

DISSERTATION

Search for SUSY with a compressed mass spectrum in the single lepton channel with the CMS experiment

Ausgeführt zum Zwecke der Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

eingereicht an der Fakultät für Physik der Technischen Universität Wien
von

MSc Priya Sajid Hussain

Matrikelnummer 11849201

ausgeführt am Institut für Hochenergiephysik (HEPHY)
der Österreichischen Akademie der Wissenschaften (ÖAW)
und am Atominstitut der Technischen Universität Wien (E141)

unter der Leitung von
Priv.-Doz. Mag. phil. Mag. Dr. rer. nat. Manfred Jeitler

Unterstützt vom Österreichischen Wissenschaftsfonds (FWF) im Rahmen des
Doktoratskollegs Particles and Interactions (DK-PI)

Wien, 25. September 2025

To Tobi, the most amazing human being I am lucky to call my husband, my incredible parents, and my marvellous sisters.

Abstract

The Standard Model (SM) describes known particles and interactions with remarkable precision; however, it does not account for dark matter, the naturalness of the Higgs mass, or the unification of forces, motivating searches for physics beyond the SM. Supersymmetry (SUSY) addresses these shortcomings by predicting a spectrum of superpartners. Although no superpartners have been observed at the LHC, the theory is not excluded: several regions of parameter space remain weakly tested, particularly those that are experimentally hard to explore rather than theoretically disfavoured. One such scenario is compressed SUSY, where the mass difference Δm between the next-to-lightest superpartner and the lightest supersymmetric particle (LSP) is small. A particularly important case is when the next-to-lightest particle is the light top squark (stop), since naturalness arguments favour light stops to cancel large contributions from the top quark to the Higgs boson mass. For even smaller Δm , the stop can become long-lived. In dark-matter coannihilation scenarios, small Δm is further motivated as it can reproduce the observed relic density. Experimentally, however, such scenarios yield final-state objects with very low transverse momenta (p_T) and modest missing transverse momentum, posing challenges for classical high-momentum searches. This thesis presents a search for pair-produced stops in supersymmetric models with Δm between the top squark and the LSP less than the W boson mass, targeting the four-body decay $\tilde{t} \rightarrow b f f' \tilde{\chi}_1^0$ in the single-lepton final state. The study uses proton–proton collision data recorded by the CMS detector at a centre-of-mass energy of $\sqrt{s} = 13$ TeV during Run 2 (2016–2018), corresponding to an integrated luminosity of 138 fb^{-1} .

To maximise sensitivity to soft kinematics, the jet thresholds are reduced to 20 GeV and a fine-grained binning in transverse mass and lepton p_T is employed, exploiting the statistical power of full Run 2 dataset. The thesis focuses on the prompt (zero-lifetime) interpretation, while also including supporting contributions to the complementary long-lived stop searches pursued in parallel. The analysis remains blinded as the long-lived component is ongoing. Within this framework, expected 95% confidence-level upper limits are obtained using Asimov datasets in a profile-likelihood approach, showing significantly improved sensitivity compared to earlier results based on 2016 data only.

Kurzfassung

Das Standardmodell der Teilchenphysik (SM) beschreibt die uns bekannten Teilchen und deren Wechselwirkungen mit bemerkenswerter Präzision. Es erklärt jedoch weder das Phänom der Dunklen Materie noch die Natürlichkeit der Higgs-Masse oder die Vereinheitlichung der elementaren Kräfte, was die Suche nach Physik jenseits des SM motiviert. Eine Erweiterung des SM ist die Supersymmetrie (SUSY), welche Antworten auf die beschriebenen Mängel des SM liefern kann. Im Rahmen der SUSY wird ein neues Spektrum von Teilchen eingeführt. Diese Teilchen werden Superpartnern genannt. Obwohl am LHC noch keine Superpartner beobachtet wurden, ist die Theorie nicht ausgeschlossen. Mehrere Bereiche des Parameterraums sind noch unzureichend getestet, insbesondere solche, die experimentell schwer zugänglich sind, obwohl sie aus theoretischer Sicht durchaus interessant wären. Ein Beispiel für ein solches Szenario ist die komprimierte (compressed) SUSY, bei der die Massendifferenz Δm zwischen dem zweit leichtesten Superpartner und dem leichtesten supersymmetrischen Teilchen (LSP) gering ist. Ein besonders interessanter Fall tritt ein, wenn das zweit leichteste Superpartner und dem leichte Top Squark (Stop) ist, weil dann große Beiträge zur Higgs-Masse vom Top Quark aufgehoben werden und die Natürlichkeit der Higgs-Masse motiviert werden kann. Bei noch kleinerem Δm kann das Stop Teilchen langlebig werden. In Szenarien der Koannihilation von Dunkler Materie ist solch ein kleines Δm zusätzlich motiviert, weil es die beobachtete Reliktdichte der Dunklen Materie reproduzieren kann. Experimentell haben die messbaren Endzustände einen sehr geringen transversalen Impuls (p_T) und nur einen mäßigem Betrag an fehlendem transversalen Impuls, was eine Herausforderung für klassische Messungen am LHC darstellt.

Diese Arbeit präsentiert eine Suche nach paarweise erzeugten Stopps in supersymmetrischen Modellen. Die Massendifferenz Δm zwischen dem Top Squark und dem LSP ist dabei kleiner als die W-Boson Masse und der Vierkörperzerfall $\tilde{t} \rightarrow b f f' \tilde{\chi}_1^0$ im Endzustand mit einem einzelnen Lepton steht im Fokus. Die Studie verwendet Daten von Proton-Proton-Kollisionen, die vom CMS-Detektor bei einer Schwerpunktsenergie von $\sqrt{s} = 13$ TeV während des Run 2 (2016–2018) aufgezeichnet wurden. Die integrierte Luminosität entspricht 138 fb^{-1} .

Um die Sensitivität für Prozesse mit niedrigen transversalen Impulsen zu maximieren, werden die Jet-Schwellenwerte auf 20 GeV reduziert und eine feinkörnige Unterteilung in transversaler Masse und Lepton p_T verwendet, wobei die statistische Aussagekraft des vollständigen Datensatzes aus dem Run 2 genutzt wird. Die Arbeit konzentriert sich auf die Prompt-Interpretation (die Lebensdauer der erzeugten Stops ist Null), enthält aber auch Beiträge zu den parallel durchgeführten ergänzenden Suchen nach langlebigen Stop Teilchen.

Die Analyse bleibt blind, da die Suche nach langlebigen Komponenten noch läuft. Es wurden erwartete 95%-Konfidenz-Obergrenzen unter Verwendung von Asimov-Datensätzen in einem Profil-Likelihood-Ansatz ermittelt, die eine deutlich verbesserte Empfindlichkeit im Vergleich zu früheren Ergebnissen zeigen, die nur auf Daten aus dem Jahr 2016 basieren.

Contents

1	Introduction	1
2	Theoretical Introduction	3
2.1	The Standard Model	3
2.1.1	Fundamental Particles in the SM	4
2.1.2	Gauge Theories of the Standard Model	6
2.2	Supersymmetry	8
2.2.1	The Minimal Supersymmetric Standard Model	10
2.2.2	Hierarchy Problem	12
2.2.3	R-Parity	13
2.2.4	Unification of Gauge Coupling	13
2.2.5	Search for compressed spectra	14
3	The CMS experiment at the LHC	17
3.1	The Large Hadron Collider	17
3.1.1	Luminosity at the LHC	18
3.1.2	Pileup	19
3.2	The Compact Muon Solenoid Detector	20
3.2.1	CMS Coordinate System	21
3.2.2	Superconducting Solenoid Magnet	22
3.2.3	Inner Tracking System	23
3.2.4	Calorimeters	25
3.2.5	The Muon System	27
3.3	Event Triggering and Data Acquisition System	28
3.4	Physics Object Reconstruction	29
3.4.1	Tracks and Clusters	30
3.4.2	Muon Identification	33
3.4.3	Electron Identification	34
3.4.4	Jet Reconstruction	35
3.4.5	Identification of b jets	35
3.4.6	Missing Transverse Momentum	36
4	Compressed SUSY search	38
4.1	Signal Models	39
4.2	Data and simulation	40
4.2.1	Data Sets	40
4.2.2	Signal and background simulation	41
4.2.3	Object Efficiencies and Corrections	43
4.3	Event Selection	47
4.3.1	Baseline Selection	48
4.3.2	Background Processes	51

4.4	Analysis Strategy	54
4.4.1	Signal Regions	57
4.4.2	Control Regions	59
4.5	Background Estimation Techniques	61
4.5.1	Estimation of main prompt backgrounds	61
4.5.2	Estimation of non-prompt backgrounds	61
4.6	Systematic Uncertainties	63
4.7	Results	64
4.7.1	Statistical analysis	64
4.7.2	Simplified Model Interpretation	66
5	Conclusion and Outlook	68
5.1	Conclusion	68
5.2	Outlook	68
5.2.1	Ongoing extensions and methods	68
5.2.2	Run 3 and HL-LHC prospects	69
	Appendices	70
	A Region plots for various eras in Run2	71
	B Hybrid isolation and impact parameter SF	73
	C Private NanoAOD production	76
	List of Figures	78
	Bibliography	83

Chapter 1

Introduction

The Standard Model (SM) of particle physics [1, 2, 3] is one of the most successful theories for explaining the fundamental forces and particles in the universe. It has accurately predicted the existence of particles like charm, bottom, top quarks, tau leptons, and the W and Z bosons, which mediate the weak force. The strong force, which holds protons and neutrons together, is described by Quantum Chromodynamics (QCD), where quarks interact by exchanging particles called gluons. A massive achievement for the SM came in 2012, when the ATLAS and CMS experiments [4, 5] at the Large Hadron Collider (LHC) discovered the Higgs boson, confirming how particles gain mass as predicted by Higgs, Englert, and Brout [6, 7].

Despite the Standard Model's spectacular achievements, it still leaves significant questions unanswered, such as why there is more matter than antimatter in the universe, the nature of dark matter [8], observation of neutrino masses and oscillations [9, 10], and how gravity fits into the picture. These limitations suggest the need for physics beyond the Standard Model (BSM). One leading idea is Supersymmetry (SUSY) [11, 12, 13], which postulates that every particle in the SM has a superpartner whose spin differs exactly by one-half unit. SM fermions thus acquire spin-0 'sfermion' partners, and SM bosons acquire spin-half 'gaugino' or 'higgsino' partners. SUSY provides possible solutions to the hierarchy problem and proposes a unification of strong and electroweak forces at extremely high energy scales, i.e. Grand Unification Theory (GUT). The lightest supersymmetric particle (LSP) is also a compelling candidate for dark matter, further driving experimental searches for SUSY at the LHC.

This thesis presents a search for top squark pair production and their decay into neutralinos via four-body process in supersymmetric models where the mass difference between the top squark and neutralino is less than the W boson mass, referred to as compressed regime. The small mass difference results in low-energy decay products, making detecting and exploring this part of the SUSY parameter space difficult. The results presented in this thesis use data from the 2016–2018 (Run 2) proton-proton collisions at the Large Hadron Collider, recorded by the CMS detector, with a centre-of-mass energy of 13 TeV and an integrated luminosity of 138 fb^{-1} . Focusing on the single-lepton channel (one electron or muon), the search targets compressed SUSY scenarios motivated by dark matter constraints, as coannihilation between top squarks and neutralinos can explain observed dark matter relic densities.

My role was pivotal in several core areas, including the implementation of the software framework used throughout the Run 2 dataset, which comprises four distinct data-taking periods; the optimisation of event-categorisation algorithms to maximise sensitivity to soft decay products. Statistical analysis workflows were refined to combine results coherently across all data-taking periods. In addition, lepton scale factors were measured and applied,

and the relevant systematic uncertainties were evaluated and incorporated into the final results.

An overview of the Standard Model (SM) and an introduction to supersymmetry is given in Chapter 2. Chapter 3 describes the CERN LHC and the CMS experiment used to gather the data for this analysis. Chapter 4 outlines the analysis strategy, background estimation techniques, systematic uncertainties, and the final results. Lastly, Chapter 5 summarises the findings and provides an outlook for future work.

Chapter 2

Theoretical Introduction

The Standard Model (SM) is a theory developed over the latter half of the 20th century through the collaborative efforts of both theorists and experimentalists. It describes three of the four fundamental forces of nature and classifies all currently known elementary particles. Despite its tremendous success, the SM has notable gaps, such as the inability to explain gravity, the fourth fundamental force. This thesis explores an extension to the SM known as supersymmetry (SUSY).

Initially, an overview of the SM will be provided, outlining its major achievements. This will be followed by a discussion of its limitations, pointing toward new physics beyond the SM, specifically focusing on the fundamental ideas of supersymmetry and its minimal supersymmetric (MSSM) extension as a potential way forward.

2.1 The Standard Model

The Standard Model (SM) of particle physics emerged from the foundations of quantum mechanics (QM), which revolutionized our understanding of atomic and subatomic phenomena in the early 20th century [14]. Pioneering concepts like wave-particle duality and the Heisenberg Uncertainty Principle shaped QM, but its limitations in addressing high-energy interactions necessitated the development of quantum field theory (QFT) [15].

QFT treats particles as excitations in underlying fields, enabling a comprehensive description of particle creation and annihilation. A landmark achievement in QFT was quantum electrodynamics (QED), developed by Richard Feynman, Julian Schwinger, and Sin-Itiro Tomonaga [16]. QED successfully described the interactions of charged particles with the electromagnetic field and was confirmed through precise experimental measurements, including the Lamb shift in hydrogen and the determination of the electron's magnetic moment. Additionally, the prediction and subsequent discovery of the positron in 1932 provided evidence for antimatter [17].

In the 1970s, the electroweak theory unified the electromagnetic and weak forces, predicting the existence of the W and Z bosons. These were experimentally confirmed in 1983 at CERN [18, 19], marking a significant validation of the theory. Concurrently, quantum chromodynamics (QCD) emerged to explain the strong force, with experimental evidence for quarks obtained from deep inelastic scattering experiments at SLAC [20] in the late 1960s.

The Higgs mechanism addressed the question of mass acquisition for particles, leading to the prediction of the Higgs boson [6, 7]. Its discovery at the Large Hadron Collider (LHC) in 2012 confirmed this crucial aspect of the SM [4, 5].

Overall, experimental discoveries, such as the positron, W and Z bosons, quarks, and

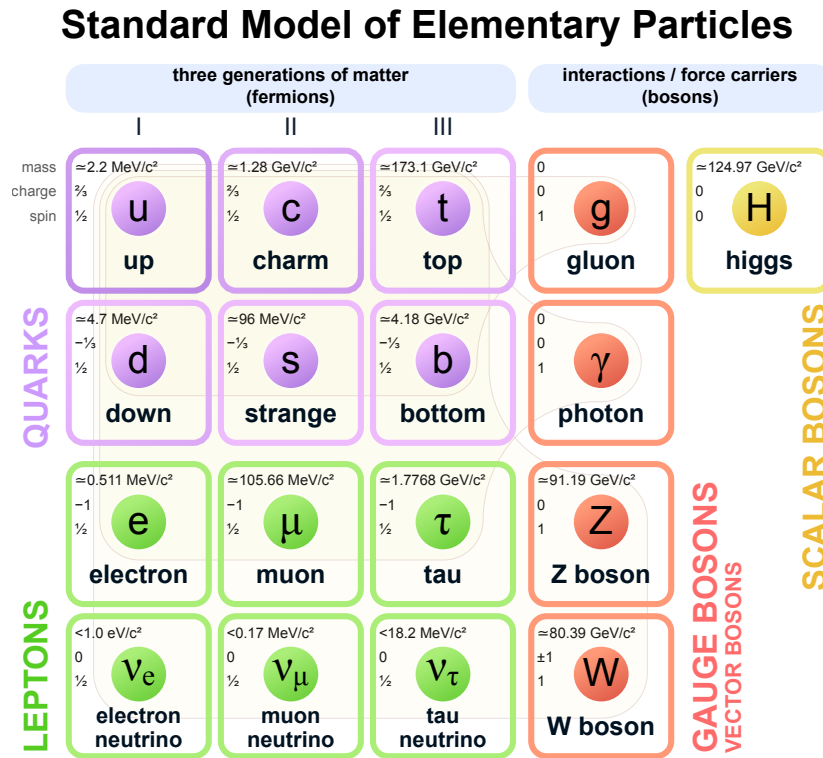


Figure 2.1: The figure shows the particle content of the SM, including mass, electric charge and spin of each elementary particle, and is taken from [21].

the Higgs boson, have validated the Standard Model as a highly successful theory. However, observations of dark matter, neutrino masses, and the matter-antimatter imbalance suggest phenomena beyond the SM, fuelling interest in extensions like supersymmetry (SUSY). The Standard Model also does not describe the gravitational force; however, due to its extreme weakness at the microscopic scale, it can be neglected.

2.1.1 Fundamental Particles in the SM

The Standard Model (SM) classifies all known elementary particles as shown in Figure 2.1 into two broad categories: fermions, which make up matter, and bosons, which mediate the fundamental forces of nature. This classification is based on the particle's spin: fermions have half-integer spin (spin- $\frac{1}{2}$) and therefore, obey the Pauli exclusion principle (PEP), while bosons have integer spin.

Fermions

Fermions, the building blocks of matter, are divided into two groups: quarks and leptons. For each fermion, the anti-particle carries the negative charge of respective particle.

Quarks are fundamental particles that interact with all fundamental forces. Each quark has a fractional electric charge, $+2/3$ (up-type) or $-1/3$ (down-type), and a colour charge in the context of QCD. They interact with each other through strong force, and can not be observed as isolated particles due to a phenomenon called colour confinement [22]. Quarks combine to form colourless bound states hadrons, as three distinctly coloured quarks, such as protons and neutrons (baryons) or quark-antiquark pairs (mesons). There are six types

(flavours) of quarks, categorized into three generations. Each generation consists of pairs that differ in rest mass and fractional electric charge. The first generation consists of the lightest and most stable particles, and less stable particles form the second and third generations.

Leptons differ from quarks in that they have no colour charge and, thus, do not experience the strong force. Similar to quarks, leptons have six flavours categorized into three generations. Each generation consists of a charged lepton (electron, muon, or tau) and an uncharged neutrino of the same flavour.

The free Lagrangian of a fermion is:

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

where ψ represents fermionic field, m mass, γ Dirac matrices and ∂_μ the four momentum derivative.

Bosons

Bosons are the particles responsible for mediating the fundamental forces. Depending on their spin, they are categorized as vector bosons (spin-1) or scalar bosons (spin-0).

Vector Bosons (spin-1) Photons are massless gauge bosons that mediate electromagnetic interactions, giving the electromagnetic force an infinite range. W and Z bosons are the massive bosons that mediate the weak force. The W^\pm bosons are charged whereas the Z boson is neutral mediating neutral weak interactions. The large masses of the W and Z bosons limit the weak force to a short range.

Scalar Boson (spin-0) The Higgs boson, the only fundamental scalar particle in the SM, was discovered in 2012 by ATLAS and CMS at the Large Hadron Collider. The massive Higgs boson gives masses to other particles.

The Klein-Gordon Lagrangian for the spin zero neutral particles is given as:

$$\mathcal{L}_{free} = \partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2, \quad (2.2)$$

And for charged particles, described as a complex field, the Lagrangian becomes:

$$\mathcal{L}_{free} = \partial_\mu\phi\partial^\mu\phi^* - m^2\phi\phi^*. \quad (2.3)$$

The above two equations are used to describe the scalar particles. For the vector field, A^μ , the Lagrangian is written as,

$$\mathcal{L}_{free} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \frac{1}{4}m^2A^\mu A_\mu. \quad (2.4)$$

If we consider a massless particle, the equation reduces to Maxwell Lagrangian:

$$\mathcal{L}_{free} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu}, \quad (2.5)$$

where $F^{\mu\nu} = \partial^\mu A_\nu - \partial^\nu A_\mu$.

2.1.2 Gauge Theories of the Standard Model

The Standard Model of particle physics is based on the gauge theory, which provides a robust mathematical framework for describing the three fundamental forces: electromagnetic, weak, and strong interactions. Gauge theory connects these forces to the behaviour of elementary particles, using symmetries to determine the rules governing their interactions.

In the Standard Model, local gauge symmetries define particle interactions by introducing gauge bosons mediating the fundamental forces. These symmetries are governed by quantum field theory (QFT), ensuring the interactions are consistent with quantum mechanics and special relativity.

Gauge symmetry requires the laws of physics not to depend on arbitrary gauges or coordinate systems, i.e. remain invariant under arbitrary local transformations of certain fields. This requirement implies that adding new interaction terms can achieve the invariance of a free quantum field theory under local transformations. These terms involve gauge fields which correspond to the generators of the gauge symmetry. For example:

- The electromagnetic force arises from $U(1)_Y$ gauge symmetry, with the introduction of the photon as its gauge boson.
- The weak force emerges from $SU(2)_L$ gauge symmetry, with W and Z as its gauge bosons.
- The strong force, governed by $SU(3)_C$ gauge symmetry, introduces gluons as its mediators.

By extending the already existing global symmetries of Lagrangian (θ) to a local space-time symmetry ($\theta(x)$), the interaction terms in SM emerge. Generally, for a given Lagrangian invariant under a symmetry group, the field transformation can be written as:

$$\psi \rightarrow \bar{\psi} \equiv \exp(ig\theta^k\tau^k)\psi ,$$

where g is the coupling constant of the field ψ for the interaction and τ^k are the generators of the group obeying commutation:

$$\left[\tau^i, \tau^j \right] = if^{ijk}\tau^k .$$

f^{ijk} is the structure constant of the Lie group. Its value is always zero for an Abelian group.

Extension of the global invariance to local one translates to:

- Adding massless gauge fields W_μ^k equal to the number of generators of the symmetry group
- The covariant derivative replaces the standard derivative:

$$D_\mu = \partial_\mu - ig\tau^k W_\mu^k ,$$

- Addition of a free Lagrangian for the vector fields as shown in Equation (2.5) :

$$F_{\mu\nu}^k = \partial_\mu W_\nu^k - \partial_\nu W_\mu^k + gf^{klm}W_\mu^l W_\nu^m$$

For non-Abelian groups the consequence of extending the symmetries is not trivial and can lead to self interaction within gauge fields. In other words, the non-Abelian gauge fields are self-coupled while the Abelian (photon) field is not. [23]

The SM is a relativistic quantum field theory (QFT) invariant under $SU(3)_C \times SU(2)_L \times U(1)_Y$.

Quantum Chromodynamics

Quantum Chromodynamics (QCD) is the $SU(3)_C$ gauge theory that describes the strong interaction, i.e. interaction between quarks and gluons. The strong interaction between the particles takes place by virtue of colour charge, with possible values of red, blue, and green. Quarks (antiquarks) each carry a single colour charge (anti-colour) charge; however, colour confinement dictates that observable states, hadrons, must be colour singlets. The $SU(3)_C$ is non-Abelian, meaning that its eight generators do not commute. As a result, the corresponding gauge bosons, the gluons, also carry colour charge and therefore self-interact. This non-Abelian structure of QCD has two key consequences. First, colour confinement binds coloured objects into colour-neutral hadrons, described in Section 2.1.1. Second, asymptotic freedom, the strong coupling $\alpha_s(Q)$ decreases with increasing momentum transfer Q , so that at short distances (or equivalently high energies) quarks behave almost as free particles [24, 25]. Finally, the gluons are massless because $SU(3)_C$ is an exact gauge symmetry.

Electroweak theory

At first glance, electromagnetic force and weak force are very different. The weak force is short-ranged, whereas the electromagnetic force is long-ranged. On top of that, the electromagnetic force is stronger than the weak force. However, these two forces are, in fact, two aspects of a single unified force: the electroweak force. The electroweak theory, which unifies these forces, is based on the $SU(2)_L \times U(1)_Y$ symmetry formalized by the Glashow-Weinberg-Salam (GWS) model [2, 3, 1]. This symmetry structure not only governs the weak interactions via the $SU(2)_L$ component but also involves $U(1)_Y$, the hypercharge group. After spontaneous symmetry breaking of $SU(2)_L \times U(1)_Y$, the unbroken symmetry is $U(1)_{EM}$, which governs electromagnetism and leaves the photon massless. The covariant derivative for the $SU(2)_L \times U(1)_Y$ symmetry group is written as:

$$D_\mu = \partial_\mu - i\frac{g'}{2}YB_\mu - i\frac{g}{2}W_\mu^k\tau^k, \quad (2.6)$$

where W_μ^k and B_μ are the $SU(2)_L$ and $U(1)_Y$ gauge fields with coupling constants g and g' , respectively. As a result of symmetry breaking, the mixing of the $U(1)_Y$ gauge field B_μ and the third component of the $SU(2)_L$ gauge field, i.e. W_μ^3 form the massless photon (A_μ) and massive Z boson (Z_μ) fields as follows:

$$A_\mu = B_\mu \cos(\theta_W) + W_\mu^3 \sin(\theta_W) \quad (2.7)$$

$$Z_\mu = -B_\mu \sin(\theta_W) + W_\mu^3 \cos(\theta_W) \quad (2.8)$$

where the mixing angle θ_W , also called Weinberg angle, relates to the coupling constants by:

$$\cos(\theta_W) = \frac{g}{\sqrt{g^2 + g'^2}} \quad \text{and} \quad \sin(\theta_W) = \frac{g'}{\sqrt{g^2 + g'^2}}$$

The electrically charged W_μ^\pm boson fields are a linear combination of the remaining two spin-1 boson fields:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2) \quad (2.9)$$

The $SU(2)_L$ part of the electroweak gauge group describes the weak interactions, and only acts on the left-handed fermion fields; right-handed fields are singlets under $SU(2)_L$. The chirality projectors P_L and P_R are used to define the left- and right-handed fields (ψ_L and ψ_R) as follows:

$$\begin{aligned} \psi_L &= P_L \psi = \frac{1 - \gamma^5}{2} \psi, \\ \psi_R &= P_R \psi = \frac{1 + \gamma^5}{2} \psi, \\ \psi &= \psi_L + \psi_R \end{aligned}$$

where $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$. The free Dirac Lagrangian in Equation (2.1) shows that the kinetic term keeps the two chiralities separate, whereas a Dirac mass term couples them,

$$-m\bar{\psi}\psi = -m(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L).$$

Because ψ_L (an $SU(2)_L$ doublet) and ψ_R (a singlet) transform differently, a bare mass term is not gauge invariant under $SU(2)_L \times U(1)_Y$. Fermion masses are therefore introduced via Yukawa interactions with the Higgs doublet and only appear after electroweak spontaneous symmetry breaking (SSB). Similarly, explicit mass terms for the electroweak gauge bosons would violate $SU(2)_L \times U(1)_Y$ gauge invariance, so in the unbroken theory the gauge bosons are massless. Through the Brout-Englert-Higgs mechanism (electroweak SSB), the W^\pm and Z acquire masses ($m_W = 80.4$ GeV, $m_Z = 91.2$ GeV) while the photon remains massless, consistent with an unbroken $U(1)_{EM}$.

2.2 Supersymmetry

The Standard Model has successfully explained many aspects of nature and predicted various experimental observations, as shown in Figure 2.2. However, there are some limitations and unanswered questions within the SM framework, including:

- Gravity: SM can only explain three out of four fundamental forces, leaving Gravity out.
- Neutrino masses: SM does not predict any masses for neutrinos or neutrino oscillations, which are experimentally observed [10, 9].
- Dark Matter and Dark Energy: SM lacks a candidate for dark matter and any explanation for dark energy, which is responsible for the accelerated expansion of our Universe.
- Fine tuning of the Higgs mass: Quantum corrections to the Higgs mass because of virtual particles are theoretically substantial due to the lack of an approximate symmetry protecting it, pushing it toward values near the Planck scale (10^{19} GeV). To match the observed Higgs boson mass of approximately 125 GeV, the bare Higgs mass must be extremely fine-tuned to cancel these corrections.

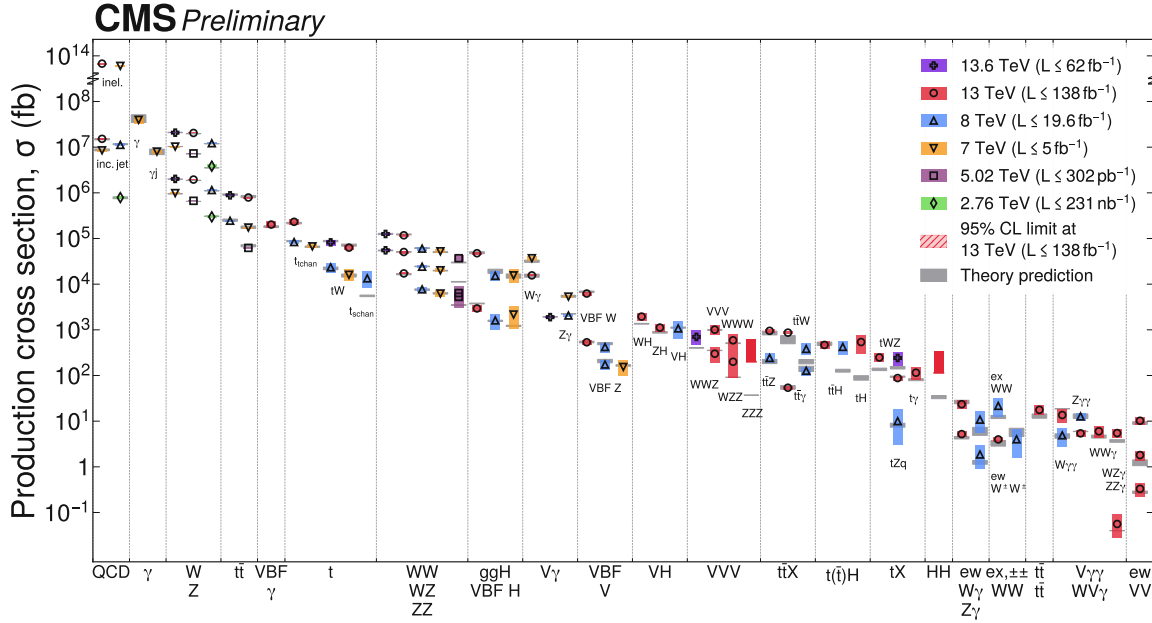


Figure 2.2: Predicted and measured Standard Model production cross sections at various centre-of-mass energies in the CMS. The measurements span over nearly thirteen orders of magnitude and are in excellent agreement with theoretical predictions. [26]

- Gauge coupling mismatch: A unification of gauge couplings associated with strong, weak and electromagnetic forces is expected to occur at a high energy scale of approximately 10^{16} GeV. Within the context of SM the couplings almost converge but fail to meet at a single point.

These shortcomings and deviations from our observations have led researchers to explore new ideas beyond the Standard Model (BSM). Various extensions to the SM have been proposed to address its unexplained aspects with Supersymmetry (SUSY) being one of the most extensively studied frameworks. SUSY remains a compelling candidate within the BSM landscape, offering solutions to critical problems such as the fine-tuning of the Higgs mass and providing a potential dark matter candidate through R-parity conservation (discussed in Section 2.2.3). Yet despite dozens of dedicated searches, no evidence for supersymmetry has been found. Significant regions of SUSY's parameter space, particularly the less conventional scenarios such as compressed and R-parity-violating scenarios, however, remain largely unexplored.

Supersymmetry introduces a spacetime symmetry that relates fermions and bosons via a supercharge operator Q , which transforms a bosonic state into a fermionic state and vice versa [27] as:

$$Q |Boson\rangle = |Fermion\rangle \text{ and } Q |Fermion\rangle = |Boson\rangle . \quad (2.10)$$

By virtue of the operator Q , every SM particle has a corresponding superpartner that differs in spin by one-half but is identical in all other quantum properties (mass, hypercharge, weak isospin, colour charge) in the case of unbroken symmetry.

Names	Spin 0	Spin 1/2	$SU(3)_C, SU(2)_L, U(1)_Y$
squarks, quarks ($\times 3$ families)	$(\tilde{u}_L, \tilde{d}_L)$	(u_L, d_L)	$(\mathbf{3}, \mathbf{2}, \frac{1}{6})$
	\tilde{u}_R^*	u_R^\dagger	$(\bar{\mathbf{3}}, \mathbf{1}, -\frac{2}{3})$
	\tilde{d}_R^*	d_R^\dagger	$(\bar{\mathbf{3}}, \mathbf{1}, \frac{1}{3})$
sleptons, leptons ($\times 3$ families)	$(\tilde{\nu}, \tilde{e}_L)$	(ν, e_L)	$(\mathbf{1}, \mathbf{2}, -\frac{1}{2})$
	\tilde{e}_R^*	e_R^\dagger	$(\mathbf{1}, \mathbf{1}, 1)$
Higgs, higgsinos	(H_u^+, H_u^0)	$(\tilde{H}_u^+, \tilde{H}_u^0)$	$(\mathbf{1}, \mathbf{2}, \frac{1}{2})$
	(H_d^0, H_d^-)	$(\tilde{H}_d^0, \tilde{H}_d^-)$	$(\mathbf{1}, \mathbf{2}, -\frac{1}{2})$

Table 2.1: The chiral supermultiplets of the MSSM, with their spins, and their gauge group representations. Adapted from [27].

Names	Spin 1/2	Spin 1	$SU(3)_C, SU(2)_L, U(1)_Y$
gluino, gluon	\tilde{g}	g	$(\mathbf{8}, \mathbf{1}, 0)$
winos, W bosons	$\tilde{W}^\pm, \tilde{W}^0$	W^\pm, W^0	$(\mathbf{1}, \mathbf{3}, 0)$
bino, B boson	\tilde{B}^0	B^0	$(\mathbf{1}, \mathbf{1}, 0)$

Table 2.2: Gauge supermultiplets of the MSSM. Adapted from [27].

2.2.1 The Minimal Supersymmetric Standard Model

The minimal supersymmetric standard model (MSSM) is a supersymmetric extension of the Standard Model (SM) that introduces the fewest new particles and matter fields. In the MSSM, each SM fermion has a superpartner with spin-0, denoted by adding an “s” prefix to their name, indicating their scalar nature. These superpartners are also represented with a tilde. For example, the superpartners of top are called the “stop”, represented as \tilde{t} . The superpartners of SM fermions are classified as “sfermions”, with the left- and right-handed sfermions denoted as \tilde{f}_L and \tilde{f}_R , respectively. Similarly, the superpartners of SM bosons take the same name as their SM counterpart, with the suffix “ino” added, such as the gluino \tilde{g} . The superpartners of Higgs bosons are called Higgsinos, represented as \tilde{H} . Table 2.1 and Table 2.2 summarise the MSSM particle content organised as chiral and gauge supermultiplets, respectively. They list the superpartners of the SM fermions, gauge bosons, and Higgs fields, including their spin and gauge group representations.

The MSSM requires two Higgs doublets, H_u and H_d , whose higgsino partners cancel gauge anomalies and which give masses to up-type and down-type fermions, respectively. The new particles introduced in the MSSM are summarised in Table 2.3 along with their gauge and mass eigenstates. While the Standard Model particles and their superpartners share the same quantum numbers and Yukawa couplings, their spins differ by $1/2$. In a scenario with unbroken supersymmetry, particles and their superpartners would have identical masses. However, experimental evidence indicates this is not the case. The mechanism responsible for breaking supersymmetry determines the mass spectrum of

Names	Spin	P_R	Gauge Eigenstates	Mass Eigenstates
Higgs bosons	0	+1	$H_u^0 H_d^0 H_u^+ H_d^-$	$h^0 H^0 A^0 H^\pm$
Squarks	0	-1	$\tilde{u}_L \tilde{u}_R \tilde{d}_L \tilde{d}_R$ $\tilde{s}_L \tilde{s}_R \tilde{c}_L \tilde{c}_R$ $\tilde{t}_L \tilde{t}_R \tilde{b}_L \tilde{b}_R$	(same)
Sleptons	0	-1	$\tilde{e}_L \tilde{e}_R \tilde{\nu}_e$ $\tilde{\mu}_L \tilde{\mu}_R \tilde{\nu}_\mu$ $\tilde{\tau}_L \tilde{\tau}_R \tilde{\nu}_\tau$	(same) $\tilde{\tau}_1 \tilde{\tau}_2 \tilde{\nu}_\tau$
Neutralinos	1/2	-1	$\tilde{B}^0 \tilde{W}^0 \tilde{H}_u^0 \tilde{H}_d^0$	$\tilde{\chi}_1^0 \tilde{\chi}_2^0 \tilde{\chi}_3^0 \tilde{\chi}_4^0$
Charginos	1/2	-1	$\tilde{W}^\pm \tilde{H}_u^\pm \tilde{H}_d^\mp$	$\tilde{\chi}_1^\pm \tilde{\chi}_2^\pm$
Gluino	1/2	-1	\tilde{g}	(same)
Goldstino (Gravitino)	1/2 (3/2)	-1	\tilde{G}	(same)

Table 2.3: MSSM particle content including gauge and mass eigenstates.

supersymmetric particles. Although there is no universally accepted model for supersymmetry breaking, both theoretical and experimental constraints suggest that a viable model must satisfy two key conditions. **First, the scale of supersymmetry breaking must be much larger than the electroweak scale (soft breaking).** Second, this breaking must be transmitted to the MSSM sector via a messenger sector. Broadly, there are two main classes of mediation, depending on the interaction through which the SUSY-breaking information is communicated, i.e. gravity mediation and gauge mediation. In gravity-mediated SUSY breaking, the breaking occurs in a hidden sector and affects the visible MSSM sector through gravitational interactions. Alternatively, in gauge-mediated SUSY breaking (GMSB) the breaking is communicated via messenger fields, which couple to the MSSM through the Standard Model gauge interactions.

In the MSSM, the total Lagrangian can be expressed as:

$$\mathcal{L}_{\text{MSSM}} = \mathcal{L}_{\text{SUSY}} + \mathcal{L}_{\text{soft}}, \quad (2.11)$$

where $\mathcal{L}_{\text{SUSY}}$ represents the supersymmetric part of the theory, containing the kinetic terms, superpotential interactions, and gauge interactions, while $\mathcal{L}_{\text{soft}}$ encapsulates the SUSY-breaking terms mediated to the MSSM sector. These soft SUSY-breaking terms are introduced to break supersymmetry explicitly but in a controlled way that avoids reintroducing large quadratic divergences. By including $\mathcal{L}_{\text{soft}}$ in the MSSM Lagrangian, the mass spectrum of SUSY particles can be adjusted without compromising the naturalness of the theory or its renormalizability.

In the MSSM, soft SUSY-breaking terms generate the mass differences between superpartners and their SM counterparts and lead to mixing between states due to the gauge structure and interactions of the theory. For example, the neutral gauginos (\tilde{B} and \tilde{W}^0) and neutral Higgsinos (\tilde{H}_u^0 and \tilde{H}_d^0) mix to form four neutral mass eigenstates called neutralinos ($\tilde{\chi}_i^0$, $i = 1, 2, 3, 4$), which are ordered by increasing mass, i.e. $i = 1$ is the

lightest. Similarly, charged Higgsinos (\tilde{H}_u^+ and \tilde{H}_d^-) and charged gauginos (\tilde{W}^+ and \tilde{W}^-) mix to form two pairs of charged mass eigenstates called charginos ($\tilde{\chi}_i^\pm$, $i = 1, 2$).

These mass eigenstates emerge after diagonalizing the corresponding mass matrices, which depend on the SUSY parameters, such as the supersymmetric Higgs-mass term (μ) and gaugino masses. The lightest neutralino ($\tilde{\chi}_1^0$) is often the lightest SUSY particle (LSP), making it a promising candidate for dark matter in many SUSY models.

Moreover, the introduction of soft SUSY-breaking terms ensures that supersymmetric particles acquire masses consistent with current experimental limits, while maintaining the natural cancellation of quadratic divergences in the Higgs boson mass. This mechanism lies at the heart of supersymmetry's solution to the hierarchy problem, providing a compelling theoretical framework for stabilizing the electroweak scale.

2.2.2 Hierarchy Problem

One of the most pressing issues in the SM is the hierarchy problem, which arises due to the large quantum corrections to the Higgs boson mass from high-energy scales. These corrections would drive the Higgs mass to the Planck scale without fine-tuning. SUSY addresses this problem by introducing superpartners that cancel the quadratic divergences from SM particles, ensuring the Higgs mass remains at the electroweak scale in a natural way.

A key motivation for supersymmetry is its ability to resolve the hierarchy problem. In the Standard Model (SM), the Higgs boson mass receives large quantum corrections from loop diagrams involving particles like the top quark, gauge bosons, and the Higgs itself [28]. These corrections are quadratically divergent, scaling as the square of the ultraviolet (UV) cutoff scale, Λ_{UV} , which could be as high as the Planck scale (10^{19} GeV). This implies that, without extreme fine-tuning, the Higgs mass would naturally be driven to a much higher value than its observed mass of approximately 125 GeV. The correction to the Higgs mass squared, Δm_H^2 , can be expressed as:

$$\Delta m_H^2 \sim \frac{\lambda}{16\pi^2} \Lambda_{\text{UV}}^2 \quad (2.12)$$

where λ is the coupling constant, for example the top Yukawa coupling for top-quark loops, $\lambda_t \sim 1$. This divergence suggests that the Higgs mass requires unnatural fine-tuning to remain at the electroweak scale.

Supersymmetry provides a natural solution by introducing superpartners whose contributions systematically cancel the quadratic divergences from SM particles. Specifically, fermionic and bosonic loops contribute terms of opposite signs, and for unbroken SUSY, these corrections cancel at all orders. However, SUSY must be a broken symmetry of nature, as we do not observe mass-degenerate SM particles and their superpartners. SUSY breaking introduces a mass difference between particles and their superpartners.

For the top quark and its superpartners, the top squarks (\tilde{t}_1, \tilde{t}_2), this implies:

$$m_{\tilde{t}_1}, m_{\tilde{t}_2} > m_t ,$$

where $m_{\tilde{t}_1}$ and $m_{\tilde{t}_2}$ are the masses of the two stop eigenstates. While this breaking reintroduces partial corrections, SUSY ensures that the quadratic divergences are replaced by weaker logarithmic corrections. The remaining correction to the Higgs mass squared can be expressed as:

$$\Delta m_H^2 \sim -m_{\text{soft}}^2 \left[\frac{\lambda}{16\pi^2} \ln \left(\frac{\Lambda_{\text{UV}}}{m_{\text{soft}}} \right) + \dots \right], \quad (2.13)$$

where m_{soft} is the largest scale of the soft SUSY-breaking terms.

To maintain naturalness, m_{soft} must not be too large, and the top squark masses should not deviate significantly from the top quark mass. Otherwise, the cancellation between fermionic and bosonic contributions becomes less effective, reintroducing fine-tuning [29]. This constraint predicts SUSY particle masses in the TeV range, which aligns with the energy scales accessible at current experiments like the LHC.

2.2.3 R-Parity

In the Minimal Supersymmetric Standard Model (MSSM), R-parity is a discrete symmetry defined as $R_p = (-1)^{3(B-L)+2s}$, where B is the baryon number, L is the lepton number, and s is spin of the particle. Under R-parity conservation, Standard Model particles have $R_p = +1$, while their supersymmetric counterparts have $R_p = -1$.

One key motivation for conserving R-parity is the prevention of rapid proton decay, which would occur if baryon and lepton number-violating terms were simultaneously present in the superpotential.

In models with R-parity conservation, the LSP is stable because it cannot decay into SM particles, making it an ideal dark matter candidate. This property naturally explains the observed relic density of dark matter, as the LSP could survive from the early universe to the present day. Observational constraints from cosmological measurements, such as those by the Planck satellite, require SUSY models to predict relic densities consistent with the observed dark matter abundance [30].

Many SUSY models predict relic densities much larger than these observed values, highlighting the need for mechanisms to reduce the LSP population. One such mechanism is co-annihilation, which becomes effective when the LSP has a nearly degenerate mass with another SUSY particle, such as the stop (\tilde{t}). The small mass gap between the LSP and the stop allows for significant co-annihilation processes, depleting the LSP density and aligning the calculated relic density with experimental observations.

Co-annihilation mechanisms become essential in compressed spectra scenarios, where the mass difference between superpartners is small [31]. In such cases, the small mass difference between the LSP and its next-to-lightest supersymmetric particle (NLSP) leads to soft decay products, presenting significant experimental challenges. However, these scenarios remain theoretically compelling because they satisfy cosmological relic density constraints and align with null results from direct SUSY searches at colliders. They highlight the importance of dedicated searches targeting compressed mass spectra to test these hypotheses in a systematic and experimentally accessible manner.

2.2.4 Unification of Gauge Coupling

Another compelling feature of supersymmetric models is the unification of gauge couplings at a high energy scale, often associated with Grand Unified Theories (GUTs). In the Standard Model, the renormalization group evolution of the gauge coupling constants suggests that they approach each other at high energies but fail to converge at a single unification scale. However, in the MSSM the presence of superpartners modifies the running of the gauge couplings, leading to unification at approximately 10^{16} GeV, as shown in Figure 2.3.

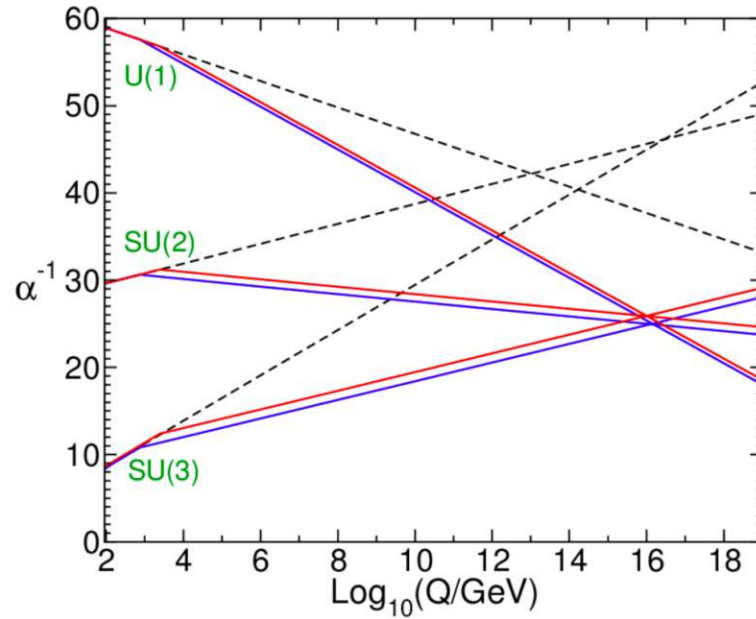


Figure 2.3: Renormalization group evolution (RGE) of the inverse gauge couplings in the Standard Model (SM) and the Minimal Supersymmetric Standard Model (MSSM): the dashed lines represent the SM, where the gauge couplings evolve separately and do not unify at a single scale. In contrast, the solid lines correspond to the MSSM, where the modified running leads to successful unification around 10^{16} GeV. [27]

2.2.5 Search for compressed spectra

In the MSSM, each SM fermion has two scalar superpartners, denoted \tilde{f}_L and \tilde{f}_R , corresponding to the superpartners of the left and right handed SM fermions. The physical sfermion mass eigenstates are mixtures of these weak-eigenstate fields, yielding \tilde{f}_1 (lighter) and \tilde{f}_2 (heavier). For light-flavour squarks, the mixing is generally negligible because the Yukawa couplings are small. However, for the third-generation fermions (top, bottom, and tau) the larger Yukawa couplings cause significant mixing, leading to substantial mass splittings between \tilde{f}_1 and \tilde{f}_2 . Since the top quark is the heaviest SM fermion, its superpartner, the stop (\tilde{t}), exhibits the strongest mixing effects, resulting in a large mass gap between its eigenstates \tilde{t}_1 and \tilde{t}_2 , often making \tilde{t}_1 the lightest squark. This makes \tilde{t}_1 searches particularly well-motivated, especially in the context of naturalness arguments. To maintain a natural solution to the hierarchy problem and avoid excessive fine-tuning, it is typically required that the stop mass be at or below the TeV scale, ensuring it is not too heavy.

A crucial parameter in stop searches is the mass difference Δm between the stop and the lightest supersymmetric particle (LSP), usually the neutralino $\tilde{\chi}_1^0$.

$$\Delta m = m_{\tilde{t}_1} - m_{\tilde{\chi}_1^0}$$

The value of Δm determines the kinematics of the stop decay channels and strongly impacts search strategies. If $\Delta m > m_t$, the stop decays predominantly via an on-shell top quark:

$$\tilde{t}_1 \rightarrow t \tilde{\chi}_1^0,$$

For $m_W < \Delta m < m_t$, the top quark is off-shell and the stop undergoes a three-body decay:

$$\tilde{t}_1 \rightarrow b W \tilde{\chi}_1^0$$

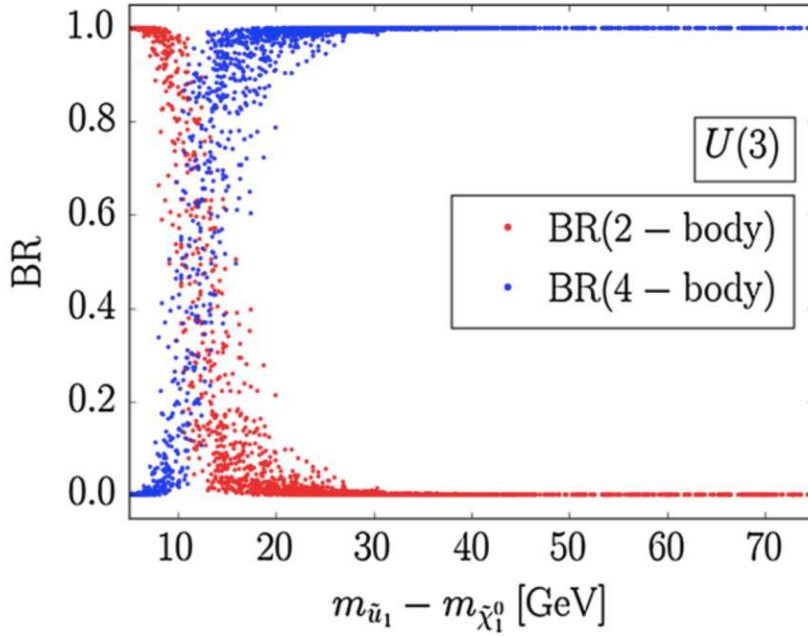


Figure 2.4: The figure is taken from [32] and shows the branching ratios for two-body FCNC (red) and four-body (blue) decays of stop. The two-body decay mode dominates for $\Delta m \lesssim 10$ GeV, while the four-body mode dominates for $\Delta m \gtrsim 20$ GeV.

For even smaller mass splittings, $\Delta m < m_W$, both the top quark and the W boson are off-shell, resulting in a four-body decay:

$$\tilde{t}_1 \rightarrow b f f' \tilde{\chi}_1^0 \quad (2.14)$$

In some scenarios, a light chargino ($\tilde{\chi}_1^\pm$) sits with an intermediate-mass between the stop and neutralino, modifying the decay chain but yielding the same four-body final state via an intermediate chargino,

$$\tilde{t}_1 \rightarrow b \tilde{\chi}_1^\pm \text{ followed by } \tilde{\chi}_1^\pm \rightarrow f f' \tilde{\chi}_1^0. \quad (2.15)$$

If all other SUSY particles are much heavier, the stop can also decay via a flavour-changing neutral current (FCNC) process [32]:

$$\tilde{t}_1 \rightarrow c \tilde{\chi}_1^0.$$

The branching ratios of the two- and four-body decay modes as a function of Δm are shown in Figure 2.4 for a specific flavour symmetry. The two-body (FCNC) decay dominates in the highly compressed regime, $\Delta m \lesssim 10$ GeV, whereas the four-body mode takes over for $\Delta m \gtrsim 20$ GeV. This parameter space where $\Delta m < m_W$, is called the compressed region, where a nearly degenerate stop and neutralino, such as the LSP, exist.

At the smallest mass splitting ($\Delta m \lesssim 30$ GeV), the total stop decay width (Γ_{tot}) becomes small, implying a long lifetime (see Figure 2.5). A $\Gamma_{tot} \approx 10^{-12}$ GeV corresponds to the lifetime of order picoseconds, i.e. a proper decay length of sub-mm, which translates to a visible displaced vertex inside the detector. The four-body decay is suppressed heavily by the phase space and its partial width falls rapidly as Δm decreases. If the flavour-changing two-body (FCNC) channel is also suppressed, the total width becomes

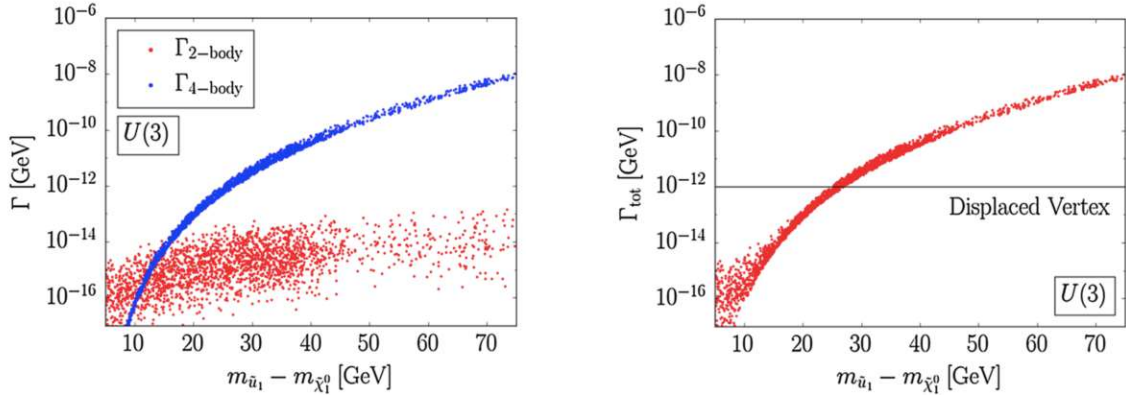


Figure 2.5: The two- and four-body decay widths (left) and the total width (right) as a function of Δm are shown, with the horizontal black line at $\Gamma_{\text{tot}} = 10^{-12}$ GeV indicates the value of total width where displaced vertices can be observed [32].

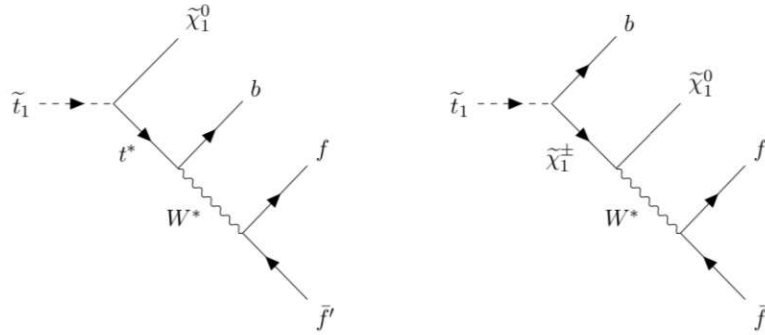


Figure 2.6: Two possible four-body decays of stop are shown in the figure, and is taken from [34].

small (Figure 2.5(left)), and the stop decay can occur over a range of millimetres to centimetres producing displaced-vertex signatures.

The search presented in this thesis focuses on the single-lepton final state, explicitly targeting the four-body stop decay; if the stop is the next-to-lightest supersymmetric particle (NLSP), it decays directly, as shown in Equation (2.14). The stop decay chains for both four-body variations, i.e. direct and chargino-mediated, are shown in Figure 2.6.

Considering the compelling theoretical features of SUSY, no superpartners have yet been observed. While Run 2 searches exclude top squarks up to 1.2 TeV for large mass splittings and cover the $m_W < \Delta m < m_t$ regime to nearly 900 GeV. The compressed band is less explored as it is a challenging phase space, but the scenario with a long-lived stop for $\Delta m \leq 30$ GeV is completely unexplored [33].

In dark-matter co-annihilation models, this small Δm is actually favoured as it can reproduce the correct relic density. Experimentally, however, this region is challenging because the visible decay products are very soft and thus easily missed by conventional search strategies optimised for energetic particles. This makes this search very sought after but challenging at the same time.

To recover sensitivity in this corner, this analysis develops “soft” techniques that improve sensitivity in the compressed regime, including strategies to identify long-lived stop decays. By unifying prompt four-body and long-lived stop searches in a single framework, this analysis exhaustively probes the compressed corners of natural SUSY. Technical details of these methods will be introduced in Chapter 4.

Chapter 3

The CMS experiment at the LHC

The results presented in this thesis are obtained using the data collected at the Compact Muon Solenoid (CMS) experiment, one of the four experiments located at the Large Hadron Collider (LHC), from the proton-proton collisions. This chapter focuses on the experimental side and sheds some light on the marvel of science and technology that LHC is. Details on the most powerful accelerator built by mankind can be found in Section 3.1, and Section 3.2 gives an overview of the CMS experiment. Unless stated otherwise, geometrical and performance parameters of the CMS subdetectors and operating conditions in this chapter refer to the Run 2 (2016–2018) CMS configuration at $\sqrt{s} = 13$ TeV with 25-ns bunch spacing.

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a 27 km long circular particle accelerator operated by CERN. The LHC is located in both France and Switzerland, about 100 m underground. The tunnel in which LHC resides now is recycled from its predecessor, Large Electron-Positron Collider (LEP) [35]. LHC is used to perform not only proton-proton (pp) collisions at very high centre-of-mass energies but also heavy ions, e.g. lead-lead (PbPb) or proton-lead (pPb) collisions. LHC is designed to reach centre-of-mass energy of 14 TeV for proton-proton collisions. However, for the data-taking period of 2015 to 2018 (Run 2), the centre-of-mass energy, denoted by \sqrt{s} , was 13 TeV. Higher centre-of-mass energies provide more energy for the production of undiscovered particles or potential new physics phenomena in existing processes. For this reason, achieving the highest possible centre-of-mass energies in discovery machines like the LHC is highly desirable.

The procurement of protons begins with a bottle of hydrogen gas, from which electrons are removed by applying an electric field. The extracted protons are then pre-accelerated through multiple stages. During Run 2, the first stage of acceleration occurred in a linear accelerator (Linac 2) at CERN; during the second long shutdown (2019-2020) it was replaced by Linac 4. The protons then pass through the Proton Synchrotron Booster (PSB) followed by the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS). In the SPS, the protons are accelerated to an energy of 450 GeV before being injected into the LHC's two separate, counter-rotating beam pipes. Within the LHC, they undergo further acceleration, completing multiple revolutions while passing through radiofrequency (RF) cavities and specialised magnet configurations. This process gradually increases their energy until they reach the final collision energy of 6.5 TeV.

Inside the LHC, the counter-rotating beams are bent by dipole magnets and focused by quadrupole magnets, respectively, so that they collide at the designated interaction points. Four detectors are placed at various interaction points (IP), as shown in Figure 3.1:

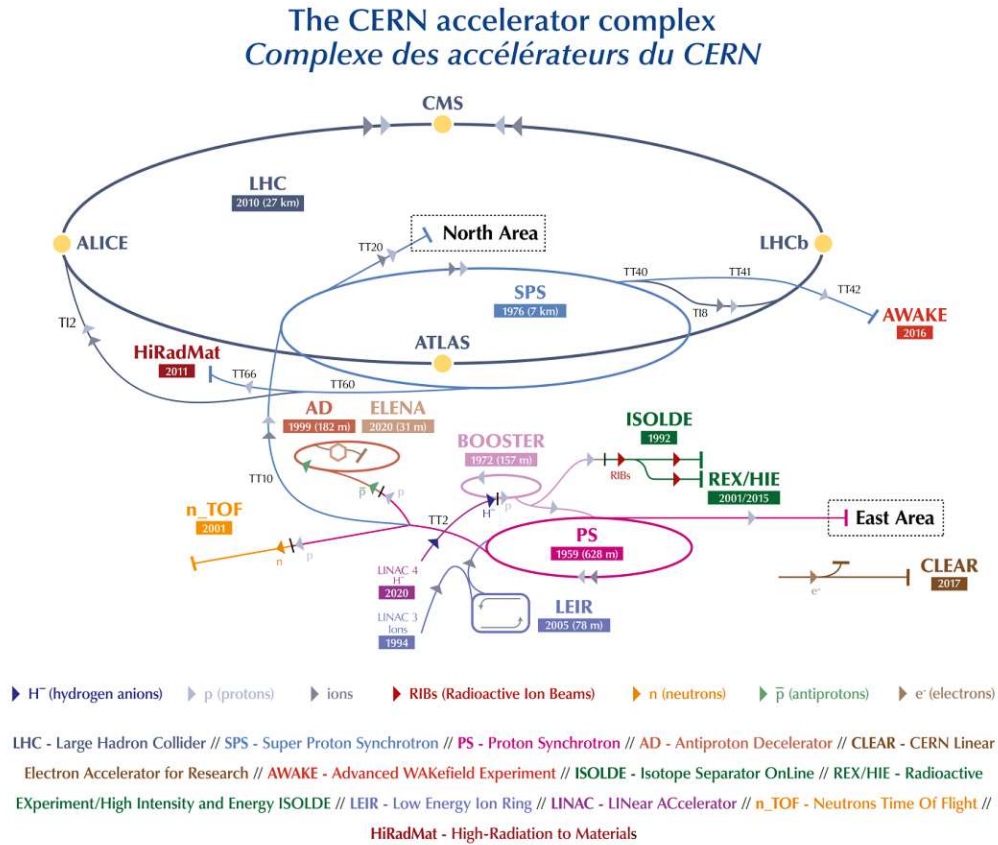


Figure 3.1: The CERN accelerator complex is shown. The figure is taken from [36].

ALICE: A Large Ion Collider Experiment with a focus on heavy-ions physics

LHCb: Large Hadron Collider Beauty Experiment mainly focuses on studying b quarks

ATLAS: A Toroidal LHC ApparatuS is a multi-purpose detector

CMS: Compact Muon Solenoid is also a multi-purpose detector

3.1.1 Luminosity at the LHC

The proton beams in the LHC are not continuous but arranged into bunches. During Run 2 the machine operated with a 25 ns bunch spacing; the nominal filling scheme provided up to 2808 bunches per beam (n_b), with each bunch containing approximately 10^{11} protons (N_p). In practice, fills typically used about 2556 colliding bunch pairs, whereas 2808 is the design maximum. When the counter-rotating beams are made to collide at one of the experiments, the given spacing between bunches results in a bunch-crossing frequency (or revolution frequency) of 40 MHz. The total number of collisions per second is determined by the instantaneous luminosity, $\mathcal{L}_{\text{inst}}$. For a process with a given cross section, σ , the event rate \dot{N} can be computed as:

$$\dot{N} = \mathcal{L}_{\text{inst}} \sigma \quad (3.1)$$

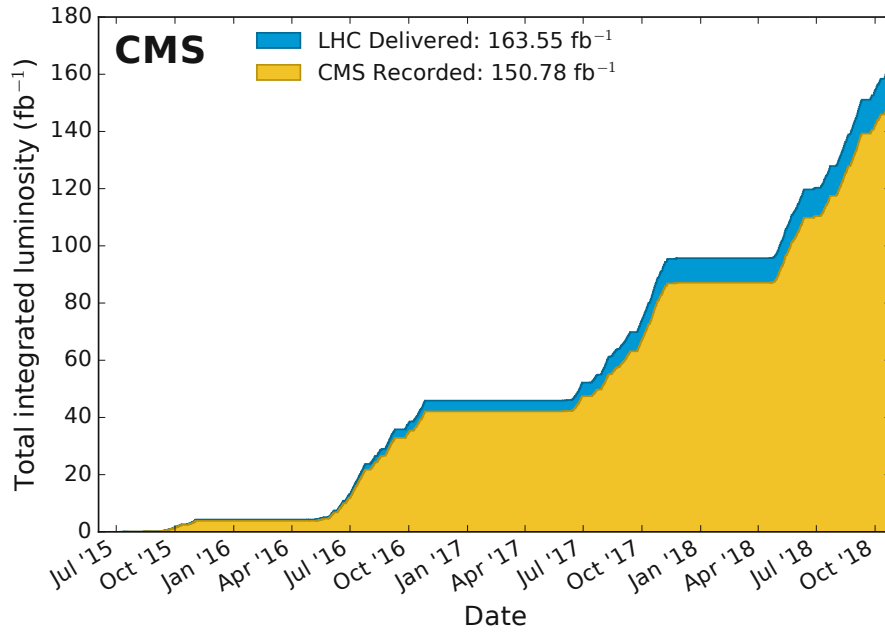


Figure 3.2: The cumulative integrated luminosity \mathcal{L} delivered by the LHC (blue) and recorded by the CMS experiment (yellow) during stable beams for pp collisions over years 2015–2018 is shown. The figure is taken from [37].

The instantaneous luminosity is given by:

$$\mathcal{L}_{\text{inst}} = \frac{\gamma f n_b N_p^2}{4\pi\epsilon_n\beta^*} F, \quad (3.2)$$

where γ is the Lorentz factor and f is the revolution frequency. The spread of the beam in the position and momentum phase space is given by the normalized transverse beam emittance, ϵ_n . β^* is the value of the beta function at the collision point, and F is the luminosity reduction factor considering the crossing angle of the collision. The integrated luminosity represents the total number of collisions over a given period. It is obtained by integrating the instantaneous luminosity:

$$\mathcal{L} = \int \mathcal{L}_{\text{inst}} dt \quad (3.3)$$

The number of events in the collected sample can be given by $N = \mathcal{L}\sigma$ for a given process. A process with a higher cross section produces a greater number of events in collisions. On the other hand, luminosity quantifies the accelerator's capacity to make particles collide. The LHC was originally designed for an instantaneous luminosity of $10^{34} \text{ cm}^{-2}\text{s}^{-1}$, which was exceeded by a factor of about two by the end of Run 2 (2017-2018). LHC delivered about 163 fb^{-1} of data in Run 2, as shown in Figure 3.2.

3.1.2 Pileup

Proton-proton interactions can be classified as either elastic collisions (protons scattering off without much interaction) or inelastic collisions (quarks and gluons of the proton strongly interact and form new particles). Among the inelastic interactions in a bunch

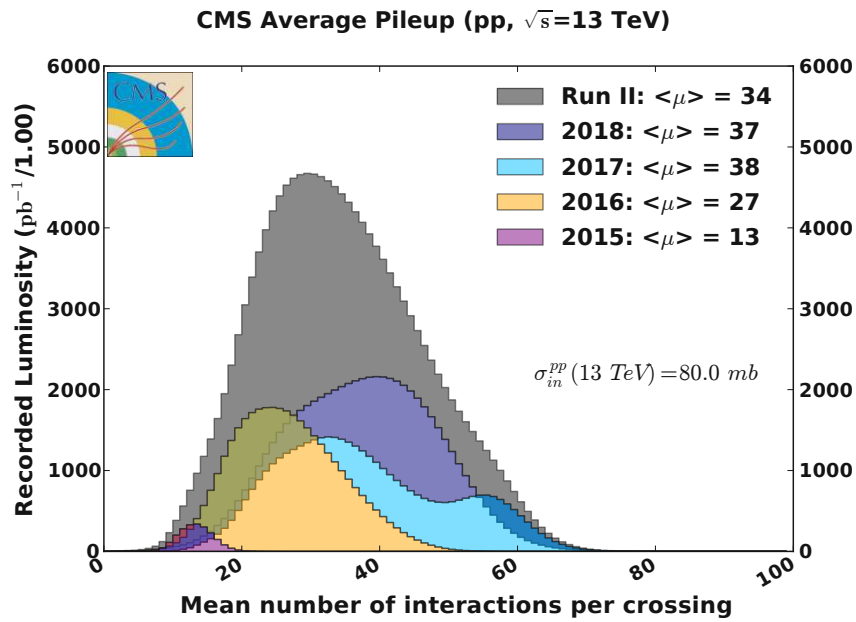


Figure 3.3: Average PU distributions for each year of data-taking are shown, together with the full Run 2 combined. The distributions of the mean number of interactions per bunch crossing, i.e., average PU, denoted by μ , are shown with the average numbers of PU quoted in the legend. The figure is taken from [37].

crossing, the interaction with the largest momentum exchange, Q^2 , is called the hard scattering interaction.

As each LHC bunch contains $\sim 10^{11}$ protons, with a bunch crossing of 25 ns (40 MHz) results in about 600 million inelastic interactions per second at the nominal luminosity.

The additional interactions to the hard scattering interaction, occurring within the same bunch crossing or even neighbouring bunch crossings are referred to as pileup (PU). Figure 3.3 shows the PU distributions from different data-taking periods during Run 2, going up to an average of 38 collisions per bunch crossing.

3.2 The Compact Muon Solenoid Detector

The CMS detector is a multi-purpose detector with several subdetectors designed to record, identify, and reconstruct different kinds of particles produced in proton-proton (pp) or heavy-ion collisions. The subdetectors are arranged in a specific order in concentric layers centred around the interaction point as shown in Figure 3.4. The silicon tracker surrounds the beamline, which can reconstruct the charged particles coming out from the interaction point. Being the innermost subsystem, it provides precise vertex (interaction point) measurements. The intense particle flux near the interaction point motivates the use of silicon due to its radiation hardness, fast response, and fine granularity. The tracker is designed with a low material budget such that most particles traverse it with minimal particle interaction such as multiple scattering, bremsstrahlung, photon conversion which would alter tracks and affect energies of particles. The electromagnetic (ECAL) and the hadronic calorimeter (HCAL), located between the silicon tracker and the solenoid magnet, are two different subsystems for measuring the energies of electrons and photons and charged and neutral hadrons, respectively. Calorimeters measure the energies of the particles by making the particles interact and deposit their energies as a cascade or shower of particles (collimated spray). These calorimeters operate inside the solenoid,

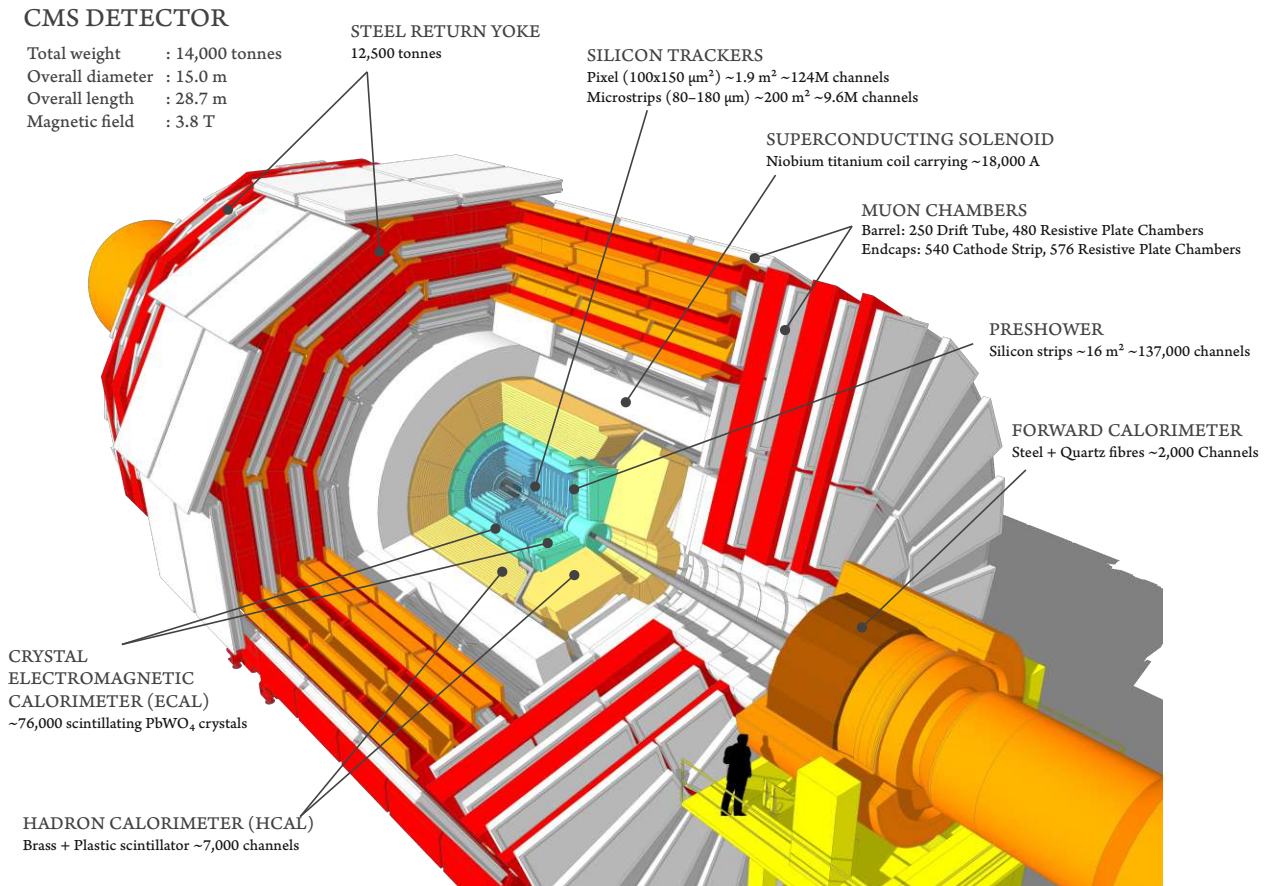


Figure 3.4: A cutaway diagram of the CMS detector along with all the subsystems is shown [38].

encapsulating all the subdetectors and providing the magnetic field necessary for particle tracking. Outside the solenoid, there is an extensive muon system.

The terms “event” and “collision” are used interchangeably. After a collision, subdetectors record information, which is processed into primitive objects. Based on this information, the trigger system decides whether to accept or reject an event. If the event is accepted and saved, offline reconstruction of events transforms raw data into particles and objects, such as leptons and missing transverse momentum, which are important for analyses. This chapter will discuss the CMS coordinate system and these subdetectors, followed by a description of event reconstruction.

3.2.1 CMS Coordinate System

The origin of the CMS coordinate system is defined at the nominal interaction point (IP), located at the center of the detector. The CMS coordinate system follows the conventional right-handed Cartesian system. The z -axis is aligned with the beamline, while the x - and y -axes define the transverse plane. As seen from above, the proton beams circulate in opposite directions (counter-rotating) within the LHC. Due to overall momentum conservation, the total momentum in the transverse plane sums to zero. As a result, the transverse momentum of individual particles, p_T , is widely used in analyses. Another important quantity in the transverse plane is the missing transverse momentum, p_T^{miss} , which accounts for the momentum imbalance caused by undetected particles, such as neutrinos. In experimental analyses, this quantity is often referred to as *missing transverse energy* (E_T^{miss}), although this terminology is somewhat misleading. Transverse mass, another key quantity in the transverse plane, is defined as:

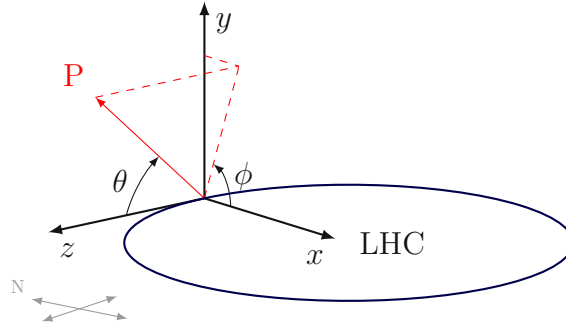


Figure 3.5: This illustration shows the CMS coordinate system. The definition of the azimuthal (ϕ) and the polar angle (θ) are depicted for a particle with momentum P , produced at the interaction point. [39]

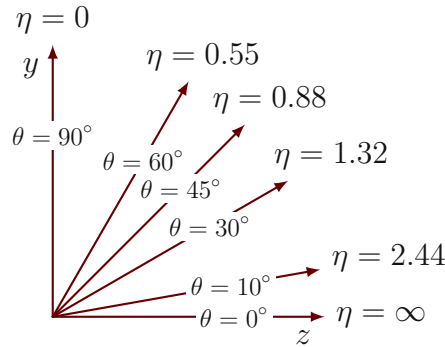


Figure 3.6: The pseudorapidity η in relation to polar angle θ is shown. [39].

$$m_T = \sqrt{m^2 + p_T^2} \quad (3.4)$$

The trajectory of a particle is characterised by two angles: the azimuthal angle (ϕ), measured from the x -axis in the transverse plane, and the polar angle (θ), measured from the z -axis, as shown in Figure 3.5. Instead of the polar angle, analyses often use pseudorapidity (η), as it remains invariant under Lorentz boosts along the beam axis, making it a more suitable variable for describing particle kinematics. Pseudorapidity is defined as:

$$\eta = -\ln \left[\tan \left(\frac{\theta}{2} \right) \right]. \quad (3.5)$$

The correlation between polar angle (θ) and pseudorapidity (η) is visualized in the Figure 3.6.

3.2.2 Superconducting Solenoid Magnet

The CMS detector's superconducting magnet is a critical component of the experiment. The large solenoid, made from reinforced Niobium-Titanium (NbTi) with an innovative four-layer winding design, has a diameter of 6 m and a length of 12.5 m. When operating at full current, it can store 2.6 GJ of energy. A steel yoke weighing approximately 12,500 tonnes surrounds the magnet to increase the central field and to improve the uniformity of the magnetic field within the tracker volume and to confine the stray field

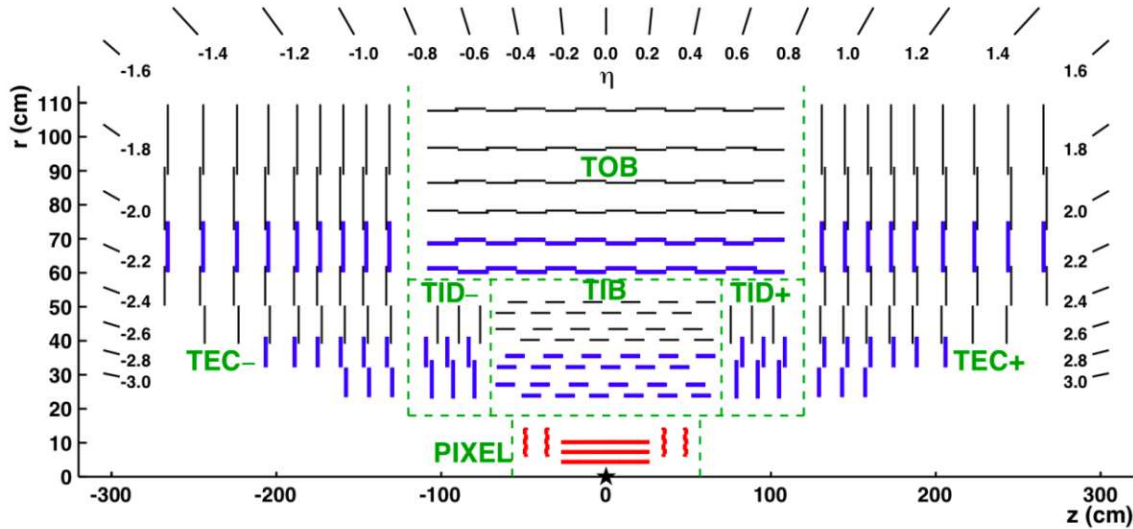


Figure 3.7: A schematic layout of the CMS tracking system before the Phase-1 upgrade, illustrating the pixel tracker and various components of the silicon strip tracker. [41]

outside the detector. The yoke comprises five cylindrical sections (barrels) and six endcap disks [40]. The magnet produces a uniform 3.8 T field in the central detector volume. Beyond its magnetic function, the steel yoke also serves as the detector’s structural support, ensuring it withstands the forces generated by the powerful magnetic field. Additionally, the steel yoke acts as an absorber before the muon layers, filtering out other particles and ensuring clean muon detection in the outer muon system.

The magnetic field bends the trajectories of charged particles (tracks) while leaving neutral particles unaffected. This separation improves particle identification in the calorimeters.

3.2.3 Inner Tracking System

Measuring the charged particle’s trajectories, resulting from the collision at the interaction point, efficiently and precisely, along with precise reconstruction of secondary vertices, is accomplished by the inner tracking system. The tracking system is located around the interaction point with a diameter of 2.5 m and a length of 5.8 m. To correctly identify the trajectories and assign them to the correct bunch crossing, the detector technology needs to be highly responsive and have fine granularity. Due to the proximity to the interaction point and the high luminosity of the LHC, the radiation damage caused by intense particle flux must be considered when designing the tracking system. All these requirements led to the choice of a fully silicon-based tracker design. The CMS tracker is the largest built silicon tracker with about 200 m² active silicon area. A schematic view of the CMS tracker, consisting of a small inner pixel detector and a larger silicon strip detector, is shown in Figure 3.7.

Pixel detector

The part of the tracking system closest to the interaction point is the pixel detector. The pixel detector covers the pseudorapidity range $|\eta| < 2.5$ with a pixel cell size of $100 \times 150 \mu\text{m}^2$. The pixel system is responsible for precise tracking points, which are important for the reconstruction of good secondary vertex, and providing seed tracks for

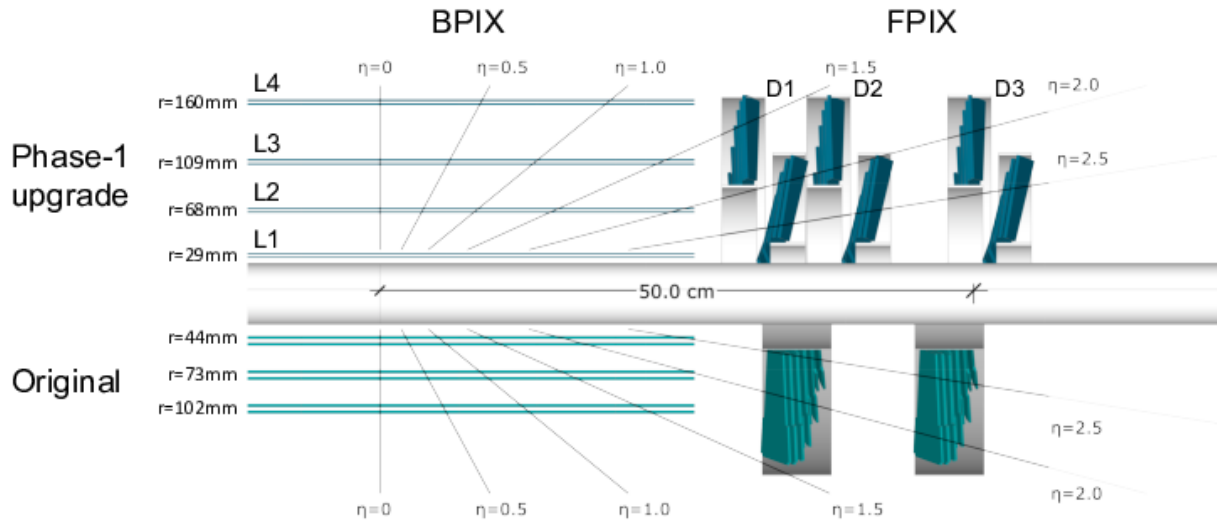


Figure 3.8: A comparison of the CMS pixel detector before and after the Phase-1 upgrade shows both versions used during Run 2 data-taking. The “Original” detector, used for data-taking in 2016, featured three barrel layers and two endcap disks. In contrast, the Phase-1 upgraded detector, installed for 2017–2018, included four barrel layers and three endcap disks, each with inner and outer rings. The figure is adapted from [42].

complete outer track reconstruction [40]. The original pixel detector consisted of three barrel layers at radii of 44, 73 and 102 mm, with two endcap disks per side at $|z| = 345$ and 465 mm. During the 2016–2017 year-end technical stop, the original pixel detector was replaced with an updated system, referred to as the CMS Phase-1 upgrade, to maintain efficient tracking at CMS, accommodating increased luminosity and pileup following the accelerator upgrade during the first long shutdown (LS1) [42]. The CMS Phase-1 pixel detector consists of four barrel layers at radii of 29, 68, 109, and 160 mm, along with three endcap disks per side, each comprising inner and outer rings, as shown in Figure 3.8.

Silicon strip detector

The Silicon Strip detector is the second closest subsystem to the beam pipe in the CMS detector. It consists of four subsystems: the Tracker Inner Barrel (TIB), Tracker Inner Disks (TID), Tracker Outer Barrel (TOB), and Tracker Endcaps (TEC). In the barrel region, the silicon strip detector comprises ten layers: four in the TIB and six in the TOB. The TIB is complemented by three TID layers at each end. The TID consists of three small disks on each end, bridging the gap between the TIB and TEC.

The Tracker Endcaps (TEC) are located in the forward region, covering $120 \text{ cm} < |z| < 280 \text{ cm}$ along the beamline. Each TEC contains nine disks per end, with each disk comprising four to seven concentric rings, transitioning gradually from the outer to the

inner region. The CMS silicon strip detector surrounds the pixel detector and together they provide efficient tracking of charged particles within $|\eta| < 2.5$.

Silicon Strip Tracker Pre-Amplifier Simulation for early 2016 data

The Silicon Strip Tracker (SST) readout system consists of a front-end APV25 readout chip, analogue optical links and an off-detector Front-End Driver (FED) processing board [43]. During late 2015 and early 2016, the Silicon Strip Tracker experienced a decrease in signal-to-noise ratio and a concurrent loss of hits on tracks. This effect resulted from saturation caused by Highly Ionizing Particles (HIPs) in the pre-amplifier of the APV25 readout chip. The operating temperature affected the drain speed of APV25 chips, leading to the slow discharge of the pre-amplifier under high occupancy conditions. Approximately 20 fb^{-1} of 2016 data was affected under these settings, referred to as “old APV settings”. These settings caused a delay in charge recovery, reducing hit efficiency. To mitigate the issue, the draining speed was later changed to allow for faster recovery of the pre-amplifier and improved hit efficiency. For the correct description of the affected 2016 data, a specific description of the APV25 chip’s dynamic behaviour under saturation conditions was included in the simulation of Silicon Strip Tracker [44]. The 2016 data-taking period was divided into two distinct periods:

- The preVFP, when the APV25 saturation issue affected data quality.
- The postVFP, after introducing the Pre-Amplifier Feedback Voltage Bias (VFP) fix, which corrected the issue.

The total recorded luminosities for these periods were 19.5 fb^{-1} (preVFP) and 16.81 fb^{-1} (postVFP).

3.2.4 Calorimeters

The Calorimeter system in CMS measures the total energy and direction of electromagnetically and strongly interacting particles by absorbing their energy. It is placed within the magnet and surrounds the tracker. CMS utilises two types of calorimeters:

- Electromagnetic Calorimeter (ECAL): Measures the energies of electrons and photons through electromagnetic showers.
- Hadronic Calorimeter (HCAL): Measures the energies of hadronic jets through hadronic showers.

Jets are reconstructed from collimated sprays of hadrons originating from quark or gluon fragmentation. When a high-energy particle interacts with the calorimeter material, it initiates a shower, producing a cascade of lower-energy secondary particles that are subsequently absorbed and detected.

The Electromagnetic Calorimeter

The ECAL is a hermetic and homogeneous calorimeter made of scintillating lead tungstate (PbWO_4) crystals. The ECAL is comprised of barrel ECAL (EB), endcap ECAL (EE) and preshower (ES) as shown in Figure 3.9. The central ECAL, including the barrel and endcap regions, consists of over 75000 scintillating crystals.

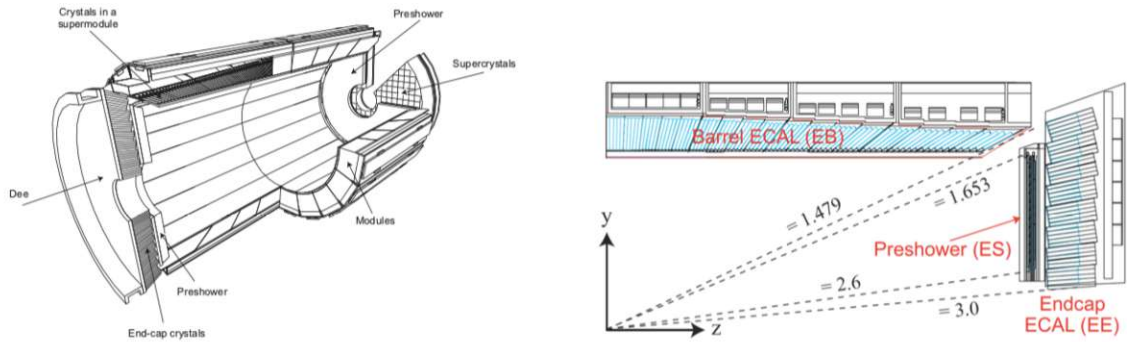


Figure 3.9: Schematic layout of the CMS Electromagnetic Calorimeter (ECAL). The left image illustrates the arrangement of the barrel supermodules, endcap crystals, and the preshower detector (ES). The right image provides a geometric view of one ECAL quadrant, highlighting its structural components. The figure is adapted from [46].

- The ECAL barrel contains 61200 crystals, arranged in 36 supermodules containing 1700 crystals each, measuring the energies of particles in the pseudorapidity range of $|\eta| < 1.479$.
- Each ECAL endcap consists of two Dees with 3662 crystals, extending the coverage to $1.479 < |\eta| < 3.0$.
- The silicon strip preshower (ES) is placed before EE, covering a pseudorapidity range of $1.653 < |\eta| < 2.6$. It consists of two layers of lead absorbers, each followed by a silicon strip sensor arranged in an interleaved structure.

The highly segmented ES, positioned in front of the EE, helps distinguish single photons from photon pairs produced in neutral pion decays ($\pi^0 \rightarrow \gamma\gamma$) [45]. Analyses typically veto the barrel-endcap transition region $1.4442 < |\eta| < 1.566$ because the reconstruction is suboptimal.

When interacting with high-energy electrons and photons ($e\gamma$), the ECAL crystals produce electromagnetic showers, i.e. a cascade of lower-energy electrons and photons from the bremsstrahlung radiation (for electrons) and pair production (for photons). This cascade eventually results in low-energy electrons and photons, which are absorbed within the crystal, emitting scintillation light that is detected. The scintillation light is read out by avalanche photodiodes (APDs) in the barrel and vacuum phototriodes (VPTs) in the endcap [40]. Although lead tungstate crystals are intrinsically radiation-tolerant, irradiation-induced colour centres are produced which absorb a fraction of transmitted light and reduce crystals' optical transmission. This transparency loss is monitored using injected laser light and accounted for during offline data reconstruction.

The Hadronic Calorimeter

The Hadronic Calorimeter (HCAL) is crucial for measuring hadronic jet energy and missing transverse momentum (p_T^{miss}). p_T^{miss} accounts for undetected neutrinos from the decays of particles produced during collision or the exotic particles from BSM interactions, leaving the detector without a trace. These requirements ensure that the HCAL provides hermetic coverage, i.e. it is able to capture nearly all hadronic energy deposits without gaps.

The HCAL, the outermost calorimeter surrounding the ECAL, provides full coverage and ensures that all interacting particles deposit their energy. The Barrel (HB) and

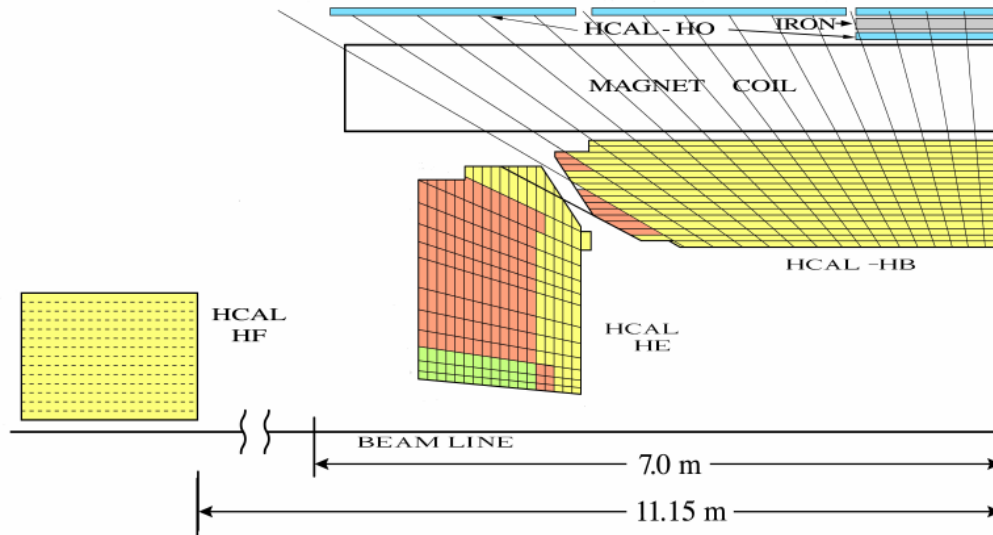


Figure 3.10: A quarter slice of the CMS Hadronic Calorimeter (HCAL) is shown. The different components include the barrel (HB), endcap (HE), outer HCAL (HO), and forward HCAL (HF), located outside the muon system. This schematic drawing illustrates their placement within CMS. The image is taken from [47].

Endcap (HE) subsystems are placed inside the solenoid magnet enclosing the ECAL. The HCAL consists of brass absorbers interleaved with plastic scintillator tiles. The inner and outer layers of the absorbers are made from stainless steel to provide more structural support to the detector. Hadrons entering the absorber induce hadronic showers, a cascade of hadronic particles, as a result of their strong interactions with the absorber nuclei. The hadronic showers are converted to a readout signal via the plastic scintillators, which are crucial for measuring their energy.

A schematic view of HCAL is shown in Figure 3.10, illustrating its subdetectors:

- HB covers the pseudorapidity range $0 < |\eta| < 1.3$,
- HE extends coverage to $1.3 < |\eta| < 3.0$,
- HO (Outer HCAL) is positioned just outside the solenoid magnet, acting as a “tail-catcher” to absorb hadronic showers escaping the main HCAL volume. HO covers the pseudorapidity region up to $|\eta| = 1.26$,
- HF (Forward HCAL) is located 11.5 m from the interaction point and covers the pseudorapidity range of $2.9 < |\eta| < 5.2$, making the HCAL the most hermetic subdetector of the CMS. Because HF is exposed to intense radiation due to its proximity to the beam pipe, it is built from radiation-hard materials. It uses steel absorbers and quartz fibres as the active materials for their construction.

3.2.5 The Muon System

As shown in Figure 3.11, a dedicated muon system is placed on the outermost layers of CMS to identify and trigger on muons and to measure their momentum. Muons, being significantly more massive than electrons, lose less energy through bremsstrahlung while traversing matter. As a result, they pass through calorimeters with minimal interaction, leaving signals primarily in the muon chambers and tracker. The muon system comprises three different types of gaseous ionisation detectors:

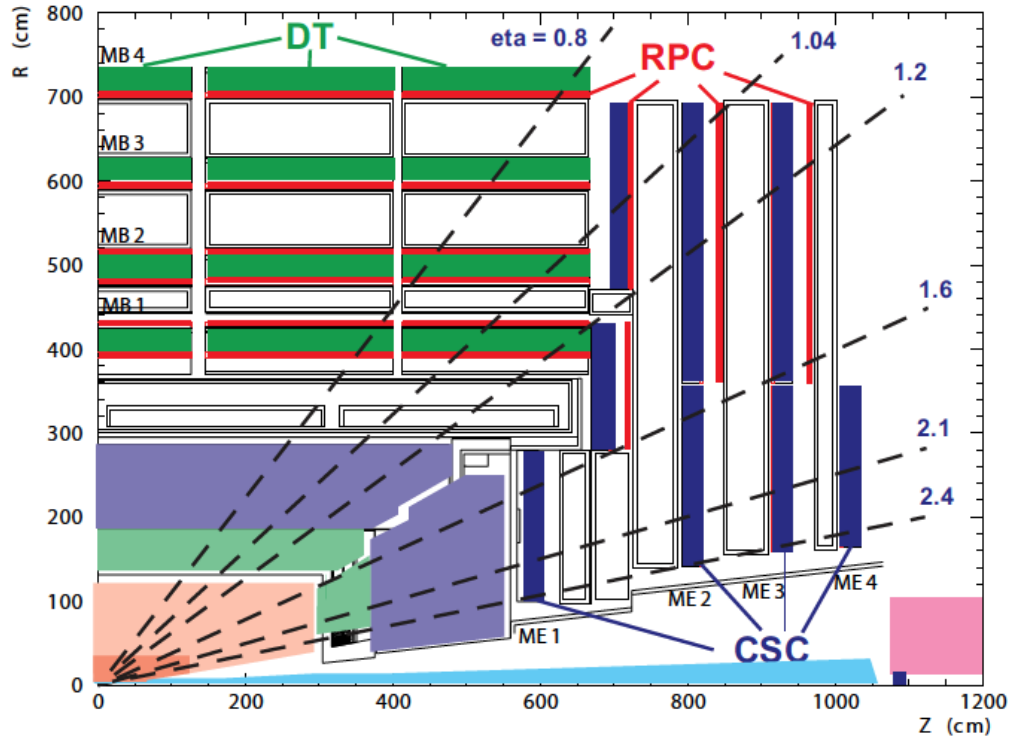


Figure 3.11: A schematic view of the CMS muon system is shown, with the various subsystems such as DTs, RPCs and CSCs in different colours, dashed lines denote constant $|\eta|$ and illustrate the pseudorapidity coverage. The schematic predates the Phase-1 muon system upgrades during LS1. Adapted from [43].

- Drift Tubes (DTs): Used in the barrel region ($|\eta| < 1.2$) where the magnetic field is homogeneous and muon flux is low.
- Cathode Strip Chambers (CSCs): Four stations of CSCs are used in the endcap region ($0.9 < |\eta| < 2.4$), where muon rates are high, and the magnetic field is non-uniform.
- Resistive Plate Chambers (RPCs): These are installed in both barrel and endcap, providing fast timing and redundancy up to $|\eta| < 1.9$, thereby improving trigger efficiency. Resistive plate chambers are gaseous parallel-plate detectors with a very fast response time.

Both DTs and CSCs measure the time and position of a muon hit with good spatial resolution, while the fast response time of RPCs enhances the timing resolution used in muon triggering.

3.3 Event Triggering and Data Acquisition System

As explained in Section 3.1.1, the LHC bunch-crossing rate is 40 MHz, meaning that proton bunches collide 40 million times per second. However, it is not feasible to read out and store every event due to hardware limitations and the sheer data volume. Furthermore, most collisions involve well-understood strong interaction processes, producing less interesting events for new physics searches. To efficiently select relevant events, CMS employs a two-tiered trigger system: the Level-1 Trigger (L1T), implemented in custom hardware, and the High-Level Trigger (HLT), based on software algorithms. The L1T is the first

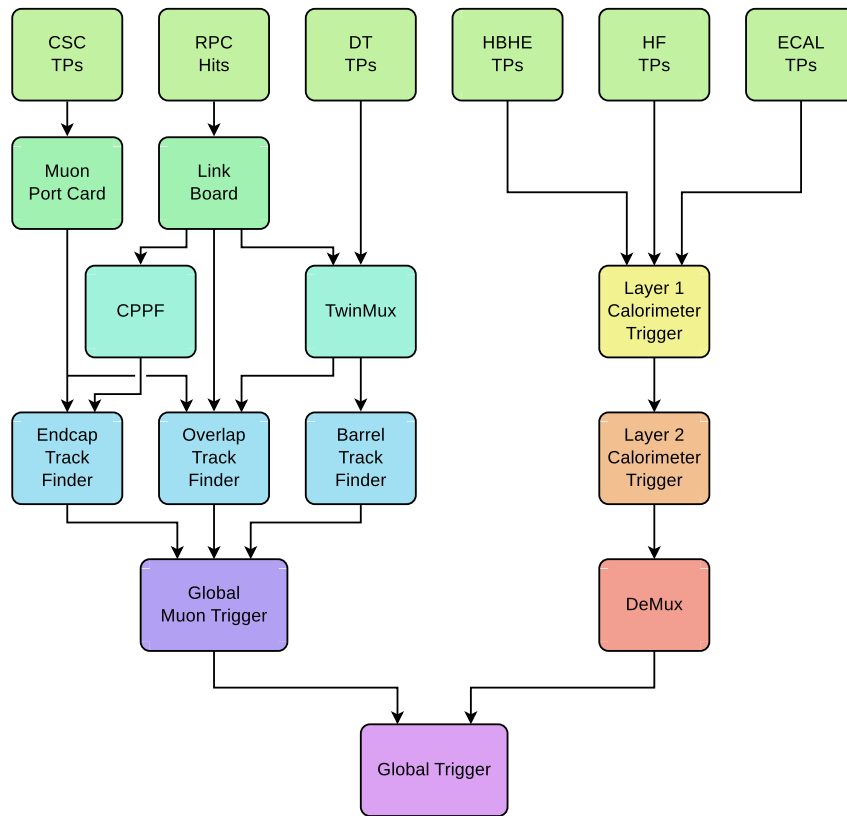


Figure 3.12: A schematic diagram of the subsystems involved in CMS Level-1 Trigger is shown. The figure is taken from [48].

stage of the CMS trigger system. It is designed to reduce the 40 MHz bunch-crossing rate to 100 kHz within a fixed latency of 4 μ s, as required by the Data Acquisition (DAQ) system. Following the Phase-1 upgrade, the L1T system could process events more efficiently at higher luminosity while maintaining this latency. Simplified physics objects are constructed from the corresponding detector signals and sent as Trigger Primitives (TPs) to the global trigger. The subdetectors involved in the Phase-1 L1 trigger data flow to the global trigger are shown in Figure 3.12. The global trigger decides to keep or throw away the event within the fixed latency. If the trigger selects the event, it is read out and forwarded to the second stage, the high-level trigger (HLT), by the DAQ system [48, 49].

The HLT further reduces the data rate to ≈ 2 kHz by using more complex reconstruction algorithms similar to the ones used to reconstruct physics objects described in Section 3.4, utilising information from all subsystems, i.e. completed with tracker information. HLT reduces the event rate by utilising track and vertex information for better background rejection in its trigger decisions. The selected events, along with all the information from detector subsystems, including L1T and HLT, are forwarded and stored by the DAQ system. The stored events undergo final reconstruction algorithms, which will be discussed in detail in the rest of the chapter.

3.4 Physics Object Reconstruction

In this section, we discuss how the data collected by the CMS detector, in the form of energy deposits and hits, is processed to reconstruct and identify physics objects using the particle-flow (PF) algorithm.

Signals (hits) from the tracker, the innermost layer of the CMS, are used to reconstruct

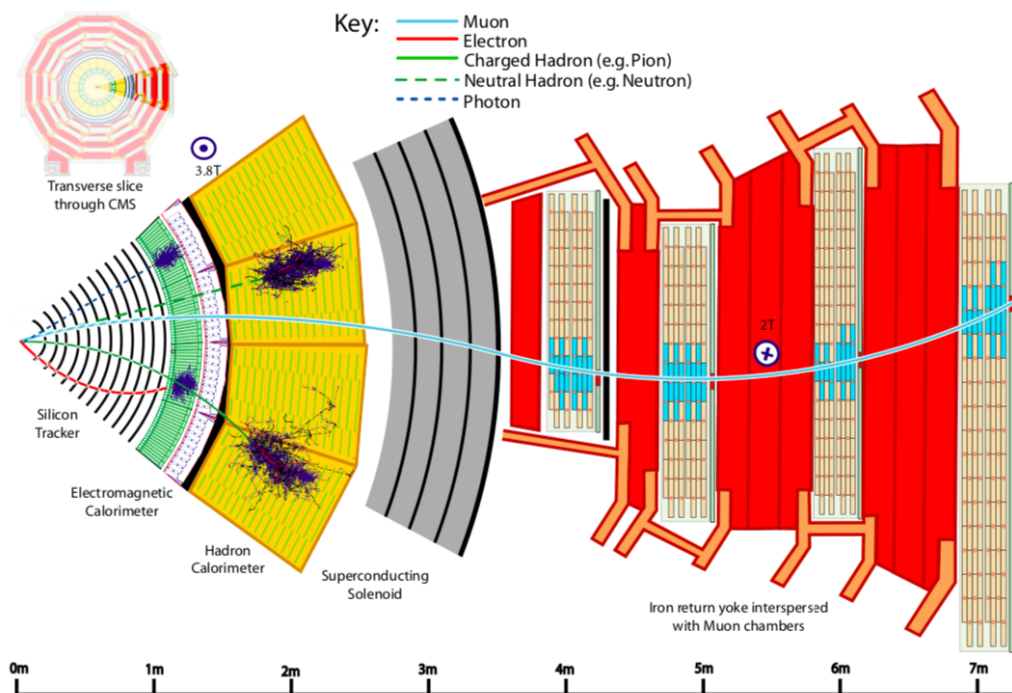


Figure 3.13: Transverse slice of the CMS detector is shown with different particle interactions in different subdetectors, where the muon and charged pion are positively charged and the electron is negatively charged. [50]

charged particle trajectories (tracks) and origins (vertices) [50]. The magnetic field around the tracker bends particle trajectories, allowing measurement of their electric charge and momentum. Electrons and photons get absorbed in the ECAL, where their energy deposits form clusters in neighbouring cells. These clusters help determine the particles' energy and direction. Similarly, charged and neutral hadrons are absorbed in HCAL, occasionally initiating a hadronic shower in ECAL, and their corresponding clusters are used to estimate the particles' energy and direction. Muons and neutrinos, however, pass through the calorimeters with almost no interaction. Muons are reconstructed from the hits they leave in the outermost subdetector, i.e. the muon system, while neutrinos leave the detector without any trace. The trajectory of SM particles as they traverse the CMS subdetectors and interact with different components is illustrated in Figure 3.13.

3.4.1 Tracks and Clusters

The first step in the particle-flow (PF) algorithm is local reconstruction, where individual subdetectors independently process their signals. Tracks are reconstructed in the silicon tracker (tracker), while energy clusters are identified in ECAL and HCAL. Muon tracks are reconstructed separately in the muon system. Following initial local reconstruction, the PF algorithm links the reconstructed objects within various subdetector systems to fully identify and reconstruct the particles later used for analyses. Reconstruction of charged-particle tracks enables proper vertex reconstruction, which helps distinguish the hard-scatter events from the pileup interactions.

Track Reconstruction

The charged particle tracks are reconstructed using the hit position in several pixel and strip detector layers. Particle properties, i.e. charge and momentum, are measured from

the curvature of tracks. The hits must be associated with a charged-particle trajectory for the track reconstruction. Track reconstruction uses a combinatorial track finder (CTF) [51] method, which takes tracker hits as input, based on Kalman filtering (KF) [52]. The CTF algorithm follows four steps:

- **Seed generation:** Initial track candidates are formed based on seeds, i.e. a few hits compatible with a charged-particle trajectory.
- **Track finding:** A pattern recognition algorithm (KF) looks for additional hits in all tracker layers following the expected path of the track candidate.
- **Track fitting:** After the complete extrapolation of the track, refitting is performed using KF, followed by smoothing and estimating the properties, such as origin, transverse momentum and direction of the final track.
- **Track selection:** If certain quality criteria pass, the track is selected. All hits in the final track are masked and not used to form any other track.

The reconstructed tracks are produced as a result of several iterations of the CTF algorithm. The tracking reconstruction efficiency increases as a function of the number of iterations incorporating different seeds and track selection criteria. The first iterations are focused on finding the tracks that are easiest to reconstruct with higher momenta and are close to the interaction point, thus requiring three to four hits in the pixel tracker. The hits affiliated with reconstructed tracks are then masked along with relaxed quality criteria to reduce the complexity for subsequent iterations. These later iterations focus on reconstructing more challenging tracks, such as displaced tracks. Final iterations incorporate the muon system by using hits from muon chambers (further discussed in Section 3.4.1), thereby increasing the muon reconstruction efficiency.

Vertex Reconstruction

Determining the interaction point in the pp collision, known as the primary vertex (PV), is a crucial part of object identification. The first step in PV reconstruction is selecting reconstructed tracks based on their closest approach to the beam spot. If the selected tracks originate from the same interaction point, they are clustered together via deterministic annealing (DA) algorithm [53]. The clusters are sorted by the sum of the squared transverse momenta of all the tracks associated with the vertex. The vertex with the highest sum is selected as the primary vertex (PV) of the hard scattering process, while the rest of the vertices are classified as pileup vertices in the event. All PV candidates containing at least two tracks are refitted using the adaptive vertex fitter (AVF) [54]. The AVF computes the best estimates of the vertex parameters, including position, fit quality indicators, and weights of the associated tracks. The AVF assigns a weight to each track $w_i \in [0, 1]$ that encodes its compatibility with the vertex ($w_i \simeq 1$ for compatible tracks, $w_i \simeq 0$ for outliers). The vertex fit quality is expressed via the number of degrees of freedom,

$$n_{dof} = 2 \sum_i w_i - 3. \quad (3.6)$$

Due to the strong correlation with the number of tracks compatible with the primary interaction region, this variable can be used to select genuine proton-proton interactions. [55] The impact parameter (IP) is defined as the distance between the closest point of the track and its corresponding particle candidate relative to the PV. The longitudinal and transverse components of the impact parameter are denoted by d_z and d_{xy} .

Muon Track Reconstruction

Track reconstruction for muons is achieved by using the reconstructed tracks from the tracker together with the hits inside the muon system. The muon system efficiently identifies muons across the full acceptance, while precise momentum measurement from the tracker greatly improves the muon reconstruction [56]. Muon track reconstruction is categorised into three types based on the sources of hits used in the track fit:

- **Standalone:** Muon tracks reconstructed exclusively from hits in the muon system (DT, CSC, and RPC) without requiring information from the tracker. These hits are clustered into track segments, which serve as seeds for the final track fit using the Kalman filter (KF) technique.
- **Tracker:** Tracks in the tracker with transverse momenta (p_T) greater than 0.5 GeV and total momentum (p) exceeding 2.5 GeV are extrapolated outward to the muon system. If one muon segment matches the tracker track within specific selection criteria, the track qualifies as a tracker-muon track. The selection criteria are based on the spatial distance between the tracker track and the muon segment.
- **Global:** A standalone-muon track is matched to a tracker track if the parameters of both propagated tracks are compatible. Global-muon tracks are formed after combining and fitting hits from both tracks, with much-improved momentum resolution for higher transverse momenta muons compared to tracker-muon tracks.

For muons with $p_T < 200$ GeV, momentum measurements rely exclusively on the tracker.

Calorimeter Clustering

A special clustering algorithm was developed for the PF event reconstruction performed separately on ECAL barrel and endcaps as well as two preshower layers and HCAL barrel and endcaps. In the forward HCAL, each calorimeter cell is used, so clustering is not required. Cluster formation starts by adding all the hits neighbouring the calorimeter cell with the highest energy deposit. The cell with the highest energy deposit serves as a seed to the PF clustering algorithm, which continues to aggregate the neighbouring cells as long as they have energies above a certain threshold. The final cluster formed is called the topological cluster.

Topological clusters usually contain more than one seeding cell. An expectation-maximisation algorithm based on a Gaussian-mixture model is then used to reconstruct final PF clusters within a topological cluster [50]. The expectation-maximisation algorithm is an iterative algorithm that first determines the expected energy deposit from a single particle by keeping the model parameters constant and then determines the position and energy of each particle inside the topological cluster by performing a maximum likelihood fit in the second step.

Electron Track Reconstruction

The electron track reconstruction was based originally on ECAL measurements to distinguish energetic and well-isolated electrons without relying on tracker information. This strategy using energetic ECAL clusters, i.e. $E_T > 4$ GeV, is referred to as the “ECAL-based” approach. In this method, an electron’s position and energy are used to deduce the position of hits in the innermost tracker. Since most electrons emit a significant amount of energy while traversing through tracker material as bremsstrahlung before getting to

the ECAL, the performance of the ECAL-based method relies on its ability to account for all the radiated energy. A supercluster (SC) is formed to recover the full electron energy by combining multiple ECAL clusters, including the electron’s own cluster and bremsstrahlung photon clusters. These are grouped in a narrow η window and a wider ϕ window to account for the possible bremsstrahlung photons due to the bending of electron tracks in the ϕ -direction as an effect of the magnetic field. However, the energy and position of superclusters associated with electrons within jets or electrons with low momentum are biased. The former is due to the overlapping contributions from other particle deposits, and the latter is due to the widespread radiated photons caused by the magnetic field’s significant bending.

However, ECAL-based reconstruction alone is not always sufficient. To improve reconstruction efficiency, a “tracker-based” seeding method was developed. In this method, electrons not forming SCs in ECAL are seeded from tracker hits. Unlike the ECAL-based approach, this method starts with tracks from the iterative tracking with $p_T > 2$ GeV. These tracks are used as electron seeds, disentangling electrons from hadrons based on their high probability of radiating in the tracker material. The corresponding track’s χ^2 is relatively well-behaved for the electrons with small energy loss. To select the track as an electron seed, the ratio of the energy of the ECAL cluster, matched to the track’s momentum, should be compatible with unity.

Radiation of energetic photons leads to poor track reconstruction with few hits. After a preselection based on the number of hits and track’s χ^2 , such tracks are fitted again, with a Gaussian-sum filter (GSF) [57]. The GSF fitting is better suited to electron tracks than the Kalman filtering (KF) method used for iterative tracking, as it accommodates the sudden and significant energy loss along the charged particle’s path.

Tracker-based seeding, in addition to ECAL-based seeding, improves the efficiency of track reconstruction by several percent.

PF candidates are reconstructed by connecting PF elements from various subdetectors via a linking algorithm. Only nearest neighbours are checked to limit computing time for a link to be formed. First, muon candidates are identified and reconstructed, and their associated tracks and clusters are excluded from further consideration. Muon reconstruction is followed by electron reconstruction, which involves linking ECAL clusters, tracker information, and bremsstrahlung photons. Isolated and energetic photons are identified in the same step, followed by removing corresponding tracks and cluster information. If a track is linked to the cluster, the remaining PF elements undergo further cross-identification of non-isolated photons, neutral hadrons, and charged hadrons originating from parton fragmentation, hadronisation, and decay within jets.

3.4.2 Muon Identification

Muon identification in the PF algorithm applies certain selection criteria to the properties of global and tracker muons. In order to distinguish muons from hadrons, for example, rejecting punch through hadrons, an isolation criterion is applied to the global muons. The isolation requirement is defined by considering the sum of p_T of tracker tracks and calorimetric energy deposits within a cone of $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} < 0.3$ around the muon to be less than 10% of the muon p_T . However, in our analysis, a custom isolation criterion, referred to as hybrid isolation (HI), is applied instead of the standard CMS isolation. The precise definition of hybrid isolation is given in Section 4.3.1.

Certain requirements specify the muon identification measurement criteria corresponding to each working point. Three identification (ID) or working points (WP) for muons [58] are defined as follows:

- **Loose ID:** The PF muon candidates pass the loose selection if they are also a tracker or a global muon. Loose ID is very effective for identifying prompt muons from the PV or decays of light and heavy quarks.
- **Medium ID:** Medium ID muons are loose ID muons with additional requirements on track fit parameters and compatibility between the tracker and standalone-muon in order to achieve high identification efficiency (99.5%) for prompt muons and muons from heavy quark decays.
- **Tight ID:** In addition to Loose and Medium ID requirements, Tight ID imposes stricter criteria such as compatibility with tracker and global muons along with good track χ^2 , a specific number of hits in different subdetectors (pixel, tracker and muon system), and PV compatibility. Tight ID is highly efficient in rejecting punch-through muons from hadronic showers and in-flight decay muons from kaon or pion decays. It also efficiently rejects cosmic muons using timing and spatial matching criteria.

Some additional IDs exist, such as soft and high momentum muon IDs [58], tailored for specific cases. Additionally, MVA-based IDs are also available.

MVA-based IDs optimise machine learning techniques to combine multiple variables related to muon identification, achieving better separation between signal and background and improving muon identification efficiency and purity. This approach is particularly useful in complex environments with high pileup conditions or distinguishing prompt muons from background sources.

3.4.3 Electron Identification

The sequential electron identification selection criteria in CMS are based on several variables related to electron reconstruction (utilising tracker and ECAL information) along with the primary vertex compatibility. Specifically, the variables include the ratio of hadronic to electromagnetic energy (H/E), ECAL shower extension in pseudorapidity direction ($\sigma_{i\eta i\eta}$), the difference between the inverse of SC energy and track momentum ($1/E - 1/p$), distance in η between SC position and the track ($\Delta\eta_{in}$) and similarly in ϕ ($\Delta\phi_{in}$), relative isolation (I_{rel}) and number of missing hits (N_{hit}^{miss}). Every working point has the conversion-veto requirement that rejects electron-positron candidates from photons.

In CMS, four working points (WPs), namely Veto, Loose, Medium and Tight, are used for the cut-based identification. The requirements for each working point are optimised separately in the barrel ($|\eta_{SC}| \leq 1.479$) and endcap ($|\eta_{SC}| > 1.479$) regions. The Veto ID working point corresponds to an average signal efficiency of 95% and rejects events where the number of reconstructed electrons exceeds the expectation from signal topology [59]. The Table 3.1 summarises veto ID selection criteria for barrel and endcap region, which is used in the work presented in this thesis.

While the standard CMS electron identification working points (WPs) are applied, the isolation criterion is removed from these WPs in this analysis and replaced by HI. Isolation is typically used to ensure that leptons do not originate from hadronic decays. Additionally, a tau veto using a very loose (VLoose) MVA-based ID is applied to reduce background contamination from hadronic tau decays. These selections are applied to ensure an accurate identification of the signal leptons while maintaining high efficiency. Detailed descriptions of the tau veto and hybrid isolation can be found in Section 4.3.1.

Variable	Barrel (Veto ID)	Endcaps (Veto ID)
$\sigma_{i\eta i\eta}$	< 0.0126	< 0.0457
$\Delta\eta_{in}$	< 0.00463	< 0.00814
$\Delta\phi_{in}$	< 0.148	< 0.022
H/E	$< 0.05 + 1.16E_{SC}^{-1}$ $+0.0324\rho E_{SC}^{-1}$	$< 0.05 + 2.54E_{SC}^{-1}$ $+0.183\rho E_{SC}^{-1}$
I_{rel}	$< 0.198 + 0.506p_T^{-1}$	$< 0.203 + 0.963p_T^{-1}$
$1/E - 1/p$	< 0.209	< 0.132
N_{hit}^{miss}	≤ 2	≤ 3
Pass conversion veto	Yes	Yes

Table 3.1: Selection criteria of cut-based electron identification for the veto working point in barrel and endcap regions. The variable ρ is the effective area correction factor for pileup subtraction in the isolation calculation.

3.4.4 Jet Reconstruction

In pp collisions at the LHC, hard-scattered quarks and gluons undergo hadronization, producing collimated sprays of particles that are reconstructed as jets. These sprays consist primarily of charged and neutral hadrons (e.g. pions, kaons, protons and neutrons); non-isolated photons and leptons from heavy-flavour hadron decays can be contained within jets but do not define them. Charged hadrons are identified using tracker information within the tracker acceptance ($|\eta| < 2.5$), distinguishing them from neutral hadrons and photons. Outside the tracker acceptance, jets from charged and neutral hadrons are indistinguishable since no tracking information is available.

ECAL and HCAL clusters without any linked tracks in the tracker are identified as non-isolated photons and neutral hadrons depending on the fraction of energy deposited in each subdetector. In the ECAL, neutral hadrons deposit only a fraction of their energy, so precedence is given to photons. Non-isolated photons are further distinguished from isolated photons using shower shape and selection criteria described in [59].

The PF jets are reconstructed by clustering all the identified hadrons using anti- k_T algorithm [60] with a radius parameter of $R = 0.4$ (AK4), where the contributions from charged hadrons which are associated with PU vertices are removed using a method called charged-hadron-subtraction [61]. The reconstructed jets are required to pass additional identification criteria based on jet energy fractions along with the number of PF candidates clustered in a jet to better discriminate noise jets from physical jets. There are three WPs defined for PF Jet ID, with “loose” and “tight” designed to remove jets originating from calorimetric noise [62]. For a more accurate description of jet properties, the jet energy scale (JES) and jet energy resolution (JER) are measured and applied as corrections.

3.4.5 Identification of b jets

Identifying heavy flavour jets originating from the hadronization of bottom quarks plays a crucial role in both precision measurements of standard model processes such as top quark and Higgs boson decays and beyond standard model searches. The relatively long lifetime (1.5 ps) of hadrons containing b quark leads to displaced tracks forming a secondary vertex (SV) at a displacement of $c\tau \approx 450 \mu\text{m}$ from primary vertex (PV). This unique characteristic is used to identify b jets. Additionally, the decay products of heavy flavour jets typically have higher transverse momentum than light-flavour jets or gluon jets. In 10-20% cases, they also contain a muon or an electron [63].

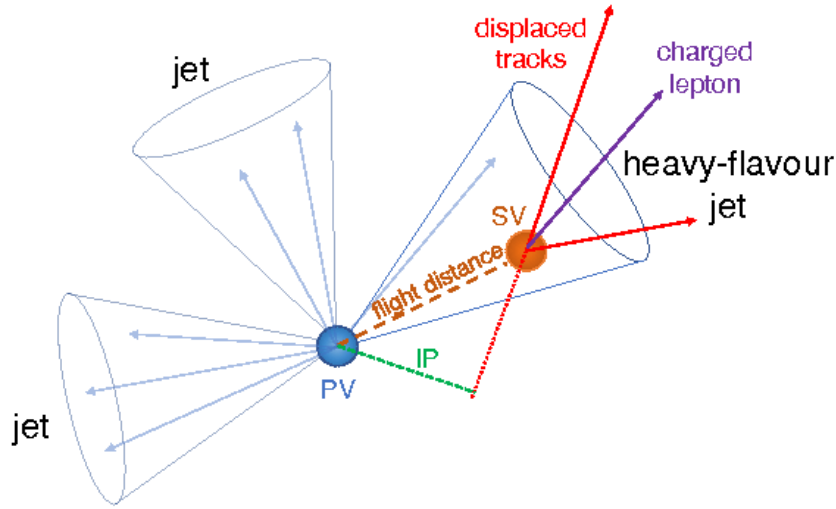


Figure 3.14: An illustration of a heavy flavour jet originating from the decay of a bottom quark hadron, showing a displaced secondary vertex (SV) and an associated charged lepton [50].

Figure 3.14 illustrates the formation of a heavy-flavour jet from a b decay, highlighting the displaced secondary vertex and a charged lepton, which are crucial for b -jet identification.

Using a deep neural network (DNN), the tagging of jets from heavy flavour hadrons is improved. The algorithm based on DNN is called DeepCSV, which uses more hidden layers and more nodes per layer compared to traditional methods, enabling it to distinguish all jet flavours more effectively. It utilises the information of SVs reconstructed using the Inclusive Vertex Finding (IVF) algorithm, along with jet constituents and jet properties, including jet p_T and η .

Selection criteria are applied to the DeepCSV b -tagging score to identify b jets. The required scores for the medium ID working point are 0.6001, 0.5847, 0.4506 and 0.4168 for the 2016-preVFP, 2016-postVFP, 2017 and 2018 data-taking periods, respectively.

3.4.6 Missing Transverse Momentum

Weakly interacting particles such as neutrinos and from various BSM models, the hypothetical exotic particles like neutralinos or other dark matter candidates do not interact with the detector, leaving it undetected. Momentum conservation is the key to inferring the momentum imbalance and consequently confirming the presence of undetected particles. The momentum imbalance is measured in the transverse plane due to missing initial momentum information and negligible transverse momentum of colliding partons. The momentum imbalance in the transverse plane is referred to as missing transverse momentum and denoted by \vec{p}_T^{miss} . It is the negative vectorial sum of all the PF candidates' momenta:

$$\vec{p}_T^{\text{miss}} = - \sum_{\text{particles}} \vec{p}_T \quad (3.7)$$

The magnitude of \vec{p}_T^{miss} is used most commonly and is denoted by p_T^{miss} .

The accuracy of p_T^{miss} measurements heavily depends on the CMS detector's ability to detect and reconstruct all particles. Any limitations in energy resolution, areas of the detector with reduced sensitivity (blind spots), or other defects can introduce biases

or inaccuracies in p_T^{miss} . Therefore, specialised corrections, such as “Type-I” corrections, are implemented to account for these effects by adjusting jet energy measurements, as discussed in Section 3.4.4. Additionally, jets are selected with specific criteria to avoid double counting of PF objects.

Chapter 4

Compressed SUSY search

This chapter presents the analysis strategy for searching compressed supersymmetric spectra with a single-lepton final state. The signal model used to design the analysis is described in Section 4.1.

The search for compressed spectra is novel and important as the final state particles are very soft, making them difficult to detect using classical searches designed for larger mass splittings between the next-to-lightest supersymmetric Particle (NLSP) and the lightest supersymmetric particle (LSP). The analysis uses specialised search strategies to overcome this, notably requiring an Initial-State Radiation (ISR) jet. The ISR jet boosts the system, enhancing the missing transverse momentum (p_T^{miss}) while maintaining the soft kinematics of the final-state particles, hence improving event selection sensitivity. The soft leptons in the final state add to the discriminating power of this signature against other SM processes like QCD multijet background and Drell-Yan production with jets where Z decays invisibly [64].

The search builds on previous analysis based on the 2016 data [65]. Various optimisations have been implemented to improve the sensitivity and extend the published results. Significant improvements include refining the signal regions, lowering jet thresholds from 30 GeV to 20 GeV to better discriminate against QCD events and lower their contribution while maintaining signal efficiency, and developing a new method to estimate misidentified leptons referred to as fake rate estimation method. A full description of the new fake rate estimation method is presented in this thesis [66].

These modifications were validated using 2016 pseudo-data, i.e. sum of all simulated SM background processes to mimic actual data, constructed from centrally produced samples of the Run 2 Legacy campaign, described below, and compared with the published 2016 data.

In CMS terminology, a campaign refers to a centrally coordinated data reprocessing or Monte Carlo (MC) production with a fixed configuration of software release, detector conditions and calibrations, centre-of-mass energy [67]. Examples include the Run 2 Legacy and Ultra-Legacy (UL) campaigns. For MC production, campaigns are often defined for individual stages (e.g. MC generator, detector response, reconstruction) and are chained so that the output of one stage becomes the input to the next.

While the primary focus of this thesis is the prompt stop scenario, complementary approaches are also being explored within the same analysis framework. These include strategies for long-lived stops, where displaced leptons are exploited, as well as techniques to incorporate very low transverse momentum electrons (~ 3 GeV). However, these extensions are beyond the scope of this thesis and are not discussed in detail.

The results presented here are obtained using the complete Run 2 dataset, corresponding to the 2016–2018 CMS data-taking period with an integrated luminosity of 138 fb^{-1}

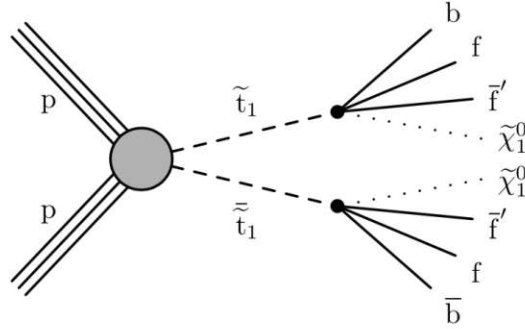


Figure 4.1: Diagram illustrating the T2tt simplified model, where a pair of top squarks (\tilde{t}_1) undergoes a four-body decay into a bottom quark, two fermions, and a neutralino ($\tilde{\chi}_1^0$).

at $\sqrt{s} = 13$ TeV. This analysis uses the Run 2 UL reprocessing of CMS data. Compared with the earlier legacy re-reconstruction, UL reprocessing includes improved reconstruction and calibrations for the full Run 2 dataset and provides matching MC campaigns (UL16preVFP, UL16postVFP, UL17, UL18). These centrally produced updates are reported by CMS to improve data–simulation agreement.

This chapter introduces the signal model, the datasets and simulated samples, and applied corrections. It also describes event filters, signal and control region definitions, and systematic uncertainties, leading to the final results.

4.1 Signal Models

Searching for physics beyond the standard model (BSM) with a complex model such as MSSM can be cumbersome. The addition of many new parameters complicates signal generation and makes interpretations highly model-specific. The Simplified Model Spectra (SMS) approach is used to simplify searches. SMS considers only a subset of relevant SUSY particles, while the rest are assumed to be heavy and beyond the reach of the LHC. This methodology reduces the number of free parameters in SMS to: the mass spectrum, production cross section, and branching fractions of light particles. The model-independent (generic) nature of these searches within SMS leaves room for reinterpretation of the search in a different BSM scenario with a similar final state. The T2tt model illustrated in Figure 4.1 and described in Section 2.2.5 represents a simplified supersymmetric scenario in which a pair of top squarks (\tilde{t}_1) is produced and undergoes a four-body decay into a bottom quark, two fermions, and a neutralino ($\tilde{\chi}_1^0$), which is assumed to be the Lightest Supersymmetric Particle (LSP). The T2tt model follows the nomenclature introduced in [68] and assumes a 100% branching fraction and zero-lifetime stop. The cross section depends on the mass of the stop.

Since the LSP is weakly interacting, it escapes detection, leading to a large amount of missing transverse momentum (p_T^{miss}). However, the presence of soft decay products makes detecting this process challenging, necessitating the special search techniques described in this chapter.

4.2 Data and simulation

Many standard model (SM) processes have the same final state as our signal, making them backgrounds to the SUSY search. A thorough understanding of SM background processes is essential to correctly compute the background and signal+background hypothesis. These background SM processes and the signal are simulated using Monte Carlo (MC) event generation methods. The MC events imitate everything from proton-proton collisions to detector responses. These simulations are also important in optimising event selection criteria, extracting correction factors, and validating background estimation methods. However, MC simulations are not sufficient, especially when estimating the contribution of falsely identified leptons (fakes) or non-prompt events. Instead, a data-driven approach is employed, relying on real collision data rather than simulations. This method removes our dependence on theoretical modelling. It mitigates potential inaccuracies from limited knowledge of complex interactions and detector effects. Various MC event generators used to simulate SM processes and signals in this analysis are discussed in Section 4.2.2.

4.2.1 Data Sets

The result of the search presented in this thesis uses the data collected by CMS detector over three years, 2016–2018, corresponding to an integrated luminosity of 138 fb^{-1} after data-quality certification. CMS organises each year into data-taking eras: 2016 B-H (seven eras), 2017 A-F (five eras), and 2018 A-D (four eras) for bookkeeping and certification, and it collected more data every year than the previous year.

As discussed in Section 3.2.3, the 2016 dataset is divided into preVFP and postVFP periods due to saturation effects in the APV25 chip caused by Highly Ionising Particles (HIPs). Highly Ionising Particles (HIPs) deposit significantly more energy than minimum-ionising particles (MIPs) as they traverse the silicon tracker, leading to increased charge deposition and saturation effects in the readout electronics. To correct for these effects in the 2016-preVFP dataset (19.5 fb^{-1}), the HIP Mitigation (HIPM) method was applied during data reconstruction. The 2016-postVFP dataset (16.8 fb^{-1}) was processed with default reconstruction settings after the problem was resolved [69].

The integrated luminosities of the data sets used in this analysis are:

- 2016-preVFP: 19.5 fb^{-1}
- 2016-postVFP: 16.8 fb^{-1}
- 2017: 41.5 fb^{-1}
- 2018: 59.8 fb^{-1}

The events used in this analysis are certified, i.e. recorded when all subdetectors are functional. For this analysis, the MET dataset is the primary dataset used, selected based on missing transverse momentum (p_T^{miss}) triggers. Additional datasets such as `SingleMuon`, `SingleElectron`, and `JetHT` are used for background estimation and various studies.

All data and simulated samples used in this analysis are reprocessed under the Ultra-Legacy (UL) campaign, ensuring uniform calibration and reconstruction across all years. This guarantees consistent physics object definitions and improved systematic uncertainty treatment, as discussed at the beginning of this chapter.

Trigger Efficiencies

The analysis is based on High-Level Trigger (HLT) paths following the syntax `HLT_PFMETX_PFMHTX_IDTight`, which select specific events with X ranging from 100 to 120 GeV for p_T^{miss} and H_T^{miss} , with H_T^{miss} defined as:

$$H_T^{\text{miss}} = \left| -\sum_i^{\text{jets}} \vec{p}_{T,i} \right|$$

The idea is to measure the offline efficiency of the p_T^{miss} trigger in the MET dataset and apply it to simulated events to take care of the difference between offline and trigger level calculations.

In order to achieve an unbiased efficiency measurement, a separate dataset recorded using uncorrelated triggers, such as the `SingleElectron`, is used. Employing triggers completely independent of the MET trigger whose efficiency is measured minimizes potential biases. The efficiency is determined as the fraction of events that pass both an offline selection and the trigger compared to events that only pass the offline selection.

The trigger efficiency is measured as a function of offline p_T^{miss} with an applied offline selection requiring a lepton with $p_T > 30$ GeV, leading jet $p_T > 100$ GeV, and $H_T > 200$ GeV. The efficiency shows a characteristic turn-on behaviour, increasing from 85% at $p_T^{\text{miss}} = 200$ to nearly 97% at $p_T^{\text{miss}} > 300$ GeV. To model turn-on fluctuations, an error function is used to fit the trigger efficiency, which is then applied to signal and background simulated samples.

The efficiency measurements are performed separately for the four UL data-taking periods: 2016-preVFP, 2016-postVFP, 2017, and 2018.

4.2.2 Signal and background simulation

As mentioned earlier, SM processes are simulated to better understand event selection, calculate corrections or scale factors to better agree with data, and measure and validate data-driven background estimations. Samples are generated for each data-taking period to match the detector conditions during data collection. Simulated events are generated in much larger quantities to decrease the statistical uncertainty of MC-based measurements than the recorded data. The samples are later weighted according to their corresponding integrated luminosity per year. The limiting factors in sample size are computational time and storage space for samples after generation.

In CMS, MC production proceeds through a chain of campaigns as shown in Figure 4.2, each with a fixed configuration; the output of one stage serves as the input to the next. The standard chain has the following stages:

GEN-SIM: event generation and full detector simulation with GEANT4. Output format of this campaign is a GEN-SIM sample.

DIGI-RECO: digitisation with pileup, trigger emulation, and offline reconstruction. Output format of this campaign is an AODSIM sample.

MiniAOD: skimming and reduction in size by saving information of analysis-level physics objects (leptons, jets, etc.). Output format is a MINIAODSIM. SIM is prefixed for simulated samples. The size of each event at this level is $\mathcal{O}(40 \text{ kB})$.

NanoAOD: further skimming to a flat, columnar n -tuple for fast analysis with standard physics object definitions. The output format is NANOAODSIM. Typical size is $\mathcal{O}(1\text{-}2 \text{ kB})$ per event.

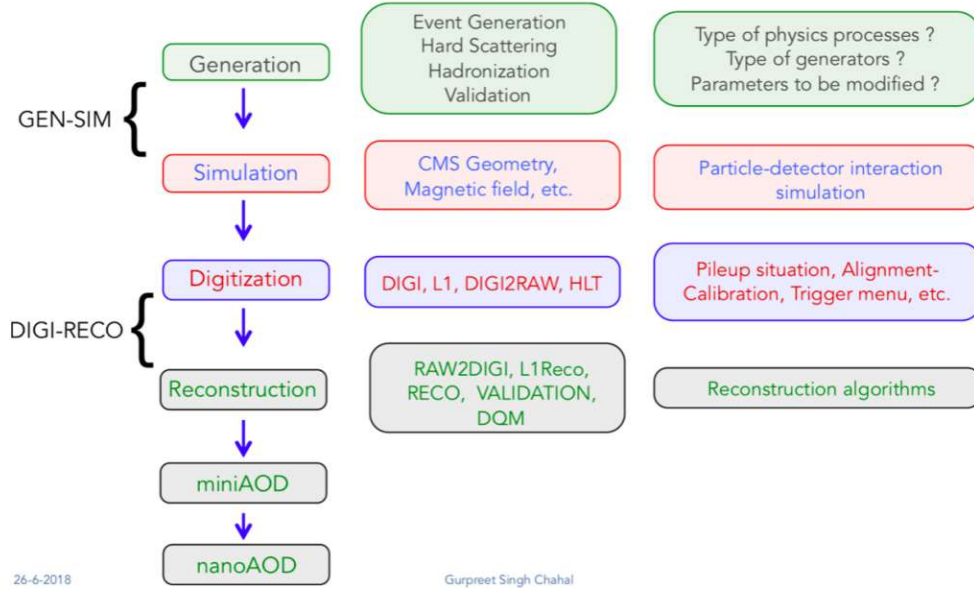


Figure 4.2: Schematic of the CMS MC production chain, reproduced from [67]. Each stage is annotated with its main inputs, processing and outputs.

A NanoAOD file contains one main Events TTree, plus auxiliary trees, e.g. Runs, LuminosityBlocks. Each branch in a tree is a column that stores quantities of interest so each row corresponds to a single collision event (real or simulated).

For this analysis we derive a private NanoAOD from the centrally produced UL MiniAOD for all analysis samples, details can be found in Chapter C. The private derivation appends production-vertex information to the generator-level particles (used only in simulations) needed for the long-lived part of the analysis, where a displaced stop leads to a lepton whose origin is expected to be offset from the primary vertex. All standard PF objects and definitions from the UL chain remain unchanged, the derivation simply restores vertex information that is not retained in the central NanoAOD stage. The private NanoAOD production is executed once, after the central MiniAOD becomes available. However, the resulting NanoAOD ntuples undergo a post-processing step where we apply the analysis object selections, skim events, drop unused branches to reduce file size, and compute a set of derived variables, i.e. observables built from NanoAOD branches such as transverse mass (M_T), hybrid isolation (HI) defined later in Section 4.3.1, $\Delta\phi^{min}$ also defined later in Section 4.3.1 and per-event weights (e.g. lepton scale factors and pileup). This step is rerun whenever definitions change (e.g. new object selection, variables, or uncertainties). In practice, we validate changes on a single year (typically one iteration per month during development) and a full Run 2 post-processing takes place once the configuration is frozen.

For MC samples production, in the UL campaign, the generators used are MadGraphMG5_aMC@NLO and POWHEG to simulate the hard scatter. The W +Jets, $t\bar{t}$, Z/γ^* +Jets, and QCD multijet processes are produced at leading-order (LO). In this thesis, the Z +Jets process refers to the Z boson decaying into neutrinos, $Z \rightarrow \nu\nu$. This process is sometimes also referred to as ZInv (Z invisible). The terms “($Z \rightarrow \nu\nu$)+Jets” and “ZInv” are used interchangeably throughout figures and text. To increase statistical power, these samples, except $t\bar{t}$, are binned in H_T where, H_T is defined as the scalar sum of the transverse momenta of all jets in the event and is calculated at the generator level. The $t\bar{t}$ process is produced separately in single- and di-leptonic channels. The single top and associated tW processes are simulated at next-to-leading order

(NLO) using POWHEG. Similarly, diboson processes are produced using either POWHEG or MadGraphMG5_aMC@NLO for various production channels at NLO. The top quark pair production, with a W, Z, or a gluon, is simulated via MadGraphMG5_aMC@NLO and categorised as ttX. Parton showering and hadronization are simulated using PYTHIA8 with the CP5 tune, derived using NNPDF3.1 NNLO [70].

After event generation and hadronization, along with the addition of parton showers, samples undergo detector simulation. SM background processes are primarily simulated using GEANT4 based Full Simulation FULLSIM. On the other hand, the signal samples are produced using CMS Fast Simulation FASTSIM [71], a parameterised detector simulation using tuned models of the tracker and calorimeter while keeping the standard CMS reconstruction. FULLSIM provides a highly detailed and computationally intensive simulation of the CMS detector response whereas, FASTSIM is a faster alternative that simplifies certain detector effects while maintaining compatibility with full simulation. Scale factors are applied to account for differences between these two approaches, ensuring consistency in event reconstruction.

For the prompt signal modelling, the lifetime of the stop pair is considered zero. The signal samples are generated at LO using MadGraphMG5_aMC@NLO with FASTSIM. This computationally efficient method makes it possible to scan the sizeable stop-LSP mass plane. The sample grid is produced in 25 GeV steps for stop masses in the range $200 \leq m_{\tilde{t}} \leq 850$ GeV. The mass splitting between the stop and the LSP (Δm) is scanned in 10 GeV steps within the range $10 \leq \Delta m \leq 80$ GeV. The production cross sections for the stop pair production are calculated at NLO [72] as a function of the stop mass alone. Certain generator-level filters are applied to select events relevant to this analysis to improve computational efficiency. The filters applied are $p_T^{\text{miss}} > 80$ GeV and $H_T > 160$ GeV, calculated at generator level. The efficiency of this filter depends on stop-LSP mass, ensuring sensitivity to the signal phase space while optimising resource usage.

4.2.3 Object Efficiencies and Corrections

Various corrections are applied to the MC simulations to be as close to the observation as possible. Some corrections are experimental, e.g. failure of some detector parts, the efficiency of measuring various quantities, or pileup information, and some can result from simulation modelling. Several of these corrections have systematic uncertainties, which are then propagated to predicted yields.

Pileup Correction

The events in simulated samples are generated using an approximate pileup (PU) profile since knowing the average pileup in data is only possible once the entire dataset is collected. As the approximation used for simulation does not match the observed value, it needs to be corrected. The correction is applied as a function of the number of interactions in the observed pileup distribution. The number of observed interactions per bunch crossing is calculated using instantaneous luminosity times the total inelastic pp cross-section of 69.2 mb [73].

Lepton efficiencies

The tag-and-probe method measures the lepton (electron and muon) identification and isolation efficiencies in data and MC at Z boson resonance. In this method, one lepton is selected based on stringent (tight) criteria and referred to as the “tag”, while the other lepton, the “probe”, is selected based on the criteria whose efficiency we aim to measure.

To ensure that the selected leptons originate from the Z boson decays, an invariant mass range of 60-120 GeV is applied.

The efficiency is determined by simultaneously fitting the invariant mass distribution of passing and failing probes, using the Crystal Ball function, i.e. Gaussian core with a power-law tail [74, 75], for the signal component and a falling exponential for the background estimation. The measured efficiency is the ratio of the signal component of the fits. This process is repeated for data and simulation across various pseudorapidity (η) and transverse momentum (p_T) ranges and separately for electrons and muons.

Two types of efficiencies are measured for leptons: identification and isolation combined with impact parameter selection. The identification (ID) efficiency of muons (electrons) is measured by selecting tracker muons (GSF tracks) as probe candidates. These probes satisfy Loose ID (Veto ID) for isolation combined with the impact parameter efficiency of muons (electrons). To derive the lepton scale factors (SF), we take the ratio of efficiency measured in data and simulation. The SFs are applied to the simulation to account for the data-simulation discrepancies as a function of p_T and η for both signal and background samples.

A dedicated measurement of the SFs is performed for all the years as the analysis has non-standard lepton identification, isolation and impact parameter requirements for the leptons. The standard ID SFs do not cover the full momentum range for this analysis, especially at low momenta. The ID scale factors for electrons for Run 2 are shown in Figure 4.3. Similarly, Muon ID scale factors are shown in Figure 4.4. To ensure compatibility with CMS recommendations, the privately derived scale factors (SFs) for lepton ID are compared to the centrally provided SFs in the overlapping momentum regions. The goal is to verify that the differences remain within the uncertainties of the central SFs. The central SFs are used directly for higher momentum leptons, as the final ID SF maps are constructed by appending the privately derived low- p_T bins. For impact parameter and hybrid isolation selections, SFs are derived separately per year, which are not centrally provided, as shown in Appendix B.

The measurements presented here use data from the end-of-year (EOY) campaign, which refers to the reconstruction of data collected at the end of each data-taking year with the best available calibrations at the time. For the final analysis, the efficiencies will be derived from the Ultra Legacy (UL) dataset, which incorporates improved calibration and reconstruction techniques, ensuring better consistency across the full Run 2.

Initial-State Radiation (ISR) model correction

Modelling Initial State Radiation (ISR) properties in simulated samples remains challenging due to discrepancies between observed and simulated ISR jet multiplicity and the transverse momentum spectrum of ISR jets (p_T^{ISR}).

In this analysis, ISR corrections are applied to W+Jets events to account for discrepancies in ISR modelling. To address these modelling effects in W+Jets events, the transverse momentum (p_T) of the W boson is reconstructed at the generator level and used as a proxy for (p_T^{ISR}). Correction factors are derived by comparing the p_T distribution of the Z boson in data and Monte Carlo (MC) simulations using dedicated samples. Additional sample-dependent factors are incorporated to ensure the total process cross section remains unchanged after applying the correction.

For the UL campaign, these correction factors were not centrally provided and were derived privately by the analysis team. Due to the challenges in ISR modelling, considerable systematic uncertainty about the size of the correction is assigned.

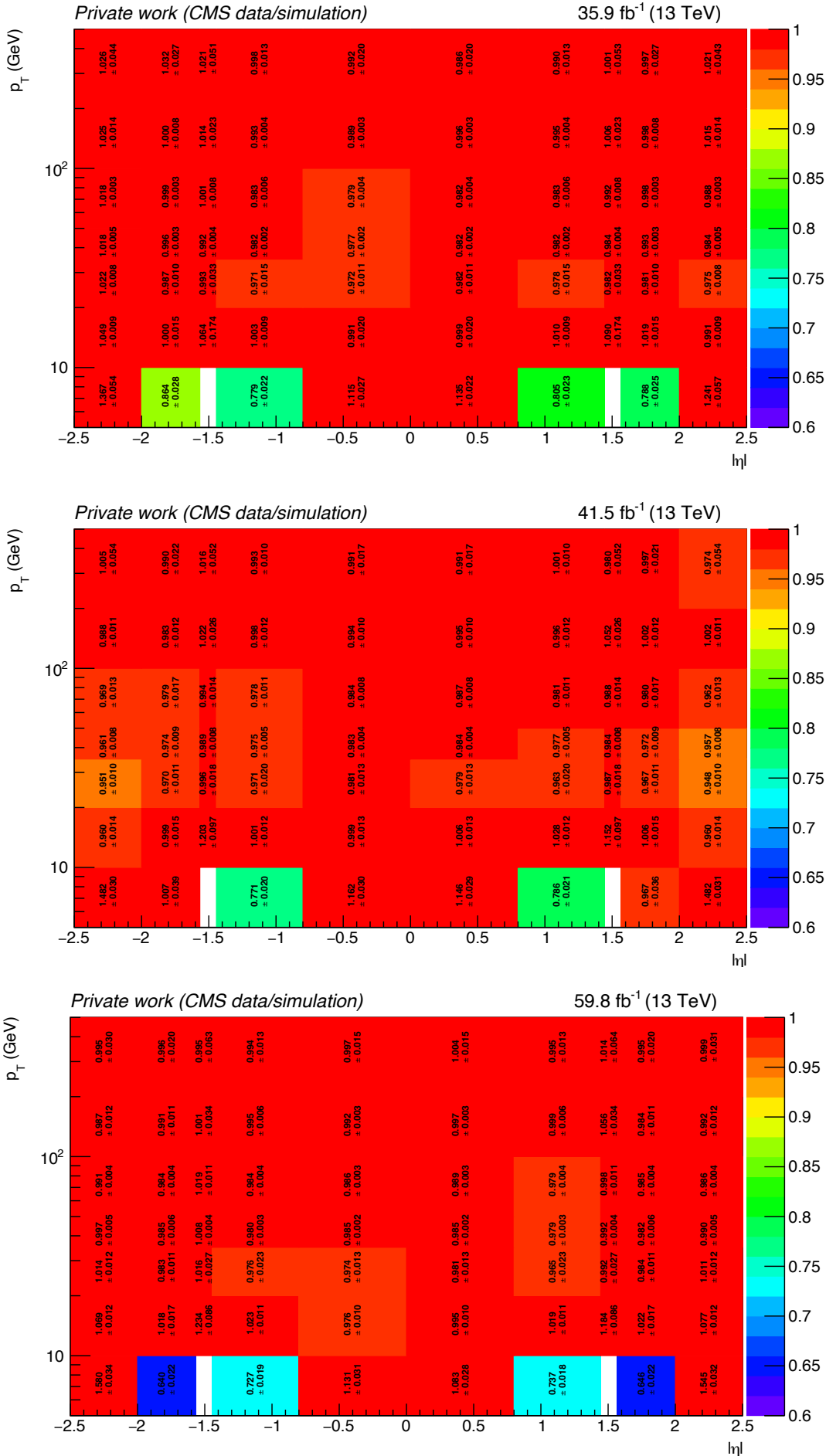


Figure 4.3: Electron identification scale factors derived privately for the 2016, 2017 and 2018 data-taking periods are shown in the figure. The low- p_T bin (5-10 GeV) is derived privately and appended to the centrally provided CMS SFs. The barrel-endcap transition region ($1.44 < |\eta| < 1.56$) is excluded from the analysis.

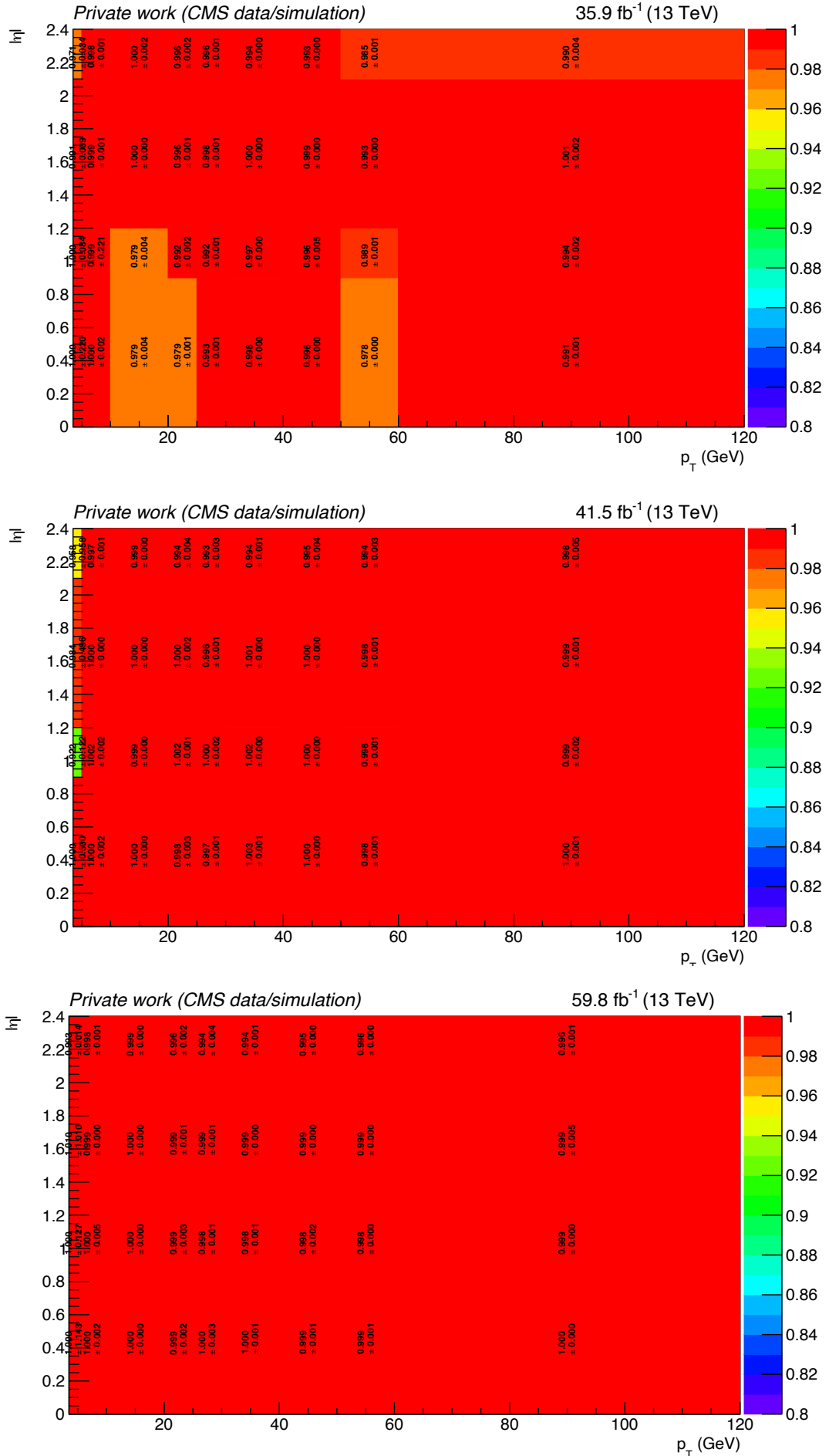


Figure 4.4: Muon identification scale factors derived for the 2016, 2017 and 2018 data-taking periods are shown. The low- p_T region (3.5–10 GeV for 2016 and up to 20 GeV for later years) is privately measured and appended to the centrally provided CMS SFs.

B-tag scale factors

The scale factors for b-tagging efficiencies of different algorithms (so-called jet-taggers) such as DEEPCSV and DEEPJET are computed for various working points (loose, medium, and tight), as discussed in Section 3.4.5. These scale factors correct for differences in b-tagging efficiencies between data and Monte Carlo (MC) simulation.

The b-tagging efficiency differs between data and MC simulations. To account for these discrepancies, scale factors (SFs) are derived by comparing data and MC as a function of jet transverse momenta (p_T), pseudorapidity (η), and jet flavour. The b-tagging efficiencies are first computed in simulation according to the analysis requirements. The SFs are then applied as event weights in simulations to correct for discrepancies between data and simulation. These weights represent the probability that a jet in the simulation is correctly tagged, ensuring consistency between data and MC.

L1 Prefiring correction

During the data-taking period from late 2016 through 2017, the ECAL endcap time alignment, an input to L1 trigger decisions, had a gradual shift for objects in the forward regions, i.e. $|\eta| > 2.5$. The L1 trigger's design does not allow consecutive bunch crossings to trigger events, and this constraint resulted in inefficiencies in L1 trigger decisions originating from ECAL trigger objects of the previous bunch crossing. This effect is referred to as L1 prefiring [48].

The CMS collaboration centrally provides correctional weights to mitigate this issue, which are applied to the MC, along with the uncertainty associated with the correction. Although the objects in the forward region do not enter the search directly, their indirect effects can result in systematic uncertainty.

Hadronic Calorimeter Endcap Minus (HEM) 15-16 Correction

During the 2018 data-taking period, two hadronic calorimeter endcap (HE) sectors, -15 and -16 , became unresponsive after a power cut in June 2018 which led to a brief voltage spike during power restoration that damaged their front-end electronics. This issue is called the Hadronic Calorimeter Endcap Minus (HEM) issue. For this analysis, the missing sectors primarily affected electrons, jets, and consequently missing transverse momenta (p_T^{miss}). An event veto is applied to take care of the issue. Events are vetoed if there are any selected electrons in the pseudorapidity range $-3.0 < \eta < -1.4$ and azimuthal range (ϕ) $-1.57 < \phi < -0.87$ or any AK4 jets in $-3.2 < \eta < -1.2$ and $-1.77 < \phi < -0.67$ region. Additionally, we remove the fraction of events corresponding to the data loss from the simulation.

4.3 Event Selection

The search discussed in this thesis focuses on stop quark pair production with a degenerate neutralino as the lightest supersymmetric particle (LSP) as described in Section 2.2.5. The stop decays into a bottom quark, a neutralino, and a fermion pair from off-shell W boson decay. As the neutralino escapes the detector unnoticed, it results in large missing transverse momentum (p_T^{miss}), which is a signature of R-parity conserving SUSY searches. The multijet background in the hadronic channel at the LHC is significantly larger than in the leptonic channel. Even though the leptonic channel is less probable, it offers better chances for the signal to be separated from the background. The hadronic channel is more

likely dominated by multijet events. Therefore, in this analysis, we search for a cleaner signature with a single lepton in the final state.

Undoubtedly, one of the main challenges in the searches of the compressed region, where the mass difference (Δm) between stop and neutralino is small, is that the final-state particles are left with minimal energy, resulting in low transverse momentum (p_T). In other words, the leptons and jets produced are soft and often fall below the trigger and reconstruction thresholds, making detecting and reconstructing them difficult. Furthermore, the missing transverse momentum in the compressed signals is lower than typical SUSY signatures. To recover sensitivity, we select events with a jet from initial-state radiation (ISR). The entire system recoils against this ISR jet, boosting it in the laboratory frame. The neutralinos take most of this momentum, yielding detectable p_T^{miss} , while the decay leptons and jets remain relatively soft. However, requiring an ISR has the drawback of low acceptance as ISR production occurs predominantly in higher-order processes. The events in this analysis are selected with at least one jet with $p_T > 100$ GeV and use the leading jet as a proxy for the ISR jet, since jets from the signal decays are typically soft in compressed spectra. The resulting topology is characterised by a leading high- p_T jet, hence large H_T (see Section 4.3.1), together with moderate p_T^{miss} and soft decay products.

Searches for new physics at the LHC are performed by carefully selecting the events that match the expected characteristics of the signal model under study. The strategy is designed to preserve sensitivity of our signal model across the wide parameter space we scan, which in the case of simplified models (SMS) corresponds to the particle masses. To achieve this, signal regions (SRs) and control regions (CRs) are constructed by applying selection based on the signal's kinematic properties to enrich signal events in SRs. The CRs are dominated by the relevant backgrounds with minimal contribution from the signal process. The exact definitions of signal and control regions will be discussed in Section 4.4.1 and Section 4.4.2, respectively. The observed data in the control region, dominated by background processes, is then used to estimate the contribution of SM processes in the signal regions.

After incorporating relevant uncertainties for signal and background processes, we use a likelihood ratio method to assess the compatibility of observed data with two hypotheses: the background-only hypothesis and the signal+background hypothesis in the SRs.

4.3.1 Baseline Selection

This analysis targets signal final states, including an energetic ISR, soft lepton, moderate missing transverse momentum, and the possibility of a soft b-tagged jet. The single lepton channel reduces some dominant SM backgrounds, such as QCD multijet events. However, we must still consider and minimise the contribution of some other SM processes with similar final states.

We apply baseline criteria to suppress these backgrounds while maximizing signal purity as best as possible. This preselection helps define a phase space before signal region optimisations. It ensures that we efficiently retain events consistent with our signal process and reject those coming from the background.

Primary Vertex

Among the reconstructed vertices, as described in Section 3.4.1, the primary vertex (PV) is chosen as the candidate that maximises the scalar sum of the squared transverse momenta of its associated tracks, i.e. $\sum_i (p_T^i)^2$. For this analysis, the chosen PV must have at least four degrees of freedom (n_{dof}) as defined in Equation (3.6), be within 24 cm in the

longitudinal plane (z -axis), and have a transverse distance of less than 2 cm from the beamline. This vertex is considered the source of the hard scattering process. The tracks associated with PV are called *prompt* tracks, and the particles associated with the tracks are hence considered to be coming from the hard scattering process.

Electron and Muon Selection

The electrons for this search are required to have transverse momentum $p_T > 5$ GeV, pass cut-based identification (ID) with **Veto** working point, as defined in Section 3.4.3, within $|\eta| < 2.5$ with the exclusion of ECAL gap, i.e. the barrel-endcap transition region at $1.4442 < |\eta| < 1.566$. For muons, the reconstruction and identification criteria are given in Section 3.4.2. Muons must have $p_T > 3.5$ GeV, $|\eta| < 2.4$ and pass the **Loose** ID working point, as defined in Section 3.4.2. Additional selection criteria are applied to reduce the contamination from non-prompt leptons, i.e. hadronic decays, based on lepton isolation and impact parameters. The impact parameter, the point of closest approach to the primary vertex of the track associated with the lepton, is essential to discriminate prompt leptons produced in the hard scattering from hadronic decays or pileup (PU). To make sure that the leptons are prompt, we require the longitudinal (d_z) and transverse (d_{xy}) component of impact parameter from PV to be less than 0.1 cm and 0.02 cm respectively.

Isolation is another important criterion for selecting prompt leptons and suppressing backgrounds from hadronic decays. The absolute isolation (I_{abs}) is calculated as the sum of transverse momenta of the particles within a cone of radius $\Delta R \equiv \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} = 0.3$ around the lepton. Since leptons with higher energy will be surrounded by more activity, we can use relative isolation (I_{rel}) defined as:

$$I_{\text{rel}} = \frac{I_{\text{abs}}}{p_T}$$

to improve our ability to select energetic leptons. A hybrid approach is adopted in this analysis to efficiently use the isolation criteria to find the prompt electrons and muons. The method switches from absolute to relative isolation as we go above a certain threshold for p_T of the lepton. The specific thresholds for hybrid isolation (HI) are defined as follows:

$$\begin{aligned} I_{\text{rel}} &< 0.2 \text{ for lepton } p_T > 25\text{GeV} \\ I_{\text{abs}} &< 5 \text{ GeV for lepton } p_T \leq 25\text{GeV} \end{aligned} \quad (4.1)$$

The hybrid isolation variable can then be expressed as:

$$\text{HI} = I_{\text{rel}} \cdot \min(p_T, 25 \text{ GeV})$$

Using the HI variable, the isolation conditions described in Equation (4.1) can be rephrased with the requirement $\text{HI} < 5$ GeV. This approach ensures that isolated leptons are selected consistently across the full transverse momenta range. Although the cut-based electron ID includes a built-in isolation requirement, we use the ID without its isolation component and instead apply the same hybrid isolation requirement to both electrons and muons.

The selection criteria for electrons and muons are summarised in Table 4.1. We look for an event with exactly one such electron or muon. To ensure this, any event with an additional electron or muon with $p_T > 20$ GeV is vetoed. Furthermore, any event with a hadronically decaying tau with $p_T > 20$ GeV is also vetoed, ensuring a more precise selection process. Tau selection criteria are also listed in Table 4.1. After applying the full signal lepton selection described above, the term lepton is limited to electrons and muons for the remainder of this thesis.

jet	b-tagged jet	physics object		
		τ	e	μ
	$p_T > 20$ GeV $ \eta < 2.4$	$p_T > 20$ GeV $ \eta < 2.3$	$p_T > 5$ GeV $ \eta < 2.5$	$p_T > 3.5$ GeV $ \eta < 2.4$
–	–	–	$ d_{xy} < 0.02$ cm	
–	–	–	$ d_z < 0.1$ cm	
–	–	–	HI < 5 GeV	($\Delta R = 0.3$)
–	medium DEEPCSV	vloose MVA ID	veto ID	loose ID

Table 4.1: Object selection criteria used to define reconstructed physics objects in this analysis.

Jets with $p_T > 20$ GeV and $|\eta| < 2.4$ are selected, a b-tagged jet is a jet passing the DeepCSV **medium** working point as defined in Section 3.4.5. Hadronic tau candidates are defined with $p_T > 20$ GeV, $|\eta| < 2.3$ and a very loose (vLoose) MVA ID. Electrons require $p_T > 5$ GeV, $|\eta| < 2.5$, and the cut-based **Veto** ID. Muons require $p_T > 3.5$ GeV with $|\eta| < 2.4$, and the **Loose** ID. For both electrons and muons we additionally require $|d_{xy}| < 0.02$ cm, $|d_z| < 0.1$ cm and apply hybrid isolation HI < 5, calculated within a cone of radius $\Delta R = 0.3$, as defined in Section 4.3.1, to ensure selection of prompt leptons.

Jets

In Section 3.4.4, we discuss how the hadronization of quarks and gluons produces showers of charged and neutral particles inside the detector, identified by the particle flow (PF) algorithm. Then, using the anti- k_T algorithm, all PF particles are clustered within a radius parameter of 0.4 to reconstruct a jet. For this analysis, jets are selected with $p_T > 20$ GeV and within tracker coverage, i.e. $|\eta| < 2.4$. The earlier 2016 analysis used a higher jet p_T threshold of 30 GeV. An improvement in jet reconstruction has made it possible to include jets with p_T as low as 20 GeV, significantly increasing the sensitivity of this analysis as shown in Figure 4.5. This increase in sensitivity is due to our signal kinematics based on soft objects that benefit from lowering the thresholds.

Since leptons can also be identified as jets, we apply a cleaning procedure to avoid double counting. Jets are selected if they are well separated from the selected lepton with an angular separation of $\Delta R > 0.4$, and if the lepton carries less than half of the jet's p_T . These criteria ensure the selected jet is not associated with a lepton.

As discussed earlier, we select events with at least one energetic jet, i.e. jet with $p_T > 100$ GeV, as a proxy for an ISR jet. To further characterise hadronic activity in the event, we compute the scalar sum of the transverse momenta of all selected jets, satisfying the selection criteria above. This hadronic energy sum is denoted by H_T . Events with $H_T > 300$ GeV are selected to distinguish the signal better from the W+Jets background, which would have lower hadronic activity.

Missing Transverse Momenta

To further efficiently suppress background and enhance signal sensitivity, we select events with high missing transverse momentum, i.e. $p_T^{\text{miss}} > 200$ GeV. This requirement targets events where the neutralinos from the signal model would escape the detection due to their weakly interacting nature, resulting in large p_T^{miss} . The quantity p_T^{miss} is described in more detail in Section 3.4.6.

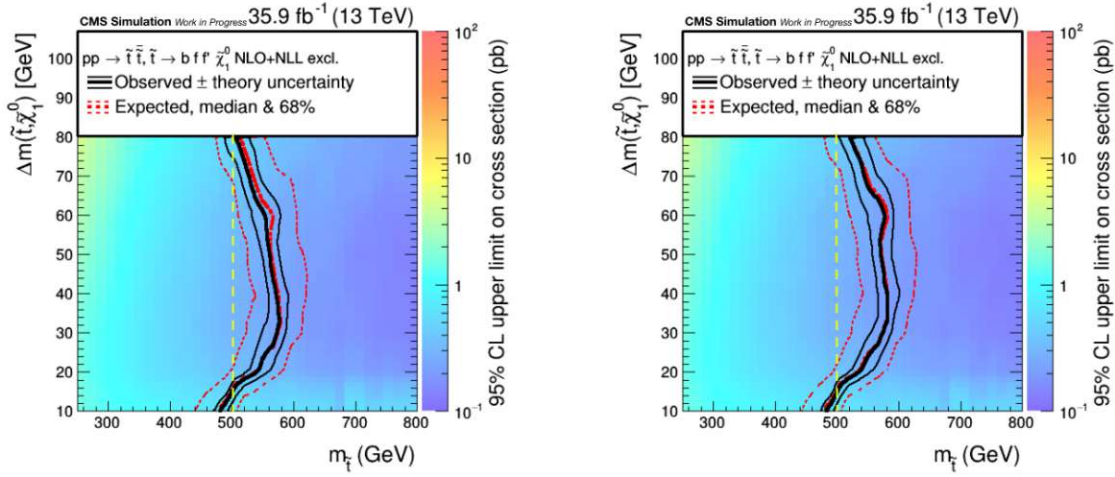


Figure 4.5: The exclusion limit on the left shows the sensitivity of the analysis using jets with $p_T > 30$ GeV. The plot on the right shows the limits after lowering the jet p_T threshold to 20 GeV. Lowering thresholds shows increased sensitivity, particularly in the higher Δm region.

Angular requirement between jets and \vec{p}_T^{miss}

QCD multijet events typically have large hadronic activity but little genuine p_T^{miss} . To mimic a SUSY signal event (with sizeable p_T^{miss}), they require an additional source, either from the mismeasurement of one of the jets or neutrinos in the heavy-flavour decay of b or c jet. In either case, the missing transverse momentum would align with the jet, which would make the angular separation between jet and missing transverse vector (\vec{p}_T^{miss}) a powerful handle on identifying QCD multijet events [76]. The Standard Model and SUSY processes are expected to give a more or less flat distribution, unlike QCD as shown in Figure 4.6. The variable is defined as:

$$\Delta\phi^{\text{min}} \equiv \min(\Delta\phi(\text{jet}_1, \vec{p}_T^{\text{miss}}), \Delta\phi(\text{jet}_2, \vec{p}_T^{\text{miss}})) \quad (4.2)$$

In order to mitigate QCD multijet background, we apply a selection on the variable $\Delta\phi^{\text{min}}$ defined in Equation (4.2). Events are vetoed if two hard jets (with $p_T > 60$ GeV) have $\Delta\phi^{\text{min}} < 0.5$. This selection for high momentum jets improves QCD rejection while maintaining the signal efficiency. To reduce the contribution from $t\bar{t}$ background, which can include many energetic jets, a veto referred to as the third jet veto, is applied on events with more than two jets if the third jet is a hard jet.

4.3.2 Background Processes

The Standard Model processes surviving the selection criteria described above, as they share the same final products with similar kinematics as our signal, become the background to the search we describe in this thesis. Requiring a single lepton in the final state reduces the contribution of the dominant QCD multijet events. However, other SM processes with a similar final state as our signal model can still contribute to the overall background composition.

The primary background is W+Jets production, where a W boson decays into a lepton and a neutrino, leading to genuine but moderate levels of missing transverse momentum, typically lower than expected from SUSY signals. There is a 25% probability of a W boson decaying into an electron or muon (including leptons originating from $\tau \rightarrow e\mu$), always accompanied by a neutrino. When such a W boson is produced along with jets,

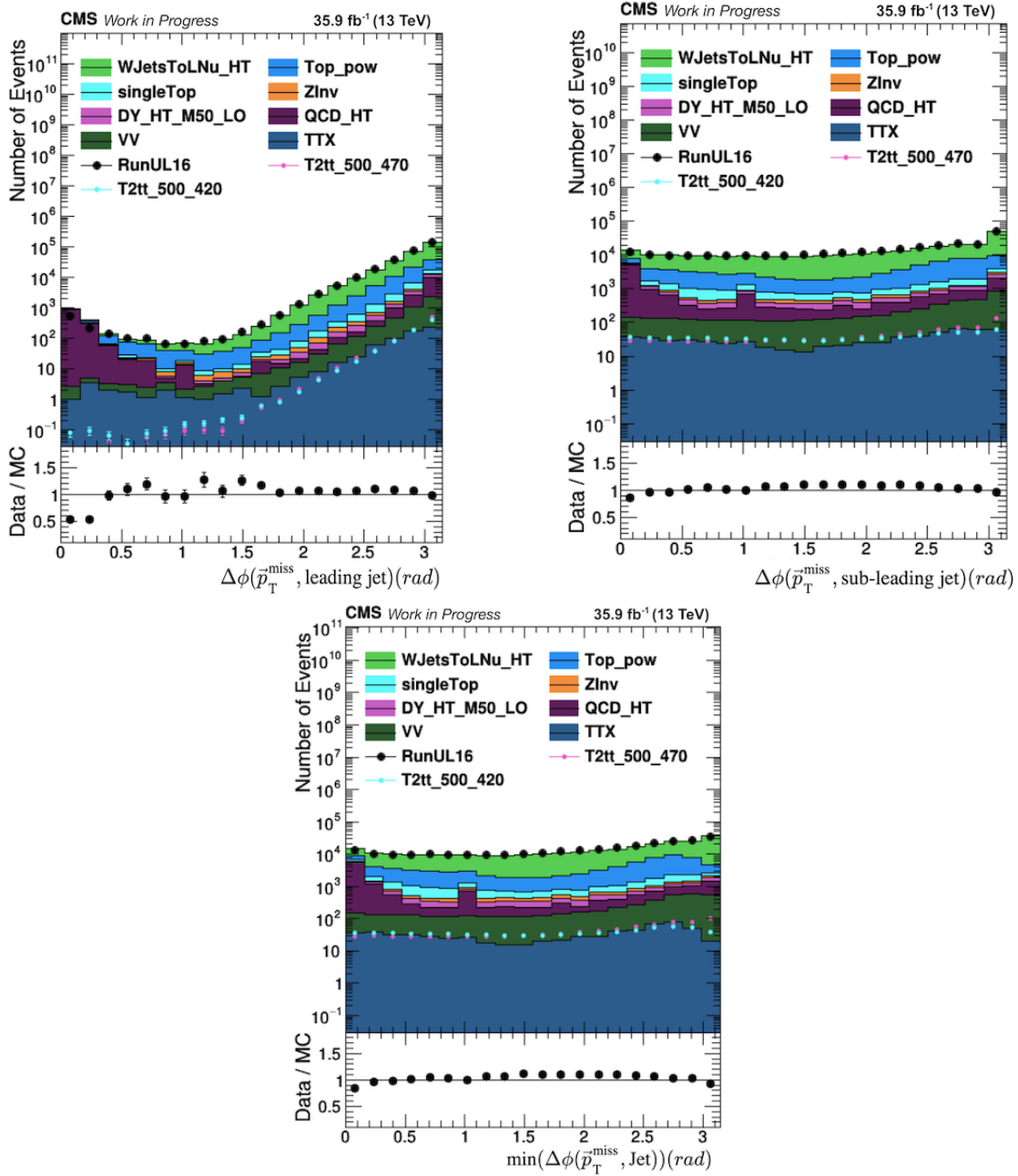


Figure 4.6: Distributions of angular separation between \vec{p}_T^{miss} and the leading (top left) and sub-leading jet (top right). The bottom plot shows the distribution of the $\Delta\phi^{\text{min}}$ variable defined in Equation (4.2). While the top row shows the individual angular separations, the bottom plot illustrates the minimum of the two. Unlike the QCD background, peaking at low values of $\Delta\phi^{\text{min}}$, the SM processes and SUSY signal points have a flat distribution.

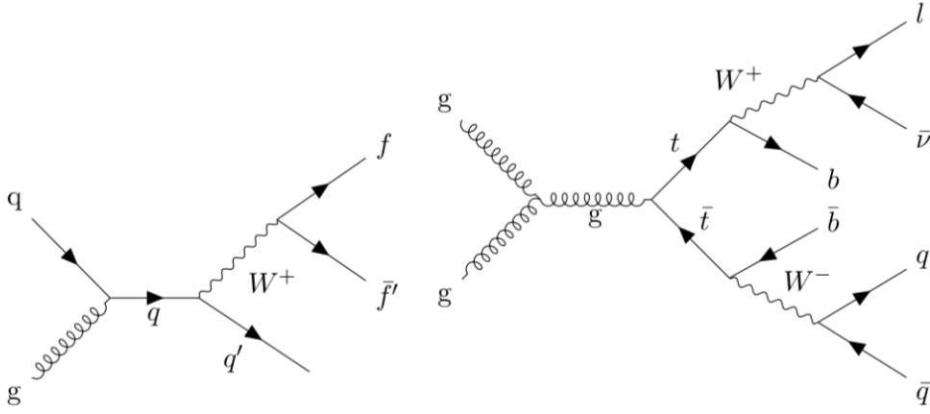


Figure 4.7: Feynman diagrams illustrating W boson production in association with jets (left) and top pair production (right). These two processes constitute the dominant backgrounds in this analysis.

the final state mimics our signal final state, particularly without a b -tagged jet. In order to discriminate against these background events, various kinematic properties, such as lepton p_T , and transverse mass M_T , can be used. Another SM process with a large cross section contributing significantly to the background is top pair production ($t\bar{t}$). Each top decays $\sim 99\%$ of the time to W bosons via $t \rightarrow Wb$. Roughly one third of $t\bar{t}$ events decay semileptonically resulting in single lepton, significant p_T^{miss} and at least a b quark in the final state. Single top events have a similar final state but a much smaller contribution. Both processes contain a b -tagged jet accompanied by a lepton and genuine p_T^{miss} from the neutrino, combined with possible mismeasurements in jet energy, resulting in additional missing transverse momentum. This similarity with signal final state kinematics makes $t\bar{t}$ process the largest contributor in background events with at least one b -tagged jet. The Feynman diagrams for W +Jets and $t\bar{t}$ are shown in Figure 4.7. These two processes are the main background of this analysis. Their estimation method will be described below in Section 4.5.1. Other relevant processes with similar final states as our signal but low in contribution due to smaller cross sections and reduced acceptance are top quark pair production associated with a boson such as $t\bar{t}W$, $t\bar{t}Z$, $t\bar{t}\gamma$, collectively called $t\bar{t}X$. In addition, there are contributions from electroweak processes such as Diboson (VV) and Drell-Yan (DY) production with prompt leptons. These smaller contributions are grouped and referred to as “rare” backgrounds in this analysis. This analysis considers the electrons and muons originating from taus as prompt leptons.

Another category of backgrounds, besides those with prompt leptons, comprises non-prompt or falsely identified (fake) leptons that can mimic the signal signature and pass our selection criteria. For this part of the analysis, where the stop is considered to have zero lifetime, the terms fake and non-prompt are used interchangeably, as their treatment is the same. The non-prompt leptons mainly originate from semileptonic decays of charm or bottom quarks within jets and hadron decays such as in-flight light meson decays. Misidentified (“fake”) leptons arise when a lepton from a hadronic jet is misidentified as a prompt lepton. Fake electrons can also arise from unidentified photon conversions or the misidentification of energetic jets in ECAL, and hadrons that punch through the calorimeter and reach the muon system can be reconstructed as fake muons. These fakes can pass the lepton selection in W +Jets or $t\bar{t}$ process, especially when the genuine prompt lepton is not reconstructed either due to its low p_T or because it lies outside the detector coverage. These non-prompt sources may have minimal contributions in comparison to the prompt. However, the large cross section of multijet processes make their contribu-

preselection requirements		
$p_T^{miss} > 200 \text{ GeV}$	$H_T > 300 \text{ GeV}$	$p_T^{ISR} > 100 \text{ GeV}$
Signal lepton: at least one electron or muon according to Table 3.1		
3 rd jet veto: no events with a third hard jet i.e. $p_T > 60 \text{ GeV}$		
anti-QCD cut : $\Delta\phi^{min} > 0.5$		
Tau veto: no events with hadronically decaying tau leptons with $p_T > 20 \text{ GeV}$		
2 nd lepton veto: no events with a second lepton with $p_T > 20 \text{ GeV}$		

Table 4.2: This table summarizes the preselection cuts applied in the search for light top squarks. These selections reduce the amount of background contamination, while retaining as much of the signal process as possible.

tions non-trivial. Additionally, jet-energy mismeasurements can produce large fake p_T^{miss} , allowing the events from these processes to pass the selection. An additional significant source of non-prompt leptons is Z boson production with jets ($Z(\rightarrow \nu\nu) + \text{jets}$), where Z boson decays invisibly, and a jet can be misreconstructed as a (fake) lepton. All non-prompt background contributions are treated together, and the method used for their estimation is described in Section 4.5.2.

The baseline selection criteria described in Section 4.3.1 are summarised in Table 4.2. After applying these selection criteria, the SM background comprises mainly W+Jets and $t\bar{t}$ events. To validate the modelling of these backgrounds and the overall agreement between data and simulation, we examine the distributions of key observables across different data-taking periods. Distributions of two such variables, i.e. leading lepton p_T and the transverse mass (M_T) after preselection are shown in Figure 4.8, Figure 4.9 for full Run 2 dataset, split into 2016-preVFP, 2016-postVFP and 2017, 2018 respectively.

4.4 Analysis Strategy

Although applying preselection cuts reduces background significantly, SM processes are still much more abundant than signal process events. To further increase the signal sensitivity, kinematical properties of the signal such as M_T , lepton transverse momentum ($p_T(\ell)$) are exploited.

In order to increase the discriminatory power against significant SM background, we use a variable called transverse mass M_T , defined as:

$$M_T \equiv \sqrt{2p_T^{miss} p_T(\ell)(1 - \cos(\Delta\phi(\vec{\ell}, \vec{p}_T^{miss})))}. \quad (4.3)$$

M_T can be interpreted as the transverse component of invariant mass calculated from the lepton momentum and the missing transverse momentum in events with semi-invisible decays.

For SM processes like W+Jets or $t\bar{t}$, the M_T distribution peaks around the W boson mass because the lepton and the neutrino, the source of the missing transverse momenta, are correlated. For the SUSY signal, however, the p_T^{miss} originates from neutralinos, uncorrelated to the lepton, and hence results in a more uniform M_T distribution. We can design a signal region above the W boson mass peak in M_T and maintain signal count while diminishing the SM W boson background.

We introduce two variables that exploit the strong correlation between p_T^{miss} and H_T , for region “1”, and p_T^{miss} and p_T^{ISR} for region “2”. These variables provide additional discrimination against the background process in their respective regions. The two variables

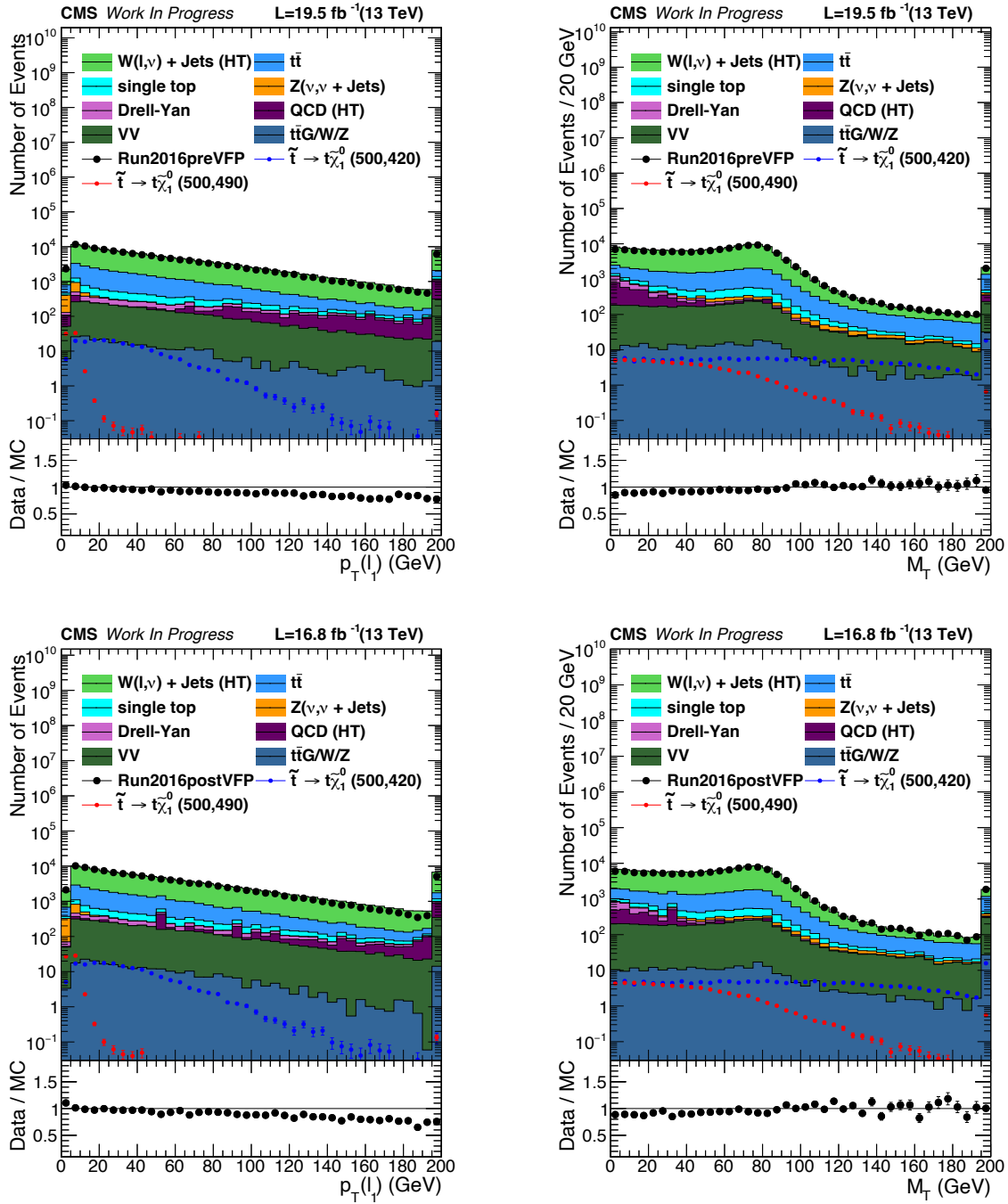


Figure 4.8: Distributions of leading lepton p_T (left) and M_T (right) for events passing the baseline selection criteria as described in Section 4.3.1, split into 2016-preVFP (top) and 2016-postVFP (bottom) data-taking periods. The 2016 dataset is split into preVFP and postVFP periods due to tracker-related issues; the affected data were reconstructed using the HIPM method. The two periods are treated separately in this analysis. Backgrounds are broken down by process, with representative signal benchmarks overlaid for comparison.

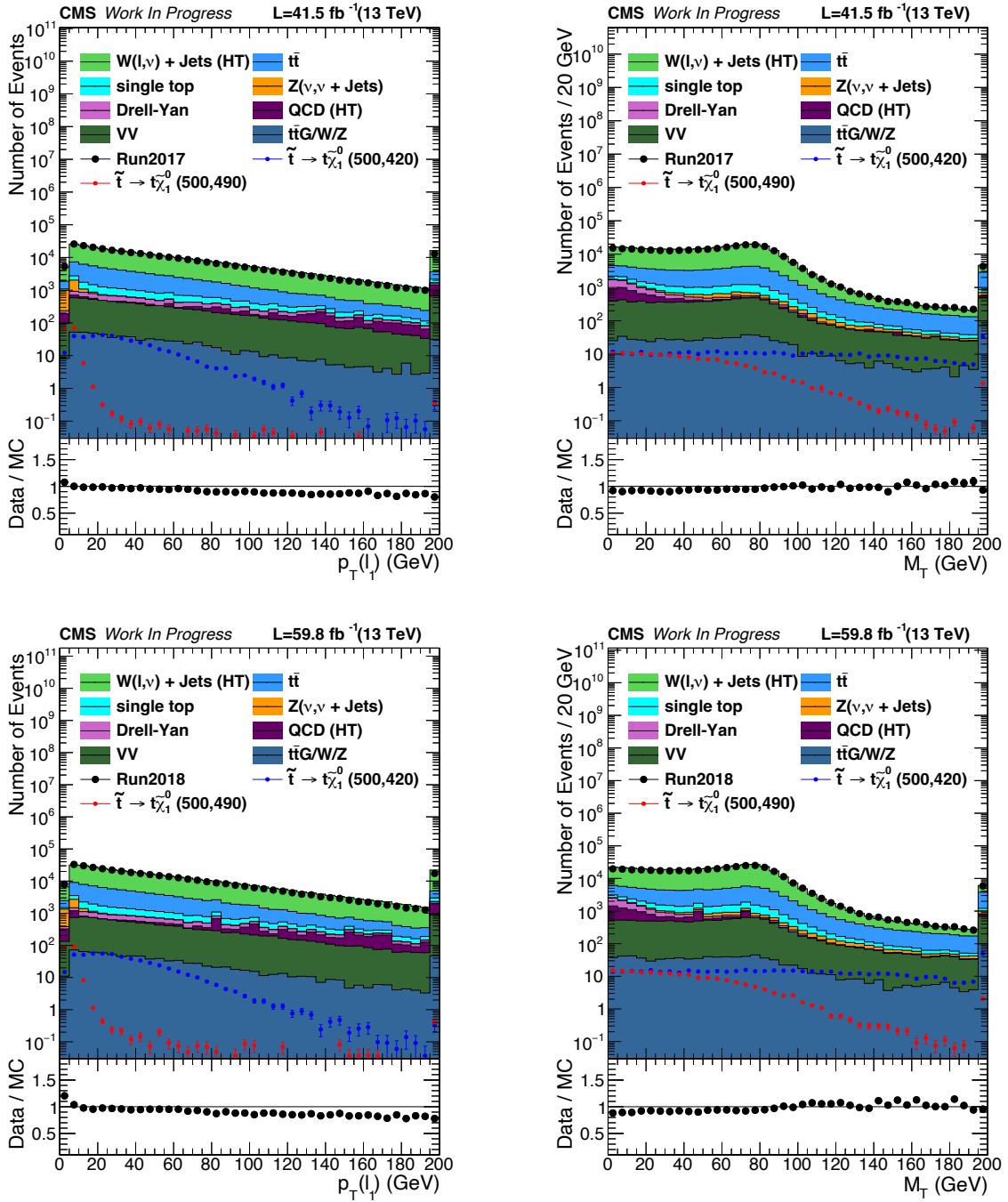


Figure 4.9: Distributions of leading lepton p_T (left) and M_T (right) for events passing the baseline selection criteria as described in Section 4.3.1, shown for 2017 (top) and 2018 (bottom) data-taking periods. These plots validate the modelling of SM background processes and assess data-MC agreement in different periods. Backgrounds are broken down by process, with representative signal benchmarks overlaid for comparison.

are defined as:

$$C_{T1} \equiv \min(p_T^{\text{miss}}, H_T - 100 \text{ GeV}), \quad (4.4)$$

$$C_{T2} \equiv \min(p_T^{\text{miss}}, p_T^{\text{ISR}} - 25 \text{ GeV}), \quad (4.5)$$

where the thresholds are chosen to maximise signal sensitivity by exploiting the correlation.

4.4.1 Signal Regions

The signal regions are designed on top of the preselection cuts to maintain the signal's sensitivity over a range of Δm , which also defines the kinematical features of the signal. Depending on the mass splitting, signal regions (SRs) are divided into two main regions, targeting small and large Δm in SR1 and SR2, respectively.

When Δm has small values, the decay products, especially the b jets, often have too little transverse momentum to be reliably reconstructed. Hence, the first signal region, SR1, vetoes events with a b-tagged jet. This selection, in turn, reduces the contribution of $t\bar{t}$ events, which typically contain b jets, making W+Jets the dominant SM background in SR1. In order to further suppress W+Jets background, the leptons are required to be within $|\eta| < 1.5$, and the acceptance threshold of H_T is raised to 400 GeV together with p_T^{miss} to 300 GeV, which can be simultaneously achieved by requiring $C_{T1} > 300$ GeV, as defined in Equation (4.4). The C_{T1} variable is therefore used to define the SR1 region, whereas C_{T2} is used for SR2.

SR2 targets signal events with larger mass splitting (Δm), so events are required to have at least one soft ($p_T > 30$ GeV) b-tagged jet. However, a veto on hard ($p_T > 60$ GeV) b-tagged jets is applied in order to maintain signal sensitivity and control the $t\bar{t}$ contribution. Because SR2 is dominated by $t\bar{t}$ background, using the p_T^{ISR} based C_{T2} variable provides better discrimination than the hadronic activity-based variable C_{T1} . In SR2, $p_T^{\text{ISR}} > 325$ GeV and $p_T^{\text{miss}} > 300$ GeV is achieved through a cut on $C_{T2} > 300$ GeV, as per definition in Equation (4.5), which further suppresses the background in SR2.

Each signal region is further split into subregions in the M_T , lepton p_T and the subsequent C_T variable. The motivations for these subdivisions are discussed in the subsections below. The regions are schematically visualized in Figure 4.10.

M_T

Each signal region is divided into four subregions based on the transverse mass M_T , exploiting its discriminatory power against the W+Jets background. The four subregions are defined as:

- a. $M_T < 60$ GeV
- b. $60 \leq M_T < 90$ GeV
- c. $95 \leq M_T < 130$ GeV
- d. $M_T \geq 130$ GeV

As discussed above, the M_T distribution for W boson decays peaks around W mass, which would concentrate all the W+Jets events in the first two regions (“a” and “b”). Additionally, only negatively charged leptons are accepted in these two subregions to better discriminate against the W+Jets process by using its charge asymmetry to our advantage.

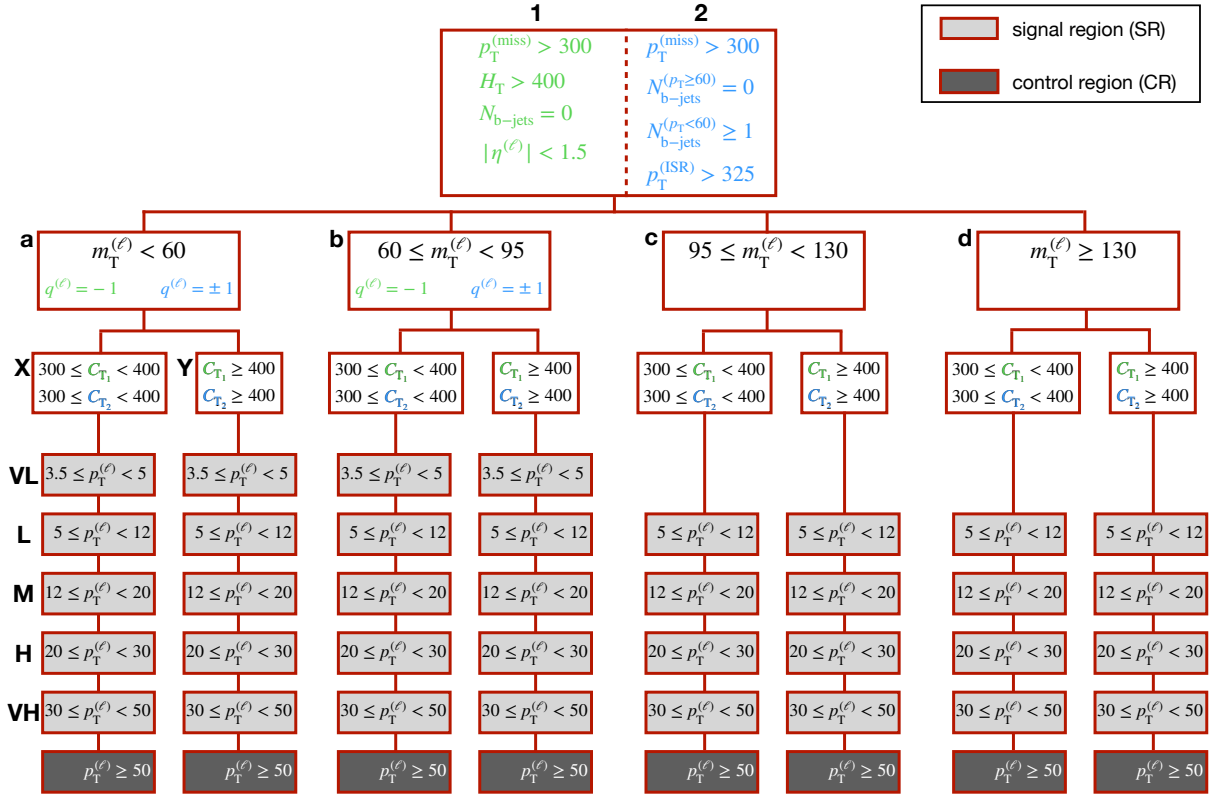


Figure 4.10: The graphic schematic visualizes the splitting of regions into 72 signal regions and 16 control regions, as described in Section 4.4.1 and Section 4.4.2, respectively. Region “1” (green) is predominantly W+Jets, whereas region “2” (blue) has large contributions from W+Jets and $t\bar{t}$. Each main region is split into four regions, labelled “a” to “d”, depending on M_T selection, and further split into C_{T_1} or C_{T_2} selections depending on whether it belongs to region “1” or “2”. These are then subdivided into six lepton p_T bins: five low $p_T(\ell)$ bins define the signal regions (“VL”, “L”, “M”, “H”, “VH”) and one high- $p_T(\ell)$ bin used to define control region.

The notations m_T^{ℓ} and p_T^{ℓ} used in the schema correspond to the variables M_T and $p_T(\ell)$ as used in this thesis. The graphic visualisation is taken from [66].

This asymmetry arises because positively charged W bosons are more frequently produced at the LHC than negatively charged ones. Since stop pair production is charge symmetric, the requirement to select only negatively charged leptons in these regions helps suppress the W+Jets background. These subregions are also adapted to the specific kinematics of the signal. Depending on the mass splitting (Δm), signal points with low Δm occupy low M_T subregions, whereas signal points with larger Δm will populate higher M_T subregions. As a result, this comprehensive subdivision based on M_T ensures that we maintain signal sensitivity over the entire range of mass splittings considered in this analysis.

C_{Ti}

Each M_T region is further divided into two regions “X” and “Y”. The thresholds for these partitions are defined below:

$$\text{X. } 300 \leq C_{Ti} < 400 \text{ GeV}$$

$$\text{Y. } C_{Ti} \geq 400 \text{ GeV}$$

where the subscript “i” in C_{Ti} refers to the signal region index (1 or 2), following the definitions discussed above.

Lepton p_T

Leptons from the signal process are usually soft, i.e. have low p_T , unlike those from background processes. Additionally, depending on the mass splitting in the signal point, there can be significant shape differences in the $p_T(\ell)$ distributions. This motivates the regions to be split further into slices of $p_T(\ell)$.

$$\text{VL. } 3.5 \leq p_T(\mu) < 5 \text{ GeV}$$

$$\text{L. } 5 \leq p_T(\ell) < 12 \text{ GeV}$$

$$\text{M. } 12 \leq p_T(\ell) < 20 \text{ GeV}$$

$$\text{H. } 20 \leq p_T(\ell) < 30 \text{ GeV}$$

$$\text{VH. } 30 \leq p_T(\ell) < 50 \text{ GeV}$$

The “VL” bin is included only in the lowest two M_T regions “a” and “b”, which covers very soft muons. This bin is included only in the low- M_T regions, since very low- p_T leptons arise mainly in small Δm signal, which populate the low- M_T phase space. The thresholds for the “VL” region only concern muons because of the momentum requirements described in Section 4.3.1.

After all subdivisions and binning in M_T , lepton p_T and C_T , a total of 72 SRs are designed. This finer splitting and additional bins, first analysed in collaboration with the work presented in [66], extend the strategy used in the published results based on 2016 data only [65] and take advantage of the full Run 2 dataset. The background composition for the defined signal and control regions for the 2016-preVFP data-taking period is shown in Figure 4.11. The corresponding distributions for the 2016-postVFP, 2017, and 2018 periods are provided in Appendix A.

4.4.2 Control Regions

The control regions (CRs) are defined by inverting the lepton p_T requirement to $p_T(\ell) > 50$ GeV. This results in one CR for four–five SRs, with common requirements on a b-tagged jet, M_T and C_T . The CRs are enriched with W+Jets and $t\bar{t}$ processes.

4.5 Background Estimation Techniques

W+Jets dominate the region “1” in both control and signal regions. Region “2” is dominated by W+Jets and $t\bar{t}$, together contributing about 80% of the background. Rare backgrounds are minor in most bins ($\approx 10 - 15\%$) and increase mainly in the highest- M_T slice, reaching $\approx 40\%$ in SR1 and $\approx 25\%$ in SR2. The contribution from non-prompt lepton sources is generally small. However, in low $p_T(\ell)$ regions, it can be non-negligible.

Different methods are implemented using data instead of relying solely on MC to estimate these backgrounds more reliably. The estimation of the main prompt backgrounds and the non-prompt component will be discussed further in the following sections.

4.5.1 Estimation of main prompt backgrounds

In order to obtain the correct normalization of the prompt component of W+Jets and $t\bar{t}$ in the signal regions (SRs), data from the dedicated control regions are used. As the lepton p_T distribution of both processes are highly correlated, and it is the kinematical variable in which SRs are further split, their normalization can be obtained simultaneously in the corresponding CR.

To achieve this, scale factors are introduced for W+Jets and $t\bar{t}$, one per control region, as freely floating parameters in the fit. This approach allows the yields of W+Jets and $t\bar{t}$ to be adjusted in the corresponding signal regions. The normalization is obtained by scaling W+Jets and $t\bar{t}$ yields to match the observed data in each control region after subtracting simulated estimates of all other background processes from the data. The resulting scale factors are then applied to the simulated yields of W+Jets and $t\bar{t}$ in the linked signal regions. As the analysis is still blinded, we replace the observed data with a background-only Asimov dataset¹, i.e. we set the content of each analysis bin to the summed Monte Carlo yields of all background processes. This “pseudo-data” is used only for validation and for expected results, the final scale factors are extracted only when the final fit is performed using the complete Run 2 dataset.

Technically, each scale factor is implemented as a freely floating parameter in the statistical model and multiplies the expected rate of the relevant process in the SRs associated with its CR.

4.5.2 Estimation of non-prompt backgrounds

The estimation of misidentified electrons or muons from heavy-flavour decays, also called fake leptons in signal regions, is performed directly from data using the tight-to-loose fake factor method [78]. This same method also estimates the contributions from processes without prompt leptons, such as QCD multijet and Z boson production with jets ($Z(\rightarrow \nu\nu) + \text{jets}$). The method employed in this analysis is described in detail in [66].

The fake rate method is based on defining an application region (AR) orthogonal to the signal region constructed by inverting the hybrid isolation requirements to have a region enriched with fakes. The selected events must have leptons passing a *loose* set of requirements on isolation and impact parameter but failing the signal requirements (*tight*). Events in the AR are extrapolated to the SR using a transfer or fake factor (F_F). The fake factor is measured in a *determination region* (DR), a fake enriched region that is non-overlapping with AR and SR, using collision data to minimize reliance on simulation and applied to collision events in the AR. The determination region is divided into SR-like

¹An Asimov dataset is a fluctuation-free (deterministic) dataset equal to the model expectation in each bin [77].

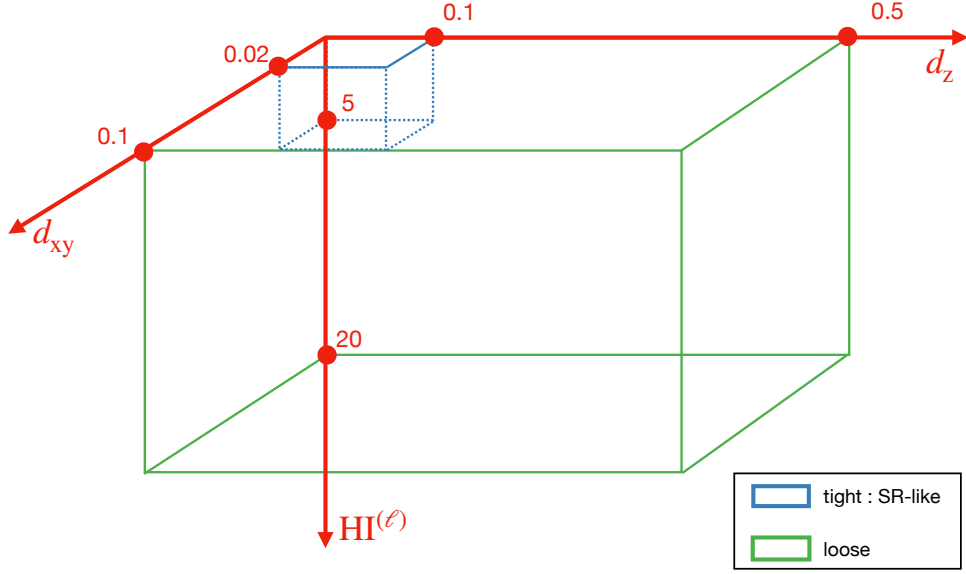


Figure 4.12: The figure visualizes loose (green) and SR-like (blue) determination region, as defined in (4.6) and (4.9), respectively. The AR-like determination region is defined by the selection $loose \wedge \neg tight$. Both regions are non-overlapping. The image is taken from [66].

and AR-like parts, with events passing the *tight* and events passing *loose* but failing the *tight* requirements, respectively.

The first step is to define the determination region, which is enriched in QCD multijet events representing non-prompt leptons and is non-overlapping with the signal and application regions. Different triggers are used to select events in the DR according to the lepton flavour. The events are selected using triggers with low- p_T lepton with a jet leg targeting soft leptons as in the SR. We use $\mu(3)PFJet(40)$ for muons across the full p_T range, and $e(8)PFJet(30)$ for electrons with $p_T \geq 12$ GeV while switching to $PFJet(40)$ for electrons with $p_T < 12$ GeV to cover the 5 GeV region and avoid the 8 GeV turn-on. At least one jet must satisfy the analysis requirements $p_T > 20$ GeV and $|\eta| < 2.4$. Events with exactly one lepton are selected.

Soft leptons in this region are selected based on the same criteria as in the main analysis, summarised in Table 4.1, except for the following selections:

$$\begin{aligned} d_{xy} &< 0.1 \text{ cm} \\ d_z &< 0.5 \text{ cm} \\ HI &< 20 \text{ GeV.} \end{aligned} \quad (4.6)$$

In order to have a QCD-dominated region for the fake factor measurement, the following selection is applied:

$$\begin{aligned} p_T^{\text{miss}} &< 50 \text{ GeV} \\ M_T &< 40 \text{ GeV,} \end{aligned} \quad (4.7)$$

The fake factor (F_F) is measured using the following formula:

$$F_F = \frac{N^{(\text{tight})} - \sum_{\text{prompt}} N_{MC}^{(\text{tight})}}{N^{(\text{loose} \wedge \neg \text{tight})} - \sum_{\text{prompt}} N_{MC}^{(\text{loose} \wedge \neg \text{tight})}} \quad (4.8)$$

The analysis requirements define the SR-like (tight) region:

$$\begin{aligned}d_{xy} &< 0.02 \text{ cm} \\d_z &< 0.1 \text{ cm} \\HI &< 5 \text{ GeV}.\end{aligned}\tag{4.9}$$

The AR-like regions select events passing the loose selection specified in Equation (4.6) but failing the tight selection given in Equation (4.9). Both regions are visualised in Figure 4.12. To ensure F_F measurement solely from non-prompt and falsely identified leptons in the data, prompt contributions are subtracted as shown in Equation (4.7). F_F is measured separately for the electron and muon channel in different p_T and η bins. Two η bins are used for the CMS detector's barrel and endcap region. The p_T bins correspond to the definitions of the analysis regions. Preliminary F_F measurement is presented in the thesis [66]. The new approach results in the more robust measurement of F_F and ensures higher purity in fake leptons than the method employed previously in [65].

The measured fake factor is further validated to ensure it is not biased by some contamination or mismodelling. A closure test uses MC in the determination region to ensure that the obtained transfer factor matches the expected yield.

Additionally, the inverted anti-QCD validation region is used to test the universality of the choice of phase space for fake factor measurement. This region is kinematically similar to the signal region but enriched in misidentified leptons by inverting the anti-QCD selection. The resulting non-closures are then used as systematic uncertainties in the F_F application.

The full Run 2 UL fake-factor measurement, being developed by our Budapest team following the 2016 legacy proof-of-concept by the Vienna team, is still in progress and therefore not included in this thesis. For this work, the fake leptons are estimated using MC simulations.

4.6 Systematic Uncertainties

ISR Modelling

A systematic uncertainty equal to the full size of the applied ISR correction is assigned to the W+Jets sample (see Section 4.2.3). No ISR correction is recommended for $t\bar{t}$ or signal samples. However, a flat 1% systematic uncertainty is applied to account for potential mismodelling.

Integrated Luminosity

The integrated luminosities for the three data-taking year are measured within uncertainties of 1.2% [79], 2.3% [80] and 2.5% [81] for 2016, 2017 and 2018, respectively. However, the results presented in this thesis were produced with the previous CMS recommendation, using a 2.5% uncertainty for 2016 [82] and a combined Run 2 uncertainty of 1.8%. A subsequent CMS update recommends a 1.6% total uncertainty for the full Run 2 combination, the improvement is an effect of the systematic correlation. This uncertainty is applied to all simulated signal and background processes, except W+Jets and $t\bar{t}$, whose normalizations are obtained from data in this analysis.

Pileup

The uncertainty associated with the PU-reweighting method is estimated using the PU profile from data according to the number of interactions at each bunch crossing, assuming a total inelastic pp cross-section of 69.2 mb [73]. In order to estimate the uncertainty, the inelastic cross section is varied up and down by 4.6%. This uncertainty is treated as fully correlated across data-taking periods.

JES and JER

Systematic uncertainties due to miscalibration of the jet energy scale (JES) are referred to as JES uncertainties. These are estimated by varying jet energy correction (JEC) within one standard deviation as a function of transverse momentum (p_T) and pseudorapidity (η).

Jet energy resolution (JER) uncertainties account for differences between the resolution observed in data and simulations. In general, the jet resolution in data is worse than in simulation, requiring the jets in MC to be smeared to describe the data better. JER uncertainties are taken as uncorrelated between data-taking periods.

b-tagged Jet Efficiency

As the number of b-tagged jets defined the signal-region categories in this analysis, any changes in the b-tagging efficiency and misidentification probability of gluons, charm, and light quarks as b jets can shift events in different signal regions. The associated systematic uncertainties on signal and background yields are estimated by varying the SFs within uncertainties separately for light and heavy-flavoured jets.

Lepton Efficiency

Lepton scale factors (SF) for identification and isolation are derived as described in Section 4.2.3 and applied to signal and background samples, along with their statistical uncertainties. Additionally, to account for any further discrepancy between low- and high- p_T leptons, a flat 1% uncertainty is applied to background processes. For signal samples, additional uncertainties are assigned to cover the differences between full detector simulation (FULLSIM) and fast simulation (FASTSIM). This uncertainty is not applied to the background process, as they are fully simulated. In the results shown here, the FASTSIM uncertainties are not yet applied to the signal samples. All lepton-related uncertainties are treated as correlated among the data-taking periods.

Trigger Efficiency

Any potential bias is expected to be very small in the trigger efficiency measurement. To account for residual effects, a flat 1% uncertainty is applied to both signal and background processes.

4.7 Results

4.7.1 Statistical analysis

The analysis constrains the SMS $m_{\tilde{\tau}}-m_{\tilde{\chi}_1^0}$ mass plane by setting 95% confidence level (CL) upper limits on the stop production. A 100% branching ratio for the four-body final state

scenario under study is assumed. The exclusion limits (upper limits on the cross section of stop pair production) are derived using a profile likelihood ratio as a test statistic based on the binned likelihood method.

The binned maximum likelihood method, at the core, has the assumption that the contents of each bin (i) follow a Poisson distribution, with n observed events when ν are expected:

$$\mathcal{P}(n_i|\nu_i) = \frac{\nu_i^{n_i} e^{-\nu_i}}{n_i!}.$$

Let bins be indexed by i (running over all SR and CR bins). The per-bin yield expectation (ν_i) is written as:

$$\nu_i(\mu, \theta, \beta) = \mu \cdot S_i(\theta) + \sum_b \beta_{k,b} B_i^{(b)}(\theta)$$

where μ is the signal strength modifier, a multiplicative factor for the SUSY signal cross section, and is the parameter of interest. The expected events can be categorised into events from signal process (S_i) and from a background process b as $B_i^{(b)}$. The statistical model also includes a set of nuisance parameters θ , that parametrize systematic uncertainties affecting both signal and background contributions. These uncertainties include luminosity, lepton efficiency, jet energy scale and resolution and are discussed in Section 4.6.

Additionally, unconstrained parameters β_k are introduced to account for the combined normalization of W+Jets and $t\bar{t}$ background processes, collectively referred to as B^{free} . Each parameter β_k corresponds to a specific control region bin k and is shared between both background processes in that bin, as described in Section 4.5.1. These parameters are floated freely in the fit, allowing the yields of W+Jets and $t\bar{t}$ to adjust in each region based on observed data. The normalisation factor is fixed to unity for all other background processes not included in B^{free} .

The full binned likelihood function for observed events across all signal and control region bins is given by:

$$\mathcal{L}(\text{data}|\mu, \theta, \beta) = \prod_i \mathcal{P}\left(n_i \middle| \nu_i(\mu, \theta, \beta)\right) \prod_j \mathcal{C}(\theta_j) \quad (4.10)$$

In the likelihood, each nuisance parameter θ_j is constrained by a prior $\mathcal{C}(\theta_j)$ that reflects our external knowledge of that systematic. For uncertainties that act multiplicatively on the event rate, such as luminosity, pileup are modelled using a log-normal distribution, which keep expected yields positive. For uncertainties that also affect the shape of the distribution, two additional input templates are provided. These templates correspond to shifts of the nuisance parameter by ± 1 standard deviation. The nominal distribution is then interpolated using a polynomial between the up and down variation and extrapolated linearly beyond that.

For a given signal strength μ we define the profile likelihood ratio as:

$$\lambda(\mu) = \frac{\mathcal{L}(\text{data}|\mu, \hat{\theta}_\mu, \hat{\beta}_\mu)}{\mathcal{L}(\text{data}|\hat{\mu}, \hat{\theta}, \hat{\beta})}, \quad (4.11)$$

where the parameters $(\hat{\theta}_\mu, \hat{\beta}_\mu)$ are the values of nuisance parameters that maximize the likelihood \mathcal{L} for a given μ , while the parameters $(\hat{\mu}, \hat{\theta}, \hat{\beta})$ maximize the likelihood globally and referred to as post-fit parameters. For exclusion, we use one-sided test statistic [77]:

$$q_\mu \begin{cases} -2\ln\lambda(\mu), & \hat{\mu} \leq \mu \\ 0, & \hat{\mu} > \mu \end{cases} \quad (4.12)$$

Setting q_μ to 0 when $\hat{\mu} > \mu$ protects against a possible upward fluctuation of data being interpreted as evidence against the signal hypothesis μ . This condition also ensures that the confidence level obtained is a one-sided interval.

The test statistic, as defined in Equation (4.12), follows a sampling distribution, $f(q_\mu)$, that depends on the hypothesis being tested. Therefore, we can find p -values for the two hypotheses, i.e. signal+background and background only:

$$\begin{aligned} p_\mu &= \int_{q_\mu^{\text{obs}}}^{\infty} f(q_\mu|\mu) dq_\mu, \\ p_b &= \int_{q_\mu^{\text{obs}}}^{\infty} f(q_\mu|0) dq_\mu, \end{aligned} \quad (4.13)$$

where q_μ^{obs} is the value of the test statistic computed with the data, and $f(q_\mu|\mu)$ and $f(q_\mu|0)$ denote the corresponding probability density distribution under the signal contribution ($\mu > 0$) or background-only ($\mu = 0$) hypothesis.

The value of CL_S is then obtained by taking the ratio of the p -values of signal+background distribution to background-only distribution denoted by p_μ and p_b respectively.

$$\text{CL}_S = \frac{p_\mu}{1 - p_b} \quad (4.14)$$

If CL_S falls below a pre-defined significance level $\alpha = 0.05$ (5%), the signal hypothesis, i.e. $\mu > 0$ is excluded at 95% confidence level ($1 - \alpha$).

4.7.2 Simplified Model Interpretation

The results for SUSY simplified models are presented in the $m_{\tilde{t}} - \Delta m$ plane to better visualize the sensitivity of the analysis across different mass splittings.

For limit calculations in the context of this thesis, observed data are replaced with the expectations from simulation. These expectations, corresponding to the sum of all background processes yield, are referred to as Asimov data.

Upper limits on the production cross section of stop quarks, assuming 100% branching ratio, are then calculated at 95% confidence level employing the asymptotic CL_S method, as described in the previous section [83, 84, 85].

The colour maps show the expected 95% upper limit on the signal cross section at each point in the $m_{\tilde{t}} - \Delta m$ plane for the four-body model. The thick red line corresponds to the expected excluded mass limits at 95% CL. Only expected limits under the background-only hypothesis are shown, as this analysis is blinded.

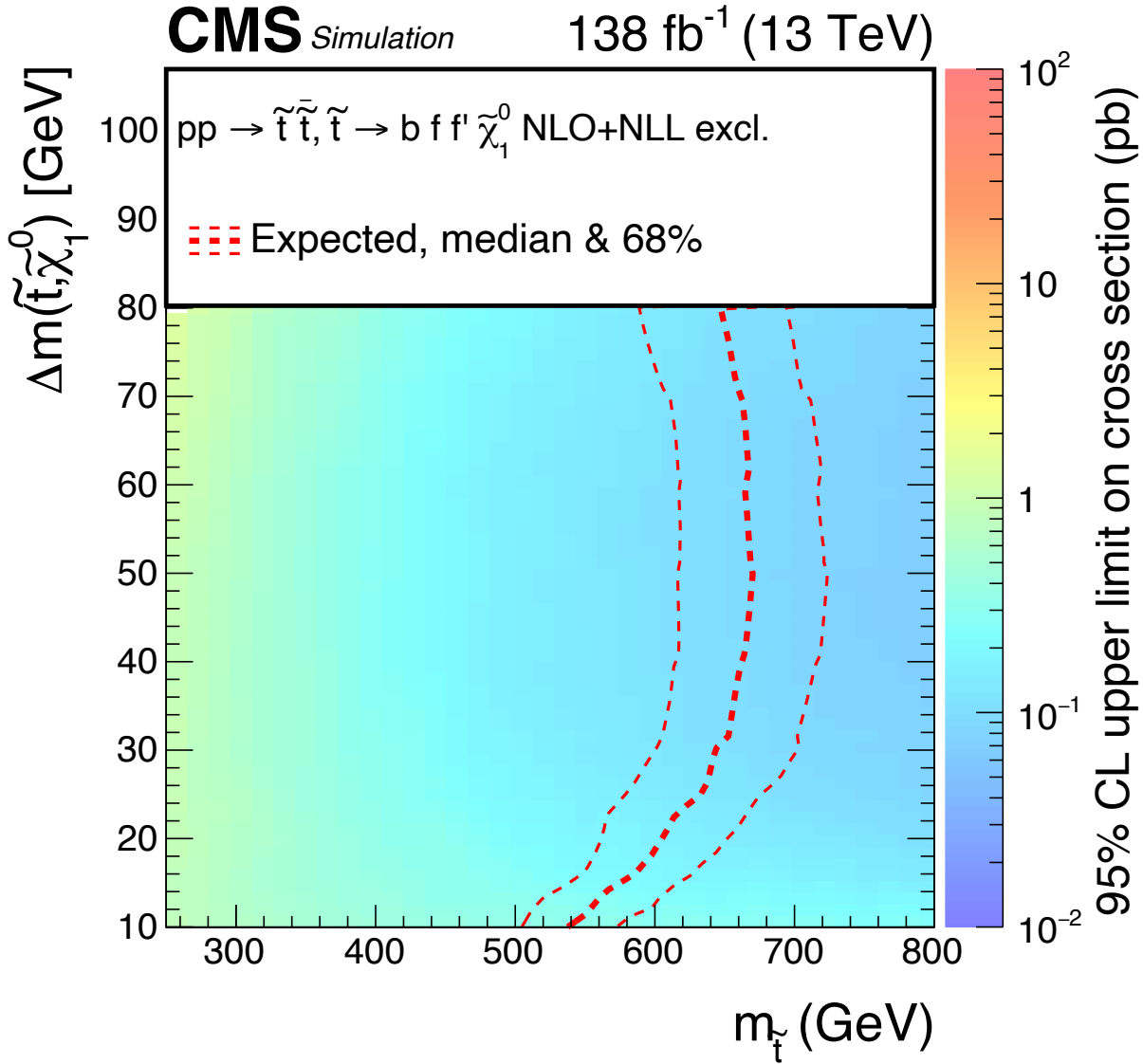


Figure 4.13:

At the 95% confidence level (CL), the exclusion limits for the four-body decay of the top squark, for the complete Run 2, are depicted on the $m_{\tilde{t}} - \Delta m(m_{\tilde{t}}, m_{\tilde{\chi}_1^0})$ plane. The color gradient illustrates the 95% CL upper limit at each point on the plane for the cross-section multiplied by the square of the branching fraction. The region below the bold dashed red line indicates the expected exclusion mass limit at 95% CL, assuming a branching fraction of 100%. The narrower red lines show the range containing 68% of the expected limit distribution.

Chapter 5

Conclusion and Outlook

5.1 Conclusion

This thesis presents a search for the pair production of supersymmetric top squarks (stops) in the compressed mass regime, where the mass difference between the stop and the lightest supersymmetric particle (LSP) is less than the W-boson mass. The focus is on the four-body decays of the stop targeting final states with a single soft lepton, missing transverse momentum, heavy-flavour (b-tagged) jets, and a hard ISR jet. The analysis is based on the full Run 2 dataset collected by the CMS detector at $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 138 fb^{-1} .

The strategy implemented for the 2016 data analysis serves as the foundation for this full Run 2 analysis, which introduces several improvements and refinements. The signal regions are defined for leptons with $p_T(\ell) \leq 50$, while events with $p_T(\ell) > 50$ are used to estimate the dominant W+Jets and $t\bar{t}$ background via in-situ normalization technique. Further splitting of the signal regions in M_T and lepton p_T improves sensitivity, especially when exploiting the full Run 2 statistics. Moreover, using low- p_T objects, such as lowering the selected jet thresholds from 30 GeV to 20 GeV, enhances the sensitivity of signals with larger mass differences (Δm). A selection on the angular separation between the jet and missing transverse vector is applied to better suppress the QCD multijet background and improve background rejection while maintaining signal efficiency. Employing the Ultra Legacy (UL) dataset, which benefits from improved reconstruction and calibration, further improves the overall performance.

Expected exclusion limits were derived using Asimov datasets, as the analysis remains blinded until all the components of the analysis are finalized. The results show significant sensitivity improvements compared to earlier studies.

5.2 Outlook

5.2.1 Ongoing extensions and methods

While this thesis focuses on the prompt four-body decays of the top squark, the full analysis considers several extensions actively under development to extend the search sensitivity. In scenarios with small mass splitting ($\Delta m \leq 30$ GeV), the stop lifetime becomes non-negligible [32], leading to displaced signatures not captured by prompt reconstruction, as described in Section 2.2.5. To encompass the long-lived stop scenario, this analysis is carried out in close collaboration with the Budapest team under the supervision of Gabriella Pásztor. The aim is to develop common strategies for both prompt and long-lived part of the analysis. Although the Budapest group pursues the long-lived

stop search, several contributions were made during the early studies and reprocessed samples to recover essential kinematic variables required for the full analysis by the Vienna group. Similarly, the full Run 2 fake-factor measurement is being finalised by the Budapest team. Moreover, further strategies are being implemented to enhance sensitivity for both prompt and long-lived scenarios. These include using secondary vertices to identify soft b quarks originating from stop decays and utilising low- p_T electron objects instead of standard reconstructed electrons to recover efficiency for signal electrons at larger impact parameters. These developments are particularly relevant for compressed scenarios that escape traditional selection criteria using standard objects reconstructed.

The Vienna and Budapest groups complement each other's efforts. While this thesis develops the prompt analysis, the Budapest group leads the long-lived interpretation, with both approaches relying on common reprocessed datasets and shared background strategies.

Together, these efforts aim to close the remaining gaps in the compressed stop search phase space and extend the sensitivity to the long-lived stops along with the prompt four-body decay model explored in this thesis.

5.2.2 Run 3 and HL-LHC prospects

Current MET triggers lose roughly 90 % of signal events because the soft decay products fall below threshold. Phase-2 (HL-LHC) CMS will introduce a Level-1 track trigger capable of selecting tracks down to $p_T \sim 2$ GeV, dramatically improving online retention of soft objects. These tools, together with new low-threshold HLT paths and upgraded tracking algorithms, position the HL-LHC era to extend sensitivity to even softer particles and metre-scale lifetimes.

While additional luminosity will indeed tighten constraints, Run 3's $\sim 300 \text{ fb}^{-1}$ will extend stop mass limits by roughly $\mathcal{O}(100 \text{ GeV})$ and further erode the large- Δm territory. It is precisely in the small- Δm compressed corners where naturalness arguments keep SUSY alive. More data not only increases statistical power but also enables the deployment of lower-threshold triggers and refined object reconstruction, pushing sensitivity into regions previously beyond reach.

Even as Run 3 and the HL-LHC progressively carve away parameter space, the compressed-spectrum and long-lived stop scenarios remain among the most theoretically motivated, and experimentally accessible frontiers for discovering (or definitively excluding) natural SUSY.

Appendices

Appendix A

Region plots for various eras in Run2

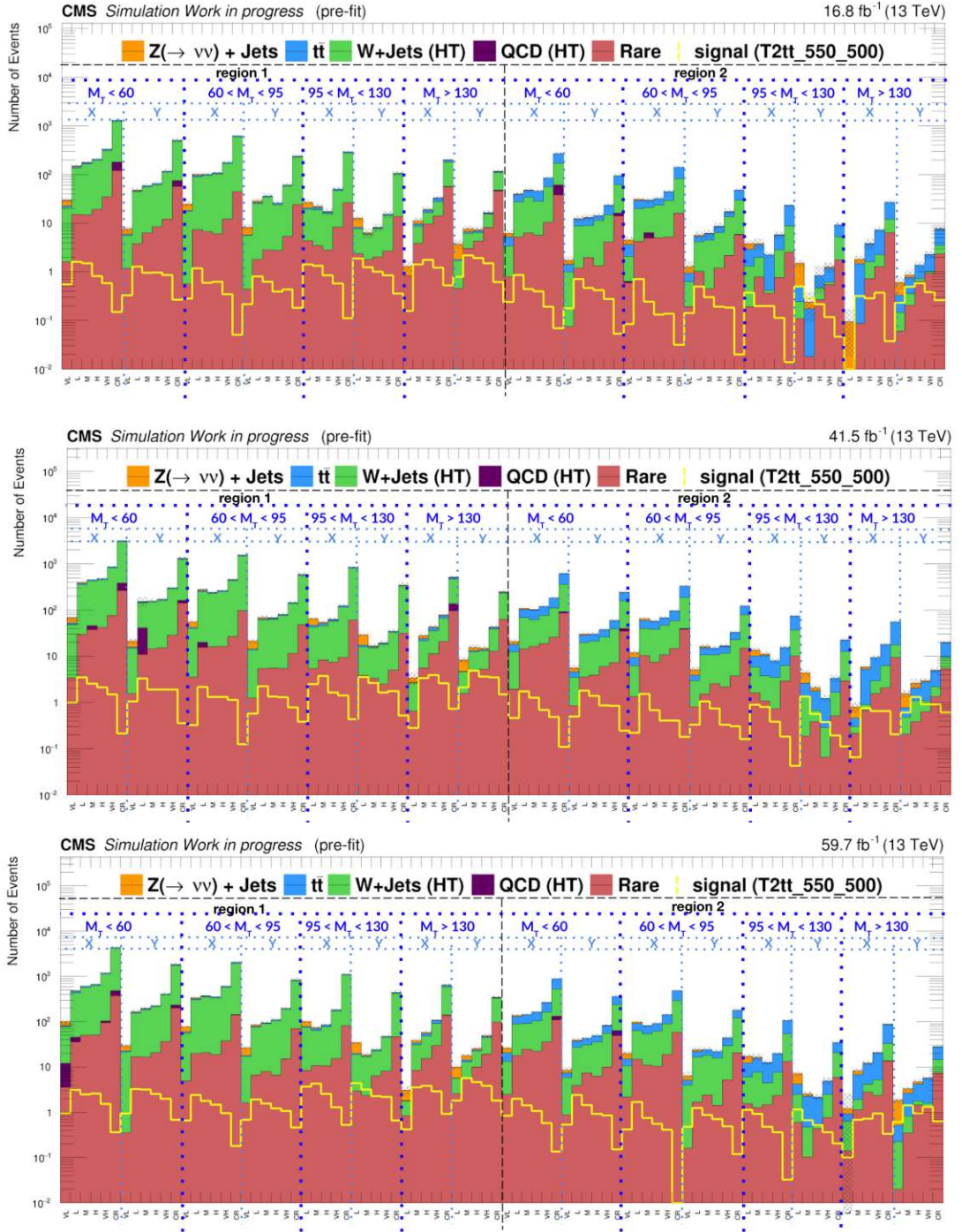


Figure A.1: The background composition of the defined signal and control regions as visualized in Figure 4.10 are presented for the 2016-postVFP (top), 2017 (middle), and 2018 (bottom) data-taking periods. The stacked histograms show contributions from W+Jets, $t\bar{t}$, Z+Jets, QCD multijet, and rare processes, estimated using MC simulation. The yellow line represents a benchmark signal model (T2tt with $m_{\tilde{t}_1} = 550$ GeV and $m_{\tilde{\chi}_1^0} = 500$ GeV). The hatched area indicates the statistical uncertainty on the total background yield.

Appendix B

Hybrid isolation and impact parameter scale factors for electrons and muons

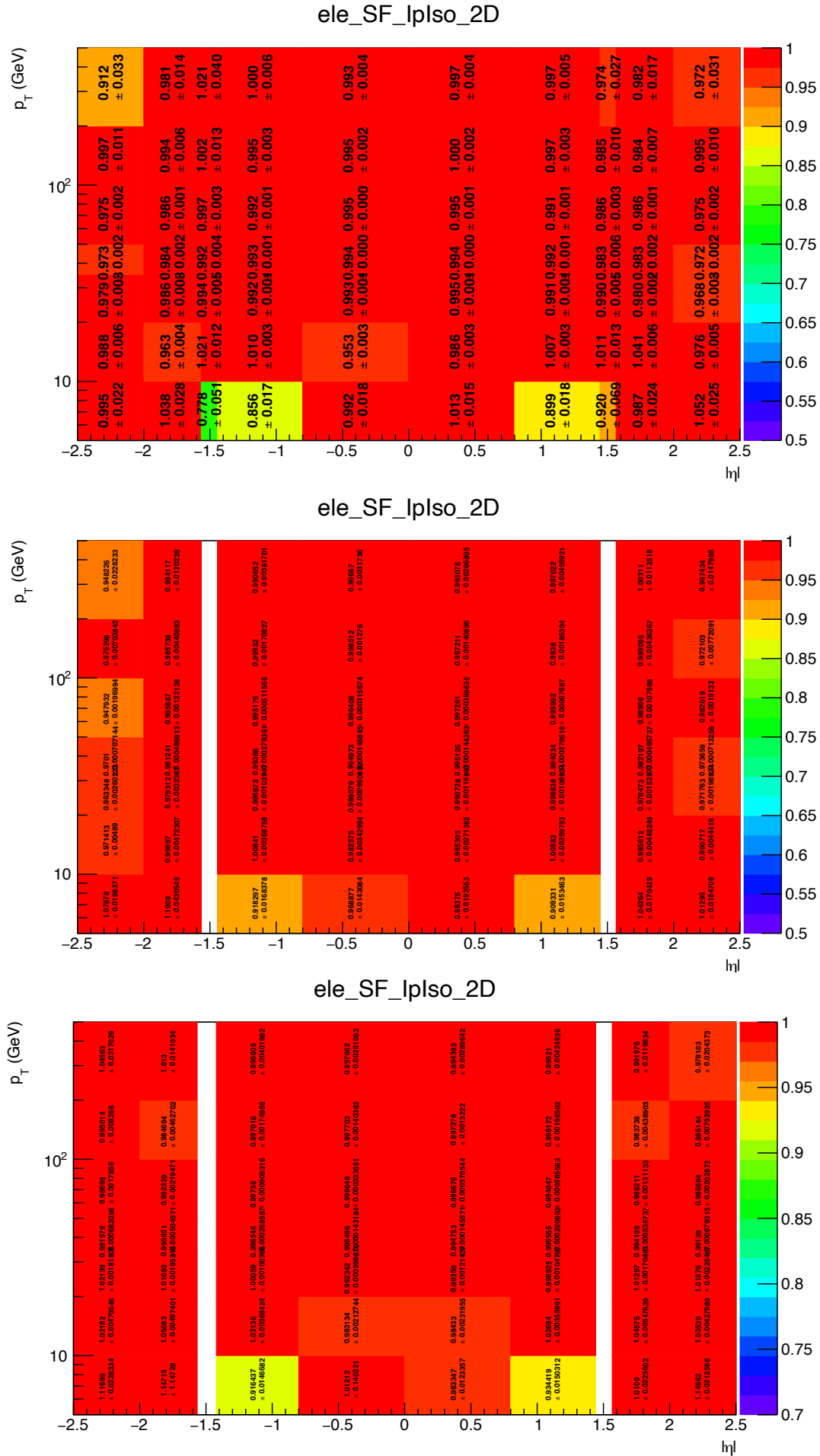


Figure B.1: Electron hybrid isolation and impact parameter scale factors derived privately for the 2016, 2017 and 2018 data-taking periods are shown in the figure. The barrel-endcap transition region ($1.44 < |\eta| < 1.56$) is excluded from the analysis.

Appendix C

Private NanoAOD production

A private NanoAOD was produced from centrally generated UL MiniAOD to restore generator–vertex information needed by the long–lived study. First step is looking up the central sample in DAS (Data Acquisition System) and finding out the CMS software (CMSSW) release using which the samples were produced.

```

cmsrel CMSSW_10_6_27
cd CMSSW_10_6_27/src/
source /cvmfs/cms.cern.ch/cmsset_default.sh
cmsenv
  
```

From DAS we can also find the “prepID”, which can be used in `mcm` [67] to find the `cmsDriver` command, which is used to produce the configuration file, for each year depending on if it is a FULLSIM or FASTSIM sample for MC or a primary dataset, for the reNanoAOD step from MiniAOD. As an example, `cmsDriver` command for a dilepton sample (FULLSIM) for 2016postVFP sample is given:

```

cmsDriver.py
--python_filename TOP-RunIISummer20UL16NanoAODv9-00143_1_cfg.py
--eventcontent NANOAOBSIM
--customise Configuration/DataProcessing/Utils.addMonitoring
--datatier NANOAOBSIM
--fileout file:TOP-RunIISummer20UL16NanoAODv9-00143.root
--conditions 106X_mcRun2_asymptotic_v17 --step NANO
--filein "dbs:/TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/
RunIISummer20UL16MiniAODv2-106X_mcRun2_asymptotic_v17-v1/MINIAOBSIM"
--era Run2_2016,run2_nanoAOD_106Xv2 --no_exec --mc -n 100
  
```

The changes per year for flags in the `cmsDriver` command are summarized in the Table C.1. Running this command generates a configuration file, which when input with the MiniAOD file would spit out NanoAOD files with same content as the central NanoAOD. In order to change the content of the NanoAOD samples, we need to alter the configuration file and give it the information we want to add.

```

from PhysicsTools.NanoAOD.common_cff import *
process.genParticleTable.variables = cms.PSet(
process.genParticleTable.variables,
vx = Var("vx", float,precision=10),
vy = Var("vy", float,precision=10),
vz = Var("vz", float,precision=10),
)
  
```

Since this is done for all the background samples of the analysis, including the signal grids, we try to automatise as much as possible. Using files including list of all miniAOD files, and all this is achieved using a wrapper in the HEPHY framework which runs all the jobs via CRAB (CMS Remote Analysis Builder) to access CMS computing resources.

Year	Condition	Era flags
2016postVFP	106X_mcRun2_asymptotic_v17	Run2_2016,run2_nanoAOD_106Xv2
2016preVFP	106X_mcRun2_asymptotic_preVFP_v11	Run2_2016_HIPM,run2_nanoAOD_106Xv2
2017	106X_mc2017_realistic_v9	Run2_2017,run2_nanoAOD_106Xv2
2018	106X_upgrade2018_realistic_v16_L1v1	Run2_2018,run2_nanoAOD_106Xv2

Table C.1: This table summarizes different flags varying per year. For FASTSIM samples, such as signal grids, we use an additional `--fast` flag. For data (primary datasets) we replace the `--mc` flag with `--data` flag and the conditions flag contains `106X_dataRun2_v35` and `--eventcontent` flag uses `NANOAOD` instead of `NANOAODSIM`.

List of Figures

2.1	Elementary particles of Standard Model and their properties.	4
2.2	CMS SM cross-section measurements	9
2.3	GUT scale within context of SM and MSSM.	14
2.4	Branching Ratios of four-body and two-body stop decays.	15
2.5	Two- and four-body decay widths and total widths as a function of Δm	16
2.6	Four-body stop decays.	16
3.1	The CERN accelerator complex.	18
3.2	Integrated luminosity of CMS during Run 2	19
3.3	CMS average pileup	20
3.4	A cutaway diagram of the CMS detector	21
3.5	Coordinate system of CMS	22
3.6	Illustration of pseudorapidity	22
3.7	Full CMS tracking system	23
3.8	CMS pixel detector	24
3.9	CMS electromagnetic calorimeter	26
3.10	CMS hadron calorimeter	27
3.11	Muon detector	28
3.12	Level-1 Trigger	29
3.13	PF algorithm	30
3.14	Illustration of a heavy flavour jet with a secondary vertex (SV).	36
4.1	T2tt SMS	39
4.2	MC sample production workflow.	42
4.3	Electron identification scale factors derived privately for the 2016, 2017 and 2018 data-taking periods are shown in the figure. The low- p_T bin (5–10 GeV) is derived privately and appended to the centrally provided CMS SFs. The barrel-endcap transition region ($1.44 < \eta < 1.56$) is excluded from the analysis.	45
4.4	Muon identification scale factors derived for the 2016, 2017 and 2018 data-taking periods are shown. The low- p_T region (3.5–10 GeV for 2016 and up to 20 GeV for later years) is privately measured and appended to the centrally provided CMS SFs.	46
4.5	Limit plots with jet p_t with different thresholds.	51
4.6	Distributions of angular separation between $p_{t\text{miss}}$ and two leading jets and $\min(\text{dphi})$	52
4.7	Feynman diagrams of W +jets and top pair production process.	53

4.8	Distributions of leading lepton p_T (left) and M_T (right) for events passing the baseline selection criteria as described in Section 4.3.1, split into 2016-preVFP (top) and 2016-postVFP (bottom) data-taking periods. The 2016 dataset is split into preVFP and postVFP periods due to tracker-related issues; the affected data were reconstructed using the HIPM method. The two periods are treated separately in this analysis. Backgrounds are broken down by process, with representative signal benchmarks overlaid for comparison.	55
4.9	Distributions of leading lepton p_T (left) and M_T (right) for events passing the baseline selection criteria as described in Section 4.3.1, shown for 2017 (top) and 2018 (bottom) data-taking periods. These plots validate the modelling of SM background processes and assess data-MC agreement in different periods. Backgrounds are broken down by process, with representative signal benchmarks overlaid for comparison.	56
4.10	Regions visualization for soft lepton analysis	58
4.11	Background composition in signal and control regions.	60
4.12	Determination region visualization.	62
4.13	Run2 expected exclusion limit.	67
A.1	The background composition of the defined signal and control regions as visualized in Figure 4.10 are presented for the 2016-postVFP (top), 2017 (middle), and 2018 (bottom) data-taking periods. The stacked histograms show contributions from W+Jets, $t\bar{t}$, Z+Jets, QCD multijet, and rare processes, estimated using MC simulation. The yellow line represents a benchmark signal model (T2tt with $m_{\tilde{t}_1} = 550$ GeV and $m_{\tilde{\chi}_1^0} = 500$ GeV). The hatched area indicates the statistical uncertainty on the total background yield.	72
B.1	Electron hybrid isolation and impact parameter scale factors derived privately for the 2016, 2017 and 2018 data-taking periods are shown in the figure. The barrel-endcap transition region ($1.44 < \eta < 1.56$) is excluded from the analysis.	74
B.2	Muon hybrid isolation and impact parameter scale factors derived privately for the 2016, 2017 and 2018 data-taking periods are shown in the figure.	75

List of Acronyms

BSM	Beyond the Standard Model
CERN	European Organization for Nuclear Research
CMS	Compact Muon Solenoid
CR	Control Region
CSC	Cathode Strip Chambers
DT	Drift Tube
ECAL	Electromagnetic Calorimeter
HCAL	Hadronic Calorimeter
HI	Hybrid Isolation
HLT	High-Level Trigger
IP	Impact Parameter
ISR	Initial State Radiation
KF	Kalman Filter
L1T	Level-1 Trigger
LHC	Large Hadron Collider
LSP	Lightest Supersymmetric Particle
MC	Monte Carlo
MET	Missing Transverse Momentum (p_T^{miss})
MSSM	Minimal Supersymmetric Standard Model
NLSP	Next-to-Lightest Supersymmetric Particle
PF	Particle Flow
PU	Pileup
PV	Primary Vertex
QCD	Quantum Chromodynamics
RPC	Resistive Plate Chambers

SF	Scale Factor
SM	Standard Model
SMS	Simplified Models
SR	Signal Region
SUSY	Supersymmetry
SV	Secondary Vertex
UL	Ultra-Legacy campaign
WP	Working Point

Acknowledgments

I would first like to express my deepest gratitude to my supervisor, Manfred Jeitler, for his constant and unwavering support throughout my PhD and for ensuring my academic success. I am also grateful to Ivan Mikulec and Wolfgang Adam for their guidance during the course of the analysis. Special thanks go to my colleagues in Budapest, especially Koushik Mandal and Gabriella Pásztor, for their invaluable collaboration.

I would also like to extend heartfelt thanks to Dietrich and Janik, whose kindness and support meant a great deal to me. I am forever indebted to all the friends who became my “Vienna family” and basically adopted me: Lukas, Claudia, Janik, Veronika, Sebastian, Marlene, Dominic, Nancy and Caro.

The DK-PI program provided me with many opportunities to attend and participate in conferences, and the opportunity to work at CERN for which I am truly grateful. Through DK-PI I also made wonderful friendships, in particular with Marlene, whom I greatly cherish and am in awe of.

Finally, I would like to acknowledge Prof. Jochen Schieck, the head of the institute for his support and Robert, our group leader, for teaching me many valuable lessons.

All my junior colleagues and friends from Hefy Tim, Markus, Maja, Tommy who most probably watched me grow (hopefully!). I was fortunate to meet and become friends with the second wave of brilliant PhD students, Mangesh, Cristina and Maryam. Special thanks to Daniel Spitzbart for his patience and guidance early on, and to the nicest postdocs, Alberto, Suman and Dennis.

I would like to thank my dearest friend Sarah for being there for me and always supporting me! To my wonderful in-laws, who welcomed me warmly, made me feel loved, and encouraged me, thank you. I also want to thank the friends I inherited from my husband, Constantin and Linda, for the amazing food, game nights, and New Year’s celebrations we shared, and I hope there are many more to come.

My journey would have looked very different had I not met all you wonderful people!

I would also like to take the opportunity to thank my external examiners Prof. Peter Križan and Prof. Feng Cunfeng for being very cooperative and kind, and especially for evaluating my thesis!

Bibliography

- [1] Abdus Salam and John Clive Ward. Electromagnetic and weak interactions. *Phys. Lett.*, 13:168–171, 1964. doi:[10.1016/0031-9163\(64\)90711-5](https://doi.org/10.1016/0031-9163(64)90711-5).
- [2] S. L. Glashow. Partial Symmetries of Weak Interactions. *Nucl. Phys.*, 22:579–588, 1961. doi:[10.1016/0029-5582\(61\)90469-2](https://doi.org/10.1016/0029-5582(61)90469-2).
- [3] Steven Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264–1266, 1967. doi:[10.1103/PhysRevLett.19.1264](https://doi.org/10.1103/PhysRevLett.19.1264).
- [4] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, 716:1–29, 2012. [arXiv:1207.7214](https://arxiv.org/abs/1207.7214), doi:[10.1016/j.physletb.2012.08.020](https://doi.org/10.1016/j.physletb.2012.08.020).
- [5] Serguei Chatrchyan et al. Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC. *Phys. Lett. B*, 716:30–61, 2012. [arXiv:1207.7235](https://arxiv.org/abs/1207.7235), doi:[10.1016/j.physletb.2012.08.021](https://doi.org/10.1016/j.physletb.2012.08.021).
- [6] Peter W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.*, 13:508–509, 1964. doi:[10.1103/PhysRevLett.13.508](https://doi.org/10.1103/PhysRevLett.13.508).
- [7] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Phys. Rev. Lett.*, 13:321–323, 1964. doi:[10.1103/PhysRevLett.13.321](https://doi.org/10.1103/PhysRevLett.13.321).
- [8] Douglas Clowe, Marusa Bradac, Anthony H. Gonzalez, Maxim Markevitch, Scott W. Randall, Christine Jones, and Dennis Zaritsky. A direct empirical proof of the existence of dark matter. *Astrophys. J. Lett.*, 648:L109–L113, 2006. [arXiv:astro-ph/0608407](https://arxiv.org/abs/astro-ph/0608407), doi:[10.1086/508162](https://doi.org/10.1086/508162).
- [9] Q. R. Ahmad et al. Measurement of the rate of $\nu_e + d \rightarrow p + p + e^-$ interactions produced by ^8B solar neutrinos at the Sudbury Neutrino Observatory. *Phys. Rev. Lett.*, 87:071301, 2001. [arXiv:nucl-ex/0106015](https://arxiv.org/abs/nucl-ex/0106015), doi:[10.1103/PhysRevLett.87.071301](https://doi.org/10.1103/PhysRevLett.87.071301).
- [10] Y. Fukuda et al. Evidence for oscillation of atmospheric neutrinos. *Phys. Rev. Lett.*, 81:1562–1567, 1998. [arXiv:hep-ex/9807003](https://arxiv.org/abs/hep-ex/9807003), doi:[10.1103/PhysRevLett.81.1562](https://doi.org/10.1103/PhysRevLett.81.1562).
- [11] J. Wess and B. Zumino. Supergauge Transformations in Four-Dimensions. *Nucl. Phys. B*, 70:39–50, 1974. doi:[10.1016/0550-3213\(74\)90355-1](https://doi.org/10.1016/0550-3213(74)90355-1).
- [12] Rudolf Haag, Jan T. Lopuszanski, and Martin Sohnius. All Possible Generators of Supersymmetries of the s Matrix. *Nucl. Phys. B*, 88:257, 1975. doi:[10.1016/0550-3213\(75\)90279-5](https://doi.org/10.1016/0550-3213(75)90279-5).
- [13] P. Ramond. Dual theory for free fermions. *Phys. Rev. D*, 3:2415–2418, May 1971. URL: <https://link.aps.org/doi/10.1103/PhysRevD.3.2415>, doi:[10.1103/PhysRevD.3.2415](https://doi.org/10.1103/PhysRevD.3.2415).

- [14] Steven Weinberg. *The Quantum theory of fields. Vol. 1: Foundations*. Cambridge University Press, 6 2005. doi:10.1017/CB09781139644167.
- [15] David Griffiths. *Introduction to elementary particles*. Wiley-VCH, 2nd, revised edition, 2008.
- [16] F. J. Dyson. The radiation theories of tomonaga, schwinger, and feynman. *Phys. Rev.*, 75:486–502, Feb 1949. URL: <https://link.aps.org/doi/10.1103/PhysRev.75.486>, doi:10.1103/PhysRev.75.486.
- [17] C. D. Anderson. The Positive Electron. *Phys. Rev.*, 43:491–494, 1933. doi:10.1103/PhysRev.43.491.
- [18] G. Arnison et al. Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c**2 at the CERN SPS Collider. *Phys. Lett. B*, 126:398–410, 1983. doi:10.1016/0370-2693(83)90188-0.
- [19] M Banner, R Battiston, P Bloch, Franco Bonaudi, K Borer, Michel Borghini, J C Chollet, A G Clark, C Conta, Pierre Darriulat, L Di Lella, J Dines-Hansen, P A Dor-saz, Louis Fayard, M Fraternali, Daniel Froidevaux, Jean-Marc Gaillard, O Gilde-meister, Giorgio V Goggi, H Grote, B Hahn, H Hänni, J R Hansen, P Hansen, Thomas M Himel, V Hungerbühler, Peter Jenni, Otto Mogens Kofoed-Hansen, E Lançon, M Livan, Sotirios S Loucatos, B Madsen, P Mani, B Mansoulié, G C Man-tovani, Livio P Mapelli, B Merkel, Michael E Mermikides, R Møllerud, B Nilsson, Christopher J Onions, G Parrou, F Pastore, Hartmute Plothow-Besch, M Polverel, J P Repellin, A F Rothenberg, A Roussarie, G Sauvage, J Schacher, J L Siegrist, H M Steiner, G Stimpfl, F Stocker, J Teiger, V Vercesi, A R Weidberg, Henri Zaccone, and W Zeller. Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN $\bar{p}p$ collider. *Phys. Lett. B*, 122:476–485, 1983. URL: <https://cds.cern.ch/record/142759>, doi:10.1016/0370-2693(83)91605-2.
- [20] E. D. Bloom, D. H. Coward, H. DeStaebler, J. Drees, G. Miller, L. W. Mo, R. E. Taylor, M. Breidenbach, J. I. Friedman, G. C. Hartmann, and H. W. Kendall. High-energy inelastic $e - p$ scattering at 6° and 10° . *Phys. Rev. Lett.*, 23:930–934, Oct 1969. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.23.930>, doi:10.1103/PhysRevLett.23.930.
- [21] Courtesy to Wikipedia: "Standard Model of Elementary Particles" by MissMJ-Own work by uploader, PBS NOVA, Fermilab, Office of Science, United States Department of Energy, Particle Data Group. Particle content of the Standard Model. [Online; accessed 20-July-2021].
- [22] Masud Chaichian and Kazuhiko Nishijima. An essay on color confinement, 1999. URL: <https://arxiv.org/abs/hep-th/9909158>, arXiv:hep-th/9909158.
- [23] C. Grojean, editor. *2012 European School of High-Energy Physics: La Pommeraye, Anjou, France 06 - 19 Jun 2012. Proceedings, 2012 European School of High-Energy Physics (ESHEP 2012). La Pommeraye, Anjou, France, June 06-19, 2012. 2012 European School of High-Energy Physics*, CERN Yellow Reports: School Proceedings, Geneva, 12 2014. CERN, CERN. URL: <https://cds.cern.ch/record/1406310>, arXiv:1504.07522, doi:10.5170/CERN-2014-008.

- [24] David J. Gross and Frank Wilczek. Ultraviolet behavior of non-abelian gauge theories. *Phys. Rev. Lett.*, 30:1343–1346, Jun 1973. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.30.1343>, doi:10.1103/PhysRevLett.30.1343.
- [25] H. David Politzer. Reliable perturbative results for strong interactions? *Phys. Rev. Lett.*, 30:1346–1349, Jun 1973. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.30.1346>, doi:10.1103/PhysRevLett.30.1346.
- [26] CMS collaboration. CMS Standard Model summary plots. [Online; accessed 24-July-2025].
- [27] Stephen P. Martin. A Supersymmetry primer. *Adv. Ser. Direct. High Energy Phys.*, 18:1–98, 1998. [arXiv:hep-ph/9709356](https://arxiv.org/abs/hep-ph/9709356), doi:10.1142/9789812839657_0001.
- [28] Savas Dimopoulos and Howard Georgi. Softly Broken Supersymmetry and SU(5). *Nucl. Phys. B*, 193:150–162, 1981. doi:10.1016/0550-3213(81)90522-8.
- [29] Gian Francesco Giudice. Naturally speaking: The naturalness criterion and physics at the lhc. In *Perspectives on LHC Physics*, page 155–178. WORLD SCIENTIFIC, January 2008. URL: http://dx.doi.org/10.1142/9789812779762_0010, doi:10.1142/9789812779762_0010.
- [30] Gianfranco Bertone, Dan Hooper, and Joseph Silk. Particle dark matter: Evidence, candidates and constraints. *Phys. Rept.*, 405:279–390, 2005. [arXiv:hep-ph/0404175](https://arxiv.org/abs/hep-ph/0404175), doi:10.1016/j.physrep.2004.08.031.
- [31] Kim Griest and David Seckel. Three exceptions in the calculation of relic abundances. *Phys. Rev. D*, 43:3191–3203, May 1991. URL: <https://link.aps.org/doi/10.1103/PhysRevD.43.3191>, doi:10.1103/PhysRevD.43.3191.
- [32] R. Gröber, Margarete M. Mühlleitner, E. Popenza, and A. Wlotzka. Light Stop Decays: Implications for LHC Searches. *Eur. Phys. J. C*, 75:420, 2015. [arXiv:1408.4662](https://arxiv.org/abs/1408.4662), doi:10.1140/epjc/s10052-015-3626-z.
- [33] Phenomenological MSSM interpretation of CMS searches in pp collisions at 13 TeV. Technical report, CERN, Geneva, 2024. URL: <https://cds.cern.ch/record/2906621>.
- [34] Navid K Rad. Search for supersymmetry partners of the top quark in models with compressed mass spectra with the CMS detector. *TU Wien Bibliothekssystem*, 2018, 2018. doi:10.34726/hss.2018.56980.
- [35] Lyndon R Evans and Philip Bryant. LHC Machine. *JINST*, 3:S08001, 2008. This report is an abridged version of the LHC Design Report (CERN-2004-003). URL: <http://cds.cern.ch/record/1129806>, doi:10.1088/1748-0221/3/08/S08001.
- [36] Esma Mobs. The CERN accelerator complex in 2019. *Complexe des accélérateurs du CERN en 2019*, 2019. General Photo. URL: <https://cds.cern.ch/record/2684277>.
- [37] CMS Collaboration. Lumi Public Results - multi year plots. [Online; accessed 28-August-2021].
- [38] Tai Sakuma and Thomas McCauley. Detector and Event Visualization with SketchUp at the CMS Experiment. *J. Phys. Conf. Ser.*, 513:022032, 2014. [arXiv:1311.4942](https://arxiv.org/abs/1311.4942), doi:10.1088/1742-6596/513/2/022032.

- [39] I Neutelings. How to draw diagrams in LaTeX with TikZ, 2020. [Online; accessed 24-August-2021].
- [40] S. Chatrchyan et al. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008. doi:10.1088/1748-0221/3/08/S08004.
- [41] CMS. Precision measurement of the structure of the CMS inner tracking system using nuclear interactions with data collected in 2018. Technical report, CERN, 2019. URL: <https://cds.cern.ch/record/2664786>.
- [42] W. Adam, T. Bergauer, D. Blöch, M. Dragicevic, et al. The CMS Phase-1 Pixel Detector Upgrade. *JINST*, 16(02):P02027, 2021. URL: <https://cds.cern.ch/record/2748381>, arXiv:2012.14304, doi:10.1088/1748-0221/16/02/P02027.
- [43] G. L. Bayatian et al. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical design report. CMS. CERN, Geneva, 2006. There is an error on cover due to a technical problem for some items. URL: <https://cds.cern.ch/record/922757>.
- [44] CMS Collaboration. Simulation of the Silicon Strip Tracker pre-amplifier in early 2016 data. [Online; accessed 09-July-2021].
- [45] Chia Ming, Kuo, and (On behalf of the CMS ECAL Group). The commissioning and first results on the performance of the cms preshower detector. *Journal of Physics: Conference Series*, 293(1):012058, apr 2011. URL: <https://dx.doi.org/10.1088/1742-6596/293/1/012058>, doi:10.1088/1742-6596/293/1/012058.
- [46] Andrea Benaglia. The CMS ECAL performance with examples. Technical report, CERN, Geneva, 2014. URL: <https://cds.cern.ch/record/1632384>, doi:10.1088/1748-0221/9/02/C02008.
- [47] S. Chatrchyan et al. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008. doi:10.1088/1748-0221/3/08/S08004.
- [48] Albert M Sirunyan et al. Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JINST*, 15(10):P10017, 2020. URL: <https://cds.cern.ch/record/2721198>, arXiv:2006.10165, doi:10.1088/1748-0221/15/10/P10017.
- [49] A Tapper and Darin Acosta. CMS Technical Design Report for the Level-1 Trigger Upgrade. Technical report, CERN, 2013. Additional contacts: Jeffrey Spalding, Fermilab, Jeffrey.Spalding@cern.ch Didier Contardo, Universite Claude Bernard-Lyon I, didier.claude.contardo@cern.ch. URL: <https://cds.cern.ch/record/1556311>.
- [50] A.M. Sirunyan, Armen Tumasyan, Wolfgang Adam, Ece Asilar, et al. Particle-flow reconstruction and global event description with the CMS detector. Particle-flow reconstruction and global event description with the CMS detector. *JINST*, 12(10):P10003, 2017. Replaced with the published version. Added the journal reference and DOI. All the figures and tables can be found at <http://cms-results.web.cern.ch/cms-results/public-results/publications/PRF-14-001> (CMS Public Pages). URL: <https://cds.cern.ch/record/2270046>, arXiv:1706.04965, doi:10.1088/1748-0221/12/10/P10003.
- [51] Serguei Chatrchyan et al. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *JINST*, 9(10):P10009, 2014. arXiv:1405.6569, doi:10.1088/1748-0221/9/10/P10009.

- [52] R. Fruhwirth. Application of Kalman filtering to track and vertex fitting. *Nucl. Instrum. Meth. A*, 262:444–450, 1987. doi:[10.1016/0168-9002\(87\)90887-4](https://doi.org/10.1016/0168-9002(87)90887-4).
- [53] K. Rose. Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. *IEEE Proc.*, 86(11):2210–2239, 1998. doi:[10.1109/5.726788](https://doi.org/10.1109/5.726788).
- [54] R. Fruhwirth, W. Waltenberger, and P. Vanlaer. Adaptive vertex fitting. *J. Phys. G*, 34:N343, 2007. doi:[10.1088/0954-3899/34/12/N01](https://doi.org/10.1088/0954-3899/34/12/N01).
- [55] Tracking and Primary Vertex Results in First 7 TeV Collisions. Technical report, CERN, Geneva, 2010. URL: <https://cds.cern.ch/record/1279383>.
- [56] A. M. Sirunyan et al. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV. *JINST*, 13(06):P06015, 2018. arXiv:[1804.04528](https://arxiv.org/abs/1804.04528), doi:[10.1088/1748-0221/13/06/P06015](https://doi.org/10.1088/1748-0221/13/06/P06015).
- [57] W. Adam, R. Fruhwirth, A. Strandlie, and T. Todorov. Reconstruction of electrons with the Gaussian sum filter in the CMS tracker at LHC. *eConf*, C0303241:TULT009, 2003. arXiv:[physics/0306087](https://arxiv.org/abs/physics/0306087), doi:[10.1088/0954-3899/31/9/N01](https://doi.org/10.1088/0954-3899/31/9/N01).
- [58] Serguei Chatrchyan et al. The Performance of the CMS Muon Detector in Proton-Proton Collisions at $\sqrt{s} = 7$ TeV at the LHC. *JINST*, 8:P11002, 2013. arXiv:[1306.6905](https://arxiv.org/abs/1306.6905), doi:[10.1088/1748-0221/8/11/P11002](https://doi.org/10.1088/1748-0221/8/11/P11002).
- [59] Albert M Sirunyan et al. Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC. *JINST*, 16(05):P05014, 2021. arXiv:[2012.06888](https://arxiv.org/abs/2012.06888), doi:[10.1088/1748-0221/16/05/P05014](https://doi.org/10.1088/1748-0221/16/05/P05014).
- [60] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti- k_t jet clustering algorithm. *JHEP*, 04:063, 2008. arXiv:[0802.1189](https://arxiv.org/abs/0802.1189), doi:[10.1088/1126-6708/2008/04/063](https://doi.org/10.1088/1126-6708/2008/04/063).
- [61] CMS. Pileup Removal Algorithms. Technical report, CERN, Geneva, 2014. URL: <https://cds.cern.ch/record/1751454>.
- [62] CMS. Jet algorithms performance in 13 TeV data. Technical report, CERN, Geneva, 2017. URL: <http://cds.cern.ch/record/2256875>.
- [63] A. M. Sirunyan et al. Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. *JINST*, 13(05):P05011, 2018. arXiv:[1712.07158](https://arxiv.org/abs/1712.07158), doi:[10.1088/1748-0221/13/05/P05011](https://doi.org/10.1088/1748-0221/13/05/P05011).
- [64] Vardan Khachatryan et al. Search for supersymmetry in events with soft leptons, low jet multiplicity, and missing transverse energy in proton-proton collisions at $\sqrt{s} = 8$ TeV. *Phys. Lett. B*, 759:9–35, 2016. Replaced with published version. Added the journal reference and DOI. All figures and tables can be found at <http://cms-results.web.cern.ch/cms-results/public-results/publications/SUS-14-021/index.html>. URL: <http://cds.cern.ch/record/2117955>, arXiv:[1512.08002](https://arxiv.org/abs/1512.08002), doi:[10.1016/j.physletb.2016.05.033](https://doi.org/10.1016/j.physletb.2016.05.033).
- [65] A. M. Sirunyan et al. Search for top squarks decaying via four-body or chargino-mediated modes in single-lepton final states in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 09:065, 2018. arXiv:[1805.05784](https://arxiv.org/abs/1805.05784), doi:[10.1007/JHEP09\(2018\)065](https://doi.org/10.1007/JHEP09(2018)065).

- [66] Janik Walter Andrejkovic. *Data-Driven Background Modeling for Precision Studies of the Higgs Boson and Searches for New Physics with the CMS Experiment*. PhD thesis, TU, Vienna, 2022. Presented 06-07-2022. URL: <https://cds.cern.ch/record/2823368>.
- [67] CMS Collaboration. Technical information about general Monte Carlo production and samples. [Online; accessed 19-February-2023].
- [68] S. Chatrchyan, V. Khachatryan, et al. Interpretation of searches for supersymmetry with simplified models. *Physical Review D*, 88(5), sep 2013. URL: <https://doi.org/10.1103/PhysRevD.88.052017>, doi:10.1103/physrevd.88.052017.
- [69] Marketa Jansova. *Search for the supersymmetric partner of the top quark and measurements of cluster properties in the silicon strip tracker of the CMS experiment at Run 2. Recherche du partenaire supersymétrique du quark top et mesure des propriétés des dépôts dans le trajectographe à pistes de silicium de l'expérience CMS au Run 2*. PhD thesis, Strasbourg, IPHC, 2018. Presented 27 Sep 2018. URL: <https://cds.cern.ch/record/2647308>.
- [70] Albert M Sirunyan et al. Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements. *Eur. Phys. J. C*, 80(1):4, 2020. arXiv:1903.12179, doi:10.1140/epjc/s10052-019-7499-4.
- [71] S Abdullin, P Azzi, F Beaudette, P Janot, A Perrotta, and (on behalf of the CMS Collaboration). The fast simulation of the cms detector at lh. *Journal of Physics: Conference Series*, 331(3):032049, dec 2011. URL: <https://dx.doi.org/10.1088/1742-6596/331/3/032049>, doi:10.1088/1742-6596/331/3/032049.
- [72] CMS Collaboration. stop antistop production cross sections in pp collisions at 13 TeV. [Online; accessed 1-January-2023].
- [73] Albert M Sirunyan et al. Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV. *JHEP*, 07:161, 2018. arXiv:1802.02613, doi:10.1007/JHEP07(2018)161.
- [74] M. Oreglia. *A Study of the Reactions $\psi' \rightarrow \gamma\gamma\psi$* . PhD thesis, Stanford University, Stanford, USA, 12 1980. URL: <https://inspirehep.net/literature/158483>.
- [75] John Erthal Gaiser. *Charmonium Spectroscopy From Radiative Decays of the J/ψ and ψ'* . PhD thesis, Stanford University, Stanford, USA, 8 1982. URL: <https://inspirehep.net/literature/183554>.
- [76] Kathrin Stoerig. *QCD background estimation for Supersymmetry searches with jets and missing transverse momentum with the ATLAS experiment at the Large Hadron Collider*. PhD thesis, Freiburg U., 2012. Presented 27 Nov 2012. URL: <https://cds.cern.ch/record/1514607>.
- [77] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, 71:1554, 2011. [Erratum: Eur.Phys.J.C 73, 2501 (2013)]. arXiv:1007.1727, doi:10.1140/epjc/s10052-011-1554-0.
- [78] Vardan Khachatryan et al. Search for new physics in same-sign dilepton events in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, 76(8):439, 2016. arXiv:1605.03171, doi:10.1140/epjc/s10052-016-4261-z.

- [79] Albert M Sirunyan et al. Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS. *Eur. Phys. J. C*, 81(9):800, 2021. [arXiv:2104.01927](https://arxiv.org/abs/2104.01927), [doi:10.1140/epjc/s10052-021-09538-2](https://doi.org/10.1140/epjc/s10052-021-09538-2).
- [80] CMS. CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV. Technical report, CERN, Geneva, 2018. URL: <https://cds.cern.ch/record/2621960>.
- [81] CMS. CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV. Technical report, CERN, Geneva, 2019. URL: <https://cds.cern.ch/record/2676164>.
- [82] CMS. CMS Luminosity Measurements for the 2016 Data Taking Period. Technical report, CERN, Geneva, 2017. URL: <https://cds.cern.ch/record/2257069>.
- [83] Thomas Junk. Confidence level computation for combining searches with small statistics. *Nucl. Instrum. Meth. A*, 434:435–443, 1999. [arXiv:hep-ex/9902006](https://arxiv.org/abs/hep-ex/9902006), [doi:10.1016/S0168-9002\(99\)00498-2](https://doi.org/10.1016/S0168-9002(99)00498-2).
- [84] Alexander L. Read. Presentation of search results: The CL(s) technique. *J. Phys. G*, 28:2693–2704, 2002. [doi:10.1088/0954-3899/28/10/313](https://doi.org/10.1088/0954-3899/28/10/313).
- [85] The ATLAS, CMS Collaborations, and The LHC Higgs Combination Group. Procedure for the LHC Higgs boson search combination in Summer 2011. Technical report, CERN, Geneva, Aug 2011. URL: <http://cds.cern.ch/record/1379837>.

PRIYA SAJID HUSSAIN

PHD PARTICLE PHYSICS

Experimental Particle Physicist working on getting a PhD on data collected at the Large Hadron Collider (CERN).

#Curious #DataAnalyst #Passionate

Mail: priasajid@gmail.com

Phone: 0049 1512 6130566

Github: <https://github.com/priyasajid>

Address: Veilchenweg 33, 51107 Köln

EDUCATION

2019 - 2025 PhD Experimental High Energy Physics

TU Wien - funded by Doktoratskolleg - Particles and Interactions
• Data analysis for the search of supersymmetric particle in a compressed scenario with LHC data collected over a span of three years. (2016-2018)
• Tool Development for visualisation of data, per pre defined categories, during data taking and offline processing.

2014 - 2016 M. Phil. High Energy Physics

University of the Punjab - Lahore
• Measurement of Muon identification and trigger efficiency using CMS data at 13 TeV.
• Intro into CERN Data Analytics Environment

2011 - 2013 M.Sc. Computational Physics

University of the Punjab - Lahore

2009 - 2011 B.Sc. Physics

University of the Punjab - Lahore

CONFERENCES & WORKSHOPS

- The DK-PI Summer School 2021 took place in Traunkirchen from September 20 to 24, 2021.
- The DK-PI Summer School 2020 took place online due to the Corona pandemic from September 21 to 25, 2020.
- The DK-PI Summer School 2019 took place in Zwettl from September 16 to 20, 2019 at Hotel Schwarz Alm.
- 05th Plenary ECFA Meeting November, 2019 at CERN
- The Third Asia-Europe-Pacific School of High Energy Physics 2016 held in Beijing, China.

TECHNICAL & SOFT SKILLS

Programming languages:

C++, python (proficient) & pandas, numpy (basic understanding)

Software Packages:

Microsoft office, Visual Studio, Visual Basic, Mathematica, ROOT, Git

Soft Skills

International collaboration, presentation & explanation of complex topics, problem solving & independent thinking

Github: <https://github.com/priyasajid>

PROFESSIONAL EXPERIENCE

April 2018 - May 2018 Insight Program

DESY - Hamburg (Germany)

Measured muon isolation efficiency, using Tag and Probe, on 2017 CMS data and documented my findings in the form of a report.

Sept. 2015 - Sept. 2016 Hosted Researcher

National Center for Physics - Islamabad (Pakistan)

July 2015 - June 2016 Teacher Assistant

Center for High energy Physics - University of the Punjab - Lahore (Pakistan)

LANGUAGES

- Urdu Native Mother tongue
- English Fluent IELTS Overall Band Score: 8.0 (Result announced on: 11 May, 2017).
- German B1