{t}}$ PS	200	1	241926	8.515e-02
DM+$\mathrm{t\bar{t}}$ PS	300	50	256765	3.987e-02
DM+$\mathrm{t\bar{t}}$ PS	300	1	251079	3.971e-02
DM+$\mathrm{t\bar{t}}$ PS	500	1	241952	5.119e-03
DM+$\mathrm{t\bar{t}}$ PS	15	10	253099	1.909e-02
DM+$\mathrm{t\bar{t}}$ PS	95	50	370018	1.093e-02
DM+$\mathrm{t\bar{t}}$ PS	10	10	250666	1.532e-02
DM+$\mathrm{t\bar{t}}$ PS	50	50	236930	2.988e-03
DM+$\mathrm{t\bar{t}}$ PS	10	50	274874	2.448e-03

Table 8.6 DM+$\mathrm{t\bar{t}}$ signal simulations using pseudoscalar (PS) couplings. The number $\mathrm{N}_{generated}$ refers to the number of events generated, and σ represents the production cross section for a specific mediator, M_Φ, and DM mass m_χ [16].

8.5.2 Uncertainties on signal models

In a method analogous to determining the systematic uncertainties on SMS models, as detailed in Section 8.4.1, the simplified DM models are also subject to systematic uncertainties.

When interpreting simplified DM models additional systematic uncertainties for the prediction and correction of the signal yield are utilised. Various uncertainties on the experimental acceptance times efficiency, $A \times \varepsilon$, are considered for which typical magnitudes are summarised in Table 8.1.

The event samples for the DM+$\mathrm{t\bar{t}}$ simplified model are generated with Powheg2 [112]. The PDF of the incoming partons is included in the event generation, described by a PDF set. The acceptance sensitivity to the PDF set is investigated by comparing results using numerous commonly used PDF sets. The evaluation of the uncertainty on the choice of PDF follows the PDF4LHC recommendation of combining the output to determine an "envelope" value combining the alternative PDF sets and a related systematic uncertainty [113]. The systematic uncertainty on the parton distribution function (PDF) is found to be in the range of 1 - 9%. In addition, the uncertainty on the renormalisation scale, found to be in range of 1 - 7%, is propagated and taken as correlated across all event categories.

8.5.3 Exclusion Limits

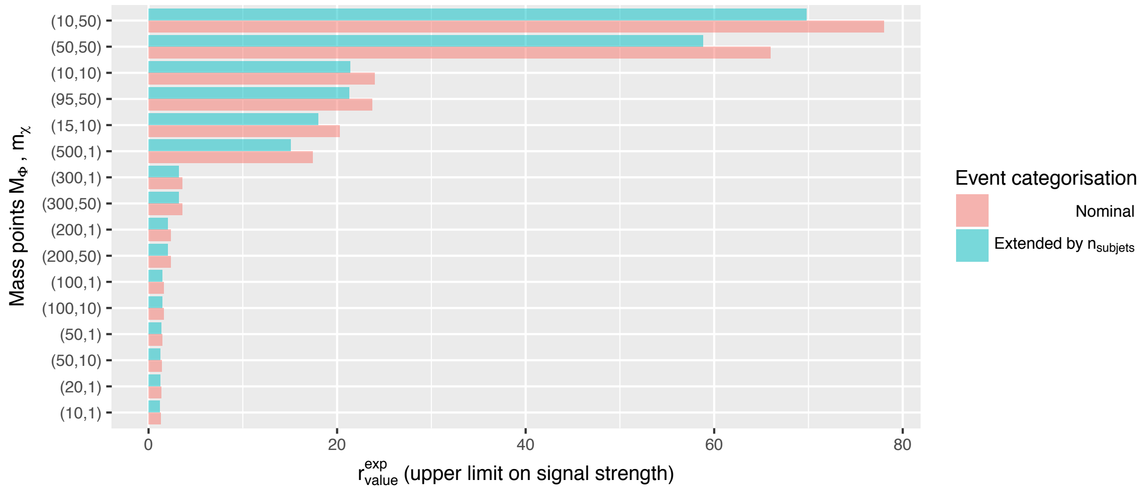
An upper limit on the cross section, for different DM and mediator masses, is computed using the CL$_s$ criterion [108], described in Section 8.3.1, and with the asymptotic formulae used to model the test statistic distribution [109].

In Table 8.7, the 95 % CL upper limits on the signal strength, r$_{value}$, are listed for different DM and mediator masses for a pseudoscalar mediated DM+t$\bar{\text{t}}$ model. The table is divided into two columns, one labelled **Nominal** to denote the results obtained with the nominal event categories, in which no t/W tagging performed. Further, an additional column labelled **Extended** is used to denote the results when event categories are further categorised according to n_{subjets}.

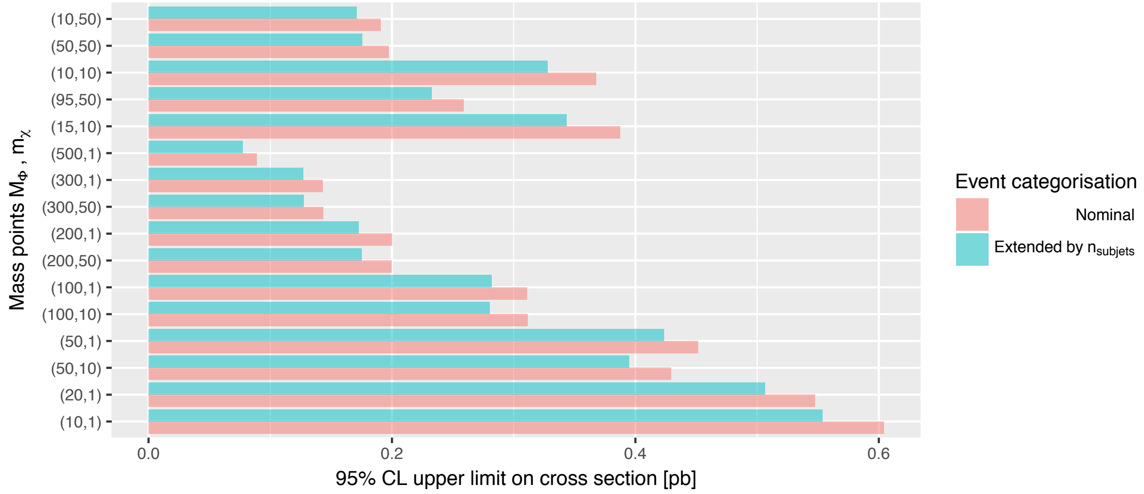
		Nominal event categories		**Extended event categories**	
M_Φ	m_χ	Expected r$_{value}$	Observed r$_{value}$	Expected r$_{value}$	Observed r$_{value}$
10	1	$1.32^{+0.77}_{-0.44}$	1.4	$1.21^{+0.71}_{-0.41}$	1.6
20	1	$1.33^{+0.77}_{-0.44}$	1.35	$1.23^{+0.71}_{-0.41}$	1.39
50	10	$1.38^{+0.81}_{-0.46}$	1.74	$1.27^{+0.73}_{-0.43}$	2.00
50	1	$1.45^{+0.84}_{-0.48}$	1.5	$1.36^{+0.75}_{-0.45}$	1.45
100	10	$1.60^{+0.93}_{-0.54}$	2.34	$1.44^{+0.87}_{-0.49}$	2.8
100	1	$1.60^{+0.94}_{-0.53}$	2.11	$1.45^{+0.86}_{-0.49}$	2.16
200	50	$2.34^{+1.40}_{-0.79}$	3.08	$2.05^{+1.26}_{-0.69}$	3.29
200	1	$2.35^{+1.41}_{-0.80}$	3.22	$2.03^{+1.26}_{-0.61}$	4.01
300	50	$3.6^{+2.2}_{-1.2}$	5.59	$3.2^{+2.01}_{-0.81}$	6.6
300	1	$3.6^{+2.2}_{-1.2}$	4.94	$3.2^{+2.01}_{-0.81}$	5.7
500	1	$17.4^{+11.0}_{-6.2}$	23.3	$15.1^{+9.2}_{-5.91}$	24.4
15	10	$20.3^{+12.0}_{-6.7}$	25.9	$18.0^{+9.7}_{-6.9}$	33.2
95	50	$23.7^{+14.0}_{-8.0}$	27.7	$21.3^{+13.1}_{-6.3}$	31.3
10	10	$24.0^{+14.1}_{-8.1}$	22.7	$21.4^{+13.2}_{-7.3}$	32.5
50	50	66^{+40}_{-22}	93.1	$58.8^{+36.2}_{-20.2}$	95.1
10	50	78^{+48}_{-26}	131.0	$69.8^{+36.2}_{-20.2}$	120.4

Table 8.7 Expected and observed limits for pseudoscalar mediated DM+t$\bar{\text{t}}$ models for both the nominal analysis with the default event categories (**Nominal**), and for the extended n_{subjets} event categories (**Extended**)

The results of Table 8.7 are also summarised in Figure 8.12. The upper limit on the signal strength, r$_{value}$, is represented as the upper limit on the production cross section, for the range of DM and mediator masses, and is additionally shown in Figure 8.12b.



(a) Expected r$_{value}$ for pseudoscalar mediated DM+t$\bar{t}$ models for both the nominal analysis with the default event categories (red), and for the extended n_{subjets} event categories (blue)



(b) Expected 95% CL upper limit on the cross section for pseudoscalar mediated DM+t$\bar{t}$ models for both the nominal analysis with the default event categories (red), and for the extended n_{subjets} event categories (blue)

Fig. 8.12 Expected r$_{value}$ (top) and 95% upper limit on the cross section (bottom) for pseudoscalar mediated DM+t$\bar{t}$ models for both the nominal analysis with the default event categories, and for the extended n_{subjets} event categories.

Further categorising events according to n_{subjets}, enhances the upper limit on the signal strength, r$_{value}$, and thus the upper limit on the production cross section for DM+t$\bar{t}$. This can be explained by the recoil of the jet system against the dark matter particle, that leads to a large momentum transfer. Such a transfer results in the jet system to acquire a boost that can result in the collimation of their decay products. In the production of DM+$t\bar{t}$, each top quark acquires a sufficiently high p_{T} that improves the sensitivity of jet substructure techniques.

CHAPTER 9

CONCLUSION AND PROSPECTS

Following the successful augmentation in centre of mass energy to $\sqrt{s} = 13$ TeV, and the deliverance of an increase in integrated luminosity, the LHC has presented a new opportunity to search for signatures of BSM theories. However, such technical changes have lead to a more extreme environment that requires a development of the individual subsystems of the detectors that are used in searching for such signatures.

During the early stages of Run 2 at the LHC, the CMS detector was subject to several upgrades that each aimed to accommodate the harsh conditions anticipated, and allow searches for signatures of BSM theories to remain sensitive to the parameter space. The efficient acceptance or rejection of events, as determined by the trigger system of the CMS detector, was maintained by the upgrade of the L1 calorimeter trigger system, particularly the replacement of the GCT with Stage 1.

On the data recorded by the CMS detector, that corresponds to an integrated luminosity of 12.9 fb^{-1}, a search for new physics has been presented. In the search, a focus on incorporating jet substructure techniques has been discussed and the results of which are interpreted under the context of SUSY and DM.

The underlying principles of the analysis rely on the suppression of the dominant QCD multijet background, made through particular requirements on the α_{T} and $\Delta\phi^{*}_{min}$ variables. The remaining backgrounds from SM processes, that contain processes with a genuine source of $\not{E}_T$ are predicted using a data-driven technique and control regions. The act of predicting the SM backgrounds in the signal region involves the extrapolation of event yields, of a given event category in a control region to that of a signal region, through the use of a transfer factor. Thereafter, the distribution of the $\not{H}_T$ variable across the event categories is investigated and a maximum likelihood fit is performed for a series of signal hypotheses.

As a means to further enhance the sensitivity, jet substructure techniques in which jets, reconstructed with a larger distance parameter and then subject to grooming algorithms, were considered. The act of tagging such jets as top quarks or W bosons, introduces an additional discrimination between events with such boosted particles and those with quark-gluon origins. The event category n_{subjets} was used to further categorise events according to the number of subjets associated to a t-tagged or a W-tagged jet.

The data observed in the signal region is statistically compatible with expected event yields from background processes. As no evidence for new physics is observed, confidence limits are determined on the masses of sparticles, in the context of supersymmetry, and the masses of mediators and dark matter candidates, in the context of dark matter. With respect to the default event categories of the analysis in which t-tagging and W-tagging are not performed, the use of jet substructure techniques, applied through the additional event category n_{subjets}, improved the upper limit on stop quark masses by 50 GeV to exclude stop masses up to 970 GeV, a limit compatible in strength across other analyses of the ATLAS and CMS collaboration. In addition, an improvement on the upper limit on the production cross section of a dark matter model is observed when further categorising events according to n_{subjets}. However, no additional sensitivity was achieved in gluino masses as categorising events according to n_{subjets}, in addition to utilising the $\not{H}_T$ dimension, did not alter the background topologies in a way that provides further discrimination to the signal.

Furthermore, a reinterpretation of the angular variable $\Delta\phi^*_{min}$ to provide a more effective means to suppress the QCD multijet background and enhance the signal acceptance shows a direct gain in the results, with respect to utilising the $\Delta\phi^*_{min}$ variable. The use of the $\min(\chi)$ variable enhances the sensitivity on gluino masses to set a limit of up to 1675 GeV. While this limit is comparable, with respect to other analyses in the CMS and ATLAS collaborations, it also demonstrates the potential effect that revisiting pre-established variables has.

Across the physics programmes at the ATLAS and CMS collaborations, no evidence has been found for BSM physics, and resultant upper limits on the parameter space set by each search are often equivalent in size. For hadronic SUSY searches in particular, an increase in the centre of mass energy and improvements to individual analyses provided significant gains in the excluded parameter space. However, the bounds on the masses of gluinos and squarks at 1 - 2 TeV questions the possibility of the LHC to explore such mass ranges for gluinos and squarks. Such limits should not discourage

SUSY, rather demand an alternative and novel approach to reach such mass ranges.

For searches targeting alternative BSM scenarios, several remain well motivated and detectable at the LHC. Examples of which in the SUSY sector are RPV models [114], which do not provide a DM candidate but evade constraints from searches which require $\not{E}_T$, and "hidden valley" models [115], which often lead to final states with multiple soft jets and little $\not{E}_T$. Furthermore, the continual search for DM and its plethora of simplified models remain well motivated and fruitful at the LHC.

While the LHC continues to analyse data up to an integrated luminosity of 3000 fb^{-1}, the best opportunity to discover a signature of BSM physics will likely come from an effective change to analysis techniques. With important developments in Run I and Run II, during which a solid foundation for the field was built, jet substructure techniques have shown to be an example of such a change.

In an analogous way in which the origin and composition of the universe is understood, the use of jet substructure tools, together with taggers, have provided discrimination between signals and backgrounds by studying the architecture of jets that in turn provide a clearer perspective of their origins. Several advances in jet substructure techniques have been made, with developments on the interior algorithms of jets and the use of jet substructure variables with multivariate techniques. Furthermore, the evolution of resolved tagging has removed the dependency of boosted topologies, often associated to BSM signatures, and in some cases has improved sensitivity to SM searches.

While future searches aim to exploit the statistical strength of updated data sets and the technical gains made in the trigger system and in analysis methodologies, the continued search for new physics at the LHC will become increasingly more challenging. However, it is evident that by introducing well established and proven methods such as jet substructure techniques, or re-optimising predefined analysis variables to provide a deeper comprehension, such future searches will have the best chance yet of detecting new physics.

Appendix A

Binning labels

$n_{\rm jet}$ categories are denoted by a four letting string that indicates the jet multiplicity and for which topology, where *a* and *j* represent asymmetric and symmetric respectively. The $n_{\rm b}$ and $n_{\rm subjets}$ categories are indicated with the letter *b* and *s* respectively.

For each category, the multiplicity is given by the prefix *eq* or *ge* which represents, $=$ and $\geq$.

The H_T category of a given event category is labelled as a character that contains the lower bin edge in GeV.

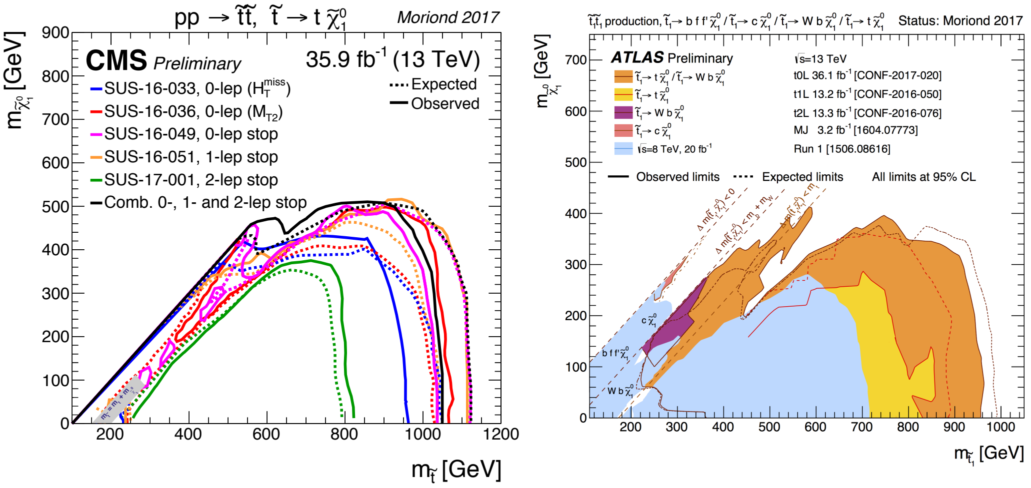
An example of an analysis event category is, *HT800 ge5j eq2b eq3s* which represents the event category $H_T \geq 800$ GeV, $n_{\rm jet} \geq 5$, $n_{\rm b} = 2$ and $n_{\rm subjets} = 3$. The full collection of binning labels and their corresponding meaning is shown in Table A.1.

$n_{\rm jet}$ (asymmetric)	$n_{\rm b}$	H_T	$n_{\rm subjets}$
ge5j(a) $\equiv \geq 5$	ge3b $\equiv \geq 3$	HT800 $\equiv H_T > 800$ GeV	ge6s $\equiv \geq 6$
eq4j(a) $= 4$	eq2b $= 2$	HT600 $\equiv 600 < H_T < 800$ GeV	eq5s $= 5$
eq3j(a) $= 3$	eq1b $= 1$	HT500 $\equiv 500 < H_T < 600$ GeV	eq4s $= 4$
eq2j(a) $= 2$	eq0b $= 0$	HT400 $\equiv 400 < H_T < 500$ GeV	eq3s $= 3$
eq1j(a) $= 1$		HT350 $\equiv 350 < H_T < 400$ GeV	eq2s $= 2$
		HT300 $\equiv 300 < H_T < 350$ GeV	eq0s $= 0$
		HT250 $\equiv 250 < H_T < 300$ GeV	
		HT200 $\equiv 200 < H_T < 250$ GeV	

Table A.1 The binning labels of the analysis.

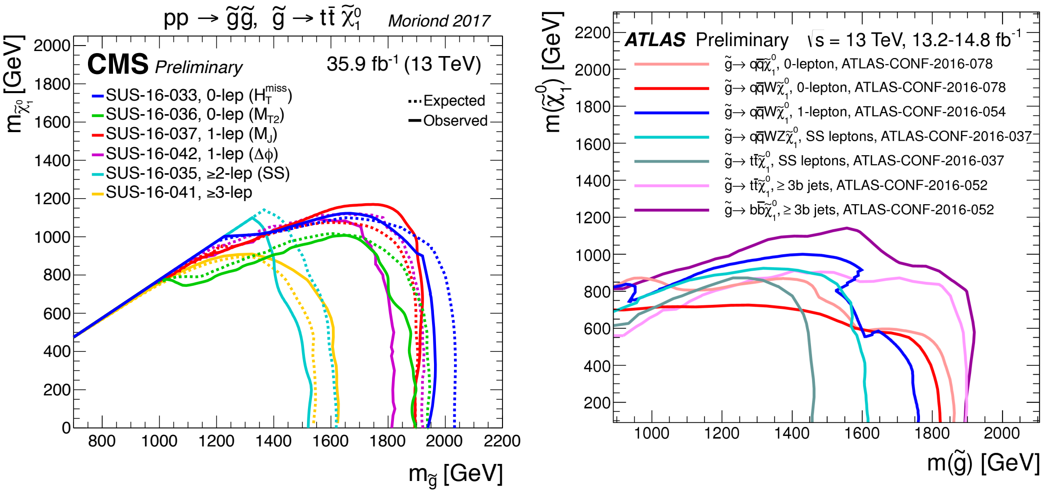
Appendix B

Current Experimental Limits



(a) Current limits of direct stop production from the CMS collaboration.

(b) Current limits of direct stop production from the ATLAS collaboration.



(c) Current limits of direct gluino production from the CMS collaboration.

(d) Current limits of direct gluino production from the ATLAS collaboration.

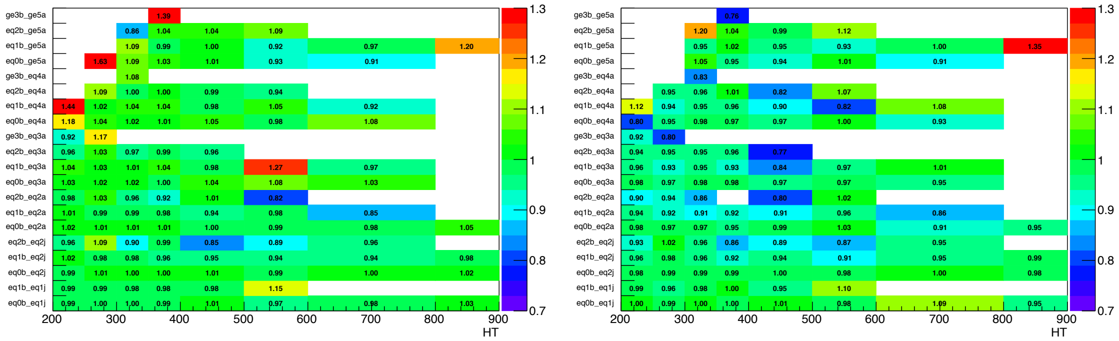
Fig. B.1 Current supersymmetry limits from the CMS [17] and ATLAS [18] collaborations.

Appendix C

Simulation driven uncertainties

The difference in the transfer factors, under variations of the known sources of systematic uncertainties are displayed for both the nominal event categories, in which events are not further categorised according to n_{subjets}, and the extended event categories, where events are categorised by n_{subjets}.

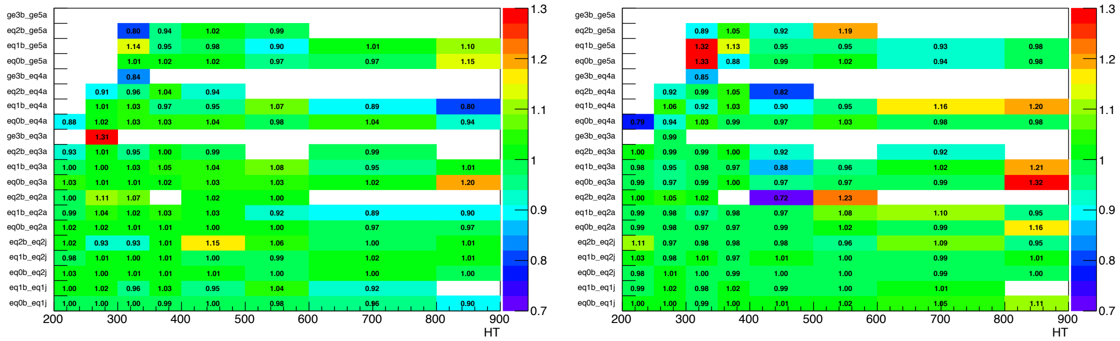
Nominal event categories



(a) $+\sigma$ variations.

(b) $-\sigma$ variations.

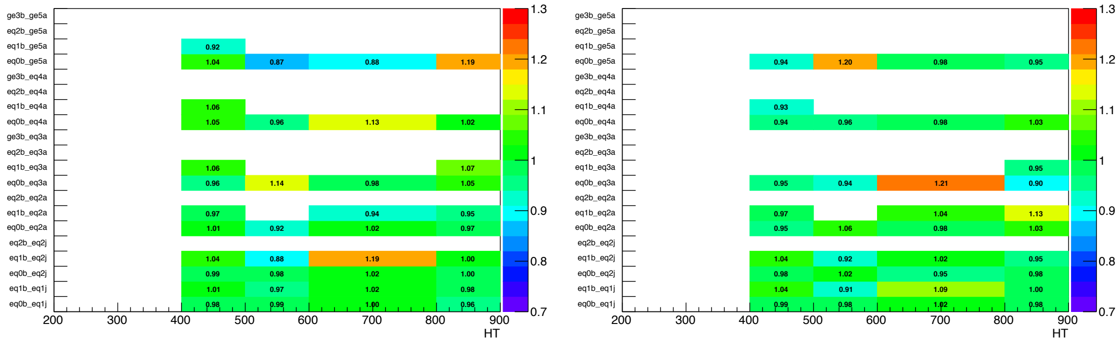
Fig. C.1 The relative change in the μ+jets $\rightarrow Z \rightarrow \nu\bar{\nu}$ transfer factors, in the nominal event categories, when varying the JEC corrections within its uncertainties, as a function of H_T. Variations according to the $+\sigma$ (left) and $-\sigma$ (right) are shown.



(a) +σ variations.

(b) -σ variations.

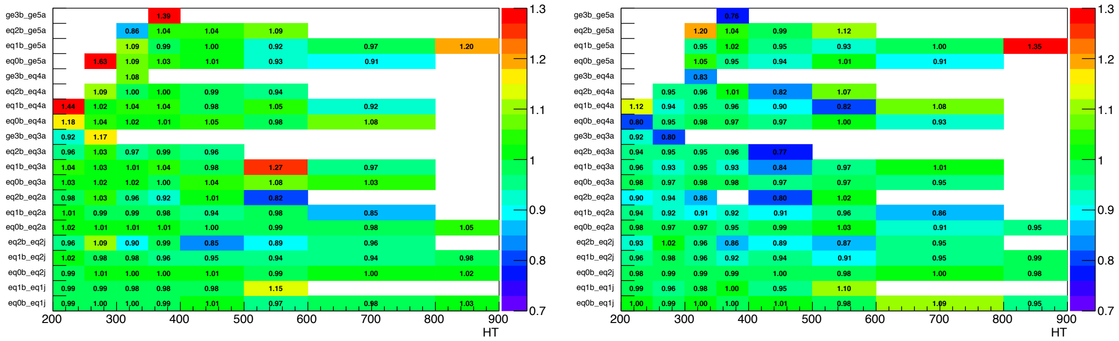
Fig. C.2 The relative change in the $\mu\mu$+jets $\rightarrow Z \rightarrow \nu\bar{\nu}$ transfer factors, in the nominal event categories, when varying the JEC corrections within its uncertainties, as a function of H_T. Variations according to the +σ (left) and -σ (right) are shown.



(a) +σ variations.

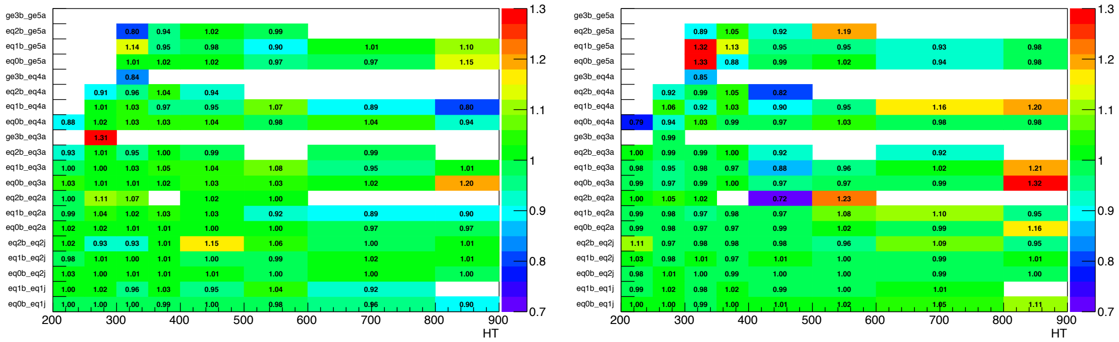
(b) -σ variations.

Fig. C.3 The relative change in the γ+jets $\rightarrow Z \rightarrow \nu\bar{\nu}$ transfer factors, in the nominal event categories, when varying the JEC corrections within its uncertainties, as a function of H_T. Variations according to the +σ (left) and -σ (right) are shown.



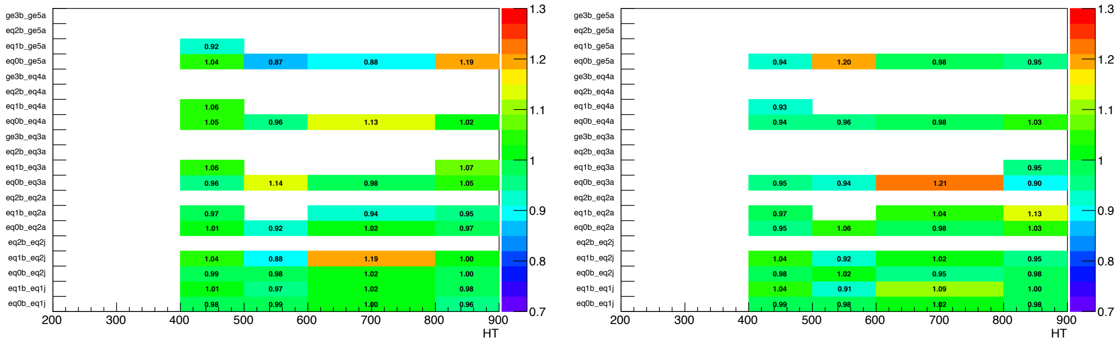
(a) +σ variations. (b) -σ variations.

Fig. C.4 The relative change in the μ+jets $\to Z \to \nu\bar{\nu}$ transfer factors, in the nominal event categories, when varying the JEC corrections within its uncertainties, as a function of H_T. Variations according to the +σ (left) and -σ (right) are shown.



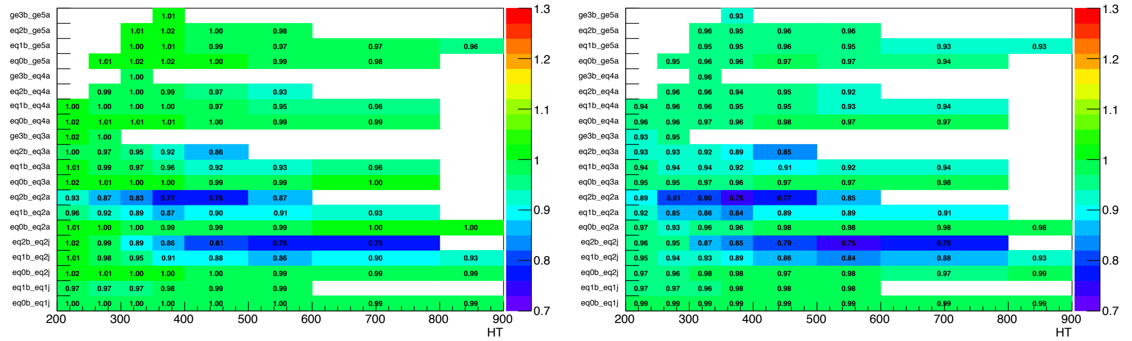
(a) +σ variations. (b) -σ variations.

Fig. C.5 The relative change in the $\mu\mu$+jets $\to Z \to \nu\bar{\nu}$ transfer factors, in the nominal event categories, when varying the JEC corrections within its uncertainties, as a function of H_T. Variations according to the +σ (left) and -σ (right) are shown.



(a) +σ variations. (b) -σ variations.

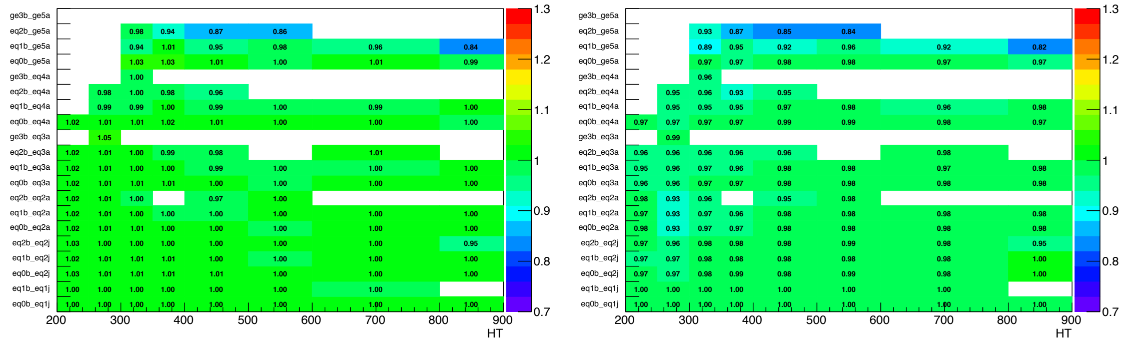
Fig. C.6 The relative change in the γ+jets $\to Z \to \nu\bar{\nu}$ transfer factors, in the nominal event categories, when varying the JEC corrections within its uncertainties, as a function of H_T. Variations according to the +σ (left) and -σ (right) are shown.



(a) $+\sigma$ variations.

(b) $-\sigma$ variations.

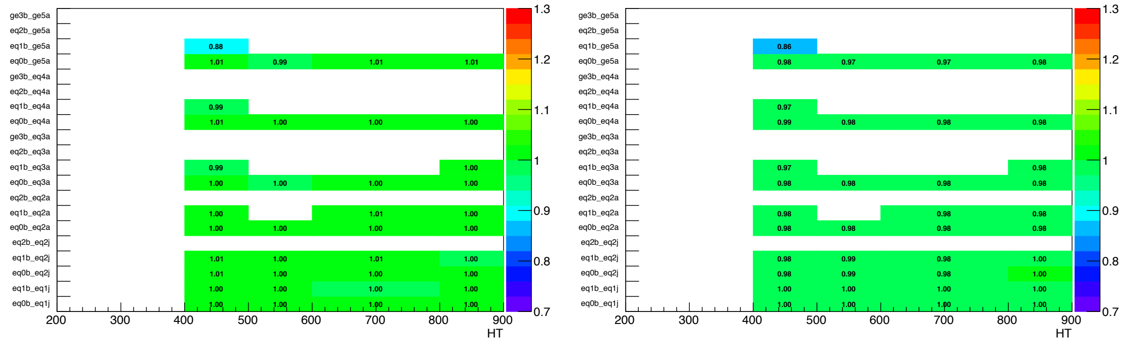
Fig. C.7 The relative change in the μ+jets $\rightarrow Z \rightarrow \nu\bar{\nu}$ transfer factors, in the nominal event categories, when varying the trigger weight in MC within its uncertainties, as a function of H_T. Variations according to the $+\sigma$ (left) and $-\sigma$ (right) are shown.



(a) $+\sigma$ variations.

(b) $-\sigma$ variations.

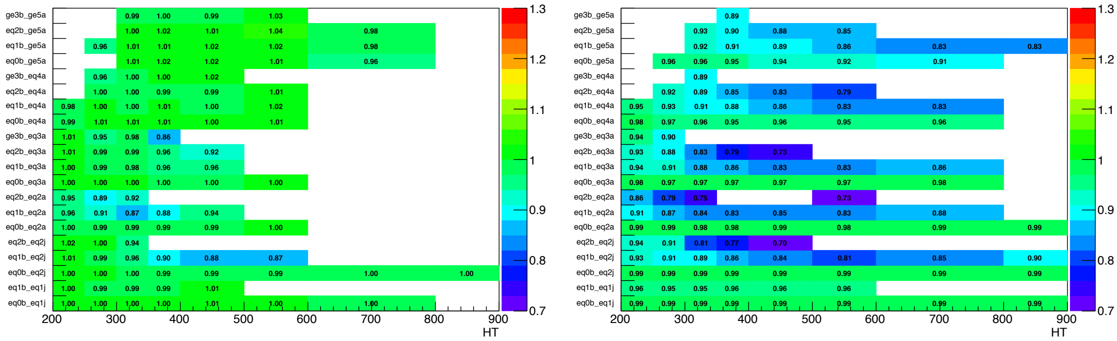
Fig. C.8 The relative change in the $\mu\mu$+jets $\rightarrow Z \rightarrow \nu\bar{\nu}$ transfer factors, in the nominal event categories, when varying the trigger weight in MC within its uncertainties, as a function of H_T. Variations according to the $+\sigma$ (left) and $-\sigma$ (right) are shown.



(a) $+\sigma$ variations.

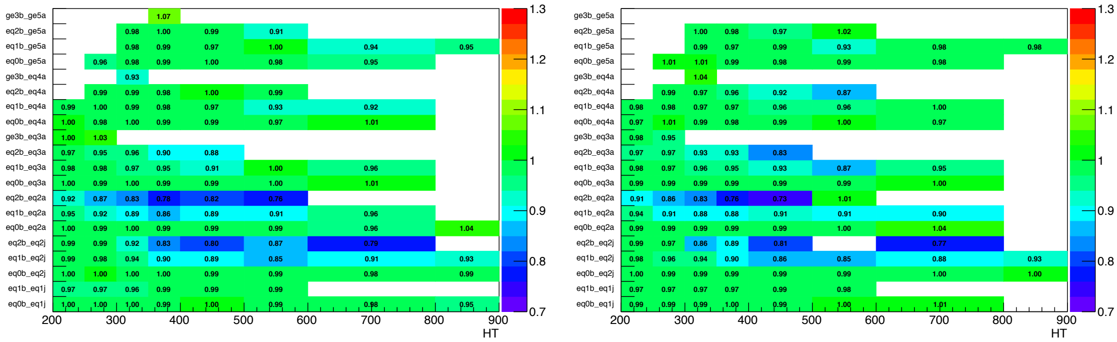
(b) $-\sigma$ variations.

Fig. C.9 The relative change in the γ+jets $\rightarrow Z \rightarrow \nu\bar{\nu}$ transfer factors, in the nominal event categories, when varying the trigger weight in MC within its uncertainties, as a function of H_T. Variations according to the $+\sigma$ (left) and $-\sigma$ (right) are shown.



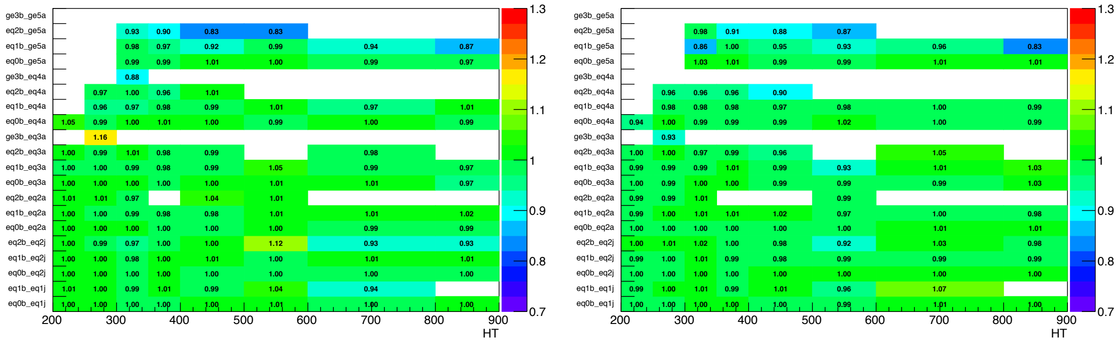
(a) $+\sigma$ variations. (b) -σ variations.

Fig. C.10 The relative change in the μ+jets $\rightarrow Z \rightarrow \nu\bar{\nu}$ transfer factors, in the nominal event categories, when varying the top p_{T} weight in MC within its uncertainties, as a function of H_T. Variations according to the $+\sigma$ (left) and -σ (right) are shown.



(a) $+\sigma$ variations. (b) -σ variations.

Fig. C.11 The relative change in the $\mu\mu$+jets $\rightarrow Z \rightarrow \nu\bar{\nu}$ transfer factors, in the nominal event categories, when varying the PU weight in MC within its uncertainties, as a function of H_T. Variations according to the $+\sigma$ (left) and -σ (right) are shown.



(a) $+\sigma$ variations. (b) -σ variations.

Fig. C.12 The relative change in the γ+jets $\rightarrow Z \rightarrow \nu\bar{\nu}$ transfer factors, in the nominal event categories, when varying the PU weight in MC within its uncertainties, as a function of H_T. Variations according to the $+\sigma$ (left) and -σ (right) are shown.

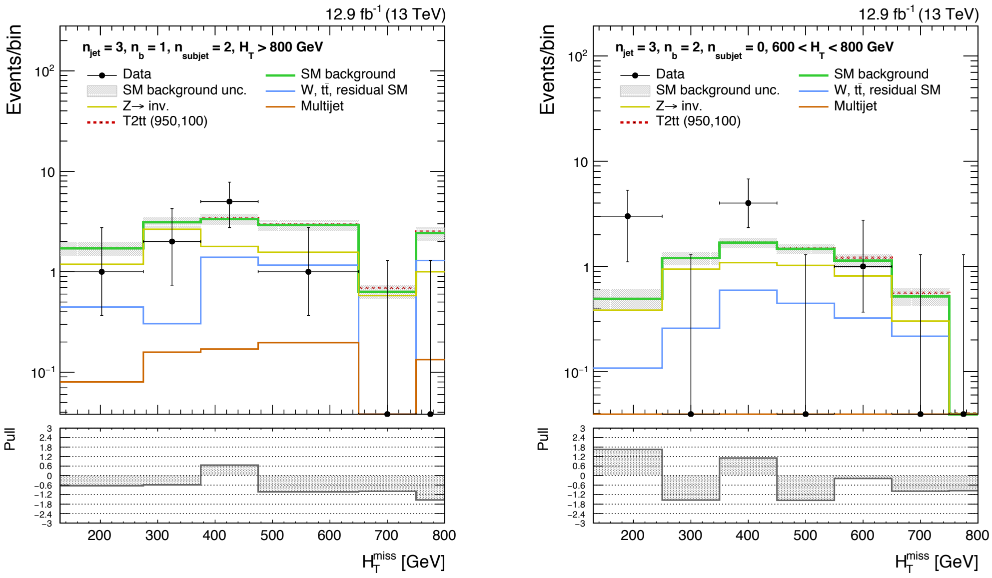
Appendix D

$\not{H}_T$ DISTRIBUTIONS

Event categories classified according to n_{subjets}

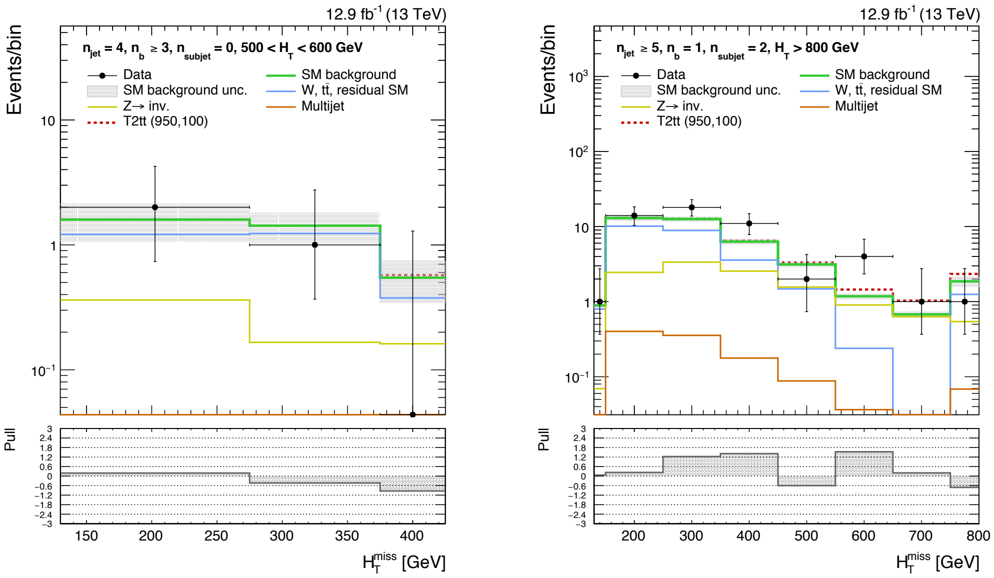
Observed and predicted event counts for events in the asymmetric and symmetric jet categories, in a collection of the H_T, n_{jet}, n_{b} and n_{subjets} event categories, across the $\not{H}_T$ dimension. Observed data counts are represented by the black filled markers and predicted backgrounds are represented by the solid histograms.

The distributions shown are associated to the signal model, T2tt, the distribution of which is shown by the dashed red line. The ratio in the bottom panel shows the pull of background predictions with respect to the observed data counts.



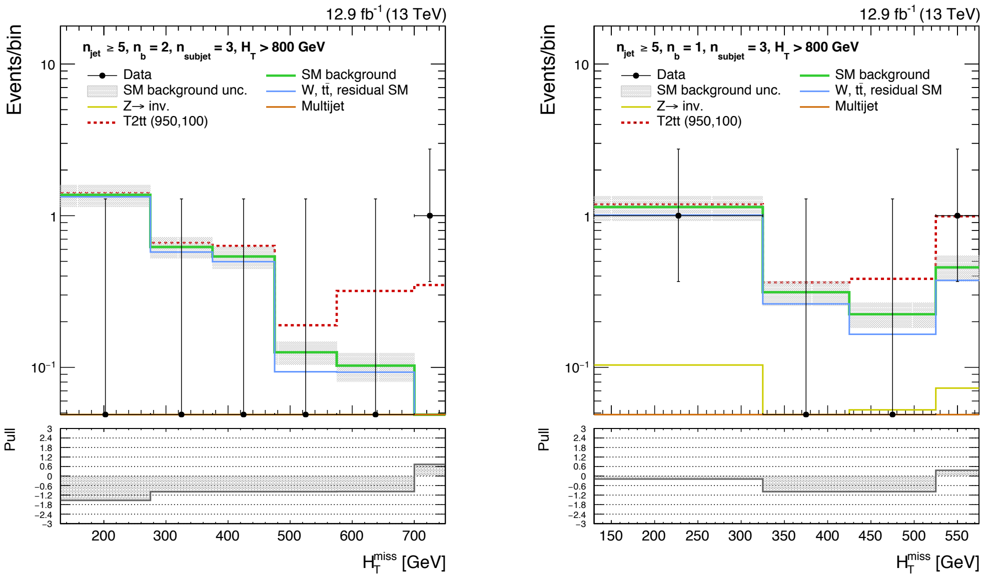
(a) $\not{H}_T$ shape in the symmetric topology for the event categories $n_{\mathrm{b}} = 1$, $n_{\mathrm{jet}} = 3$, $n_{\mathrm{subjets}} = 3$ and $H_T > 800$ GeV.

(b) $\not{H}_T$ shape in the symmetric topology for the event categories $n_{\mathrm{b}} = 2$, $n_{\mathrm{jet}} = 3$, $n_{\mathrm{subjets}} \geq 0$ and $600 < H_T < 800$ GeV.



(a) $\not{H}_T$ shape in the symmetric topology for the event categories $n_\mathrm{b} \geq 3$, $n_\mathrm{jet} = 4$, $n_\mathrm{subjets} \geq 0$ and 500 $< H_T <$ 600 GeV.

(b) $\not{H}_T$ shape in the symmetric topology for the event categories $n_\mathrm{b} = 1$, $n_\mathrm{jet} \geq 5$, $n_\mathrm{subjets} = 2$ and $H_T > 800$ GeV.



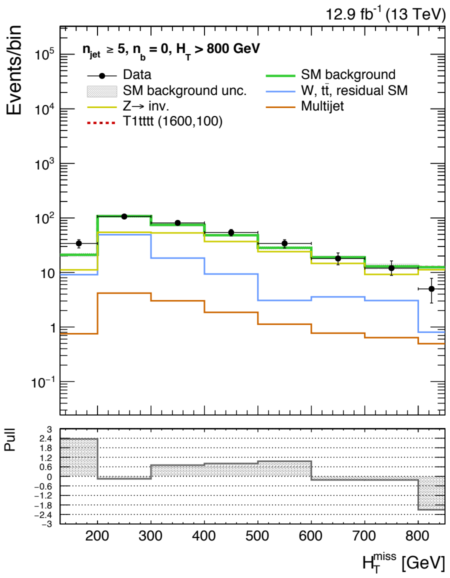
(c) $\not{H}_T$ shape in the symmetric topology for the event categories $n_\mathrm{b} = 2$, $n_\mathrm{jet} \geq 5$, $n_\mathrm{subjets} = 3$ and $H_T > 800$ GeV.

(d) $\not{H}_T$ shape in the symmetric topology for the event categories $n_\mathrm{b} = 1$, $n_\mathrm{jet} \geq 5$, $n_\mathrm{subjets} = 3$ and $H_T > 800$ GeV.

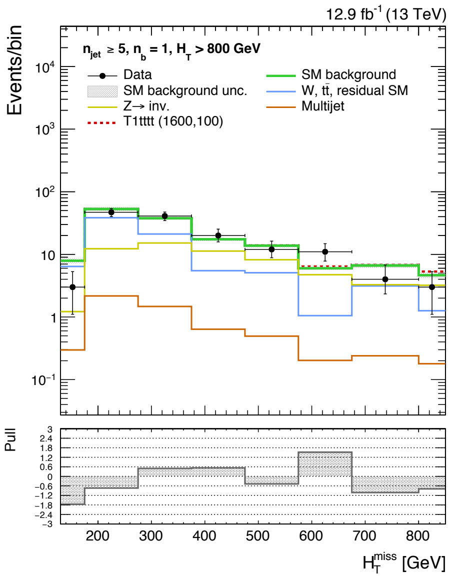
Fig. D.2 The total event yields in data and the SM expectation with their associated uncertainties as a function of $\not{H}_T$ for events in the signal region further categorised according to n_subjets. The distributions below represent the significance of deviations, denoted as a pull, observed in data with respect to the SM expectations.

Nominal event categories

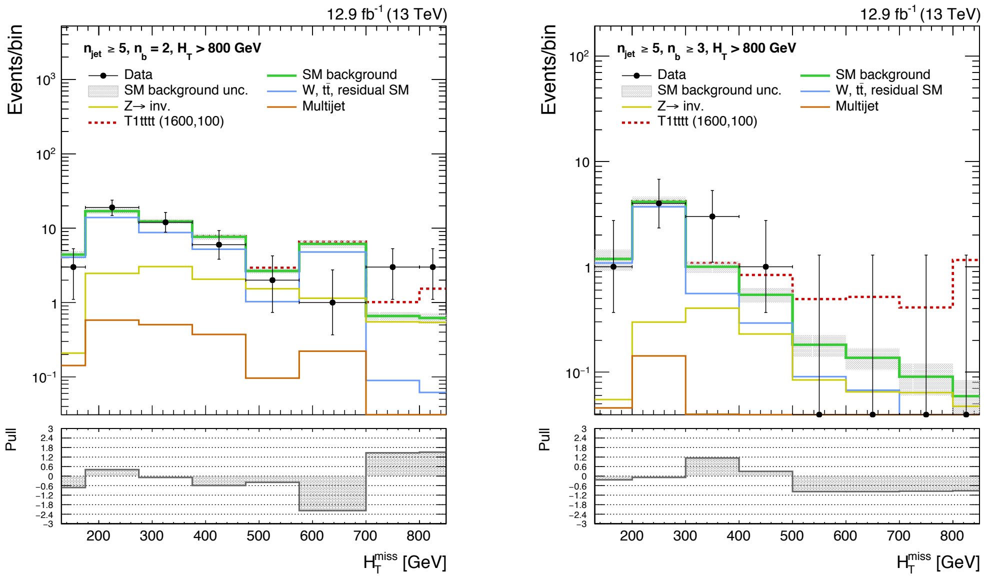
Observed and predicted event counts for events in the symmetric jet categories, in sensitive H_T, n_{jet} and n_{b} event categories, across the $\not{H}_T$ dimension. The distributions shown are associated to the signal model, T1tttt, the distribution of which is shown by the dashed red line.



(a) $\not{H}_T$ shape in the symmetric topology for the event categories $n_{\mathrm{b}} = 0$, $n_{\mathrm{jet}} \geq 5$ and $H_T > 800$ GeV.

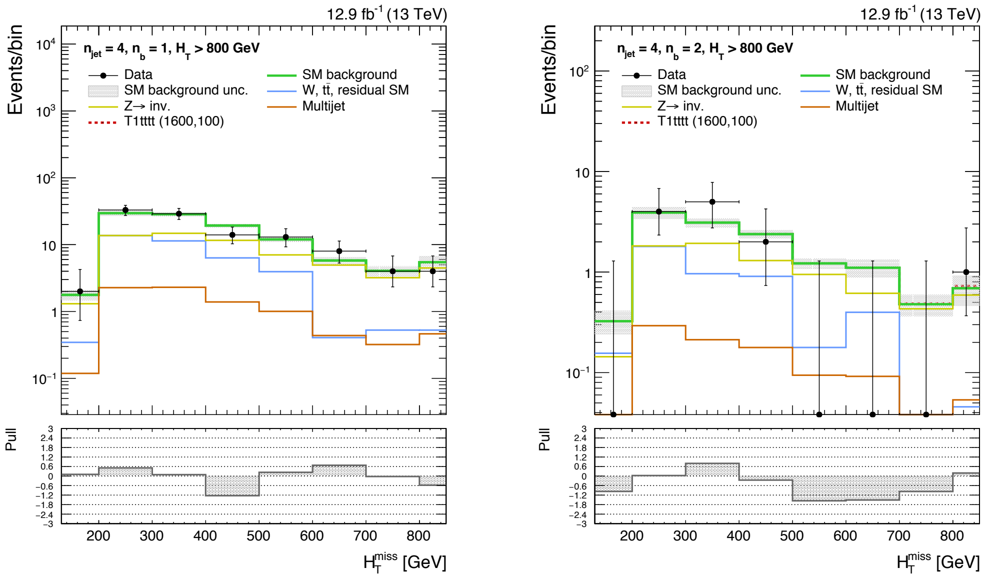


(b) $\not{H}_T$ shape in the symmetric topology for the event categories $n_{\mathrm{b}} = 1$, $n_{\mathrm{jet}} \geq 5$ and $H_T > 800$ GeV.



(a) $\not{H}_T$ shape in the symmetric topology for the event categories $n_\mathrm{b} = 2$, $n_\mathrm{jet} \geq 5$ and $H_T > 800$ GeV.

(b) $\not{H}_T$ shape in the symmetric topology for the event categories $n_\mathrm{b} \geq 3$, $n_\mathrm{jet} \geq 5$ and $H_T > 800$ GeV.



(c) $\not{H}_T$ shape in the symmetric topology for the event categories $n_\mathrm{b} = 1$, $n_\mathrm{jet} = 4$ and $H_T > 800$ GeV.

(d) $\not{H}_T$ shape in the symmetric topology for the event categories $n_\mathrm{b} = 2$, $n_\mathrm{jet} = 4$ and $H_T > 800$ GeV.

Fig. D.4 The total event yields in data and the SM expectation with their associated uncertainties as a function of $\not{H}_T$ for events in the signal region. The distributions below represent the significance of deviations, denoted as a pull, observed in data with respect to the SM expectations.

Appendix E

Likelihood fit modelling

Pulls of the observation compared to the prediction from the control region only fit, discussed in Section 8.1, for all the event categories used in the analysis.

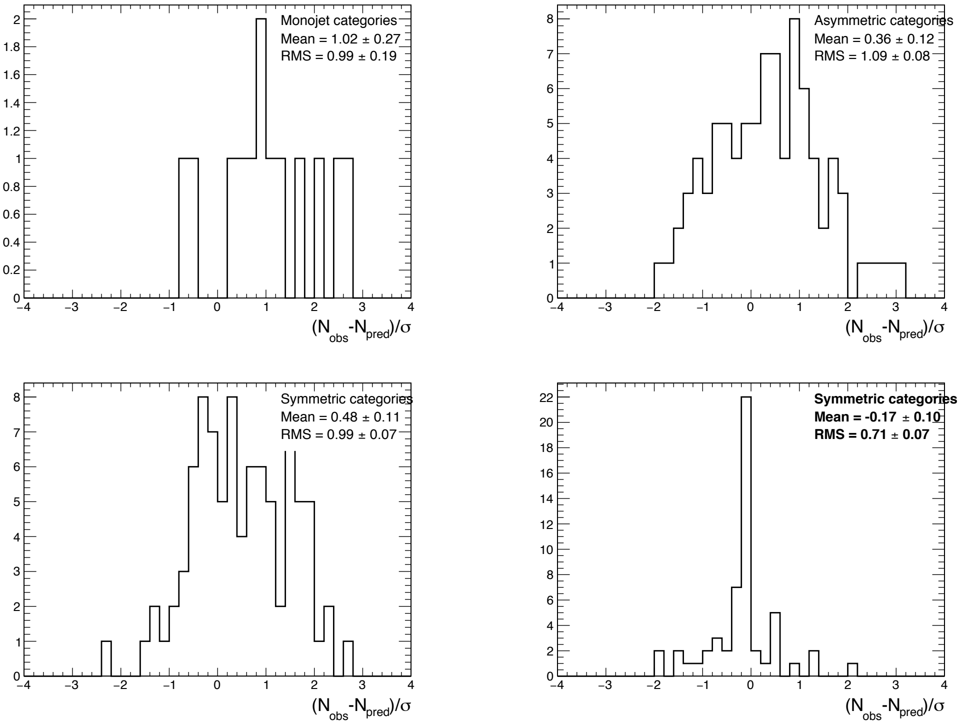


Fig. E.1 Pulls of the observation compared to the prediction from the CR-only fit in the monojet categories (top-left), asymmetric categories (top-right), symmetric categories (bottom-left) and n_{subjets} categories (bottom-right). The pulls are normalised to the total prediction uncertainty, including both the systematic and statistical component.

.

BIBLIOGRAPHY

[1] S. P. Martin, “A Supersymmetry primer”, *Adv.Ser.Direct.High Energy Phys.* **21** (2010) 1–153, `doi:10.1142/9789814307505_0001`, `arXiv:hep-ph/9709356`.

[2] C. Borschensky et al., “Squark and gluino production cross sections in pp collisions at $\sqrt{s} = 13, 14, 33$ and 100 TeV”, *Eur. Phys. J.* **C74** (2014), no. 12, 3174, `doi:10.1140/epjc/s10052-014-3174-y`, `arXiv:1407.5066`.

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

[4] T. Sakuma and T. McCauley, “Detector and Event Visualization with SketchUp at the CMS Experiment”, *J.Phys.Conf.Ser.* **513** (2014) 022032, `doi:10.1088/1742-6596/513/2/022032`, `arXiv:1311.4942`.

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

[6] CMS Collaboration, G. L. Bayatian et al., “CMS Physics: Technical Design Report Volume 1: Detector Performance and Software”. Technical Design Report CMS. CERN, Geneva, 2006.

[7] S. Dambach, U. Langenegger, and R. Horisberger, “CMS pixel module readout optimization and study of the lifetime in the semileptonic decay mode”,.

[8] Brooke.J, Cussans.D, Frazier.R and Ali Shah.A, “Hardware and Firmware for the CMS Global Calorimeter Trigger”, 2014.

[9] B. Kreiset al., “Run 2 upgrades to the CMS Level-1 calorimeter trigger”, *Journal of Instrumentation* **11** (2016), no. 01, C01051.

[10] CMS Collaboration, “Identification of b and c jets at the CMS Experiment in the LHC Run 2”, CMS-PAS-BTV-16-002, CERN, Geneva, (2017). Manuscript, 2017.

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

[12] J. Thaler and K. Van Tilburg, “Identifying Boosted Objects with N-subjettiness”, *JHEP* **03** (2011) 015, `doi:10.1007/JHEP03(2011)015`, `arXiv:1011.2268`.

[13] CMS Collaboration, “CMS Luminosity - Public Results”. https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults. Accessed: 2017-06-10.

[14] 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”, CMS-PAS-SUS-16-016, CERN, Geneva, (2016).

[15] T. Sakuma, H. Flaecher, and D. Smith, “Alternative angular variables for QCD multijet background event suppression in SUSY searches at the LHC.”,. Manuscript, 2017.

[16] D. Abercrombie et al., “Dark Matter Benchmark Models for Early LHC Run-2 Searches: Report of the ATLAS/CMS Dark Matter Forum”, `arXiv:1507.00966`.

[17] CMS Collaboration, “CMS Supersymmetry Physics Results”. https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSUS. Accessed: 2017-08-29.

[18] ATLAS Collaboration, “ATLAS Supersymmetry (SUSY) searches”. https://twiki.cern.ch/twiki/bin/view/AtlasPublic/SupersymmetryPublicResults. Accessed: 2017-08-29.

[19] Particle Data Group Collaboration, “Review of Particle Physics”, *Chin.Phys.* **C38** (2014) 090001, `doi:10.1088/1674-1137/38/9/090001`.

[20] CMS Collaboration, “W/Z-tagging of Jets”. https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetWTagging. Accessed: 2017-04-16.

[21] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys.Lett.* **B716** (2012) 1–29, `doi:10.1016/j.physletb.2012.08.020`, `arXiv:1207.7214`.

[22] 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`.

[23] ALEPH, CDF, D0, DELPHI, L3, OPAL, SLD, LEP Electroweak Working Group, Tevatron Electroweak Working Group, SLD Electroweak and Heavy Flavour Groups Collaboration, "Precision Electroweak Measurements and Constraints on the Standard Model", `arXiv:1012.2367`.

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

[25] S. L. Glashow, "Partial-Symmetries of Weak Interactions", *Nuclear Physics* **22** (1961), no. 4, 579 – 588, `doi:http://dx.doi.org/10.1016/0029-5582(61)90469-2`.

[26] A. Salam, "Weak and Electromagnetic Interactions", *Conf.Proc.* **C680519** (1968) 367–377.

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

[28] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys. Rev. Lett.* **13** (Aug, 1964) 321–323, `doi:10.1103/PhysRevLett.13.321`.

[29] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (Oct, 1964) 508–509, `doi:10.1103/PhysRevLett.13.508`.

[30] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global Conservation Laws and Massless Particles", *Phys. Rev. Lett.* **13** (Nov, 1964) 585–587, `doi:10.1103/PhysRevLett.13.585`.

[31] Gfitter Group Collaboration, "The global electroweak fit at NNLO and prospects for the LHC and ILC", *Eur.Phys.J.* **C74** (2014) 3046, `doi:10.1140/epjc/s10052-014-3046-5`, `arXiv:1407.3792`.

[32] A. Riotto and M. Trodden, “Recent progress in baryogenesis”, *Ann.Rev.Nucl.Part.Sci.* **49** (1999) 35–75, `doi:10.1146/annurev.nucl.49.1.35`, `arXiv:hep-ph/9901362`.

[33] W.-S. Hou, “Source of CP Violation for the Baryon Asymmetry of the Universe”, *Chin.J.Phys.* **47** (2009) 134, `arXiv:0803.1234`.

[34] Y. Fukuda et al., “Evidence for Oscillation of Atmospheric Neutrinos”, *Phys. Rev. Lett.* **81** (Aug, 1998) 1562–1567, `doi:10.1103/PhysRevLett.81.1562`, `arXiv:hep-ex/9807003`.

[35] L. E. Strigari, “Galactic Searches for Dark Matter”, *Phys. Rept.* **531** (2013) 1–88, `doi:10.1016/j.physrep.2013.05.004`, `arXiv:1211.7090`.

[36] M. Markevitch 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`.

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

[38] WMAP Collaboration, “Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Data Processing, Sky Maps, and Basic Results”, *Astrophys. J. Suppl.* **180** (2009) 225–245, `doi:10.1088/0067-0049/180/2/225`, `arXiv:0803.0732`.

[39] ATLAS Collaboration and CMS Collaboration, “Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments”, *Phys. Rev. Lett.* **114** (May, 2015) 191803, `doi:10.1103/PhysRevLett.114.191803`, `arXiv:1503.07589`.

[40] Y. A. Gol’fand and E. P. Likhtman, “Extension of the Algebra of Poincaré Group Generators and Violation of p Invariance”, *JETP Lett.* **13** (1971) 323.

[41] J. Wess and B. Zumino, “Supergauge transformations in four dimensions”, *Nucl. Phys. B* **70** (1974) 39, `doi:10.1016/0550-3213(74)90355-1`.

[42] H. P. Nilles, “Supersymmetry, Supergravity and Particle Physics”, *Phys. Reports* **110** (1984) 1, `doi:10.1016/0370-1573(84)90008-5`.

[43] H. Haber and G. Kane, “The Search for Supersymmetry: Probing Physics Beyond the Standard Model”, *Phys. Reports* **117** (1987) 75, `doi:10.1016/0370-1573(85)90051-1`.

[44] R. Barbieri, S. Ferrara, and C. A. Savoy, “Gauge Models with Spontaneously Broken Local Supersymmetry”, *Phys. Lett. B* **119** (1982) 343, `doi:10.1016/0370-2693(82)90685-2`.

[45] S. Dawson, E. Eichten, and C. Quigg, “Search for Supersymmetric Particles in Hadron - Hadron Collisions”, *Phys. Rev. D* **31** (1985) 1581, `doi:10.1103/PhysRevD.31.1581`.

[46] 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`.

[47] G. R. Farrar and P. Fayet, “Phenomenology of the production, decay, and detection of new hadronic states associated with supersymmetry”, *Physics Letters B* **76** (1978), no. 5, 575 – 579, `doi:http://dx.doi.org/10.1016/0370-2693(78)90858-4`.

[48] 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`.

[49] J. Alwall, P. C. Schuster, and N. Toro, “Simplified models for a first characterization of new physics at the LHC”, *Phys. Rev. D* **79** (Apr, 2009) 075020, `doi:10.1103/PhysRevD.79.075020`, `arXiv:0810.3921`.

[50] 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`.

[51] C. Boehm, A. Djouadi, and Y. Mambrini, “Decays of the lightest top squark”, *Phys.Rev.* **D61** (2000) 095006, `doi:10.1103/PhysRevD.61.095006`, `arXiv:hep-ph/9907428`.

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

[53] Bruce, R. et al, “LHC Run 2: Results and Challenges ”, CERN-ACC-2016-0103, CERN, Geneva, (Jul, 2016).

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

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

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

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

[58] S. Abdullin et al., “The CMS barrel calorimeter response to particle beams from 2 to 350 GeV/c”, *The European Physical Journal C* **60** (Apr, 2009) 359–373, `doi:10.1140/epjc/s10052-009-0959-5`.

[59] CMS Collaboration, “CMS. The TriDAS project. Technical design report, vol. 1: The trigger systems”,.

[60] K. Compton et al., “The MP7 and CTP-6: Multi-hundred Gbps processing boards for calorimeter trigger upgrades at CMS”, *JINST* **7** (2012) C12024, `doi:10.1088/1748-0221/7/12/C12024`.

[61] G. Hall, “A time-multiplexed track-trigger for the {CMS} HL-LHC upgrade”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **824** (2016) 292 – 295, `doi:https://doi.org/10.1016/j.nima.2015.09.075`. Frontier Detectors for Frontier Physics: Proceedings of the 13th Pisa Meeting on Advanced Detectors.

[62] CMS TMTT Team Collaboration, “A time-multiplexed track-trigger for the CMS HL-LHC upgrade”, *Nucl. Instrum. Meth.* **A824** (2016) 292–295, `doi:10.1016/j.nima.2015.09.075`.

[63] A. Sveteket al., “The Calorimeter Trigger Processor Card: the next generation of high speed algorithmic data processing at CMS”, *Journal of Instrumentation* **11** (2016), no. 02, C02011.

[64] CMS Collaboration, “CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project”, , Geneva, (2002).

[65] C. Eck et al., “LHC computing Grid: Technical Design Report. Version 1.06 (20 Jun 2005)”. Technical Design Report LCG. CERN, Geneva, 2005.

[66] CMS Collaboration, “Particle Flow Event Reconstruction in CMS and Performance for Jets, Taus and MET”, CMS Physics Analysis Summary CMS-PAS-PFT-09-001, (2009).

[67] CMS Collaboration, “Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV”, *JINST* **7** (2012) P10002, `doi:10.1088/1748-0221/7/10/P10002`, `arXiv:1206.4071`.

[68] 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`.

[69] CMS Collaboration, “Electron and photon measurement with the CMS detector”, *Conf. Proc.* **C0908171** (2009) 448–450, `doi:10.3204/DESY-PROC-2010-04/P16`. [,448(2010)].

[70] CMS Collaboration, “Performance of tau-lepton reconstruction and identification in CMS”, *JINST* **7** (2012) P01001, `doi:10.1088/1748-0221/7/01/P01001`, `arXiv:1109.6034`.

[71] The CMS collaboration, “Reconstruction and identification of lepton decays to hadrons and at CMS”, *Journal of Instrumentation* **11** (2016), no. 01, P01019.

[72] M. Cacciari, G. P. Salam, and G. Soyez, “The anti-k_T jet clustering algorithm”, *JHEP* **04** (2008) 063, `doi:10.1088/1126-6708/2008/04/063`, `arXiv:0802.1189`.

[73] CMS Collaboration, “A Cambridge-Aachen (C-A) based Jet Algorithm for boosted top-jet tagging”, CMS-PAS-JME-09-001, CERN, Geneva, (Jul, 2009).

[74] G. P. Salam, “Towards Jetography”, *Eur. Phys. J.* **C67** (2010) 637–686, `doi:10.1140/epjc/s10052-010-1314-6`, `arXiv:0906.1833`.

[75] CMS Collaboration, “Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS”, *JINST* **6** (2011) P11002, `doi:10.1088/1748-0221/6/11/P11002`, `arXiv:1107.4277`.

[76] M. Cacciari and G. P. Salam, “Pileup subtraction using jet areas”, *Physics Letters B* **659** (2008) 119 – 126, `doi:10.1016/j.physletb.2007.09.077`, `arXiv:0707.1378`.

[77] M. Cacciari, G. P. Salam, and G. Soyez, “The catchment area of jets”, *Journal of High Energy Physics* **2008** (2008), no. 04, 005, `arXiv:0802.1188`.

[78] 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`.

[79] CMS Collaboration, “Identification of b-quark jets with the CMS experiment”, *JINST* **8** (2013) P04013, `doi:10.1088/1748-0221/8/04/P04013`, `arXiv:1211.4462`.

[80] T. C. collaboration, “Identification of b-quark jets with the CMS experiment”, *Journal of Instrumentation* **8** (2013), no. 04, P04013.

[81] CMS Collaboration, “Identification of b-quark jets with the CMS experiment”, *JINST* **8** (2013) P04013, `doi:10.1088/1748-0221/8/04/P04013`, `arXiv:1211.4462`.

[82] CMS Collaboration, “Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8$ TeV”, *JINST* **10** (2015), no. 02, P02006, `doi:10.1088/1748-0221/10/02/P02006`, `arXiv:1411.0511`.

[83] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, “Recombination algorithms and jet substructure: Pruning as a tool for heavy particle searches”, *Phys. Rev. D* **81** (May, 2010) 094023, `doi:10.1103/PhysRevD.81.094023`.

[84] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, “Soft drop”, *Journal of High Energy Physics* **2014** (2014), no. 5, 146, `doi:10.1007/JHEP05(2014)146`.

[85] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, “Jet Substructure as a New Higgs-Search Channel at the Large Hadron Collider”,

Phys. Rev. Lett. **100** (Jun, 2008) 242001, `doi:10.1103/PhysRevLett.100.242001`.

[86] D. Krohn, J. Thaler, and L.-T. Wang, “Jet trimming”, *Journal of High Energy Physics* **2010** (2010), no. 2, 84, `doi:10.1007/JHEP02(2010)084`.

[87] S. D. Ellis and D. E. Soper, “Successive combination jet algorithm for hadron collisions”, *Phys. Rev.* **D48** (1993) 3160–3166, `doi:10.1103/PhysRevD.48.3160`, `arXiv:hep-ph/9305266`.

[88] CMS Collaboration, “Identifying Hadronically Decaying Vector Bosons Merged into a Single Jet”, CMS-PAS-JME-13-006, CERN, Geneva, (2013).

[89] CMS Collaboration, “CMS Luminosity Measurements for the 2016 Data Taking Period”, CMS-PAS-LUM-17-001, CERN, Geneva, (2017).

[90] J. Alwall et al., “MadGraph 5 : Going Beyond”, *JHEP* **06** (2011) 128, `doi:10.1007/JHEP06(2011)128`, `arXiv:1106.0522`.

[91] T. Sjostrand, 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`.

[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] S. Agostinelli et al., “Geant4—a simulation toolkit”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** (2003), no. 3, 250 – 303, `doi:http://dx.doi.org/10.1016/S0168-9002(03)01368-8`.

[94] A. Giammanco, “The Fast Simulation of the CMS Experiment”, *Journal of Physics: Conference Series* **513** (2014), no. 2, 022012.

[95] CMS Collaboration, “Utilities for Accessing Pileup Information for Data”. https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupJSONFileforData#Pileup_JSON_Files_For_Run_II. Accessed: 2017-05-12.

[96] 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`.

[97] 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`.

[98] 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`.

[99] 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`.

[100] 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.

[101] CMS Collaboration, K. Stifter, “The CMS Beam Halo Monitor Detector System”, in *Proceedings, Meeting of the APS Division of Particles and Fields (DPF 2015): Ann Arbor, Michigan, USA, 4-8 Aug 2015.* 2015. `arXiv:1511.00264`.

[102] 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`.

[103] K. Rehermann and B. Tweedie, “Efficient Identification of Boosted Semileptonic Top Quarks at the LHC”, *JHEP* **03** (2011) 059, `doi:10.1007/JHEP03(2011)059`, `arXiv:1007.2221`.

[104] CMS Collaboration, “Boosted Top Jet Tagging”. https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetTopTagging. Accessed: 2017-04-16.

[105] CMS Collaboration, “Measurement of differential top-quark pair production cross sections in *pp* colisions at $\sqrt{s} = 7$ TeV”, *Eur. Phys. J.* **C73** (2013), no. 3, 2339, `doi:10.1140/epjc/s10052-013-2339-4`, `arXiv:1211.2220`.

[106] 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), no. 11, 542, `doi:10.1140/epjc/s10052-015-3709-x`, `arXiv:1505.04480`.

[107] 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`.

[108] A. L. Read, "Presentation of search results: the CL s technique", *Journal of Physics G: Nuclear and Particle Physics* **28** (2002), no. 10, 2693.

[109] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J.* **C71** (2011) 1554, `doi:10.1140/epjc/s10052-011-1554-0,10.1140/epjc/s10052-013-2501-z`, `arXiv:1007.1727`. [Erratum: Eur. Phys. J.C73,2501(2013)].

[110] C. Collaboration, "LHC SUSY Cross Section Working Group". https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SUSYCrossSections. Accessed: 2017-06-10.

[111] CMS Collaboration, "Search for supersymmetry in the all-hadronic final state using top quark tagging in pp collisions at $\sqrt{s} = 13$ TeV", CMS-PAS-SUS-16-030, CERN, Geneva, (2016).

[112] S. Alioli, P. Nason, C. Oleari, and E. Re, "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`.

[113] M. Botje, J. Butterworth, A. Cooper-Sarkar et al., "The PDF4LHC Working Group Interim Recommendations", (2011). `arXiv:1101.0538`.

[114] R. Barbier et al., "R-Parity-violating supersymmetry", *Physics Reports* **420** (2005), no. 1, 1 – 195, `doi:https://doi.org/10.1016/j.physrep.2005.08.006`.

[115] M. J. Strassler and K. M. Zurek, "Echoes of a hidden valley at hadron colliders", *Physics Letters B* **651** (2007), no. 5, 374 – 379, `doi:https://doi.org/10.1016/j.physletb.2007.06.055`.

List of Acronyms

CMS Compact Muon Solenoid

LHC Large Hadron Collider

SM Standard Model

SUSY Supersymmetry

LSP Lightest Supersymmetric Particle

GR General Relativity

VBF Vector Boson Fusion

ggH Gluon Fusion

VH Vector Boson Associated Production

DM Dark Matter

CL Confidence Level

ISR Initial State Radiation

FSR Final State Radiation

EFT Effective Field Theory

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

QFT Quantum Field Theory

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

oRSC Optical Regionl Summary Card

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

WLCG Worldwide LHC Computing Grid

FPGA Field Programmable Gate Array

PV Primary Vertex

CTF Combinatorial Track Finder

DA *Deterministic Annealing*

PF Particle Flow

GSF Gaussian Sum Filter

BDT Boosted Decision Tree

MC Monte Carlo

HPS Hadron Plus Strips

JES Jet Energy Scale

CHS Charged Hadron Subtraction

CHF Charged Hadron Fraction

SIT Single Isolated Track

EM Electromagnetic

CJV Central Jet Veto

MVA Multi-Variate Analysis

JER Jet Energy Resolution

UES Unclustered Energy Scale

CSV Combined Secondary Vertex

CSVv2 Combined Secondary Vertex version 2

4FS Four-Flavour Scheme

5FS Five-Flavour Scheme

DAQ Data Acquisition

LHCHXSWG LHC Higgs Cross Section Working Group

LO Leading Order

MPF Missing Transverse Energy Projection Fraction

MPI Multi-Parton Interaction

MSSM Minimal Supersymmetric Standard Model

SMS Simplified Model Spectra

NLO Next-To-Leading Order

NNLO Next-To-Next-To-Leading Order

NNLL Next-To-Next-To-Leading Logarithmic

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

SSV Simple Secondary Vertex

TEC Tracker Endcaps

TIB Tracker Inner Barrel

TID Tracker Inner Disks

TOB Tracker Outer Barrel

UE Underlying Event

TF Transfer Factor