Precision measurement of the Z invisible width with the CMS experiment

Shane Davy Breeze
Imperial College London
Department of Physics

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

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Abstract

A precision measurement of the partial width of the Z boson decay to an invisible final state is reported. Events from proton-proton collisions provided by the Large Hadron Collider at a centre-of-mass energy of 13 TeV are collected by the Compact Muon Solenoid experiment with the invisibly decaying Z bosons inferred from a hadronic recoil. The data collected corresponds to a total integrated luminosity of $35.9\,\text{fb}^{-1}$. The invisible width is measured to be 512^{+16}_{-15} MeV, consistent with the standard model. This is a competitive result to previous measurements performed at the Large Electron Positron collider. The measurement is systematically dominated by uncertainties associated with the measurement of the energy scale of the recoiling hadronic system.

Declaration

I hereby declare that the work contained within this thesis is my own. It is produced on top of the existing work and ideas of several individuals and the Compact Muon Solenoid collaboration as a whole. The work performed by others is reviewed by the author, as key aspects in the later stages, in chapters 1–4. Although author contributions have been made to the hardware trigger, they have been omitted in this thesis. Chapter 5 details the event selection and categorisation designed by the author, although inputs are based on work performed by others. Chapter 6 consists of work performed by CMS collaborators and theoreticians, apart from the subsections discussing the $p_{\mathrm{T}}^{\mathrm{miss}}$ trigger efficiency and the $p_{\mathrm{T}}^{\mathrm{miss}}$ calibrations. All remaining chapters details the work done solely by the author.

In addition to the work presented, the author made key contributions to the analysis known as the α_{T} search [1].

The analysis is currently in the internal CMS review process, with a plan to publish this measurement. All results shown are presented as the work performed by the author and do not represent approved results by the CMS collaboration.

Shane Davy Breeze.

Acknowledgements

This thesis represents the accumulation of the work that I have done in the past four years. None of this would have been possible without the support, guidance and encouragement I have received from many people, before and during my PhD. To all of you, I am truly thankful.

Firstly, I would like to acknowledge the STFC for providing the funding to allow me to pursue my ambitions. This includes the two terrific years spent out at CERN. And to the Imperial HEP group for accepting me within their research group for, what I believe, is a successful PhD.

I would like to thank my supervisor, Sarah Malik, for her guidance throughout my PhD and, in particular, her patience and encouragement pushing me further. Although the outcome may differ from our original plan, I still believe it is an exceptional result that would not have been possible without you.

I would also like to thank the CMS collaborators and colleagues I have worked closely with throughout the past four years. Although this thesis does not include the work I performed on the hardware trigger or the α_T analysis, I am still grateful to both the Calo layer 2 and α_T groups. In particular, I would like to thank Alex Tapper for his support and guidance which resulted in my work being published early on during my PhD. I would also like to thank Rob Bainbridge for his continual support during the height of the α_T analysis, and after. And to Nick Wardle, thank you for the countless discussions that has helped my understanding of statistics, as well as our more general discussions.

Thank you to all my friends, old and new, who I have relied heavily upon not only with my work, but in keeping me grounded and helping me through the stressful times. I could not wish for a better group of friends.

To my Mum, thank you for all the support and encouragement throughout my whole life. You have been inspirational and have always believed in me. And thank you to my whole family; my dad, step-dad and siblings, for your support and interest in my work.

Contents

List of figures xiii

List of tables xvii

1. Introduction 1

2. The Standard Model of Particle Physics 3
2.1. Introduction . . . 3
2.2. Elementary particles . . . 3
2.3. Gauge theories . . . 4
2.3.1. Quantum Electrodynamics . . . 5
2.3.2. Quantum Chromodynamics . . . 6
2.3.3. Electroweak physics . . . 7
2.3.4. Spontaneous symmetry breaking . . . 8
2.4. Event simulation and generation . . . 9
2.4.1. Hadron collider physics . . . 10
2.5. Summary . . . 11

3. The LHC and CMS detector 13
3.1. Introduction . . . 13
3.2. The LHC . . . 13
3.2.1. Accelerator Complex . . . 13
3.2.2. LHC Main Ring . . . 14
3.3. Compact Muon Solenoid (CMS) Experiment . . . 15
3.3.1. Silicon Tracker . . . 16
3.3.2. ECAL . . . 18
3.3.3. HCAL . . . 21
3.3.4. Solenoid Magnet . . . 22
3.3.5. Muon Chambers . . . 23
3.3.6. Hardware Trigger . . . 25
3.3.7. Software Trigger . . . 27
3.3.8. Worldwide LHC Computing Grid (WLCG) . . . 27
3.4. Summary . . . 27

4. Event reconstruction **29**
4.1. Introduction . . . 29
4.2. Tracking . . . 29
4.2.1. Electron tracking . . . 32
4.2.2. Muon tracking . . . 32
4.3. Vertex reconstruction . . . 32
4.4. Calorimeter clustering . . . 33
4.5. The particle flow algorithm . . . 34
4.5.1. Muons . . . 35
4.5.2. Electrons and isolated photons . . . 35
4.5.3. Hadrons and non-isolated photons . . . 36
4.5.4. Jets . . . 37
4.5.5. Missing transverse momentum . . . 37
4.6. Summary . . . 38

5. Event Selection and Categorisation **41**
5.1. Introduction . . . 41
5.2. Samples . . . 41
5.2.1. Data collection . . . 41
5.2.2. Simulation . . . 43
5.3. Physics objects . . . 44
5.3.1. Muons . . . 44
5.3.2. Electrons . . . 44
5.3.3. Photons . . . 45
5.3.4. Jets . . . 47
5.4. Missing transverse momentum . . . 50
5.5. Event categorisation . . . 50
5.5.1. Baseline selection . . . 51
5.5.2. Categorisation . . . 52
5.6. Summary . . . 52

6. Simulation corrections **55**
6.1. Introduction . . . 55
6.1.1. Tag and probe method . . . 56
6.2. Trigger Efficiency . . . 56
6.2.1. Muon trigger efficiency . . . 56
6.2.2. Electron trigger efficiency . . . 58
6.2.3. $p_{\mathrm{T}}^{\mathrm{miss}}$ trigger efficiency . . . 59
6.3. Object-based corrections . . . 62
6.3.1. Muons . . . 62

6.3.2. Electrons and Photons . . . 64
6.3.3. Jets . . . 67
6.4. Luminosity . . . 70
6.5. Pileup reweighting . . . 70
6.6. ECAL timing degradation . . . 70
6.7. Theory corrections . . . 72
6.8. $p_{\mathrm{T}}^{\mathrm{miss}}$ calibration . . . 72
6.9. Summary . . . 80

7. Background estimation **81**
7.1. Introduction . . . 81
7.2. $\mathrm{W} \to \ell\nu$ prediction . . . 82
7.2.1. μ + jets control region . . . 82
7.2.2. e + jets control region . . . 83
7.2.3. τ_h + jets validation region . . . 85
7.2.4. Systematic uncertainties . . . 86
7.2.5. Transfer factor method . . . 86
7.2.6. Validation . . . 93
7.3. QCD multijet estimation . . . 99
7.4. γ^* contribution . . . 102
7.5. Remaining backgrounds . . . 103
7.6. Summary . . . 104

8. Invisible width measurement **105**
8.1. Introduction . . . 105
8.2. Signal regions . . . 105
8.2.1. $\ell\ell$ + jets region . . . 105
8.2.2. $p_{\mathrm{T}}^{\mathrm{miss}}$ + jets region . . . 107
8.3. Likelihood fit . . . 108

9. Conclusions **115**

A. The likelihood formalism **119**
A.1. Definition . . . 119
A.2. Limited sample size of alternative templates . . . 121
A.3. Alternative templates . . . 122

Bibliography **127**

List of figures

2.1. Summary of the standard model particles. 4

3.1. Schematic view of the LHC complex. 14

3.2. Schematic view of the CMS detector. 16

3.3. Cross-section view of the silicon tracker. 18

3.4. Front cross-section of the ECAL subdetector 20

3.5. Front cross-section view of the four HCAL subdetectors 21

3.6. Isometric view of the CMS solenoid encompassing the HCAL. 23

3.7. Isometric view of an octant of the CMS muon systems. 24

4.1. Coordinate system to describe the curved path of a charged particle within an magnetic field. 31

6.1. Muon trigger efficiency measurements. 57

6.2. Electron trigger efficiency. 58

6.3. Efficiency of the missing energy triggers in data and simulation. 59

6.4. Missing energy trigger efficiency measurements with the isolated track selection. 60

6.5. Systematic uncertainties associated with the missing energy trigger efficiency measurements. 61

6.6. Corrections to simulated muon identification efficiencies. 63

6.7. Corrections to simulated muon isolation efficiencies. 64

6.8. Corrections to simulated electron and photon identification efficiencies. 65

6.9. Corrections to simulated electron reconstruction efficiency. 66

6.10. Corrections to photon pixel veto efficiency. 66
6.11. Corrections to τ-tagged jet efficiency. 69
6.12. Pileup distribution in data and simulation for reweighting. 71
6.13. Kinematics diagram of a $Z/\gamma^*(\rightarrow \ell\bar{\ell}) + \text{jets}$ events. 74
6.14. Recoil distributions in dilepton final states. 75
6.15. Dilepton p_T distribution. 76
6.16. Fit result to extract the recoil scale and resolution in data and simulation. 77
6.17. Recoil scale and resolution in the dimuon final state. 78
6.18. Recoil scale and resolution in the dielectron final state. 79

7.1. Single muon final state kinematics. 83
7.2. Single electron final state kinematics. 84
7.3. Single hadronic τ-lepton final state kinematics. 85
7.4. Muon and electron transfer factors into the signal region 88
7.5. Theoretical uncertainty on the transfer factors 89
7.6. Object-based uncertainties on the transfer factors 90
7.7. Theoretical uncertainty on the transfer factors 91
7.8. Object-based uncertainties on the transfer factors 92
7.9. Nuisance parameters from a simultaneous fit to the muon and electron control regions 94
7.10. Recoil distributions in the muon and electron control regions after applying the transfer factors 95
7.11. Nuisance parameters from a fit with the transfer factor method from the muon and electron control regions to the τ-lepton validation region. 96
7.12. Recoil distributions corrected by the transfer factor from the muon and electron control regions to the τ-lepton validation region. 96
7.13. Nuisance parameters from a fit to the positive and negative muon control regions. 98
7.14. Recoil distributions after a fit to the positive and negative muon control regions 98

7.15. QCD-enriched sideband kinematics. 100
7.16. QCD multijet estimation within the enriched sideband 101
7.17. QCD multijet scaling factor for the signal region. 101
7.18. Invariant mass distributions of the $Z/\gamma^* \to \ell\bar{\ell}$, $Z \to \ell\bar{\ell}$ and $\gamma^* \to \ell\bar{\ell}$ processes. 103

8.1. Dimuon kinematics. 106
8.2. Dielectron kinematics. 107
8.3. $p_{\mathrm{T}}^{\mathrm{miss}}$ + jets kinematics. 108
8.4. Recoil distributions after a fit to the data between the signal and control regions. 109
8.5. Nuisance parameters after a fit to the data between the signal and control regions. 110
8.6. Likelihood profile in the r_{inv} dimension. 111
8.7. Two-dimensional likelihood profile in the r_{inv}–$r_{\ell\ell}$ plane. 112
8.8. Summary of the invisible width measurements. 113

A.1. Example of MC alternative template smoothing. 122
A.2. Alternative templates for the W + jets process. 123
A.3. Alternative templates for V + jets processes within the signal region 124

List of tables

1.1. Summary of the measurements of the Γ_{inv} from the LEP experiments for the direct measurements (with the statistical and systematic uncertainties), its combination, and the indirect measurement. 2

5.1. Requirements placed on muons. 45

5.2. Criteria to select and veto electrons. 46

5.3. Photon requirements to veto events. 46

5.4. Jet selection requirements. 47

5.5. Branching fractions of the τ-lepton decay modes. 48

5.6. Analysis event categorisation. 53

6.1. Summary of the uncertainties associated to theoretical predictions. 73

7.1. Recoil bins. 82

7.2. Contributions to the $Z/\gamma^* \to \ell\bar{\ell}$ from the $Z \to \ell\bar{\ell}$ and $\gamma^* \to \ell\bar{\ell}$ processes. 103

Chapter 1.

Introduction

The classification and description of fundamental particles and their interactions by the standard model (SM) of particle physics has been successfully confirmed by countless experiments. However, it lacks a candidate for dark matter [2, 3], a neutrino oscillation mechanism [4, 5, 6], or incorporating the gravitational interaction. At the LHC, the SM gained further success upon the discovery of the Higgs boson [7, 8]. Although numerous direct searches for beyond the SM (BSM) have been performed, no clear signal of BSM physics has been observed. As these searches continue to probe the extremities of distributions for BSM physics, an alternative direction proceeds through precision measurements of the SM, as presented in this thesis.

The SM predicts that about 20% [9] of the time a Z boson decays into a neutrino-antineutrino pair. This decay constitutes the Z invisible width, a signal not discernible in collider-based experiments. Instead, it is inferred from the missing momentum associated with the invisible decay of the Z as it recoils against a detectable system. A measurement of the Z invisible width is a key test of the SM and can be translated into constraints on the number of neutrino species coupling to the Z boson. Deviations from the SM expectation may point towards the existence of new particles beyond the SM which contribute to the Z invisible width. Therefore, a precise measurement of this quantity could reveal signs of new physics

The experiments at LEP measured the invisible width of the Z using two methods. The direct method measures the rate of an energetic photon recoiling against an invisible system attributed to the s-channel Z production and subsequent invisible decay with an associated initial state radiated (ISR) photon: $e^+e^- \to \gamma Z \to \gamma\nu\bar{\nu}$. The indirect method measures the total Z width, by studying the lineshape of the Z boson from e^+e^- collisions with a centre of mass energy near the Z mass, and subtracts the partial widths to all visible final state. This method provides the most precise measurement, with a sensitivity to BSM particles. A summary of the measurements performed by the LEP experiments L3 [10], OPAL [11] and ALEPH [12] for the direct method, a combination and the indirect measurement is shown in Tab. 1.1.

Experiment	Γ_{inv} (MeV)
L3	$498 \pm 12 \pm 12$
OPAL	$539 \pm 26 \pm 17$
ALEPH	$450 \pm 34 \pm 34$
LEP combined	503 ± 16
LEP indirect	499.0 ± 1.5

Table 1.1.: Summary of the measurements of the Γ_{inv} from the LEP experiments for the direct measurements (with the statistical and systematic uncertainties), its combination, and the indirect measurement.

To date, there is no published measurements of the Z invisible width at a hadron collider. However, a precise measurement at the LHC complements the LEP measurements through an independent test of the SM at a different energy scale, exposing contributions to the invisible width to TeV scale contributions. These additional contributions may come from couplings of the Z with dark matter particles, such as the lightest supersymmetric candidate, or a mixing between the neutrino and sterile neutrino states [13], to which both the direct and indirect measurements are sensitive. However, the direct measurement is also sensitive to contributions to the invisible final state without coupling to the Z boson, such as a heavier mediator coupling to neutrinos or other exotic invisible particles, or a four fermion interaction vertex [13]. Therefore, both measurements are valuable tests of the SM with the direct measurement a candidate for the LHC to provide a complementary measurement.

The experiment and analysis presented by the author is a direct measurement of the invisible width of the Z boson with the CMS detector using data from proton-proton collisions at $\sqrt{s} = 13$ TeV in 2016, corresponding to an integrated luminosity of 35.9 fb^{-1}. The measurement targets the invisible decay of the Z bosons recoiling against a hadronic system. The rate of the invisible final state is compared to the reference signal of the partial widths to leptons (both electrons and muons) $\Gamma(\mathrm{Z} \to \ell\ell)$, since these decays are kinematically similar. The invisible width is then given by

$$\Gamma(\mathrm{Z} \to \mathrm{inv}) = \frac{\sigma(\mathrm{Z}+\mathrm{jets})\mathcal{B}(\mathrm{Z} \to \mathrm{inv})}{\sigma(\mathrm{Z}+\mathrm{jets})\mathcal{B}(\mathrm{Z} \to \ell\bar{\ell})}\Gamma(\mathrm{Z} \to \ell\bar{\ell}) \,, \tag{1.1}$$

where $\sigma(\mathrm{Z}+\mathrm{jets})$ is the cross section for the $\mathrm{Z}+\mathrm{jets}$ process, modified by either the invisible branching fraction $\mathcal{B}(\mathrm{Z} \to \mathrm{inv})$ or dilepton branching fraction $\mathcal{B}(\mathrm{Z} \to \ell\bar{\ell})$.

Chapter 2.

The Standard Model of Particle Physics

2.1. Introduction

The standard model (SM) of particle physics is a quantum field theory description of the fundamental particles of matter and their interactions. The SM has been widely successful in its predictions, falling short in providing a dark matter candidate, including the existence of neutrino oscillations and incorporating gravitational interactions, among other possible discrepancies. Nevertheless, it is the current best minimal description unifying the electromagnetic, weak and strong forces. In the quantum field theory, the equations of motion are determined from the Lagrangian through the principle of least action. Gauge theories provide the mathematical basis to embed physical symmetries within the model.

The SM has had many distinguishing successes since its conception. One of the most notable is the prediction of the Higgs boson [14, 15, 16] and its subsequent discovery by the ATLAS [17] and CMS [18] experiments. The Higgs field is crucial in providing a mechanism for the apparent masses of the weak bosons, quarks and charged leptons whilst maintaining a gauge invariant theory.

This chapter will include an overview of the properties of the SM with particular detail on the quantum electrodynamic, quantum chromodynamic and electroweak sectors, leading to a discussion of the spontaneous symmetry breaking in the SM, known as the Higgs mechanism. This is followed by a discussion of the Monte Carlo techniques used to simulate the SM.

2.2. Elementary particles

The set of fundamental particles and interaction mediators within the SM is shown in Fig. 2.1. This includes the quarks — fermion constituents of protons and neutrons — with three generations of up and down-type quarks with a charge of $+2/3$ and $-1/3$ (in units of the elementary charge). A quark also takes one of three colour states, the strong force equivalent of the

electric charge. Through these colour states the quarks interact with the gluon colour octet via the strong interaction. In addition, the lepton sector consists of charged fermions which, in addition to the quarks, interact through the photon field leading to the electromagnetic interaction. Furthermore, all fermions — quarks, charged leptons and neutral leptons (neutrinos) — interact via the weak force, mediated by the W and Z bosons. A set of antiparticles mirror all the fermions leading to 12 fermion fields and an equal number of anti-fermion fields, split into 6 quarks and 6 leptons.

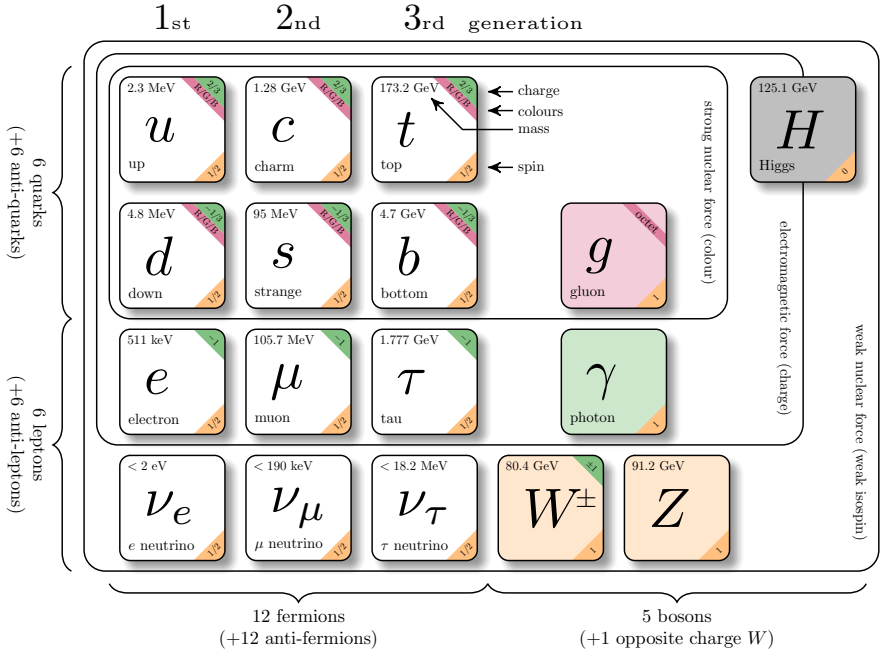


Figure 2.1.: Diagram of the particles within the SM including their quantum numbers such as charge, colour and spin, along with their measured masses. The left and right-handed chiral states are omitted and left for discussion in the main text.

2.3. Gauge theories

In quantum field theory, the fundamental fields are unobservable. Instead, properties deriving from these fields, such as the cross section or rate of an interaction or decay, are measured. Many configurations of the underlying fields can give identical observable quantities. The transformation between these configurations is known as a gauge transformation and leaves the physical observables unchanged, hence these transformations are known as gauge invariant. Under Noether's theorem, a change in the field configuration which leaves the action, and hence the Lagrangian density, invariant is associated with conserved currents [19].

2.3.1. Quantum Electrodynamics

Quantum electrondynamics (QED) is a gauge theory that sets out to describe the electron and the photon along with their interactions. The QED Lagrangian density consists of a Dirac field, the electron; a vector field, the photon, derived from Maxwell's equations; and an interaction term between the electron and photon fields:

$$\mathcal{L}_{\text{QED}} = \mathcal{L}_{\text{Dirac}} + \mathcal{L}_{\text{Maxwell}} + \mathcal{L}_{\text{int}} \,. \tag{2.1}$$

Maxwell's equations of electromagnetism are obtained from the Euler-Lagrange equations applied to the Lagrangian

$$\mathcal{L}_{\text{Maxwell}} = -\frac{1}{4} F^{\mu\nu} F_{\mu\nu} \,, \tag{2.2}$$

where $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$ [20] for the 4-potential A^μ with an electric scalar potential time-like and magnetic vector potential space-like components.

The Dirac Lagrangian describes the equations of motion, through application of the Euler-Lagrange equations [21], of a spin-half fermion field ψ. This field is represented as a four-component spinor which may be interpreted as the spin-up and down of the left and right-handed helicity states [20]. To form Lorentz invariant quantities, required in the Lagrangian, the adjoint field is defined as $\bar{\psi} = \psi^\dagger \gamma^0$ where $\psi^\dagger$ is the Hermitian conjugate of ψ and γ^0 is the time-like component of the Dirac gamma matrices. These matrices obey the anticommutation relation $\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}$, where $g^{\mu\nu}$ is the Minkowski space metric, leading to the Lorentz invariant quantities such as $\bar{\psi}\psi$ and $\bar{\psi}\gamma^\mu\psi$ [20]. These quantities form the basis of the Dirac Lagrangian of a free spin-half fermion:

$$\mathcal{L}_{\text{Dirac}} = i\bar{\psi}\gamma^\mu \partial_\mu \psi - m\bar{\psi}\psi \,, \tag{2.3}$$

where $i\partial_\mu$ is the position-space representation of the 4-momentum p_μ, and m is the mass of the particle [22]. $U(1)$ group transformations of the field

$$\psi \mapsto e^{i\alpha}\psi \tag{2.4}$$

are not physical observables since they cancel in the magnitude of the matrix element used to determine the rate or cross section of a process. The Lagrangian is already invariant to such global transformations. However, local transformations where $\alpha \mapsto \alpha(x)$ introduces a term to the Dirac Lagrangian given by

$$\mathcal{L}_{\text{Dirac}} \mapsto \mathcal{L}_{\text{Dirac}} - \bar{\psi}\gamma^\mu(\partial_\mu \alpha(x))\psi \,, \tag{2.5}$$

To enforce gauge invariance, this additional term must be removed by introducing an additional field to the Dirac Lagrangian with a $U(1)$ transformation which opposes the additional term. Therefore, a $U(1)$ gauge invariant electron field must include an interaction with this new field, which is the photon field. This interaction term is realised through gauge invariance as

$$\mathcal{L}_{\text{int}} = -g\bar{\psi}\gamma^{\mu}\psi A_{\mu} \, , \tag{2.6}$$

where A_{μ} transforms into $A_{\mu} - (\partial_{\mu}\alpha(x))/g$ under a local phase transformation with a coupling constant g between the electron and photon interaction. The interaction term is absorbed into the definition of a gauge covariant derivative $D_{\mu} \equiv \partial_{\mu} + igA_{\mu}$ to give the full QED Lagrangian as

$$\mathcal{L}_{\text{QED}} = i\bar{\psi}\gamma^{\mu}D_{\mu}\psi - m\bar{\psi}\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} \, . \tag{2.7}$$

Additional Dirac terms, with associated photon interactions, are included for all spin-half charged fermions of the SM.

2.3.2. Quantum Chromodynamics

The QED Lagrangian does not include self interaction terms with the photon field A_{μ}. The extension to include interactions of the form $A^{\mu}A^{\nu}\partial_{\mu}A_{\nu}$ and A^4 introduces us to the theory of the strong force — quantum chromodynamics (QCD). QCD sets out to explain the quarks, gluons and their interactions. The quark model explains the plethora of mesons and baryons as bound states of two and three quarks, respectively. These quarks have fractional electric charges of $+2/3$ and $-1/3$ with six flavours. Furthermore, to explain the baryon spectrum with the Fermi exclusion principle, one of three colour states is assigned to quarks with equivalent anti-quark colour states [20]. The minimal symmetry of which the quark colour states may be invariant to belongs to the $SU(3)$ group. Including the Dirac Lagrangian for the quark states and enforcing gauge invariance requires countering field interaction terms, similar to the interaction between the electron and photon. An $SU(3)$ rotation is represented by eight 3×3 matrices, hence eight counter terms with an independent field are required. These represent the eight gluon states $A_{\mu,a}$ with $a \in \{1, 2, \ldots, 8\}$. Therefore, the QCD Lagrangian is given by

$$\mathcal{L}_{\text{QCD}} = i\bar{\psi}_f\gamma^{\mu}D_{\mu}\psi_f - m_f\bar{\psi}_f\psi_f - \frac{1}{4}F_a^{\mu\nu}F_{\mu\nu,a} \, , \tag{2.8}$$

where ψ_f is the quark spinor state with flavour f. The covariant derivative and field strength tensor are extended to gluon equivalents as

$$D^{\mu} = \partial^{\mu} - ig_s A_a^{\mu} t_a \tag{2.9}$$

and

$$F_a^{\mu\nu} = \partial^\mu A_a^\nu - \partial^\nu A_a^\mu + g_s f_{abc} A_b^\mu A_c^\nu \,, \tag{2.10}$$

where g_s is the quark-gluon coupling strength, t_a are the generators of the $SU(3)$ group and f_{abc} appears from the non-commutative property of the generators: $[t_a, t_b] = i f_{abc} t_c$, leading to gluon self-interaction vertices [20].

As a result of the QCD interactions the strong coupling constant g_s decreases with the energy scale leading to two notable effects. The first is weak strength of the strong force at high energies, allowing perturbative calculations. Whereas at low energies the increasing strength leads to the confinement of quarks and gluons, preventing the existence of isolated color charged particles in normal conditions. This property is important in high energy particle collisions where outgoing quarks and gluons form into showers of hadrons, detected as a cone of hadronic activity [20].

2.3.3. Electroweak physics

Another realised symmetry is the invariance between up and down-type quarks. However, this symmetry is broken by the weak force where left and right-handed fermions are treated differently. Instead the left-handed up and down-type quarks are symmetric under the weak interaction, grouped into doublets with an $SU(2)$ symmetry, and similarly for right-handed antiparticles. However, right-handed fermions and left-handed antifermions do not interact with the weak force, and hence form a singlet with a $U(1)$ symmetry. Therefore, the electroweak interaction is symmetric under $SU(2) \times U(1)$ transformations. The modified covariant derivative requires three generators of the $SU(2)$ group and a single generator of the $U(1)$ group:

$$D_\mu = \partial_\mu + i\frac{g}{2} W_\mu^i \sigma^i + g' Y_W B_\mu \,, \tag{2.11}$$

where i labels the generators of the $SU(2)$ group, 2×2 Pauli matrices σ^i, with associated field W_μ^i and coupling g; and Y_W is the generator of the $U(1)$ group with the field B_μ and coupling g'. The W_μ^i and B_μ fields represent the physical photon, $W^\pm$, Z^0 bosons through the combinations [20]

$$\begin{aligned} W_\mu^\pm &= \frac{1}{\sqrt{2}} \left(W_\mu^1 \mp i W_\mu^2 \right) \,, \\ Z_\mu^0 &= \frac{1}{\sqrt{g^2 + g'^2}} \left(g' W_\mu^3 + g B_\mu \right) \quad \text{and} \\ A_\mu &= \frac{1}{\sqrt{g^2 + g'^2}} \left(g' W_\mu^3 - g B_\mu \right) \,. \end{aligned} \tag{2.12}$$

However, introducing mass terms for the W and Z bosons breaks the gauge invariance of the Lagrangian. Similarly, the broken symmetry between left and right-handed particles leads to gauge invariance breaking terms: $\hat{\psi}_L\psi_R$ and $\hat{\psi}_R\psi_L$, preventing quark mass terms. Instead, a process known as spontaneous symmetry breaking gives rise to these mass terms whilst remaining as a gauge invariant theory.

2.3.4. Spontaneous symmetry breaking

The spontaneous symmetry breaking in the electroweak sector provides mass terms for the W and Z bosons through the Higgs mechanism [14, 15, 16]. A Lorentz scalar, isospin-doublet field ϕ obeying the $U(1)$ gauge symmetry is introduced with a quartic potential

$$\mathcal{L}_{\text{Higgs}} = (D_\mu\phi)^*(D^\mu\phi) + \mu^2|\phi|^2 - \lambda|\phi|^4 \,. \tag{2.13}$$

The covariant derivate includes the weak fields for a gauge invariant Lagrangian. However, the vacuum expectation value of the field ϕ is non-zero and given by

$$\langle\phi\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\v\end{pmatrix} \tag{2.14}$$

with $v^2 = \mu^2/2\lambda$. The ϕ field can be rewritten in terms of two fields, h and χ, with no vacuum expectation values:

$$\phi = \frac{1}{\sqrt{2}}(v+h)e^{i\chi/v} \,. \tag{2.15}$$

The vacuum state is symmetric in a $U(1)$ transformation, allowing for any χ terms to be absorbed by a redefinition of the $U(1)$ gauge transformation. Expanding out the Lagrangian with this redefinition introduces h as a real scalar field with mass terms for the h, W and Z bosons of $\sqrt{2}\mu$, $gv/2$, and $(v/2)\sqrt{g^2+g'^2}$ respectively. The photon does not acquire a mass term, as required for the SM.

Since the ϕ field is an $SU(2)$ doublet a gauge invariant coupling with the left-handed lepton doublet E_L and the right-handed singlet e_R is included in the Lagrangian as

$$-\lambda_e(\bar{E}_L\phi)e_R + \text{h.c.} \,, \tag{2.16}$$

where h.c. represents the Hermitian conjugate of the terms already present and λ_e is a new dimensionless coupling constant [20]. Again, expanding out the Lagrangian with Eq. 2.15 leads to lepton mass terms with $m_e = \lambda_e v/\sqrt{2}$. Similarly, the equivalent $SU(2)$ invariant coupling for the left-handed quark doublet Q_L and right-handed singlets d_R and u_R is included in the

Lagrangian as

$$-\lambda_d(\bar{Q}_L\phi)d_R - \lambda_u\epsilon^{ab}(\bar{Q}_L\phi_b^\dagger)u_R + \text{h.c.}\,, \tag{2.17}$$

where ϵ^{ab} is a two-dimensional Levi-Civita tensor [20]. In a similar manner to the lepton fields, the quark fields acquire mass terms with $m_d = \lambda_d v/\sqrt{2}$ and $m_u = \lambda_u v/\sqrt{2}$. Additional generations of quarks introduce mixing terms between the various generations of up and down-type quarks through W boson interactions with couplings determined by the 3×3 unitary Cabibbo-Kobayashi-Maskawa (CKM) mixing matrix [23, 24].

Neutrinos do not have a right-handed chiral state and hence no couplings to the ϕ field can be included without breaking a gauge symmetry. Therefore, the neutrinos are massless in the SM. However, the neutrino flavours mixing has been measured experimentally and included ad hoc into the SM through the Pontcorvo-Maki-Nakagawa-Sakata matrix[25, 26], analogous to the CKM matrix.

In addition to the mass terms introduced through spontaneous symmetry breaking, the h field couples to the W and Z with a coupling proportional to the fourth-power of their mass, to fermions to the square of their mass, along with a trilinear and quartic self coupling.

2.4. Event simulation and generation

The complexity involved with modern particle accelerators, colliders and detectors require the use of simulated events to interpret the results and validate our understanding of the SM and beyond. To avoid needlessly generating all possible events from a collision, only the processes of interest are simulated through Monte Carlo integration of multidimensional integrals to determine differential cross sections, and hence the scattering probabilities. Additional Monte Carlo techniques allow the sampling of these distributions to generate events.

Event simulation is based on perturbation theory where higher order diagrams are neglected in calculations to avoid expensive computations. However, ignoring the entire perturbation series leads to infrared and ultraviolet divergences [27]. These are avoided through renormalisation techniques and introducing the factorisation μ_F and renormalisation μ_R scales. Physical observables should not depend on these quantities, however, limitations in the calculations lead to some dependence.

2.4.1. Hadron collider physics

In a proton-proton collider, the hard interaction cross section between parton a and parton b in each proton is factorised into [27]

$$\sigma = \sum_{a,b} \int_0^1 dx_a dx_b \int f_a(x_a, \mu_F) f_b(x_b, \mu_F) d\hat{\sigma}_{ab\to n}(\mu_F, \mu_R) \,, \tag{2.18}$$

where x_a is the proton momentum fraction associated to parton a and x_b for parton b in the other proton. The $f(x, \mu)$ are the parton distribution function which describe the momentum fraction for a particular parton within a proton. The parton cross section $\hat{\sigma}_{ab\to n}$ denotes the interaction of partons a and b to form a specific n-particle final state. This is determined from a matrix element calculation integrated over the available phase space for the outgoing particles, averaged over initial spin and colour states and scaled by the parton flux. The number of Feynman diagrams used in the matrix element calculation is determined from the accuracy required. Leading order (LO) refers to the lowest power in the strong and electroweak coupling constants, which often refers to tree-level diagrams. Next-to-leading order (NLO) generators include the next lowest power, with the inclusion of higher-order virtual and real-emission corrections. Although the electroweak coupling constant is relatively small, virtual gauge boson loops result in logarithms [28] that scale with the ratio of the centre of mass energy to the weak scale (100 GeV), and hence become large corrections at a TeV colliders and beyond. The order to which these logarithms are calculated in the perturbation series proceeds from the leading logarithm (LL) to the next-to-leading logarithm (NLL), and so on.

The choice of μ_F and μ_R are typically set by the scale of the interaction, such as the mass of an s-channel resonance, or transverse momentum of the primary outgoing particle. The scale of the interaction may also have further meaning on the initial and final-state parton showers, where care is taken to avoid double-counting particles between the matrix element calculation and parton showering. The impact of the choice of scale is determined by varying both μ_F and μ_R up and down by a factor, both uncorrelated and fully correlated. This rescaling is provided as an alternative event weight to propagate as a systematic uncertainty.

The parton distribution functions are typically determined from a fit between data and theory predictions. The evolution of these functions from low momentum scale measurements to energy frontiers is given by the DGLAP equations [29, 30, 31]. The uncertainties associated with these measurements are propagated through to the generated events.

The fixed order calculations discussed are not sufficient at encapsulating the full picture of the scattering process. Higher order effects are included through parton showering where the fragmentation of partons is simulated, down to the hadronization at 1 GeV scales. This hadronization process confines coloured particles into bound neutral final states, typically

resulting in a shower of hadrons from each energetic parton. These clusters of hadrons are grouped into jets as a physical observable in-place of the parton. Multiple jet algorithms define the set of rules to group final state particles into jets. These typically involve a requirement on a distance parameter between a pair of particles as well as the momentum assigned to each particle. The goal is to cluster particles whilst remaining insensitive to non-perturbative effects associated with collinear splitting and soft emission [32]. The most prominent algorithm used by the CMS experiment is known as the anti-k_t algorithm [33]. This algorithm iterates over all final state particles, applying a condition on the distance parameter

$$d_{ij} = \min\left(p_{ti}^{-2}, p_{tj}^{-2}\right) \frac{\Delta R_{ij}^2}{R^2} \,. \tag{2.19}$$

where p_{ti} and p_{tj} are the transverse momentum of particles i and j, ΔR_{ij}^2 is the distance between particles i and j in the rapidity-azimuthal plane and R is a chosen scale for the jets, typically set to 0.4. Particles are combined into a jet if they satisfy this condition.

2.5. Summary

The SM is a gauge theory of the fundamental particles and their interactions, unifying the electromagnetic, weak and strong forces into a single consistent theory. The matter particles consist of the coloured spin-half quarks and leptons. Through the spontaneous symmetry breaking of the electroweak sector, the quarks and charged leptons, along with the W and Z bosons, acquire a mass with the introduction of a scalar field, the Higgs field.

Simulation and generation of events within the SM framework, and possible extensions, involve Monte Carlo integration and sampling techniques to solve multidimensional integrals. The cross section of a proton-proton collision is factorised into the constituent parton interaction and parton distribution functions of the proton. After the matrix element calculation is performed, all final state partons undergo a showering procedure and hadronisation, forming into a jet of particles. These jets are clustered with the anti-k_t algorithm.

Chapter 3.

The LHC and CMS detector

3.1. Introduction

The LHC was built with the design goal of leading the energy frontier of high energy physics. This led to the successful discovery of the Higgs bosons by the ATLAS [7] and CMS [8] experiments. Moreover, BSM searches with these experiments continue to push stringent limits on the possible parameter-space of theorised new physics scenarios. Such discoveries and searches are only possible by the hugely complex detectors and the LHC accelerator complex, built and maintained by thousands of Engineers and Physicists.

3.2. The LHC

3.2.1. Accelerator Complex

Proton collisions at the LHC [34] start from the ionisation of hydrogen gas to liberate the proton nuclei from their bound atomic states. Subsequently, the protons are accelerated by the radio-frequency (RF) cavity based accelerator LINAC2 and injected into the PS Booster — four superimposed synchrotron rings accelerating protons from 50 MeV to 1.4 GeV — to enhance the intensity of the proton injected into the PS. The PS is a synchrotron with a radius of 72 m bending the proton beam into a ring with energies up to 25 GeV before being injected into the SPS. Prior to this injection the proton beam is bunched into discrete packets of protons about 4 ns long with an equal spacing of 25 ns required by the LHC for a bunch crossing rate of 40 MHz. The SPS is another synchrotron with a larger radius of 6.9 km to accommodate the acceleration of protons up to 450 GeV for the extraction by the LHC. The whole accelerator complex, shown in Fig. 3.1, allows the LHC to accelerate protons bunches up to energies of 6.5 TeV with a bunch crossing period of 25 ns and luminosities of the order of $10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ [34].

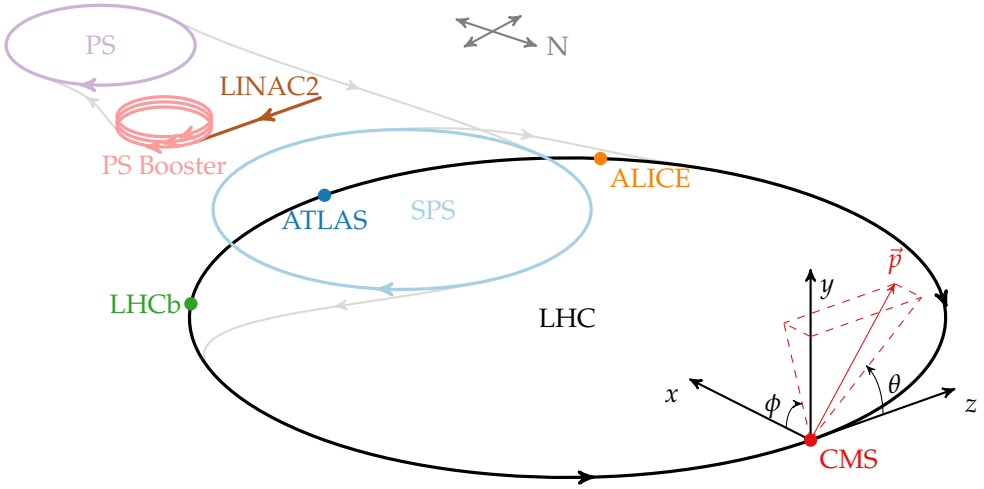


Figure 3.1.: An isometric schematic view of the LHC complex chain delivering proton-proton collisions at the four main detectors: LHCb, ATLAS, ALICE and CMS. The coordinate system used by the CMS detector is shown in both cartesian and polar coordinates. The pseudorapidity η, defined as $-\ln(\tan(\theta/2))$, typically replaces the θ angle because of its convenient properties under Lorentz transformations. This diagram is not to scale and the relative positions of the rings are guides to the physical positions.

3.2.2. LHC Main Ring

The main physics goals of the LHC project was the discovery of the Higgs boson, to observed new physics and rare SM processes. The expected rate of a particular process with a cross section σ at an integrated beam luminosity L is

$$\mathcal{N} = L\sigma \,. \tag{3.1}$$

The event rate at the LHC is directly controlled by the luminosity. For a Gaussian distributed beam, the luminosity can be given in terms of beam parameters as:

$$L = \frac{N_b^2 n_b f_{\text{rev}} \gamma_r}{4\pi\epsilon_n \beta^*} F \,, \tag{3.2}$$

where N_b is the number of particles per bunch, n_b is the number of bunches per beam, f_{rev} is the revolution frequency, γ_r is the relativistic gamma factor, ϵ_n is the normalised transverse beam emittance, β^* is the beta function at the collision point (related to the transverse size of the beam) and F is the geometric luminosity reduction factor due to the crossing angle of each beam at the crossing point.

At the LHC, the luminosity is maximised by tuning all parameters in Eq. (3.2). This necessitated the use of proton beams circulating in separate vacuum chambers and merging

at insertion points to the detectors. The beams nominally circulate in 2808 bunches with a spacing of 25 ns inside twin bore superconducting magnets — two sets of coils and beam channels within the same structure and cryostat — achieving a peak dipole field of 8.33 T to bend the beam. RF cavities apply the potential gradient to accelerate the protons from 450 GeV to 6.5 TeV. Quadrupole magnets focus the beam to suppress dispersion to maintain the beam luminosity. Additional quadrupoles focus the beam as they enter the collision sections and defocus as they exit [34].

3.3. Compact Muon Solenoid (CMS) Experiment

The conditions provided by the LHC during 2016 data taking resulted in an average number of collisions per bunch crossing of approximately 25, leading to 1000 charged particles every 25 ns bunch crossing. The CMS detector [35] was designed to account for particle multiplicities of this order by focusing on high-granularity subdetectors with good time resolutions. This required a large number of detector channels and millions of synchronised detector electronics, all with a high radiation tolerance.

In addition to the conditions provided by the LHC, the CMS detector's design was driven by four main physics requirements to achieve a broad range of precision measurements and new physics searches, namely: searching for the Higgs Boson, supersymmetric particles, new massive vector bosons, extra dimensions; precision studies of the SM; and heavy-ion physics. The physics requirements are summarised as follows:

1. **Muon identification:** ability to clearly identify muons up to $|\eta| = 2.4$ and dimuon mass resolutions of about 1% for transverse momenta of 100 GeV, with an unambiguous charge up to momenta of 1 TeV.

2. **Charged particle reconstruction:** good momentum and reconstruction efficiency, particularly for the inner tracker, allowing for the identification of secondary vertices for the triggering and offline tagging of τ's and b-jets. This requires a tracker component with a few cm of the interaction point.

3. **Electromagnetic energy resolution:** measure the diphoton and dielectron mass with a resolution of 1% for transverse momenta of 100 GeV up to $|\eta| = 2.4$. In addition, efficiently reject backgrounds from π^0 decays into two photons and maintain prompt lepton isolation at high luminosities.

4. **Transverse missing momentum and dijet mass resolution:** near complete reconstruction of the proton interaction to measure the transverse missing momentum. This requires a

hermetic detector with hadron calorimeters extending up to $|\eta| = 5.0$ and fine component segmentation in the η-ϕ plane of less than 0.1×0.1.

The full design of the CMS detector is shown in Fig. 3.2 which achieves the requirements outlined above. The main distinguishing feature of the CMS detector is a 3.8 T superconducting solenoid magnet enclosing the tracking and calorimetry subdetectors with an iron return yoke to maintain a 2 T field outside the solenoid for the muon subdetectors. All subdetectors are segmented into various regions in η known as the barrel (covering the central η range) and endcap (covering up to $|\eta| = 3.0$, with a slight overlap with the barrel). The hadronic calorimeter has an additional segment in the forward region, (covering $3 < |\eta| < 5$) as required for the missing momentum reconstruction. Further subdetector dependent segmentation is done within the barrel, endcap and forward regions; discussed in further detail in the following subsections. This gives the CMS detector a total length of 21.6 m and a diameter of 14.6 m and weighing 12 500 tons.

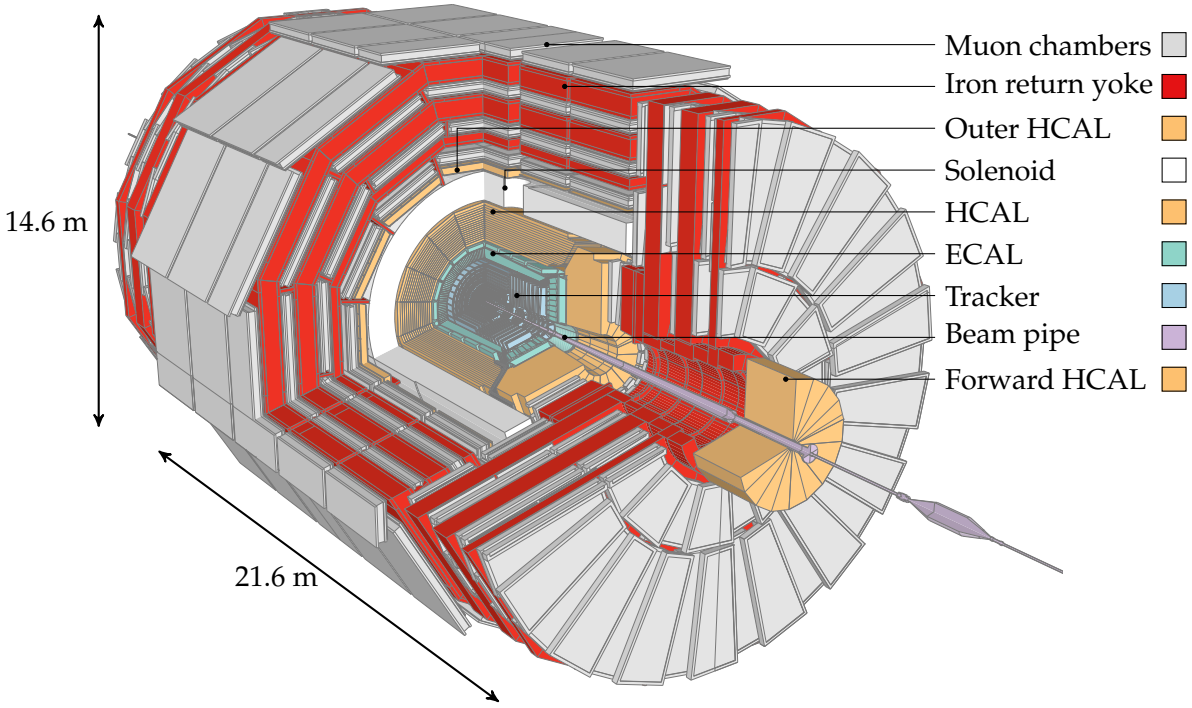


Figure 3.2.: A schematic view of the CMS detector and subdetectors with a section removed to view the internal components. Acronyms are explained in the text. The diagram was produced using the tools outlined in [36].

3.3.1. Silicon Tracker

The inner most subdetector, the silicon tracker [35], measures the position of charged particles from ionisation deposits (known as hits) across a silicon-based reverse-biased p-n junction. A

collection of hits is used to reconstruct the curved trajectory, with a radius of curvature r, of the particle with a charge q within the magnetic field B to determine the momentum of the particle transverse to the field, p_{T}, given by:

$$p_{\mathrm{T}} = rqB \,. \tag{3.3}$$

A collection of tracks may intersect at a common point leading to the reconstruction of the primary interaction, pileup and secondary vertices. All these vertices are typically a few centimetres apart, therefore, the tracker components are placed as near as 4.4 cm to the beam pipe. The silicon tracker, shown in Fig. 3.3, consists of two main components: pixels and strips. The pixels measure 3-dimensional positional coordinates of hits for environments of high particle flux (up to 10 million particles per second) near the beam. Pixel modules are arranged into three layers placed in the barrel region (BPIX) at radii of 4.4 cm, 7.3 cm and 10.2 cm with a length of 53 cm, centred on the z-axis. The pixel modules are staggered in the ϕ-direction to achieve an overlapping configuration. Two additional pixel layers are placed in both endcap regions (FPIX) with a turbine-like geometry with the blades rotated by 20° for overlapping blades. The overlap of pixel modules ensures the charged particle passes through at least one module in each layer. If two modules are hit the charge sharing between these modules provides additional positional information on the hit leading to improved position resolutions.

Silicon strips are used further away from the beam pipe as the particle flux reduces. Strips provide an accurate 2-dimensional position for hits, hence layers are typically oriented in a stereo-configuration[1] for the full 3-dimensional position. The silicon strips are separated into an inner and outer component. The inner component is further split by the barrel and endcap regions. The tracker inner barrel (TIB) component consists of four layers positioned at a radius of 20–55 cm and covering $|z| < 65$ cm with cell sizes of 10 cm $\times$ 80 µm. The first two layers use the stereo-configuration with an angle of 100 mrad. The tracker outer barrel (TOB) component has six layers extending the coverage to $|z| < 124$ cm and 55 cm $< r <$ 116 cm. Again the first two layers are placed in a stereo-configuration with an angle of 100 mrad. The inner endcap (TID) component provides coverage for the gap between the TIB and TOB with three small disks oriented perpendicular to the z-axis to avoid excessive shallow track crossing angles. The first two disks have a stereo-configuration. The tracker outer endcap (TEC) consists of nine disks extending into 124 cm $< |z| <$ 282 cm with stereo-configuration for the first two and the fifth disks [37]. The entirety of the silicon detector consists of 66 million pixels and 9.6 million silicon strips with near hermetic coverage up to $|\eta| = 2.4$.

To process the data from the pixels and strips an on-detector chip processes and buffers the analogue signal while waiting on a decision to accept or reject the data (detailed in Sec. 3.3.6).

[1] A stereo-configuration has layers rotated at an angle with respect to each other to form a crosshatch pattern.

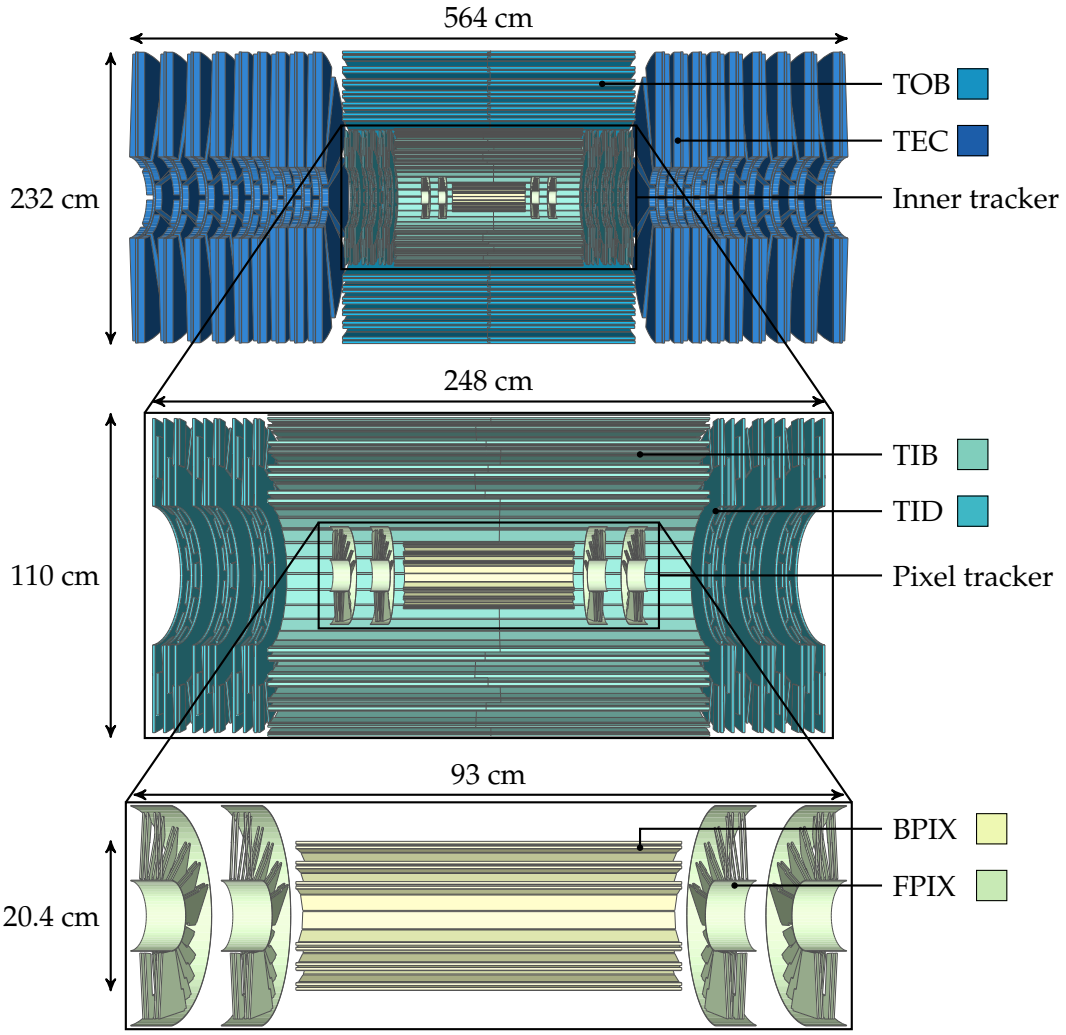


Figure 3.3.: A cross-sectional view, in the y-z plane, of the silicon tracker. The full tracker is shown in the upper part of the figure with a zoomed inset of the inner tracker in the central section and a further zoomed inset of the pixel tracker in the lowest part. The acronyms labelling various sub-components are defined in the text. The diagram was produced using the tools outlined in [36].

Upon an accept signal being received the data is transmitted over optical links to off-detector chips for further processing, such as digitisation and data formatting.

3.3.2. ECAL

The ECAL [35] is a homogeneous calorimeter made of lead tungstate ($PbWO_4$) crystals designed to measure electromagnetic deposits (typically electrons and photons) up to $|\eta| = 3.0$. The choice of lead tungstate scintillating crystals was driven by their short radiation lengths,

X_0 (0.89 cm), and Moliere radius[2] (2.2 cm) with fast responses (80% of the light emitted is within 25 ns) and have a satisfactory level of radiation resistance. These properties allow a compact design fully enclosed within the solenoid with sufficient containment of electron and photon showers both laterally and along the length of the crystals. The drawbacks of the crystal include low light yield (30 photon/MeV) requiring a coupling to photodetectors with intrinsic gains operable in the magnetic field, and deterioration from continuous irradiation monitored throughout.

The ECAL is divided into the barrel section (EB) and the two endcap sections enclosing the forward regions consisting of the endcap crystals (EE) and the preshower (ES), diagrammatically shown in Fig. 3.4. The EB has an inner radius of 129 cm with the crystals titled at a 3° angle in ϕ and η with respect to the nominal vertex position to avoid particles travelling fully between two crystals. The crystals cover a 0.0174 (1°) angle in ϕ and η with the front face cross-section of $22 \times 22\,\text{mm}^2$ and a length of 230 mm corresponding to $25.8X_0$. A submodule consists of 5 pairs of crystals held together by a thin-walled glass-fibre structure. The submodules are assembled into groups of 40–50 including the readout electronics and thermal screen. Four modules are placed side-by-side to create a supermodule that covers half the length of the barrel and 20° in ϕ, two halves of 18 supermodules make the EB, covering a range of up to $|\eta| = 1.479$. The front face of the EE are placed at $|z| = 314$ cm from the nominal vertex and extends the ECAL coverage in the range $1.479 < |\eta| < 3.0$. Each endcap consists of two dees — semi-circular aluminium plates with structural support units — with crystals arranged in an x-y grid of 5×5 groups known as supercrystals. Partial supercrystals are placed on the inner and outer boundaries of the EE. Each EE crystal has a frontal cross-section of $28.6 \times 28.6\,\text{mm}^2$ and a length of 220 mm (corresponding to $24.7X_0$). The ES is placed in front of the EE, covering the range $1.653 < |\eta| < 2.6$, and consists of two planes of silicon strip detectors behind disks of lead absorbers at depths of $2X_0$ and $3X_0$, respectively. The purpose of the ES is to initiate the electron or photon showers with granular tracking to distinguish the individual photons from the decay of $\pi^0 \to \gamma\gamma$ for rejection and to provide directional information on the electron or photon. A preshower placed in barrel was rejected in favour of accommodating a larger tracker.

The on-detector readout electronics for the ECAL amplify and shape the signal from the photodiodes and digitise at the bunch crossing rate of 40 MHz. This data is buffered until an accept signal is received to transfer the data to off-detector electronics for further processing. In addition to this processing, the on-detector electronics run fast algorithms on data from the 5×5 clusters of crystals to aid in the accept decision, detailed in Sec. 3.3.6.

[2] A radiation length is the distance travelled by a particle through a material where, on average, the particle has lost 63.2% of it's original energy. Similarly, the Moliere radius is the radius of a cylinder which captures, on average, 90% of the particle's shower, transverse to the initial direction of travel.

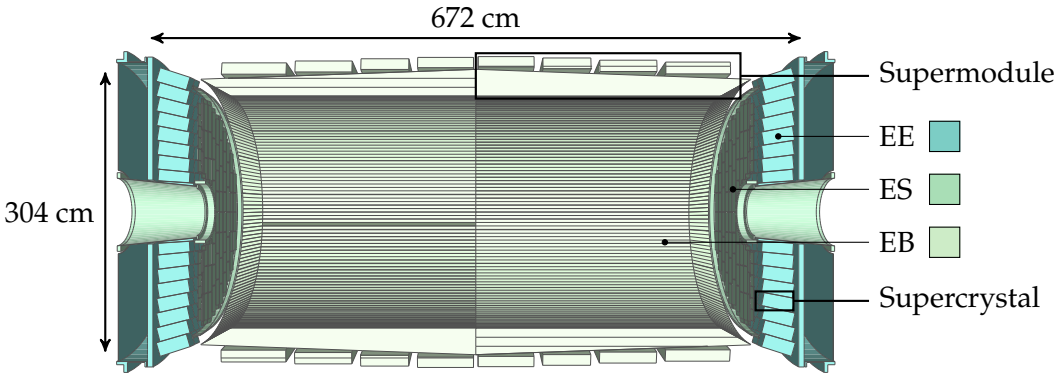


Figure 3.4.: Front cross-sectional view of the ECAL subdetector. Individual crystals are not shown, instead modules and supercrystals are displayed. The diagram was produced using the tools outlined in [36].

The energy of a shower determined from photodiode signals is corrected to reproduce the true energy of the particle initiating the shower. Several factors result in the need for a correction factor, namely: an imperfect clustering algorithm, energy loss from Bremsstrahlung, shower containment variations, imperfections within the crystals, crystal-to-crystal variations, misalignment of the crystals and limitations of the photodiodes. The reconstructed energy of a shower, E, is given by

$$E = G\mathcal{F}\sum_i c_i A_i \,, \tag{3.4}$$

where G is the global absolute scale in converting GeV, determined from in situ measurements of photons in $Z \to \mu\mu\gamma$. The function $\mathcal{F}$ is a correction dependent on the type of particle; its position and momentum; and the clustering algorithm used, determined from simulation validated by test beam and in situ measurements of $Z \to ee$ and $Z \to \mu\mu\gamma$. The product of c_i — the intercalibration coefficients obtained from laboratory measurements, test beam precalibrations, comic ray measurements and in situ $W \to e\nu$ measurements, and $\pi^0 \to \gamma\gamma$ and $\eta \to \gamma\gamma$ mass reconstruction — and A_i, the signal amplitude, are summed over each clustered crystal labelled by i.

The corrected energy resolution of a supermodule is measured in a test beam by fitting a gaussian function to the reconstructed energy distributions and parameterised as

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \,, \tag{3.5}$$

where $S = 2.8\,\%$ is a stochastic term, $N = 12\,\%$ is a noise term and $C = 0.3\,\%$ a constant term. The total energy resolution ranges between 1.50–0.35 % for 10–250 GeV, beyond which the constant term dominates at 0.3 % [38].

3.3.3. HCAL

The HCAL [35] is designed to provide full coverage of hadronic activity from a proton-proton interaction, built from four distinct subdetectors, shown in Fig. 3.5 found in the barrel (HB), endcaps (HE), beyond the solenoid (HO) and far in the forward region (HF) [39]. The HB and HE subdetectors are sampling calorimeters, fully immersed within the 3.8 T field, with a brass absorber and plastic scintillator tiles in a sampling ratio[3] of about 3 : 40. The HB and HE are separated by a gap at about 57° designed to be non-projective with the interaction point to avoid particles traversing an uninstrumented gap.

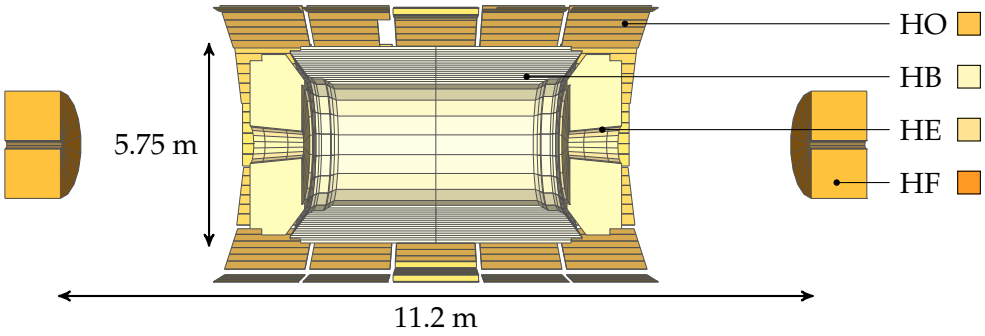


Figure 3.5.: Front cross-sectional view of the four HCAL subdetectors. The magnet between the HB and HO, along with the other subdetectors have been omitted. The HF is placed beyond the muon endcaps extending the coverage of the HCAL in the forward region. The diagram was produced using the tools outlined in [36].

The HB covers the central region up to $|\eta| = 1.4$ with a inner radius of 1777 mm and outer radius of 2876.5 mm, made of two half-barrels in z of 18 identical wedges covering 20° in ϕ with further segmented into 5° sectors. The wedges consist of towers: an alternating stack, parallel to the beam axis, of 17 active plastic scintillator tiles and brass plates, with the first and last absorber replaced by steel for structural strength, forming a tower. Constraints on the space between the ECAL and solenoid restricted the HB to 5.7 hadronic interaction lengths at $\eta = 0$ and up to 9 at $\eta = 1.4$, necessitating the HO placed beyond the solenoid to improve the longitudinal containment of hadronic showers. Scintillation photons are produced from the interaction of hadronic showers with the active scintillation tile, collected by wavelength shifting (WLS) fibres embedded within the plastic spliced to clear fibres running along the tower. The optical signal is summed across the depth of a tower in three groups: the first, the next 4 and the final 12 scintillator tiles ($\{1 + 4 + 12\}$). The clear fibres are coupled to by Silicon photomultiplers (SiPM). Full scintillator depth-segmentation would allow hadronic shower discrimination, however, the segmentation is limited by signal-to-noise, number of readout

[3] The sampling ratio is the depth ratio of sampling to absorbing elements.

channels, power and cooling constraints. Therefore, a balance of a three-depth configuration is in-place [39].

The HE subdetector in the endcap provides coverage in the region $1.3 < |\eta| < 3.0$, overlapping the HB for improved coverage across the uninstrumented gap. The absorber-scintillator towers follow the same design as the HB towers with wedges covering 20° in ϕ, segmented to 5° for towers in $|\eta| < 1.74$. The towers in $|\eta| > 1.74$ are instead segmented into 10° to accommodate the bending radius of the WLS fibres. The η segmentation also increases from 0.087 for the towers at $|\eta| < 1.74$ to 0.09–0.35 this η range. The coarse azimuthal and polar segmentation allows for a finer depth segmentation with configurations of four or five-depths in-place, grouped by $\{1+2+3+5+7\}$. The HE provides sufficient material interaction lengths hence requires no outer calorimeter [39].

The HO consists of plastic scintillator tiles placed beyond the magnet to catch energy leakage from the HB. Instead of brass, the solenoid and iron return yoke act as the absorber extending the hadronic interaction lengths by 2–3. Two scintillator layers are placed at $\eta = 0$, either side of a 19.5 cm thick piece of iron, where the interaction length of the HB is the lowest. WLS fibres fed to SiPM detectors readout the signal from the HO. The segmentation follows that of the muon system with 30° wedges divided into 5° sectors in ϕ to overlap with the HB segmentation and five η sectors up to $|\eta| = 1.2$. Partial plates are placed around the cryogenics and power cables for the magnet.

High intensity radiation falls upon the HF, placed at $|z| = 11.2$ m downstream of the interaction point. Consequently, an alternative design of steel absorbers embedded with radiation hard quartz fibres to produce Cherenkov radiation from hadronic showers, detected by multi-anode photomultiplier tubes [39]. The steel absorbers are segmented in ϕ by 20° wedges subdivided into two 10° sectors, apart from the two towers nearest to the beam pipe which retain the 20° segmentation. The steel absorber towers are placed parallel to the beam axis with an η segmentation of approximately 0.175, varying with z. The fibres also run parallel to the beam axis placed 5 mm apart in an x–y grid with alternating long (full length of 10 interaction lengths) and short (starting after 1.2 interaction lengths). The two fibre lengths provide shower depth information similarly to the depth segmentation in the HB and HE [40].

3.3.4. Solenoid Magnet

The muon identification requirement of the CMS detector requires a large bending power provided by the solenoid magnet with a field strength of 3.8 T. Such a high field is obtained by a superconducting magnet, shown in Fig. 3.6 maintained at operating temperatures by superfluid helium cooling. To extend the bending power the return field is focused by an iron return yoke spread throughout the muon chambers to about 2 T.

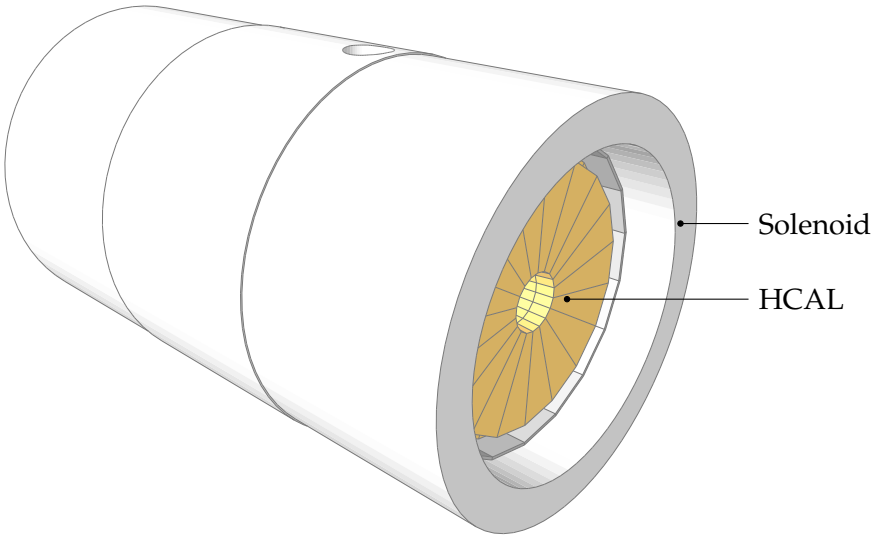


Figure 3.6.: Isometric view of the solenoid and HCAL with the muon chambers and HO removed. The diagram was produced using the tools outlined in [36].

3.3.5. Muon Chambers

Muons from proton-proton collision, or secondary decays, travel through the detector leaving tracks in the tracker and energy deposits in the calorimetry system, however, most muons continue unscathed. The muon system [35] is designed to extend the 3-dimensional position measurements of these muons several metres away from the interaction point. Three types of gaseous detectors are used with the design driven by the large area to cover and different radiation environments, particularly between the barrel and endcap.

The layout of the three gaseous detectors in the CMS muon system is shown in Fig. 3.7. In the barrel, anode wires centred in gas chambers, known as drift tubes (DT), collects the ionisation from passing charged particles for 2-dimensional position measurements. DTs perform poorly in the larger background and muon rate environment of the endcaps, hence cathode stripe chambers (CSC) are in-place. A charged particle ionises the atoms of the gas volume of a CSC with anode wires to collect the liberated electrons and perpendicular copper cathode strips where a mirror charge is induced allowing a full 3-dimensional position measurement. Resistive plate chambers (RPC) accompany both the DTs and CSCs to provide fast responses with good timing resolution at the cost of coarser positional resolution. The gas volume within an RPC is ionised by charged particles with the liberated electrons swept by an electric field induced from a large potential difference across plastic tiles either side of the gas volume. The electrons are collected by metal strips behind the plastic and the signal read out. The CMS RPCs consist of a double-gap separating 2 mm thick bakelite tiles by a 2 mm gas gap.

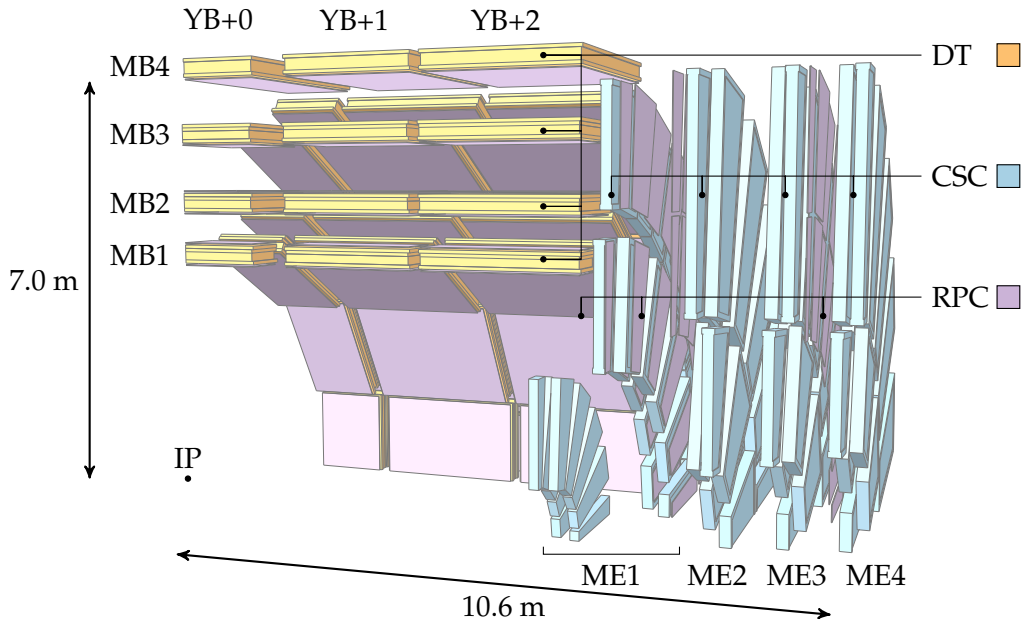


Figure 3.7.: Isometric view of the rear-top-right octant of the muon system, omitting the iron return yoke. The interaction point (IP) is shown in the lower left. The division of the muon system into stations (MB), wheels (YB) and disks (ME) are discussed in the text. The diagram was produced using the tools outlined in [36].

The muon barrel section (MB) cover $|\eta| < 1.2$ and consists of four concentric stations labelled MB1–MB4 position at about 4.0, 4.9, 5.9 and 7.0 m away from the beam axis. Each station is subdivided into 12 sectors in ϕ, with subsequent sectors overlapping to avoid uninstrumented gaps. Two partial sectors are placed around the magnet cryogenic lines and power cables. The MB is divided into five wheels along η labelled YB $-$ 2 to YB $+$ 2. Stations MB1 and MB2 consist of two RPCs placed either side of a DT chamber whereas MB3 and MB4 have one RPC placed on the inside of a DT chamber. These DT chambers consist of four layers of drift tubes, known as a superlayer (SL). Two SLs have their anode oriented along the beam axis to measure the r-ϕ coordinates of muon tracks, SL_ϕ, and the third SL rotated by 90° measures the z coordinate (omitted in MB4), SL_θ. The SLs are separated by a honeycomb (HC) structure in the order: SL_ϕ, HC, SL_θ, SL_ϕ; where the final SL_ϕ is omitted in MB4.

The muon endcaps (ME) enclose the solenoid both ends of the CMS detector with sufficient overlap with the barrel: $0.9 < |\eta| < 2.4$. Both endcaps are divided into four disks, labelled ME1–ME4, perpendicular to the beam axis, subdivided into two concentric rings of 36 chambers (apart from ME2–ME4's innermost rings with 18 chambers each). The chambers within a disk overlap to avoid uninstrumented gaps. Each CSC is trapezoidal in shape with 6 gas gaps to measure the z position, cathode strips projecting radially outwards to measure the ϕ position and anode wires closely following the ϕ direction to measure the radial position from the beam axis. RPCs accompany each CSC up to $|\eta| = 2.1$.

3.3.6. Hardware Trigger

The 40 MHz bunch crossing frequency provided by the LHC at 2016 luminosities yields on average 25 proton-proton interactions per crossing leading to about 1 billion events per second. Storage capacities restrict full detector archiving of about 100 crossings per second. Therefore, an online[4] triggering system applies a rejection factor of the order of 1 in 10 thousand with minimal impact to the broad physics program. This system is realised in two-levels: the first level a hardware trigger (also known as the Level-1 trigger, L1T) [41, 42] with custom programmable processors, and a second level a software trigger [43] using general-purpose commercially available processors (discussed in Sec. 3.3.7).

The detector data for each bunch crossing is held in on-detector buffers whilst waiting on a decision from the L1T to further process the crossing with the software trigger. The L1T processes up to 128 bunch crossings simultaneously, allowing a total time of 3.2 µs on reaching a decision. The allocated time includes data transit time to the electronics, located in an adjacent cavern to the detector for radiation protection, which reduces the total processing time to about 1 µs. The L1T is sent a new bunch crossing to process every 25 ns and after the allocated time a decision must be made, also every 25 ns. These restrictions prevent data fetching and result in most operations consisting of simple arithmetic or querying memory in lookup tables[5]. Furthermore, information from the tracker and HO is omitted from the L1T decision owing to their slow response. The other subdetectors provide coarse information, to reduce processing time, and are divided into the calorimeter and muon triggers with the candidates sent to the global trigger for the final decision.

The L1T decision is based on reconstruction of individual objects such as photons, electrons, muons, jets, transverse energy sum or missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$). The physics requirements can be summarised, with examples of targeted physics, as follows:

1. Triggering of events with a single lepton of $|\eta| < 2.4$ and $p_{\mathrm{T}} > 40\,\mathrm{GeV}$ with an efficiency above 95% to target events with W decays.
2. Dilepton triggers with the same kinematical requirements and an efficiency above 95% to collect Z decays.
3. Single photon and diphoton triggers with similar thresholds for processes involving a Higgs boson.
4. $E_{\mathrm{T}}^{\mathrm{miss}}$ threshold about 100 GeV for sensitivity to new weakly interacting particles.

[4] Online refers to tasks or processes performed in real-time with strict timing constraints whilst data taking. Conversely, offline refers to processing after the storage of data without strict timing constraints.

[5] A lookup table stores expensive computations in a pre-computed table for significant savings in processing time and complexity.

Calorimeter trigger

ECAL (EB and EE) and HCAL (HB, HE and HF) signals are clustered into trigger towers: 5×5 arrays of crystals in the EB and EE; single physical tower for the HE and HB up to $|\eta| = 1.74$ and twice the number of physical towers in ϕ beyond this, whereby the energy of the physical tower is divided equally between the trigger towers; and physical towers grouped into pairs in ϕ and threes in η for the HF. The transverse energy of each trigger tower is determined by on-detector electronics and the result is sent to the L1T electronics along with flexible pass-fail information to, for example, encode depth structure. This information is further processed with each bunch crossing sent to one of ten processing nodes (with two redundant nodes) to scan across the full η–ϕ plane and identify physics objects: jets, τs, electrons and photons; as well as calculating the transverse energy sum and $E_{\mathrm{T}}^{\mathrm{miss}}$. The processing nodes run concurrently with a new bunch crossing occupying the first node upon completion of the previous crossing. This is known as a time multiplexed system where the output must be reassembled into the original bunch crossing order (de-multiplexed) for the global trigger.

Muon trigger

The muon trigger system primarily consists of track finding from successive chamber hits by a muon. This is divided into regional track finders using a combination of the DTs and RPCs in the barrel (up to $|\eta| = 0.8$), CSCs and DTs in the endcaps (from $|\eta| = 1.25$) and all three chambers in the overlap region between the barrel and endcaps. Additionally, the system is tasked with assigning hits to the correct bunch crossing, achieved by the high timing precision of the RPCs and of the CSC's anode wire response. Timing is also performed with the DTs, albeit with a poor resolution due to long drift times of 400 ns. The tracks from these regional triggers is sent to the global muon trigger to merge and remove duplicate muon candidates, along with combining the isolation information from the calorimeter trigger, and finally sorts the muons for the global trigger.

Global trigger

The trigger objects and variables received from the calorimeter and muon triggers is collected at the global trigger for synchronisation and correlation of relevant trigger candidates. The union of a set of requirements on the trigger objects is determined and forms the basis of the L1T decision. However, evolving conditions and the collection of events from high cross-sectional processes requires the global trigger to reject random events to reduce the trigger rate. This process is known as prescaling and reduces the trigger efficiency by the prescale factor: e.g. a prescale factor of 10 will accept 1 out of 10 events regardless of decision formed from the

trigger candidates resulting in a rate reduction of a factor of 10 and a drop efficiency of 90%. After the prescaling is applied the L1T decision is communicated to the data buffers to discard or transmit the bunch crossing to the software trigger.

3.3.7. Software Trigger

The software trigger (also known as the high-level trigger, HLT) accepts the full detector granularity and all subdetector information from the on-detector readout. All bunch crossings filtered by the L1T are distributed among processing nodes in a processor farm, each node running a copy of the software code to reconstruct and reject events. To meet the timing constraint, the HLT performs a partial reconstruction of the event at first, guided by the trigger objects from the L1T, gradually performing the full reconstruction until the event fails the trigger criteria. Furthermore, the HLT must be efficient at selecting physics objects of interest and inclusive to avoid rejecting exotic physics. Therefore, the HLT reconstruction is closely aligned with offline reconstruction (discussed in Chap. 4). However, calibrations and other run conditions available for offline analysis cannot be used. After accepting an event, the data is archived and made readily available for further analysis by the computing grid.

3.3.8. Worldwide LHC Computing Grid (WLCG)

Despite the data reduction achieved by the CMS trigger system, a vast amount of data was accepted and requires prompt full reconstruction with possible revisions in light of improved object reconstruction, calibrations or code errors. Moreover, simulations of proton-proton collisions within the SM and possible BSM physics, including a simulation of the detector, the trigger and pileup with full reconstruction demands a significant amount of processing power. To achieve this the WLCG was established [44]. The WLCG is a highly distributed computing cluster divided into tiers situated at: the main CERN site, national laboratories and Universities worldwide. The computing provides the resources required for data reconstruction and simulations, and supports in the design, evaluation, construction and calibration of the detector and future upgrades.

3.4. Summary

The LHC provides proton-proton bunch crossings at a rate of 40 MHz, a centre-of-mass energy of 13 TeV, and the CMS detector centred on one of these crossings. The CMS detector is designed to detect and reconstruct the position and momenta of outgoing particles from the collisions whilst mitigating backgrounds such as noise and pileup, and optimising the

geometry and readout channels with modern technology. The main feature is a 3.8 T solenoid enclosing the silicon tracker and calorimetry with an outer muon system. A two-level trigger system reduces the data output by selecting events of interest for a manageable output rate for archiving. The immense processing power required for further analysis of the data and production of simulations is mediated by the WLCG.

Chapter 4.

Event reconstruction

4.1. Introduction

The reconstruction and identification of particles with the CMS detector starts from the hits of ionising particles and energy deposits of particles showers. The granularity of the CMS detector allows muons, electrons and photons, and charged and neutral hadrons to be distinguished. Instead of simply detecting, reconstructing and identifying particles with their respective dedicated system, information from all aspects of the detector are correlated to extract the best estimate for the reconstructed set of particles. Each particle is distinguished individually by a technique known as particle flow (PF).

The initial step in reconstruction, prior to the PF method, consists of forming charged particle tracks from hits in the silicon tracker, tracks in the muon chambers and calorimeter clusters from energy deposits in the calorimetry system. The basic building blocks form the input into the PF algorithm which attempts to link the various blocks into a coherent picture of a particle candidate. For example, tracks are linked to an ECAL deposit to form an electron candidate. These are further linked to ECAL deposits consistent with Bremsstrahlung photons from the original electron track. These secondary photons may undergo further interactions resulting in displaced electrons tracks also linked to the primary electron. Furthermore, the reconstruction of all particles results in a simple $p_{\mathrm{T}}^{\mathrm{miss}}$ calculation from the negative vector sum of all particle momenta.

This chapter will describe in detail the process of tracking, clustering and the PF method in the full reconstruction of a proton-proton interaction.

4.2. Tracking

Charged particles originating from the interaction point travel outward through the magnetic field in helical trajectories. The hits in the silicon tracker are reconstructed into the track by a

fitting procedure known as the combinatorial track finder. Hits are formed within the pixel tracker by clustering deposits, above a certain threshold, in adjacent pixels. The position of the hit for single pixel clusters is taken at the centre of the pixel while multi-pixel clusters have a charge-weighted position, correcting for the drift of electrons in the magnetic field. A more sophisticated, albeit, computationally demanding method is used for the final track fit which involves a χ^2-fit of expected charge deposits from a large number of simulated particles through the pixels to the observed charge deposit, accounting for pixel deterioration. Strip tracker hits are formed from clusters of adjacent strips with charge deposits exceeding a noise-level. The hit position is determined from a charge-weighted average of the cluster, correcting for the electron drift in the magnetic field and known inefficiencies with the charge collection. All hits have an associated uncertainty propagated to the track fitting procedure [45].

The combinatorial track finder involves a series of Kalman filters (KF) [46] — a recursive parameter estimator. A Kalman filter starts with an initial state (a seed), typically from a guess or estimation, and iterates across measurements updating the initial state parameters by comparing the prediction to the observations. For track fitting, the fit is performed to five parameters which describe points along a helical track, known as the perigee parameterisation:

$$\left(\frac{q}{p_{\mathrm{T}}}, x_t, y_t, \theta_t, \phi_t\right) \tag{4.1}$$

where q/p_{T} is the ratio of the charge to track transverse momentum to give the signed curvature scale of the trajectory. The parameters x_t and y_t are the distances from the interaction point to a point along the trajectory in the $\vec{x}_t$ and $\vec{y}_t$ axis, shown in Fig. 4.1 along with the definitions of the angles θ_t and ϕ_t.

The trajectory building with a Kalman filter starts with the seed generation stage by combining a set of pixel hits, where the resolution is the greatest and occupancy the lowest. Each iteration of the Kalman filter extends to the next tracker layer where each hit candidate, within a χ^2 window, is kept to form independent track candidates. Fake tracks will typically have a layer where no hits are within the window and can be dropped. Further checks are performed on the consistency of hits with the p_{T} of the track, in addition to vertex constraints. Finally, a cleaning process assigns unique tracks to a seed and vice versa. After determining the full set of tracks, the Kalman filter procedure is repeated with the precise hit reconstruction, accounting for inter-layer material and field inhomogeneities, with inflated uncertainties to avoid biasing the initial result. Furthermore, the Kalman filter is repeated, seeded by hits from the outermost layer iterating inwards. The best fit tracks are taken from an average of the Kalman filter starting from the inner layers and the reverse.

The trajectory building is performed ten times with different seeds to target various classes of tracks with hits associated with candidate tracks removed from subsequent iterations:

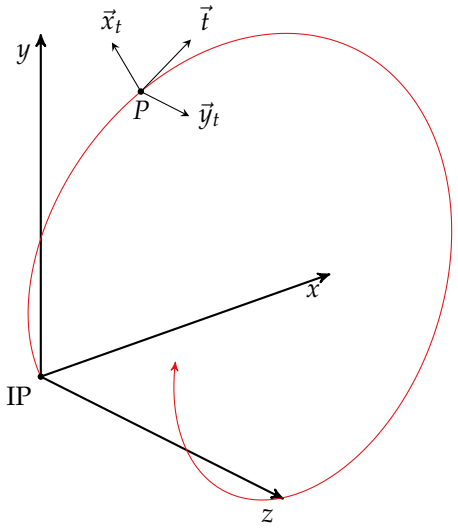


Figure 4.1.: Diagram of a track (shown in red) from the interaction point (IP) with the conventional CMS coordinate system as well as a local coordinate system along the trajectory, an example of which is shown at point P where $\vec{t}$ is the tangential vector, $\vec{x}_t$ is perpendicular to both the z-axis and $\vec{t}$ and $\vec{y}_t$ is the remaining vector to create a right-handed coordinate system. The angle between the x-axis and the projection of $\vec{t}$ onto the x-y plane is ϕ_t. Similarly, the angle between the y-axis and the projection of $\vec{t}$ onto the y-z plane is θ_t.

- Prompt and high p_{T} tracks, and b-hadron decays are targeted by the first three iterations, seeded by three hits in the pixel detector with constraints on the track p_{T} and distance of nearest approach to the beam axis.
- The fourth and fifth iteration targets tracks with one or two missing hits in the pixel detector as a result of particle interactions and decays, or detector inefficiencies. These are seeded by two pixel hits or a combination of three pixel and strips hits.
- The next two iterations are seeded by strip hits to target displaced tracks.
- The eighth iteration targets high p_{T} jets with merged constituent tracks, seeded by pairs of hits in the pixel and strips, with the initial track compatible with a high energy calorimeter deposit.
- The final two iterations are designed to track muons missed by prior iterations by using information from the muon chambers to determine the seeds.

After all iterations are complete a full set of tracks within the tracker volume are defined, known as inner tracks.

4.2.1. Electron tracking

Electron tracks may be missed by the combinatorial track finder as a result of significant Bremsstrahlung and high energy photon emission. To recover these tracks another fitting procedure is performed, based on a gaussian-sum filter (GSF). The GSF method models hits by a weighted sum of gaussian distributions for a more accurate representation of hits with sudden and substantial radiative losses. The results of the fit to the observed hits is fed into a boosted decision tree[1] (BDT) to optimise electron track reconstruction efficiency while minimising fake tracks. This procedure is also effective at tracking electron-positron pairs from tracker converted photons.

4.2.2. Muon tracking

A track fitting procedure is performed with hits in the muon chambers to form muon track candidates. Various quality muons are defined based on the inputs and results of the track fitting:

- **standalone muon:** track fitting seeds are formed of track segments from hits within the DT or CSC detectors. The fit itself uses hits within all muon chambers.
- **global muon:** a standalone muon track matched to an inner track.
- **tracker muon:** one muon segment matched to a unique inner track with requirements on the muon's momentum and its compatibility with an origin at the beam axis.

4.3. Vertex reconstruction

The conditions provided by the LHC during 2016 data-taking yielded on average about 25 proton-proton interactions per bunch crossing. The low cross section of hard interactions between protons typically results in a single interaction point of interest, known as the primary vertex. The other vertices are classified as pileup which result in additional particles leading to misreconstruction of the primary event. Therefore, the reconstruction of these vertices is critical in mitigating the impact of pileup interactions, particularly in determining the $p_{\mathrm{T}}^{\mathrm{miss}}$ of the primary event.

The vertex reconstruction is performed by a fit to a selection of tracks. The collection of tracks are selected by their track quality parameters (sufficient number of hits and a satisfactory

[1] A boosted decision tree is a form of supervised artificial learning with an output from a weighted linear sum of many decision trees, determined by optimising a function of the truth and prediction with a regularisation term to penalise tree depth.

track fit χ^2) and consistency with the beam spot, a 3-dimensional profile of the luminous region determined from an average over multiple bunch crossings. From these track collections a set of vertex seeds are found by searching for convergence points of multiple tracks [47]. These seeds initialise the adaptive vertex fitter [48], a similar procedure to the Kalman filter although all tracks are considered in turn weighted by the compatibility of the track's nearest approach to the vertex seed. This is followed by updating the vertex position from a method of least squares to the weighted tracks. This procedure is iterated until convergence resulting in a collection of vertex candidates. The primary vertex is selected from this collection by the greatest $\sum p_{\mathrm{T}}^2$ over all outgoing tracks [49].

4.4. Calorimeter clustering

A clustering algorithm groups energy deposits within the calorimeter towers to encapsulate the whole extent of a particle shower, or multiple overlapping showers. These clusters are sent to the PF algorithm to link the cluster with a track, or multiple tracks. The clustering is performed separately in the ECAL barrel and endcaps, HCAL barrel and endcaps, and the two preshower layers. No clustering algorithm is necessary in the HF since forward jets typically have large momenta and hence collimated showers. Seeds initialise the clustering and are identified as cells with an energy above a threshold and greater than the nearest neighbouring cells. Topological clusters are formed by continually merging neighbouring cells, with an energy twice the noise level, into the cluster. Individual clusters within a topologial cluster are distinguished by a gaussian-mixture model: a topological cluster in M cells arise from N gaussian energy deposits. The deposits are parameterised by the amplitude of the i-enumerated gaussian, A_i; the position of the gaussian centre in η-ϕ, $\vec{\mu}_i$; and the width σ of the gaussian, fixed to values assigned for each calorimeter system. The initial values are taken from the cell seeding the topological cluster and an expected energy fraction measured in cell j is determined from

$$f_{ji} = \frac{A_i \exp\left(-(\vec{c}_i - \vec{\mu}_i)^2/(2\sigma^2)\right)}{\sum_{k=1}^{N} \exp\left(-(\vec{c}_j - \vec{\mu}_k)^2/(2\sigma^2)\right)}, \tag{4.2}$$

where $\vec{c}_i$ is the position of element i. The parameters are estimated from an analytical maximum likelihood fit with

$$A_i = \sum_{j=1}^{M} f_{ji} E_j \tag{4.3}$$

and

$$\vec{\mu}_i = \sum_{j=1}^{M} f_{ji} E_j \vec{c}_j \,, \tag{4.4}$$

where E_j is the energy deposited in cell j. This fit is iterated until convergence and the position and energy of the gaussian functions are taken as the cluster parameters.

The energy of the clusters are calibrated to accurately reflect the true energy deposited. The calibrations are applied independently to electromagnetic and hadronic deposits, reflecting the difference in the corresponding showers. Calibrations are determined from test beam data, radioactive sources, cosmic ray measurements, in situ collision data and simulated events.

4.5. The particle flow algorithm

The PF algorithm attempts to resolve all particles from the interaction point by exploiting the high granularity of the CMS detector. The algorithm must deal with particle interactions within the detector leading to crooked tracks (bent at a 'kink') and secondary particles. Furthermore, to avoid a quadratic computation with the number of particles only the nearest neighbouring PF blocks in η-ϕ are considered as candidates in the linking algorithm.

The linking of inner tracks to muon tracks has been discussed in Sec. 4.2.2. Calorimeter clusters are linked to inner tracks by extrapolating the track to the ECAL or HCAL and a link formed if these positions lie within a topological cluster's cells, allowing sufficient margins to account for gaps in calorimeter coverage, shower profile uncertainties and multiple scattering. A link distance quantifies the distance between the extrapolated track position and the cluster position in η-ϕ with the link of minimum distance kept. Bremsstrahlung photons are linked to electron GSF tracks if tangents, at each tracker layer, to the tracks are consistent with ECAL clusters. A dedicated conversion finder creates a link between two tracks and a photon if the tracks are compatible with the characteristics of a photon conversion. Clusters are linked together (preshower to ECAL and ECAL to HCAL) if the more granular calorimeter cluster is within the less granular calorimeter cluster, retaining the link forming the minimum distance between the two clusters. Finally, tracks are linked together by a common secondary vertex, attributed to nuclear-interactions. These tracks must contain at least three tracks: one incoming from the primary vertex and at least two outgoing, or at least three outgoing. All links are formed and the subsequent part of the PF algorithm recovers inefficiencies and removes fake particle candidates by targeting particular particle classes, as follows.

4.5.1. Muons

Global muons are classified as isolated to avoid the misidentification of charged hadrons, whereby the total energy in a cone of radius $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.3$ around the muon candidate must not exceed 10% of the muon p_{T}:

$$I_{\mathrm{track,calo}} = \frac{1}{p_{\mathrm{T}}} \left(\left| \sum_{\substack{i \in \mathrm{tracks} \\ \Delta R < 0.3}} \vec{p}_{\mathrm{T},i} \right| + \sum_{\substack{j \in \mathrm{clusters} \\ \Delta R < 0.3}} E_{\mathrm{T},j} \right) < 0.1 \,, \tag{4.5}$$

where $I_{\mathrm{track,calo}}$ is known as the track-calorimeter isolation, i sums the transverse momenta $\vec{p}_{\mathrm{T}}$ of all tracks within $\Delta R = 0.3$ and j sums the transverse energies E_{T} of all clusters within $\Delta R = 0.3$. Non-isolated muons are also identified without the isolation requirement, however, inner tracks must match at least three track segments in the muon detectors or calorimeter clusters compatible with the muon hypothesis to avoid high p_{T} charged hardon misidentification.

The p_{T} of PF muon candidates is reconstructed from the tracker for $p_{\mathrm{T}} < 200\,\mathrm{GeV}$, otherwise the p_{T} is taken from the track fit resulting in the lowest χ^2 from the following combinations of information: tracker only, tracker and the first muon detector plane, global muons, and global muons without high occupancy muon detector planes. After all PF muon candidates are identified and reconstructed the associated PF blocks are removed from further candidate identification. Note that under certain conditions charged hadrons may later be revised as muons.

4.5.2. Electrons and isolated photons

Electron and isolated photon reconstruction must collect and associate Bremsstrahlung photons and electrons from photon conversions to accurately measure the true energy of these particles. These secondary particles are associated to a seed candidate: an ECAL supercluster above a transverse energy threshold of 10 GeV with no GSF track link for photons, or GSF tracks associated to ECAL clusters with fewer than three additional linked tracks. Both classes of candidates must not have HCAL clusters within $\Delta R = 0.15$ of the candidate with an energy deposit exceeding 10% of the supercluster energy. The assigned energy of the candidate is determined from the sum of all energy deposits linked to the seed, this includes Bremsstrahlung photons at tangents to GSF tracks at tracker layers and GSF tracks found by the conversion finder. This energy is calibrated and taken as the final energy for isolated photon candidates. Electron candidate's energy is determined from a combination of the calibrated ECAL energy and GSF track momentum.

To further reduce misreconstruction the candidates must pass additional criteria. A BDT is trained separately for the barrel and endcaps, and for isolated and non-isolated electrons

with a requirement placed on the output determined by the following input features: energy radiated from the GSF track, distance between the GSF track extrapolation and the ECAL seeding cluster, ratio of the HCAL to ECAL energy, KF and GSF track χ^2 and the number of hits in the tracker. Photons, on the other hand, are required to be isolated from other tracks, calorimeter clusters and have HCAL to ECAL energy ratios consistent with typical photon showers.

Finally, as with muons, the PF blocks associated with the electron and isolated photon candidates are removed from further processing.

4.5.3. Hadrons and non-isolated photons

The remnants of jet fragmentation and hadronisation with the subsequent decay of heavy hadrons include: charged hadrons such as $\pi^{\pm}$, $K^{\pm}$ and protons; neutral hadrons such as K_L^0 and neutrons; non-isolated photons typically from π^0 decays; and rarely additional muons from early charged hadron decays. Discrimination of hadrons, other than charged and neutral, is difficult with high misidentification rates, therefore, the PF algorithm makes no attempt at this distinction.

From the remaining PF blocks, all ECAL clusters within the tracker acceptance ($|\eta| < 2.4$) with no linked tracks are considered as photons. Similarly, HCAL clusters with the same acceptance with no linked tracks are treated as neutral hadrons. This assumption holds well within the tracker acceptance where charged and neutral hadrons are distinguishable and 25% of the jet's energy is deposited in the ECAL as photons while only 3% as neutral hadrons (1% for τ-lepton decays as a result of Cabibbo-suppression). However, beyond $|\eta| = 2.4$ up to the ECAL acceptance ($|\eta| < 3.0$) the precedence to photons leads to a high misidentification rate of charged and neutral hadrons, which can account for 25% of the jet's energy deposited in the ECAL. Instead, ECAL clusters linked to HCAL clusters are assigned as hadron showers with an unknown charge, while ECAL clusters without this link are considered photons. Beyond $|\eta| = 3.0$, within the HF, electromagnetic and hadronic activity are distinguished by exploiting depth and lateral shower profile information.

Misidentification and misreconstruction may leave inconsistencies in the calibrated calorimetric energy and tracker momenta for charged hadrons. These are dealt with by reassigning the identification or refining the reconstruction, since an excess infers the presence of photons or neutral hadrons. An excess, δ_{T}, in the calibrated cluster energy, consisting of the ECAL and HCAL energies $E_{\mathrm{T}} = E_{\mathrm{T}}^{\mathrm{ECAL}} + E_{\mathrm{T}}^{\mathrm{HCAL}}$, is defined as

$$\delta_{\mathrm{T}} \equiv E_{\mathrm{T}} - |\sum \vec{p}_{\mathrm{T}}| \,, \tag{4.6}$$

where $|\sum \vec{p}_{\mathrm{T}}|$ is the sum of linked track momenta. If the excess is smaller than E_{T}^{ECAL} but larger than the expected resolution of the hadronic deposit above 500 MeV then the excess is revised as a photon with an energy of δ_{T} with the energy calibration repeated under this hypothesis. However, if the excess is greater than E_{T}^{ECAL} then the ECAL deposit is still associated to a photon but the remaining excess, if above 1 GeV, is interpreted as a neutral hadron. If the excess is smaller than 500 MeV and still inconsistent with the expected energy resolution, a χ^2 fit to the tracker and calorimeter information is performed to refine the energy of the charged hadron. Finally, in the rare case where the track momenta are in excess of E_{T}, i.e. $\delta_{\mathrm{T}} < 0$, a relaxed search for muon candidates is performed. Any remaining excess is associated to misreconstructed tracks with a momentum resolution greater than 1 GeV. The leading[2] misreconstructed track candidate is repeatedly removed until no excess or no candidates remain.

Displaced tracks within the tracker volume predominantly from nuclear interactions are yet to be identified. These involve secondary charged particle tracks linked to a secondary vertex. The charged particles are replaced by a single charged hadron, conserving energy and momentum, with a mass of a charged pion. If a primary charged particle track exists, with a well-defined momentum, then any deficit in the outgoing energy is attributed to undetected secondary particles and the energy of the primary charged hadron is updated to account for these particles.

4.5.4. Jets

The fragmentation and hadronisation of outgoing quarks form a cone of hadrons known as a jet. The anti-k_{T} algorithm [50, 51] discussed in Sec. 2.4.1 clusters PF candidates into cones of radius $R = 0.4$ in the η-ϕ plane. The minimum clustering threshold is set to 15 GeV, below which jets are unreliable at measuring the original parton momentum. Seeding jets by PF candidates, instead of solely relying on calorimeter deposits, improves the energy resolution by complimentary tracker measurements for charged hadrons affected by the poorer hadron calorimeter resolution. However, as with any other jet seed, neutrinos from hadron decays within the jet are not accountable.

4.5.5. Missing transverse momentum

The PF method allows a simple calculation of the uncorrected (or raw) $p_{\mathrm{T}}^{\mathrm{miss}}$ from:

$$\vec{p}_{\mathrm{T,raw}}^{\,\mathrm{miss}} = -\sum_i \vec{p}_{\mathrm{T},i} \,, \tag{4.7}$$

[2] Leading refers to the candidate in a collection with the largest p_{T}.

where i sums over the PF candidates. However, residual misidentification and misreconstruction can lead to a large $p_{\mathrm{T}}^{\mathrm{miss}}$ similar to physics beyond that expected from the SM. To control the artificially large $p_{\mathrm{T}}^{\mathrm{miss}}$ all particles with a significant correlation with the direction and magnitude of the $p_{\mathrm{T}}^{\mathrm{miss}}$ are selected. The particles are typically high p_{T} muons and their identification and reconstruction revised as follows:

- **Genuine cosmic ray muons** in coincidence with an LHC beam crossing, positioned further than 1 cm from the beam axis, are removed from the PF candidate collection if this leads to a reduction in the $p_{\mathrm{T}}^{\mathrm{miss}}$ of at least a half. Semi-leptonic decays of b-hadrons are protected as their removal would increase the $p_{\mathrm{T}}^{\mathrm{miss}}$.
- **Severely misreconstructed muon momentum** as a result of wrong track association, steel yoke interactions, in-flight decay or significant synchrotron radiation. This typically leads to inconsistent inner tracks and outer muon chamber tracks resulting in a misreconstructed muon p_{T} from the PF algorithm. Instead the p_{T} of the muon is taken from the track which reduces the $p_{\mathrm{T}}^{\mathrm{miss}}$ by at least a half, otherwise the p_{T} from the PF algorithm is used.
- **Particles misidentified as muons** typically from charged-hadronic punch-through into the muon systems. In this case, neutral hadron deposits within the calorimeters are wrongfully included as independent particles leading to a double-counting of the charged hadron as a muon and a neutral hadron. Provided that the $p_{\mathrm{T}}^{\mathrm{miss}}$ is reduced by a half and the muon momentum and neutral hadron energies are larger than 100 GeV, then the neutral hadron is removed and the muon is revised as a charged hadron.
- **Muon and neutral hadron overlap** with similar energies which are misidentified as a charged hadron. The charged hadron is converted into a muon and a neutral hadron if the associated $p_{\mathrm{T}}^{\mathrm{miss}}$ is reduced by a half.

After these revisions the $p_{\mathrm{T}}^{\mathrm{miss}}$ is updated. The approach validated with in situ measurements and studies in simulated events which show minimal impact on processes with real $p_{\mathrm{T}}^{\mathrm{miss}}$, such as the pair-production of $t\bar{t}$.

4.6. Summary

The CMS software reconstruction incorporates the whole detector, correlating detector elements to improve particle identification and reconstruction with an attempt to reconstruct all detectable particles in a technique known as particle flow. The PF algorithm starts from two basic building blocks: tracks and clusters. Tracks are formed from a series of Kalman filters applied to hits: charge deposits within the tracks. An alternative GSF track finder attempts to reconstruct the tracks of electrons subject to radiative losses. Vertex reconstruction

is performed from a fit to high-quality track candidates consistent with the beam spot and the primary interaction point identified as the vertex with the greatest $\sum p_{\mathrm{T}}^2$ over outgoing tracks. On the other hand, clusters are formed from energy deposits within calorimeter cells, grouped and refined by a gaussian-mixture model.

The PF algorithm links the building blocks (tracks and clusters) to form a coherent picture of each individual particle with complementary measurements from the various subdetectors. Muon candidates have links between inner and muon chamber tracks. Electron and isolated photon candidates are formed from ECAL clusters, linked to GSF tracks for electrons, with further linking to clusters consistent with Bremsstrahlung photons or GSF tracks compatible with photon conversions. Hadrons and non-isolated photons have possible links between tracks, HCAL and ECAL clusters. Further refinement is performed on irregular candidates. After all PF candidates are determined, jets are clustered from all candidates with the anti-k_{T} algorithm and the $p_{\mathrm{T}}^{\mathrm{miss}}$ determined from a negative sum over all candidate $\vec{p}_{\mathrm{T}}$ with possible identification revisions to avoid large spurious $p_{\mathrm{T}}^{\mathrm{miss}}$.

Chapter 5.

Event Selection and Categorisation

5.1. Introduction

The measurement of the $Z \to$ inv width requires the collection of $Z(\to \nu\bar{\nu}) +$ jets signal events with the key signature of jets recoiling against an invisible system, also known as $p_T^{miss} +$ jets. The $Z/\gamma^*(\to \ell\bar{\ell}) +$ jets process is used as a reference to extract the invisible width, under the assumption of a purely SM $Z \to \ell\bar{\ell}$ interaction. In addition, samples of $W(\to \ell\nu) +$ jets provide a data-driven estimate of large backgrounds with a similar signature to the signal. A combination of the large QCD multijet cross-section and inherent jet misreconstruction leads to additional backgrounds and the need of a sideband[1] enriched with QCD multijet events for a data-driven estimate of this background.

This chapters details the process of data collection and reduction to target specific processes in various categories required for this analysis. Included is the detail of particle identification refinement from the PF algorithm results for an improved signal acceptance whilst maintaining a high background rejection.

5.2. Samples

5.2.1. Data collection

Detector measurements associated to a proton-proton bunch crossing are stored by various sets of HLT requirements, known as HLT paths. The dataset under consideration was collected during 2016 with proton-proton collisions at a centre-of-mass energy of 13 TeV. Issues during detector live-time, such as problems with a particular subdetector or beam conditions, requires

[1] A sideband is an independent set of events to a region obtained by inverting or somehow altering a requirement which defined the original region.

the rejection of problematic data resulting in a dataset with an integrated luminosity of $35.9\,\text{fb}^{-1}$.

The relevant process can be classified into events with missing transverse momentum or more than one lepton. Events with a significant missing transverse momentum are selected by the HLT paths requiring two conditions or a third to be satisfied. The first is

$$p_{\mathrm{T},\not\mu}^{\mathrm{miss}} = \left| -\sum_{i \notin \text{muons}} \vec{p}_{\mathrm{T}} \right| \geq \hat{p}_{\mathrm{T}}^{\mathrm{miss}} \,, \tag{5.1}$$

where the sum is over all PF candidates apart from muons and $\hat{p}_{\mathrm{T}}^{\mathrm{miss}}$ is a particular threshold. The muons act as invisible particles for $p_{\mathrm{T},\not\mu}^{\mathrm{miss}}$ hence this collects $\mathrm{Z} \to \nu\bar{\nu}$, $\mathrm{W} \to \mu\nu$ and $\mathrm{Z}/\gamma^* \to \mu^+\mu^-$ events. The second condition is

$$p_{\mathrm{T,had}}^{\mathrm{miss}} = \left| -\sum_{i \in \text{jets}} \vec{p}_{\mathrm{T}} \right| \geq \hat{p}_{\mathrm{T}}^{\mathrm{miss}} \,, \tag{5.2}$$

summed over all PF jets passing the tight jet identification selection with $p_{\mathrm{T}} > 20\,\text{GeV}$ and no muon candidates. This selection is similar to Eq. 5.1, however, the combination of both maintains a trigger rate for lower $\hat{p}_{\mathrm{T}}^{\mathrm{miss}}$ thresholds by removing events with inconsistent visible (excluding muons) and hadronic recoils, primarily due to misreconstructed soft ($p_{\mathrm{T}} < 20\,\text{GeV}$) hadronic activity. The threshold $\hat{p}_{\mathrm{T}}^{\mathrm{miss}}$ varies from 90–120 GeV depending on the LHC beam conditions to avoid prescaling. The third condition recovers small inefficiencies by applying the selection

$$p_{\mathrm{T}}^{\mathrm{miss}} = \left| -\sum_{i \in \mathrm{PF}} \vec{p}_{\mathrm{T}} \right| \geq 170\,\text{GeV} \,, \tag{5.3}$$

summed over all PF candidates, with various beam halo or HCAL noise cleaning requirements (discussed in Sec. 5.5.1) to maintain this threshold as beam conditions change.

A complementary trigger path collects events with at least one muon rather than the momentum imbalance of an event. Events pass the trigger requirement if there exists a global or tracker muon with $p_{\mathrm{T}} > 24\,\text{GeV}$ which is isolated from other activity (total energy within a cone of $\Delta R = 0.3$ is below 9% relative to the muon p_{T}). To collect events with at least one electron an alternative trigger path is used with the requirement of at least one GSF electron passing a tight identification selection and a $p_{\mathrm{T}} > 27\,\text{GeV}$. Both lepton trigger thresholds are driven by a goal to maintain data collection efficiency and control trigger rates.

5.2.2. Simulation

A vast array of simulated proton-proton collisions attempt to fully replicate the events selected as signal candidates, including backgrounds with similar experimental signatures to the signal. Vector boson production; namely $Z \to \nu\bar{\nu}$, $W \to \ell\nu$, $Z/\gamma^* \to \ell\bar{\ell}$ and prompt photon; in association with jets, V + jets, are produced with the matrix element (ME) event generator MADGRAPH5_AMC@NLO V2.2.2 [52] with restricted boson decays and Feynman diagrams up to $\mathcal{O}(\alpha^2\alpha_s^2)$, hereby known as next-to-leading order (NLO) in the strong coupling constant. The production is split into boson p_T regions to populate the low cross-section high-p_T regions. These generated events are coupled to the PYTHIA8.2 package [53] to mediate the parton showering (PS), hadronisation and particle decays with the CUETP8M1 underlying event description [54]. Double-counting of jet multiplicity final states with the ME and PS is removed by the FXFX [55] matching and merging technique. Another set of processes similar to $Z/\gamma^* \to \ell\bar{\ell}$ are generated where, in turn, the Z and γ^* interactions are turned off to model the Z, γ^* and interference terms of the $Z/\gamma^* \to \ell\bar{\ell}$ process. Furthermore, the processes involving a γ^* have an invariant mass cut-off below 50 GeV to avoid the large cross-section away from the Z peak.

Minor background processes include the pure QCD interactions between partons forming a multijet final state with jet misreconstruction leading to spurious p_T^{miss}, faking the signal signature. The QCD multijet process is produced at leading order (LO) accuracy in the strong coupling constant by the PYTHIA8.2 package, split by the interaction scale Q^2 to populate the low cross-section regions. Top production leads to real p_T^{miss} with zero, one or two leptons in the final state with the pair production, $t\bar{t}$ + jets, produced using a similar technique as the V + jets processes, albeit inclusive in p_T, and single top production — via the s-channel, t-channel or associated W production — are produced at NLO by a various combination of the MADGRAPH5_AMC@NLO V2.2.2, MADSPIN [56], PYTHIA8.2 and POWHEG [57, 58, 59] packages. Diboson processes (combination of Z, W and prompt γ from the hard scatter) may have a similar signature to the signal for particular decays, or particles beyond the detector's acceptance or misidentified. Combinations of MADGRAPH5_AMC@NLO V2.2.2, MADSPIN, PYTHIA8.2 and POWHEG are used to generate the diboson process at NLO. Finally, vector boson scattering (VBS) involves the production of a vector boson (W or Z) from VV scattering, in association with two jets, hence has a signature similar to the V + jets processes. The VBS sample is produced at LO accuracy by the MADGRAPH5_AMC@NLO V2.2.2 package coupled to PYTHIA8.2, with the boson decay restricted to charged leptons or neutrinos. An overlap between the phase-space of the diboson and VBS samples is removed by an ad hoc requirement of one generated boson from the hard scatter for VBS.

All samples are generated with NNPDF3.0 [60] LO and NLO parton distribution functions, where the LO and NLO generators are used. After the relevant process has been generated the

full detector response to such an interaction, along with pileup vertices, is simulated with the GEANT4 [61] package.

5.3. Physics objects

Analysis specific physics objects are derived from PF candidates to refine and target the selection towards the key signature. These include muons; electrons; photons; jets with tagging as hadronic τ-lepton decays or b-hadrons; and the global event descriptor $p_{\mathrm{T}}^{\mathrm{miss}}$, redefined in terms of the recoil parameter.

5.3.1. Muons

The analysis definition of a muon is split into two categories: veto and selection muons. The selection defining these categories is shown in Tab. 5.1. Muons from in flight hadron decays are suppressed by a requirement on the particle-based isolation $I_{\Delta\beta}$ with a pileup mitigation term, known as the $\Delta\beta$ correction

$$I_{\Delta\beta} = \frac{1}{p_{\mathrm{T}}} \left| \sum_{i \in \mathrm{ch}} \vec{p}_{\mathrm{T},i} + \max\left(0, \sum_{i \in \mathrm{nh}} \vec{p}_{\mathrm{T},i} + \sum_{i \in \gamma} \vec{p}_{\mathrm{T},i} - \frac{1}{2} \sum_{i \in \mathrm{PU\ ch}} \vec{p}_{\mathrm{T},i}\right)\right| , \tag{5.4}$$

for a muon with transverse momentum p_{T}; charged hadrons (ch), neutral hadrons (nh) and photons (γ) from the primary vertex in a cone of $\Delta R = 0.4$ around the muon; and charged hadrons from pileup vertices (PU ch) also within this cone. This correction mitigates the contribution from neutral hadrons from pileup vertices by subtracting the momentum of charged hadrons from these vertices, with a factor of a half to account for the average charged-to-neutral particle ratio in the isospin limit. Additional selection is placed on the quality of selection muon tracks, including the impact parameter with the primary vertex split by the transverse, d_{xy}, and longitudinal, d_z, components.

5.3.2. Electrons

Similar to muons, the electrons are classified as veto or selection based on criteria detailed in Tab. 5.2. The first requirements primarily reduce fake electrons from jets with criteria on $\sigma_{i\eta i\eta}^{5\times5}$, the extent of the shower along the ECAL crystals in the η direction within a 5×5 crystal cluster, typically smaller for electrons and photons than hadronic showers; and $E_{\mathrm{HCAL}}/E_{\mathrm{ECAL}}$, the ratio of the energy deposited in HCAL towers located behind the ECAL cluster to the ECAL cluster energy, significantly smaller for electron and photons compared to jets. Subsequent

	Veto	Selection
Muon seed	Global or tracker	Global
PF muon	True	True
Track fit normalised χ^2	-	< 10
Number of matched stations	-	> 1
Number of muon chamber hits	-	> 0
d_{xy} (cm)	-	< 0.2
d_z (cm)	-	< 0.5
Pixel hits	-	> 0
Tracker layer hits	-	> 5
$I_{\Delta\beta}$	< 0.25	< 0.15
p_{T} (GeV)	> 10	> 30
$\lvert\eta\rvert$	< 2.5	< 2.4

Table 5.1.: Requirements for muons passing the veto and selection identification. Most track quality criteria are omitted for veto muons to retain a high efficiency at the loss of fake muon rejection [62]. The p_{T} requirement on selection muons is driven by trigger requirements for the single muon triggers used for the $p_{\mathrm{T}}^{\mathrm{miss}}$ trigger efficiency measurements.

requirements check the consistency of the GSF track with the ECAL cluster, namely $|\Delta\eta_{\mathrm{In}}|$, the difference in η of the energy-weighted supercluster centre and the GSF track extrapolated to the ECAL; $|\Delta\phi_{\mathrm{In}}|$, similarly defined in the ϕ direction; $|(1/E_{\mathrm{ECAL}} - 1/p_{\mathrm{T,track}})|$; and the impact parameters d_{xy} and d_z. The conversion veto rejects electrons with tracks consistent with photon conversions identified by the conversion finder. The isolation, I_{EA}, requirement on electrons reduces contamination from in flight hadron decays, with a pileup mitigation technique known as effective area:

$$I_{\mathrm{EA}} = \frac{1}{p_{\mathrm{T}}} \left| \sum_{i \in \mathrm{ch}} \vec{p}_{\mathrm{T},i} + \max\left(0, \sum_{i \in \mathrm{nh}} \vec{p}_{\mathrm{T},i} + \sum_{i \in \gamma} \vec{p}_{\mathrm{T},i} - \rho A_{\mathrm{eff}}\right) \right| , \tag{5.5}$$

which is similar to $I_{\Delta\beta}$, however, the subtracted quantity is the product of the energy density ρ, defined by the total energy of the event per unit area in η–ϕ, of the event with the (η-dependent) effective area of the electron, a pre-computed average over multiple events.

5.3.3. Photons

The selection for photons is closely aligned with electrons, although with a single veto category and without tracker based requirements and an explicit rejection of candidates linked to tracks. Furthermore, the isolated requirements are split by particle candidates: charged hadrons,

	Veto		Selection	
	Barrel	Endcaps	Barrel	Endcaps
GSF track	True			
$\sigma^{5\times5}_{i\eta i\eta}$	0.011	0.0314	0.00998	0.0292
$E_{\text{HCAL}}/E_{\text{ECAL}}$	0.298	0.101	0.0414	0.0641
$\lvert\Delta\eta_{\text{In}}\rvert$	0.00477	0.00868	0.00308	0.00605
$\lvert\Delta\phi_{\text{In}}\rvert$	0.222	0.213	0.0816	0.0394
$\lvert(1/E_{\text{ECAL}} - 1/p_{\text{T,trk}})\rvert$ (1/GeV)	0.241	0.14	0.0129	0.126
d_{xy} (cm)	0.05	0.1	0.05	0.1
d_z (cm)	0.1	0.2	0.1	0.2
Missing tracker hits	2			
Conversion veto	True			
I_{EA}	0.0994	0.107	0.0588	0.0571
p_{T} (GeV)	> 10		> 30	

Table 5.2.: Criteria for the electrons identified as veto or selection for the analysis. These values are tuned to obtain a pre-defined efficiency and background rejection. All parameters must be less than the value unless stated otherwise. The criteria are split by barrel and endcap regions to retain similar electron identification efficiency in both [63].

neutral hadrons and photons; with a p_{T}-dependence to remain efficient across a broad range of momenta. The selection is detailed in Tab. 5.3.

	Barrel	Endcaps
Linked track	False	False
$\sigma^{5\times5}_{i\eta i\eta}$	0.01031	0.03013
$E_{\text{HCAL}}/E_{\text{ECAL}}$	0.0597	0.0481
I_{ch}	$1.295/p_{\text{T}}$	$1.011/p_{\text{T}}$
I_{nh}	$10.910/p_{\text{T}} + 0.00148 + 1.7\times10^{-5}p_{\text{T}}$	$5.931/p_{\text{T}} + 0.0163 + 1.4\times10^{-5}p_{\text{T}}$
I_{γ}	$3.630/p_{\text{T}} + 0.0047$	$6.641/p_{\text{T}} + 0.0034$
p_{T} (GeV)	> 25	

Table 5.3.: Photon requirements for the veto collection with all parameters lower than the values in the table unless stated otherwise. The criteria are split by barrel and endcap regions as with electrons [63]. Only veto photons are defined for this analysis.

5.3.4. Jets

Requirements on jets, detailed in Tab. 5.4 depend on the region which they fall into with a veto and a selection category defined below $|\eta| = 2.4$ and inclusive in η, respectively. Pileup suppression is implemented by removing charged hadrons associated to pileup vertices from the jet cone. Leptons or photons within their respective veto collection which reside within a cone of a jet results in the removal of the jet from the event.

	Central	Endcaps	Forward
Charged hadron energy fraction	> 0	-	-
Charged EM energy fraction	< 0.99	-	-
Charged particle multiplicity	> 0	-	-
Constituent multiplicity	> 1	> 1	-
Neutral hadron energy fraction	< 0.99	< 0.99	-
Neutral EM energy fraction	< 0.99	< 0.99	< 0.9
Neutral particle multiplicity	-	-	> 10
p_{T} (GeV)		> 40	

Table 5.4.: Jet selection based on position: central ($|\eta| < 2.4$), endcaps ($2.4 < |\eta| < 3$) and forward ($|\eta| > 3$). Distinction between charged and neutral particles is lost beyond the tracker acceptance of $|\eta| < 2.4$ and forward jets are required to have a significant neutral particle number due to their high radiation environment.

Additional selection applied to jet structure allows the tagging of hadronic τ-lepton decays or jets originating from b-hadrons.

Hadronic τ-lepton tagging

The τ-lepton decays into hadrons and a neutrino, labelled as τ_h, 64.8% of the time, with the remaining decays into an electron or muon and neutrinos. The production of a W boson and its subsequent decay $\mathrm{W} \to \tau_h \nu$ leads to the signal signature of $p_{\mathrm{T}}^{\mathrm{miss}} +$ jets, therefore, tagging and vetoing events with a τ_h-lepton is vital. The tagging is accomplished with the hadrons-plus-strips (HPS) [64] algorithm.

The branching fractions of the τ-lepton are shown in Tab. 5.5 with a significant τ_h branching fraction into one or three charged hadrons and up to two π^0, detected from their decay into two photons which typically convert into e^+e^- pairs within the tracker. Therefore, the HPS algorithm is seeded by jets with $p_{\mathrm{T}} > 14\,\mathrm{GeV}$ and $|\eta| < 2.5$ and matches charged hadrons to η-ϕ strips of photon and electron constituents. The strips are generated by an iterative procedure starting from the highest p_{T} electron or photon candidates and merging the next highest

p_{T} candidate which lies in an $\eta \times \phi$ window of 0.05×0.20, enlarged in ϕ to accommodate electron bending, around the strip centre. After merging the strip centre is updated from an energy-weighted average of the merged candidates and the iteration proceeds until no further candidates are available. Upon generating all strips, only those with a p_{T} greater than 2.5 GeV are kept as π^0 candidates.

Decay mode	$\mathcal{B}$ [%]
$\tau^- \to \ell\bar{\nu}_\ell\nu_\tau$	35.2
$\tau^- \to h^-\nu_\tau$	11.5
$\tau^- \to h^-\pi^0\nu_\tau$	25.9
$\tau^- \to h^-\pi^0\pi^0\nu_\tau$	9.5
$\tau^- \to h^-h^+h^-\nu_\tau$	9.8
$\tau^- \to h^-h^+h^-\pi^0\nu_\tau$	4.8
Remaining hadronic decays	3.3
All hadronic decays	64.8

Table 5.5.: Branching fractions $\mathcal{B}$ for τ-lepton decay modes without particular distinction between charged hadron, $h^\pm$, type and highlighting the decay modes targeted by the HPS algorithm. The two hadron final state is typically mediated by $\rho(770)$ resonance and the three hadron final state by $a_1(1260)$. The table is adjusted [64] with approximate values based on the particle data group (PDG) [9].

The τ_h-lepton candidates are formed by matching strips to charged particle candidates within the seeding jet and consistent with the vertex nearest to the leading track: nearest approach to the vertex of at most 0.4 cm in z and 0.03 cm in the x-y plane to remove spurious tracks and mitigate pileup effects without a significant loss in efficiency. All τ_h-lepton candidates with one or three charged hadrons and up to two strips, within a cone $\Delta R = (3\,\mathrm{GeV}/p_{\mathrm{T}})$, are classified into a particular decay mode if the invariant mass of all the hadrons is consistent with the τ-lepton mass, with some margin to account for the neutrino. The τ_h-lepton is assigned the decay with the highest p_{T} from the remaining modes for a unique classification.

Misidentification of quark or gluon jets as a τ_h-lepton may result in a loss in signal efficiency for $\mathrm{Z}(\to \nu\bar{\nu}) + \text{jets}$ when vetoing τ_h events. These are recovered by discriminating τ_h-leptons and quark or gluon jets with an isolation-based BDT. The input features vary upon the classified decay mode and include: the lead track d_{xy} and its significance, the distance between the τ-lepton production and decay vertices and its significance, charged and neutral particle isolation sums and the $\Delta\beta$ correction value, an integer encoding the decay modes, a success flag upon finding a decay vertex for the τ-lepton, along with the p_{T} and η of the τ-lepton. This BDT is trained and validated on simulated events with a selection on the discrimination variable defining various signal efficiency and background rejection, adjusted by the τ-lepton p_{T}, working points. A very loose working point with a high efficiency is important for this

analysis for an efficient rejection of τ_h-lepton backgrounds. Further discrimination on electron or muon misidentification as the charged hadron candidate are omitted in favour of higher selection efficiencies.

The reconstructed τ_h-leptons are required to have $p_T > 20\,\text{GeV}$, $|\eta| < 2.3$ and pass the very loose working point to be classified as a veto τ_h-lepton, with a more stringent selection of $p_T > 30\,\text{GeV}$, $|\eta| < 2.3$ and passing the tight working point for the selection classification.

b-jet tagging

The jets associated to the $Z(\to \nu\bar{\nu})$ production are typically quark or gluon jets with processes involving neutrinos in the final state along with b-jets resulting in a similar signature, such as top-quark production. The $Z(\to \nu\bar{\nu}) + b$ jets process is not targeted in this analysis and hence are discarded. Therefore, discrimination of b-jets from quark or gluon jets and the subsequent veto of events with b-jets is vital for this analysis.

The b-jet tagging is performed on analysis-level jets by applying criteria on the jet's constituent tracks and vertices to identify b-hadron decays with displacements of about 1 cm, hence only valid for jets within the tracker's acceptance $|\eta| < 2.4$. First, a set of secondary vertices are found by an inclusive vertex finding (IVF) algorithm from all tracks within an event with $p_T > 0.8\,\text{GeV}$ and a longitudinal impact parameter less than 3 mm. The vertex finder is initialised by track seeds defined by a 3-dimensional impact parameter of at least 50 µm and a 2-dimensional impact parameter significance above 1.2. From these seed tracks the algorithm follows [65]:

- **Track clustering:** grouping the seed track with compatible tracks based on separation and opening angle.
- **Secondary vertex fitting and cleaning:** fitting the position of the secondary vertex from the clustered tracks with the adaptive vertex fitter (discussed in Sec. 4.3). Identified vertices must have 2 and 3-dimensional flight distances consistent with b or c-hadron decays and vertices sharing 70% of tracks with another nearby vertex are removed.
- **Track arbitration:** tracks associated with both primary and secondary vertices are removed if they are more compatible with the primary vertex.
- **Refitting and cleaning:** positions of the secondary vertices are refitted and duplicate vertices are removed.

Resulting set of secondary vertices are associated to jets if the angular distance between the jet axis and the vertex is within $\Delta R = 0.3$. The jets seeding the b-tag fall into three independent categories based on the number of secondary vertices and tracks: at least one secondary vertex,

no secondary vertex but at least two tracks, and the remaining set of candidates. An artificial neural network[2] is trained on each category with simulated events to discriminate b, c and light (u, d, s or gluon) jets. The artificial neural network is fed a large set of input features based on the jet and its constituent secondary vertices and tracks with parameters based on distance, invariant mass and multiplicities along with the jet p_{T} and η [65]. Subsequently, a jet is b-tagged if the resulting discriminating variable is above a threshold defined by a medium working point where the misidentification probability is about 1%, resulting in a b-jet tagging efficiency of about 63% [65]. This working point defines a set of b-tagged jets, subsequently used to veto events.

5.4. Missing transverse momentum

The transverse momentum imbalance of an event is reconstructed from the negative sum of transverse momenta of all PF candidates. This is used as an estimate for the vector boson p_{T} in $\mathrm{Z}(\rightarrow \nu\bar{\nu}) + \text{jets}$ events, where it cannot be determined directly. Therefore, in the collection of events targeting vector boson decays with partial or fully reconstructed decay products a proxy is used for the vector boson $\vec{p}_{\mathrm{T}}$, known as the recoil $\vec{U}$, defined as

$$\vec{U} = \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} + \sum_{i \in \ell} \vec{p}_{\mathrm{T},i} \,, \tag{5.6}$$

where i sums over all electrons and muons passing the analysis selection. In effect, this parameter determines the missing transverse momentum of an event treating electrons and muons as neutrinos.

5.5. Event categorisation

A set of categories are defined and used throughout this analysis for various purposes including signal collection, background prediction, and validation. First, a baseline selection is placed on all regions furthered by region-specific criteria to define independent categories. The baseline selection is driven by the desired event topology, trigger thresholds and rejection of spurious $p_{\mathrm{T}}^{\mathrm{miss}}$ backgrounds.

[2] An artificial neural network is a form of machine-learning, inspired by biological neural networks, with interconnected nodes typically with a non-linear function output of a weighted sum of its inputs. The weights are adjusted in the learning process to optimise a loss function of the output nodes and the truth category or value of a set of inputs.

5.5.1. Baseline selection

The first set of requirements, known as the $p_{\mathrm{T}}^{\mathrm{miss}}$ filters, remove scarce events which typically have spurious $p_{\mathrm{T}}^{\mathrm{miss}}$. All events without a reconstructed primary vertex within acceptable parameters of the beam spot are discarded since the offset of the origin biases the $p_{\mathrm{T}}^{\mathrm{miss}}$ distribution and few simulated events predict this contribution. Although infrequently an issue, machine induced particles, specifically high energy muons, may interact with the calorimeters with a significant energy deposit leading to spurious $p_{\mathrm{T}}^{\mathrm{miss}}$. Such events are removed by searching for atypical calorimeter shower shapes matching to CSC detectors. Further noise in the HCAL subdetectors readout system may give rise to spurious $p_{\mathrm{T}}^{\mathrm{miss}}$, hence events with HCAL readout channels consistent with significant noise leading to a biased $p_{\mathrm{T}}^{\mathrm{miss}}$ are discarded [66]. Also as a result of noise ECAL crystals may be masked during reconstruction, however, this may lead to a significant loss of energy if, for example, a jet is incident upon the masked crystal leading to spurious $p_{\mathrm{T}}^{\mathrm{miss}}$. These events are removed by inspecting L1T towers to determine if a significant amount of energy was lost. A final source of spurious $p_{\mathrm{T}}^{\mathrm{miss}}$ is a result of high p_{T}, low resolution PF muons which lead to few unaccounted issues in the reconstruction, removed by inspecting candidate events in depth.

The next criteria target the $p_{\mathrm{T}}^{\mathrm{miss}}$ + jets topology with thresholds driven by the missing energy trigger paths. To remain efficient in collecting events with these trigger paths the selection $\mathcal{U} = |\vec{\mathcal{U}}| > 200\,\mathrm{GeV}$ (discussed in Sec. 6.2.3) is applied with the recoil in place of the $p_{\mathrm{T}}^{\mathrm{miss}}$ to target similar topologies between events with missing energy from neutrinos and those with leptons in the final state. The topology is further targeted by requiring a lead jet $p_{\mathrm{T}} > 200\,\mathrm{GeV}$ within tracker acceptance ($|\eta| < 2.4$), avoiding dijet-like events and resulting in the jet as a primary recoil against the vector boson. An additional requirement on the charged hadron energy fraction between 0.1–0.95 for this lead jet reduces spurious $p_{\mathrm{T}}^{\mathrm{miss}}$ events arising from detector noise and machine-induced particles.

Furthermore, events are rejected if there are additional veto objects than selection objects for the jet, b-tagged jet, or photon collections. These vetoes primarily remove events with VBS, top production and γ + jets processes.

The final baseline requirement is

$$\frac{\left|p_{\mathrm{T}}^{\mathrm{miss}} - p_{\mathrm{T,calo}}^{\mathrm{miss}}\right|}{\mathcal{U}} < 0.5\,, \tag{5.7}$$

which tests the compatibility of the particle-based $p_{\mathrm{T}}^{\mathrm{miss}}$ and the muon-corrected calorimeter-based $p_{\mathrm{T,calo}}^{\mathrm{miss}}$, relative to the recoil. This inconsistency is typically due to an issue in the reconstruction and does not significantly impact the signal.

5.5.2. Categorisation

The event categorisation is shown in Tab. 5.6 with lepton selections and vetoes defining each region and possible requirements to enhance the signal. The ℓ + jets ($\ell = e, \mu$) regions have an accompanied transverse mass selection on the W boson, $M_{\mathrm{T},\ell}$, approximated as

$$M_{\mathrm{T},\ell} = \sqrt{2 p_{\mathrm{T},\ell} p_{\mathrm{T}}^{\mathrm{miss}} \left(1 - \cos(\Delta\phi)\right)}\,, \tag{5.8}$$

where $p_{\mathrm{T},\ell}$ is the transverse momentum of the leptons and $\Delta\phi$ is the opening angle between the lepton and the $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ vectors. The window placed on $M_{\mathrm{T},\ell}$ surrounds the W mass to enhance these regions in W + jets events and reduce combinatorial backgrounds. Similarly, the $\ell\ell$ + jets regions are accompanied by an invariant mass $M_{\ell\ell}$ window centred on the Z mass, significantly reducing the combinatorial background and contributions from γ^* + jets. A remaining large contribution remains from the QCD multijet process whereby spurious $p_{\mathrm{T}}^{\mathrm{miss}}$ arises from jet mismeasurement resulting in an overlap between the $p_{\mathrm{T}}^{\mathrm{miss}}$ and a jet. To avoid significant signal rejection only the overlap between the leading four jets are considered with the selection:

$$\Delta\phi_{\mathrm{min}} = \min_{i \in \{j_0, j_1, j_2, j_3\}} \Delta\phi\left(\vec{p}_{\mathrm{T}}^{\mathrm{miss}}, \vec{p}_{\mathrm{T},i}\right) > 0.5\,, \tag{5.9}$$

where i represents the set of the leading four jets and $\Delta\phi(\vec{a}, \vec{b})$ represents the opening angle between vectors $\vec{a}$ and $\vec{b}$. The selection on $\Delta\phi_{\mathrm{min}}$ is inverted to obtain a QCD multijet enriched sideband. The e + jets region has an additional requirement placed on the $p_{\mathrm{T}}^{\mathrm{miss}}$ to further avoid the QCD multijet process with jets misidentified as electrons.

5.6. Summary

The data relevant for this analysis was collected during 2016 with the $p_{\mathrm{T}}^{\mathrm{miss}}$, muon and electron-based triggers. The data is complemented by a vast array of simulated events, predominantly from V + jets, generated at NLO accuracy in the strong coupling constant. Minor background processes are produced with a variety of event generators at LO and NLO accuracy. The initiating parton in all generated events are sampled from parton distribution functions from the NNPDF3.0 group at LO and NLO accuracy where applicable. Generated events are typically coupled to a package to mediate parton showering, hadronisation and particle decays. An overlay of pileup vertices are placed within the generated events. Subsequent detector simulation and signal digitisation is performed with the GEANT4 package.

Analysis level physics objects are defined with a veto and selection working point. Leptons and photons are required to pass a pileup-corrected isolation requirement. Furthermore,

Category	Requirements
$p_{\mathrm{T}}^{\mathrm{miss}}$ + jets	No additional veto electrons, muons or τ_h-leptons
μ + jets	A single selected muon
	No additional muons, or veto electrons or τ_h-leptons
	$30 < M_{\mathrm{T},\mu} < 125\,\mathrm{GeV}$
$\mu\mu$ + jets	Two selected and oppositely charged muons
	No additional muons, or veto electrons or τ_h-leptons
	$71 < M_{\mu\mu} < 111\,\mathrm{GeV}$
e + jets	A single selected electron
	No additional electrons, or veto muons or τ_h-leptons
	$30 < M_{\mathrm{T},e} < 125\,\mathrm{GeV}$
	$p_{\mathrm{T}}^{\mathrm{miss}} > 100\,\mathrm{GeV}$
ee + jets	Two selected and oppositely charged electrons
	No additional electrons, or veto muons or τ_h-leptons
	$71 < M_{ee} < 111\,\mathrm{GeV}$
τ_h + jets	A single selected τ_h-lepton
	No additional τ_h-leptons, or veto electrons or muons

Table 5.6.: Definitions of each analysis category with the addition of the baseline selection and $\Delta\phi_{\mathrm{min}} > 0.5$, and an additional set of mirroring categories with the same definition but in the $\Delta\phi_{\mathrm{min}}$ sideband, $\Delta\phi_{\mathrm{min}} < 0.5$. A description of the variables is provided in the main text.

leptons are ignored from the $p_{\mathrm{T}}^{\mathrm{miss}}$ calculation to define the recoil of an event. Jets are subcategorised in tags as originating from a hadronic τ-lepton or b-hadron decays. The objects are used in the categorisation of each event with a baseline selection to filter rare spurious $p_{\mathrm{T}}^{\mathrm{miss}}$ events, apply trigger thresholds and a lead jet selection to target the topology of a primary jet recoiling against a vector boson. Vetoes are placed on events containing forward ($\eta > 2.4$) jets, to reduce VBS contamination; b-jets and selected photons. The final baseline condition removes misreconstructed events with an inconsistent $p_{\mathrm{T}}^{\mathrm{miss}}$ and $p_{\mathrm{T,calo}}^{\mathrm{miss}}$. Subsequent categorisation falls into $p_{\mathrm{T}}^{\mathrm{miss}}$ + jets, ℓ + jets ($\ell = e, \mu$ or τ_h) or $\ell\ell$ + jets regions, with mirroring regions in a QCD multijet enriched sideband. Mass windows are placed on the combinations of the leptons and the $p_{\mathrm{T}}^{\mathrm{miss}}$ to enrich the signal in all applicable regions. The resulting regions are used throughout the analysis.

Chapter 6.

Simulation corrections

6.1. Introduction

This chapter details the corrections applied to the Monte Carlo (MC) simulated events aimed at improving the description of observed events (data) by the MC. These corrections are typically understood in terms of a scale multiplied to a quantity, also known as a scale-factor, hence a scale-factor of one represents no deviation from the nominal MC prediction. There are two main categories of corrections used in this analysis. The first corresponds to changing the weighting of a MC event, known as an event-based correction. The weight given to a MC event is the product of the cross-section of the particular process and the integrated luminosity, $\sigma\mathcal{L}$. This event weight is later modified to improve the agreement between data and MC. The cross-section σ is calculated to the highest order available in both the strong and electroweak coupling constants, whereas the luminosity $\mathcal{L}$ is measured by the CMS and surrounding detectors. Event-based corrections will scale this weight on an event-by-event basis, however, the sum of the weights over all events should remain unchanged to avoid altering the cross-section. The second set of corrections represents a shift in a parameter, such as altering the momentum of an object, hereby known as an object-based correction, which inherently keeps the cross-section constant.

Typical aggregation of information is in the form of histograms where data in a particular bin is determined from the total number of events within the bin's criteria, whereas the MC contribution is a sum over the weights for the same criteria. Therefore, the event-based corrections simplify to a scale up or down to an event's contribution in the sum. However, object-based corrections alter which bin the event may fall in. Both corrections, even if unity, are associated with an uncertainty to alter the correction up or down resulting in an alternative histogram under different variation hypotheses. These hypotheses are used later in the statistical interpretation of the results.

An event-based correction is determined by measuring the efficiency of a selection or requirement in data, $\varepsilon_{\text{data}}$, and again in MC, ε_{MC}, parameterised in terms of observables $\vec{x}$

resulting in a scale-factor, $\hat{f}$, of

$$\hat{f}(\vec{x}) = \frac{\varepsilon_{\text{data}}(\vec{x})}{\varepsilon_{\text{MC}}(\vec{x})} \,. \tag{6.1}$$

This reweighting procedure attempts to have MC replicate the observed efficiency for a particular selection. The efficiencies are measured with a few different techniques, discussed in detail below.

6.1.1. Tag and probe method

A common technique to measure the efficiency in data is known as the tag and probe method. This technique involves the selection of events containing two candidates: a tag which must pass a tight set of restrictions T to firmly avoid misidentification, and a probe which passes very loose requirements P. The events are such that the tag infers knowledge of the probe, for example, in $Z/\gamma^* \to \ell\bar{\ell}$ where one lepton is a tag and the other a probe. The efficiency of a selection S, given the probe requirements P, is determined by counting events:

$$\varepsilon(S|P) = \frac{N(S \cap T \cap P)}{N(T \cap P)} \,, \tag{6.2}$$

where $N(R)$ is the number of events passing some set of requirements R. For a tag and probe from a Z boson decay the event counting is taken from the signal by performing a maximum likelihood fit to the Z peak with a falling background. A similar procedure can be performed in MC.

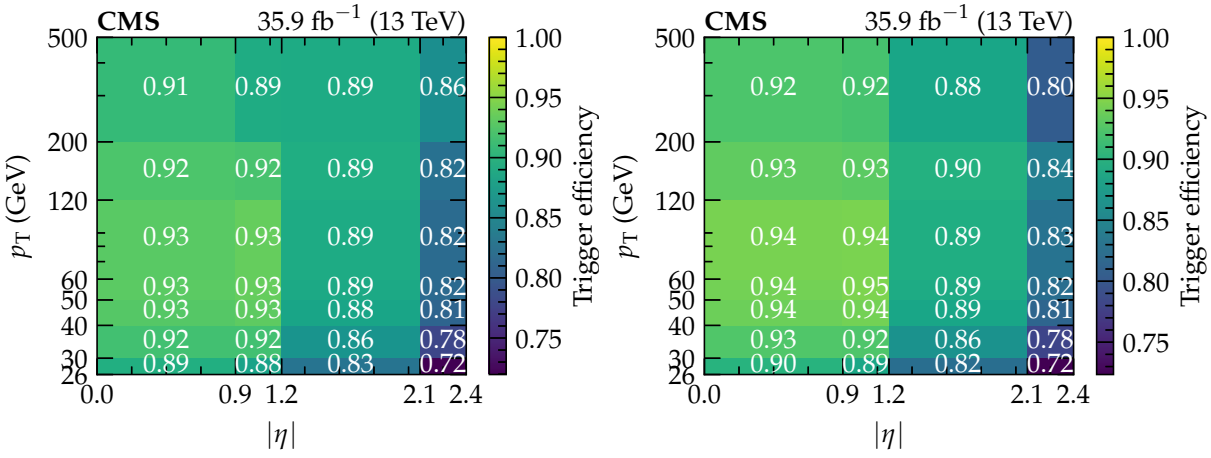
6.2. Trigger Efficiency

The trigger selection discussed in Sec. 5.5.1 collect a particular set of events for the analysis; however, limitations in the reconstruction available to the L1T and HLT results in a loss of events. To replicate the data collection in simulated events the efficiency of a trigger selection is measured. The trigger selections are split into the muon, electron and $p_{\text{T}}^{\text{miss}}$ categories.

6.2.1. Muon trigger efficiency

The muon trigger efficiency [67] is measured with the tag and probe method where the tag selection T and probe selection P must pass the full set of analysis requirements apart from the p_{T} and η criteria for the probe. This allows the parameterisation of the muon efficiency with p_{T} and η as the trigger performance is expected to be highly dependent on these variables.

The tag is also required to cause the event to pass the muon trigger selection such that the event is stored. The events are subsequently split into all probes and those passing the muon triggers with further division into independent p_T and η bins. In each category, the maximum likelihood fit is performed to the invariant mass of the two muons to extract the signal and background contributions in each bin. The signal is modelled as a Breit-Wigner convoluted by a gaussian distribution and the background is modelled as a falling exponential. The efficiency is determined in each bin from the ratio of signal events where the probe passes the muon trigger selection to signal events without this requirement. The statistical uncertainty on the signal events is propagated through to the efficiencies. Additional systematic uncertainties are determined by performing the fit varying, in turn, the number of invariant mass bins, invariant mass range, signal and background models, tag selection and probe multiplicity. During the 2016 data-taking period, saturation issues in the pre-amplifier chips for the tracker resulted in a loss of hits. This effect was mitigated in the later part of this period and further reduced by a re-processing of the data. Nevertheless, the muon trigger efficiencies are split into the initial part and final part of the data-taking period, as shown in Fig. 6.1.



(a) Initial data-taking period

(b) Final data-taking period

Figure 6.1.: Muon trigger efficiencies parameterised by its p_T and $|\eta|$, split by the data-taking period. To maintain a consistent background rejection, the trigger efficiency is lower in the end-caps where the muon background rate is larger and for low p_T muons where the track reconstruction is poorer. Inefficiencies for high p_T muons result from misidentified charged hadrons or issues in reconstructing the p_T for straight tracks. Most of the trigger efficiency arises from the L1T decision which is heavily constrained by the timing budget. The differences between the data-taking periods are at most 2%, with significant overlap within the uncertainties.

This efficiency, ε_i, is associated to a single muon; however, events collected with higher muon multiplicities were triggered by any of the muons, hence the efficiency associated with

the event, ε, for any number of muons is

$$\varepsilon = 1 - \prod_{i \in \text{muons}} (1 - \varepsilon_i) \,, \tag{6.3}$$

as follows from a binomial distribution where at least one muon must pass the trigger selection. This efficiency scales the MC event weight in regions collected by the muon triggers with a luminosity-weighted average of the initial and final data-taking periods.

6.2.2. Electron trigger efficiency

The electron trigger efficiency measurement [63] follows a similar procedure to the muons. The analysis selection is applied to the tag and the probe, with relaxed kinematic criteria. The pair of electrons are formed into a Z boson candidate with a fit to the invariant mass distribution to extract the signal contribution and determine the efficiency of the electron trigger selection. Systematic uncertainties are determined with the same method detailed for the muon trigger efficiencies. The efficiencies for the electron trigger selection is shown in Fig. 6.2 and follow the same event-based efficiency in Eq. 6.3.

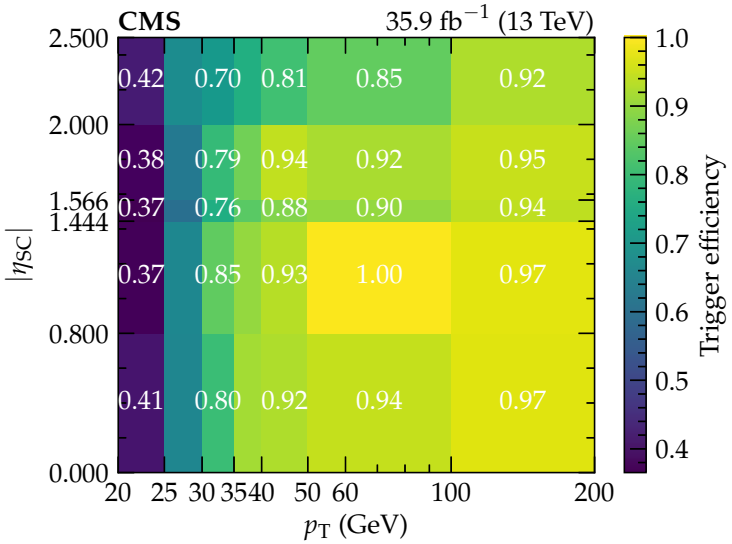
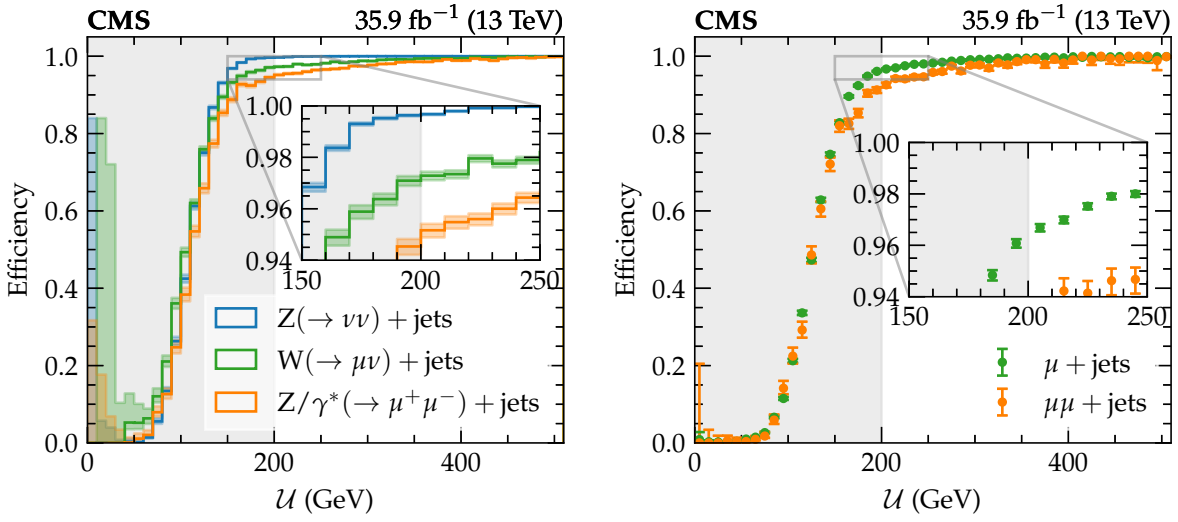


Figure 6.2.: Electron trigger efficiency with a threshold of $p_T > 27\,\text{GeV}$, parameterised with the p_T and absolute supercluster position in η, $|\eta_{SC}|$. The low p_T inefficiency is driven by the trigger threshold and the efficiency in the partially uninstrumented gap in the ECAL is shown in the bin $1.444 < |\eta_{SC}| < 1.566$.

6.2.3. $p_{\mathrm{T}}^{\mathrm{miss}}$ trigger efficiency

The $p_{\mathrm{T}}^{\mathrm{miss}}$ trigger paths collect events for the $p_{\mathrm{T}}^{\mathrm{miss}}$ + jets, μ + jets, $\mu\mu$ + jets, τ_h + jets and the equivalent QCD multijet enriched sidebands. To measure the efficiency in data, an independent set of events are collected using the single muon triggers which may pass or fail the $p_{\mathrm{T}}^{\mathrm{miss}}$ trigger requirements. Furthermore, the $p_{\mathrm{T}}^{\mathrm{miss}}$ triggers primarily place a selection on the recoil which ignores the muons, allowing events to populate the tail of the recoil distribution. No reference trigger is required in MC and the efficiency is extracted from the trigger emulation. The results for $\mathrm{Z}(\to \nu\bar{\nu})$ + jets, $\mathrm{W}(\to \ell\nu)$ + jets and $\mathrm{Z}/\gamma^*(\to \ell\bar{\ell})$ + jets shown in Fig. 6.3, alongside the measurements from data in the μ + jets and $\mu\mu$ + jets regions.

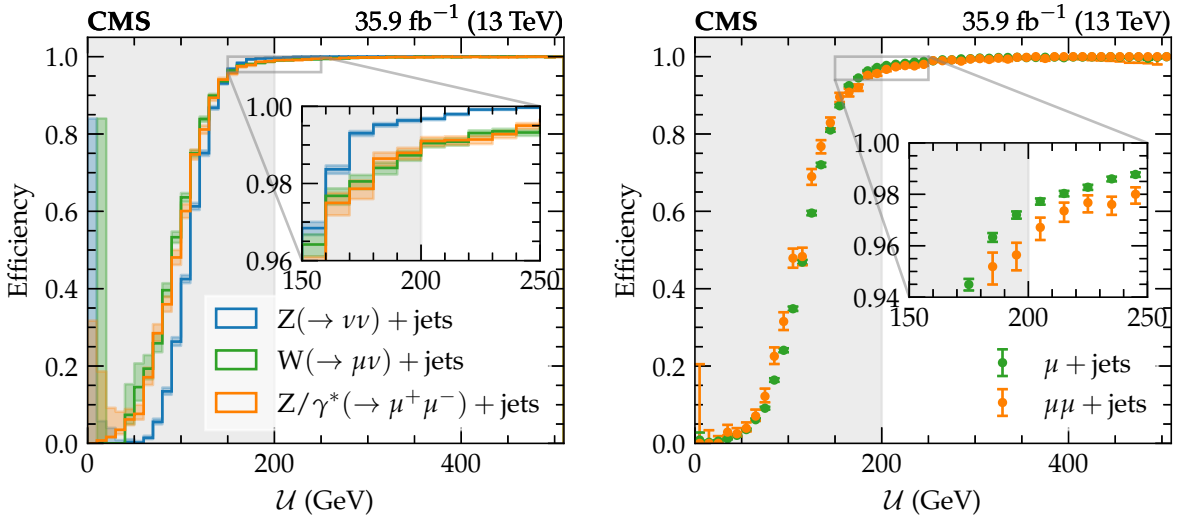


(a) Simulation **(b)** Data

Figure 6.3.: The dependence of the $p_{\mathrm{T}}^{\mathrm{miss}}$ trigger efficiency on the offline recoil $\mathcal{U}$. The efficiency below 200 GeV (grey band) is shown, however, these events are removed in the final analysis. The uncertainty bands represent the 68.2% asymmetric confidence interval on a binomial distribution for a given total number of event with a fraction passing the selection. This band only includes the statistical uncertainty in both simulation and data.

The lower efficiency measured in data is a result of the evolution of trigger paths to higher thresholds as the beam luminosity increased throughout the 2016 data-taking period, whereas the simulated trigger efficiency is primarily controlled by the lowest threshold. The impact of the evolving trigger paths in data is replicated in MC by applying a scale factor correction determined from the ratio of data to MC efficiency, parameterised with the recoil. However, a clear efficiency difference is observed between the various muon multiplicity regions. In fact, the HLT muon collection is not perfectly aligned with the offline reconstructed collection causing some muons to be included in the HLT recoil calculation leading to values below the thresholds. This is significantly mitigated by the addition of a $p_{\mathrm{T}}^{\mathrm{miss}}$ trigger path which selects events with a $p_{\mathrm{T}}^{\mathrm{miss}}$ above 75 GeV and an isolated track with $p_{\mathrm{T}} > 50$ GeV. This recovers events

where the muon was misidentified at the HLT level, however, still classified as an isolated track. The corresponding trigger efficiency measurements are shown in Fig. 6.4, with a significant improvement between the muon multiplicity regions.

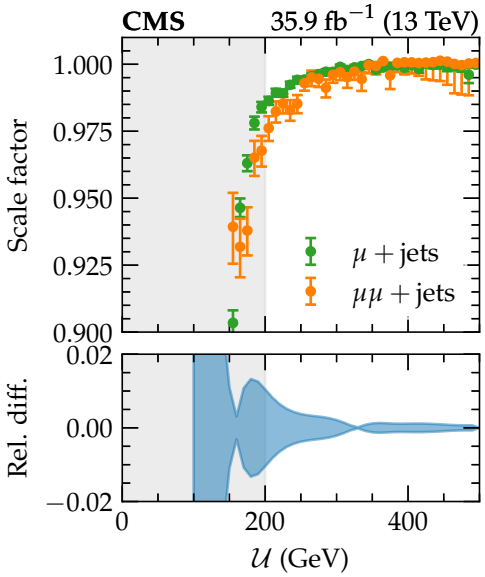


(a) Simulation **(b)** Data

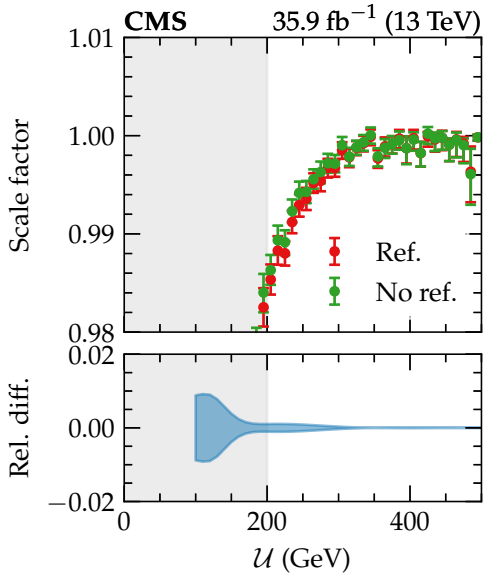
Figure 6.4.: The $p_{\mathrm{T}}^{\mathrm{miss}}$ trigger efficiency measurements similar to Fig. 6.3, with the addition of the isolated track selection.

The $p_{\mathrm{T}}^{\mathrm{miss}}$ trigger scale factors are taken from the μ + jets region where the statistical uncertainty is the lowest. The remaining difference in the scale factors between the μ + jets and $\mu\mu$ + jets regions, shown in Fig. 6.5a is taken as a systematic uncertainty to cover residual effects from the muon mismatch between the HLT and offline. The effect of using the muon triggers as a reference in data is emulated in simulation and the difference between the nominal scale factor is taken as a systematic uncertainty. A final systematic uncertainty is determined by the difference in the scale factors in an independent region, namely the μ + jets QCD enriched sideband. The scale factors and the associated systematic uncertainties are shown in Fig. 6.5. The reference trigger introduces a small bias of less than 0.5%, whereas the muon multiplicity systematic covers a 1% effect and the QCD sideband a 2% effect at $\mathcal{U} = 200\,\mathrm{GeV}$ and drops off to a negligible level above 500 GeV where the efficiencies and scale factors plateau at 1. The scale factors associated with the μ + jets region in Fig. 6.5a is applied to all events collected by $p_{\mathrm{T}}^{\mathrm{miss}}$ triggers to primarily correct for the changing conditions during data taking by up to 2%, with three systematic uncertainties propagated as alternative templates for the final statistical analysis to cover biases in the scale factor measurement.

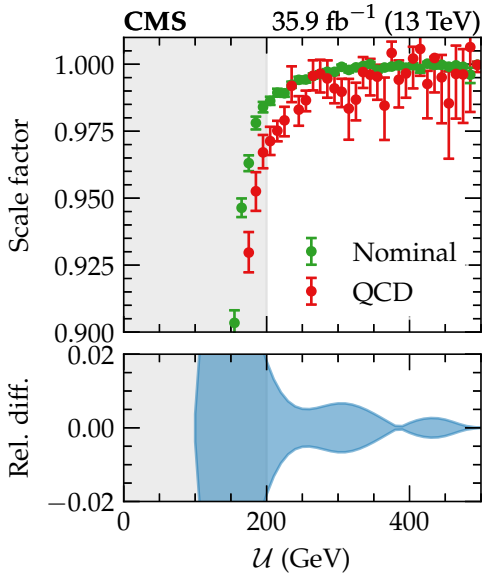
The muon multiplicity systematic is decorrelated between the $p_{\mathrm{T}}^{\mathrm{miss}}$ + jets, μ + jets and $\mu\mu$ + jets, allowing the systematic effect to pull each in separate directions. The reference trigger systematic is correlated among all regions. Similarly to the muon multiplicity, the



(a) Muon multiplicity



(b) Reference trigger



(c) Independent sample

Figure 6.5.: The $p_{\mathrm{T}}^{\mathrm{miss}}$ trigger scale factors measured for various regions to determine the central correction and relative difference between them for systematic uncertainties. The statistical uncertainties in the scale factors would artificially inflate the scale factors due to correlations between the bins, therefore, it is not used any further in the analysis.

independent sample systematic is decorrelated among all regions. This correlation scheme is a conservative estimate of the underlying uncertainty with the estimation of the $p_{\mathrm{T}}^{\mathrm{miss}}$ trigger scale factors.

6.3. Object-based corrections

Corrections based on the physics objects of an event may alter the weight of the event and will be discussed here, along with corrections to particle kinematics. Furthermore, the selection discussed in Sec. 5.5.2 are not strict requirements on MC events. In particular, events with an object and an associated scale-factor f will contribute to the object selection regions by a weight f. However, this event will also contribute to object veto regions by a weight $1-f$ such that the total contribution of the event across all regions becomes 1, leaving the overall cross-section unchanged. When $f=1$ the strict veto requirements are recovered and the event does not enter the object vetoing regions. The uncertainty on the efficiency of a selection is encoded in the scale-factor, δf, such that the weight where the event is vetoed is

$$1-f = 1-\mathcal{N}_f(f_0,\delta f_1) = 1-(f_0+\mathcal{N}_f(0,1)\delta f_1)\,, \tag{6.4}$$

where f_0 is the central value of the scale-factor, δf_1 is its 1σ variation and $\mathcal{N}_f(\mu,\sigma)$ is a gaussian distributed random variable with a mean μ and width σ. Therefore, MC events vetoed may still contribute to a region, albeit weighted down or possibly negatively. Such a procedure is performed with muons, electrons, photons, τ_h-leptons and b-tagged jets for both the veto and selection definitions.

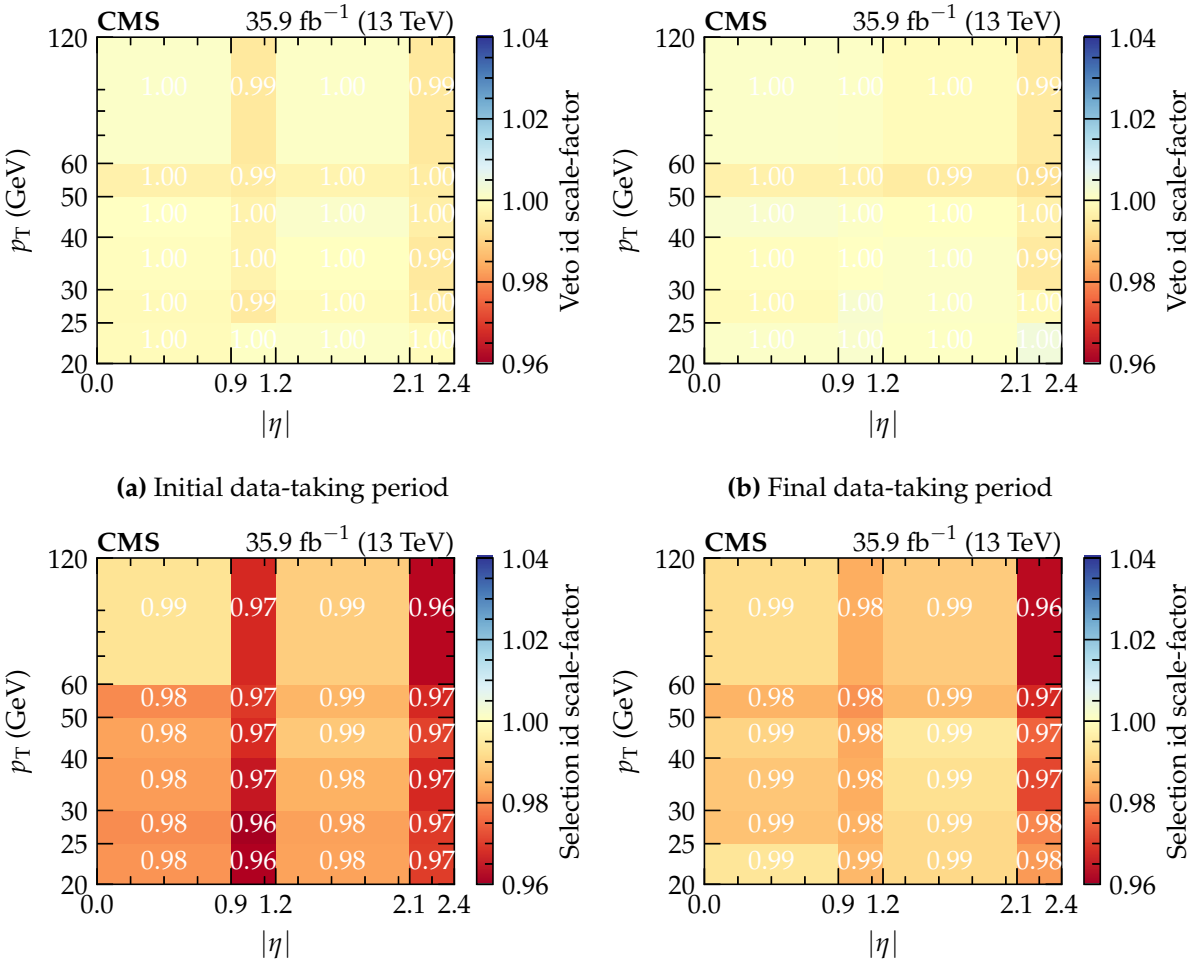
6.3.1. Muons

Muons are formed from tracks with identification and isolation requirements to reduce misidentification. The efficiency in selecting muons is factorised into

$$\varepsilon_\mu = \varepsilon_\mu(\text{ID}|\text{track})\varepsilon_\mu(\text{Iso}|\text{ID})\,, \tag{6.5}$$

where the track collection efficiency is near 100% and hence omitted. The parameter $\varepsilon(\text{ID}|\text{track})$ is the identification efficiency for a given set of tracks and $\varepsilon(\text{Iso}|\text{ID})$ is the isolation efficiency given a collection of muons passing identification requirements. Since the analysis uses two muon identification and isolation requirements the efficiency must be measured for both the veto and selection criteria. The measurement is performed using the tag and probe technique with the tag passing a tight set of requirements to ensure minimal misidentifications and the probe defined as a set of tracks or muons with an isolation requirement [62]. Similarly to the muon trigger efficiency measurements, the events are split into bins of lepton p_T and η. A likelihood fit to the invariant mass distribution of each bin is independently performed to extract the signal contribution for the numerator and denominator of the efficiency: probes passing the isolation and identification requirements and all probes, respectively. A similar approach is performed in MC.

During the 2016 data-taking period, saturation issues in the pre-amplifier chips for the tracker resulted in a loss of hits. This effect was mitigated in the later part of this period and further reduced by a re-processing of the data. Nevertheless, the identification and isolation efficiencies are split into the initial part and final part of the data-taking period. The ratio of the efficiencies measured in data and MC for each period is shown in Fig. 6.6 and 6.7 and subsequently the product of the identification and isolation scale-factors is applied as a correction to MC.



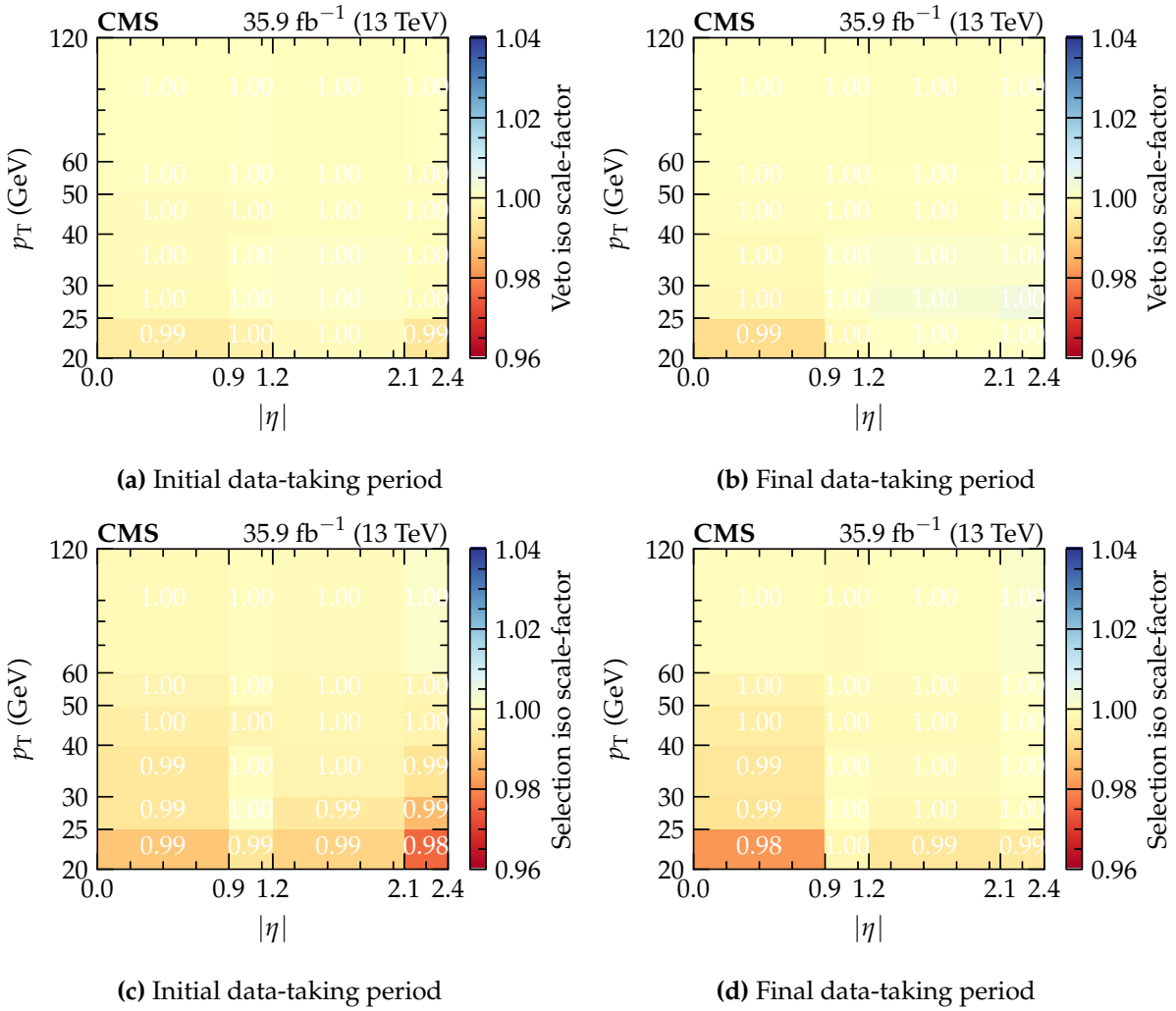
(a) Initial data-taking period

(b) Final data-taking period

(c) Initial data-taking period

(d) Final data-taking period

Figure 6.6.: Muon identification (id) scale corrections applied to MC, parameterised with the muon p_T and $|\eta|$. The top row shows the scale-factors for veto muons with corrections less than 1% as a result of the basic selection. The bottom row shows the scale-factors for selection muons with corrections from 1–4% with the largest discrepancies on the edges of the endcaps ($0.9 < |\eta| < 1.2$ and $2.1 < |\eta| < 2.4$) where the muon identification modelling does not perform as well in MC.



(a) Initial data-taking period

(b) Final data-taking period

(c) Initial data-taking period

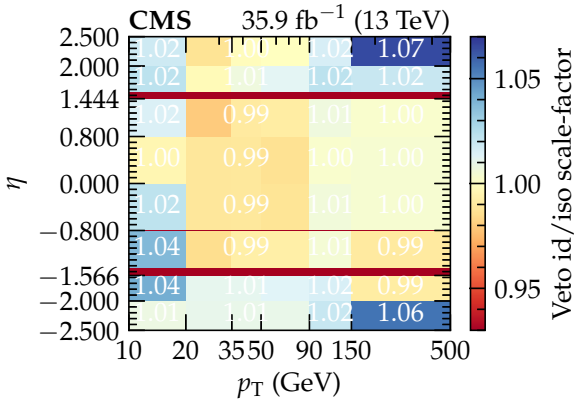
(d) Final data-taking period

Figure 6.7.: Muon isolation (iso) scale corrections applied to MC, parameterised with the muon p_T and $|\eta|$. The modelling of the isolation in MC is better than the identification with at most 2% difference for selection muons with discrepancies only in the low p_T region. The z-axis range is the same as the muon identification scale factor in Fig. 6.6.

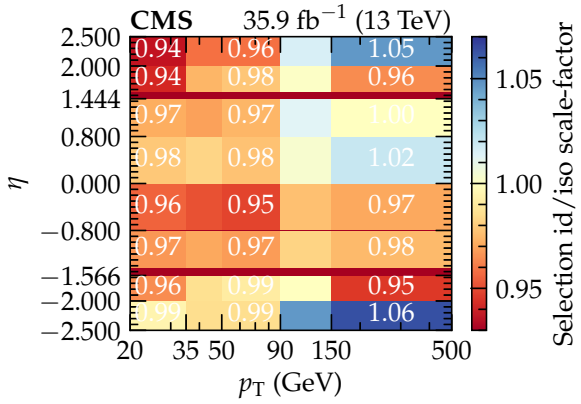
6.3.2. Electrons and Photons

The efficiency of the identification and isolation of electromagnetic showers are measured with the tag and probe technique in a similar manner to the muon-based efficiencies [63]. A fit to the invariant mass distribution of electron pairs is performed to extract the signal yield in various probe p_T and η bins. This allows the efficiencies to be determined with analogous measurements in MC with the counting method. Systematic uncertainties are determined from the variation in efficiencies from altering the tag selection, signal model (template from MC or fully parametric), background model, MC generator (LO against NLO), and additional sources for the photon identification efficiency to account for differences in the electron and

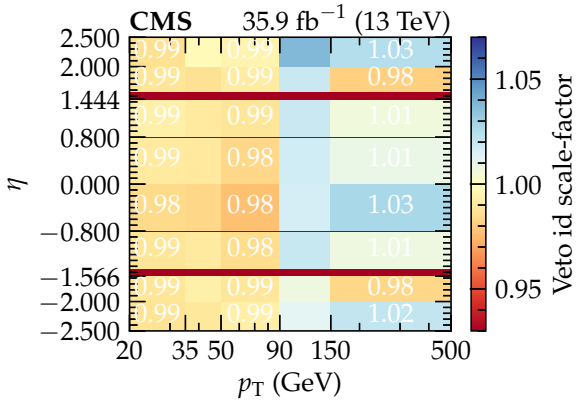
photon showers. The ratio of the efficiency in data and MC is determined, shown in Fig. 6.8 and applied as a correction to MC.



(a) Electron veto



(b) Electron selection



(c) Photon veto

Figure 6.8.: Electron and photon identification scale-factors for the various analysis-level object categorisations. The bands at $1.444 < |\eta| < 1.566$ correspond to the partially uninstrumented gap in the ECAL. The corrections typically arise from the mismodelling of the interactions of electromagnetic particles with the inner detector material resulting in photon conversions and Bremsstrahlung, particularly in the endcaps.

In addition, the efficiency for the reconstruction of GSF tracks matched to ECAL deposits is measured with the same method and shown in Fig. 6.9. Similarly, the efficiency from vetoing photons with consistent track in the pixel detector is measured with a similar method in $Z/\gamma^* \to \mu^+\mu^-\gamma$, where the tag consists of a muon pair and the probe a final state radiated photon. The ratio of the efficiency in data and MC for the GSF track veto is shown in Fig. 6.10.

Finally, the difference in the scale and resolution of electrons and photons in data and MC is measured from the impact on the mean and width of the invariant mass distribution in $Z/\gamma^* \to e^+e^-$ events. A fit to the Z lineshape is performed to extract the mean and width in data and MC. The observed difference is used as a correction to electron and photon energies

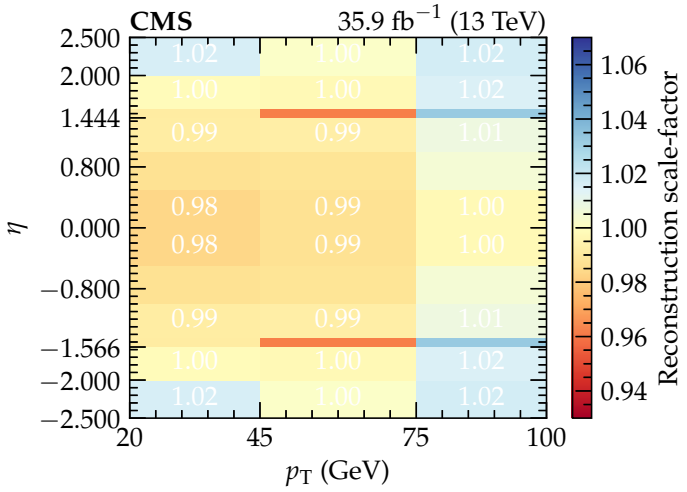


Figure 6.9.: Electron scale-factors associated with the reconstruction of electrons from tracks and ECAL deposits. The ECAL partially uninstrumented gap at $1.444 < |\eta| < 1.566$. Mostly a good agreement between data and MC with deviations for low p_T and endcap electrons, again affected by material interactions.

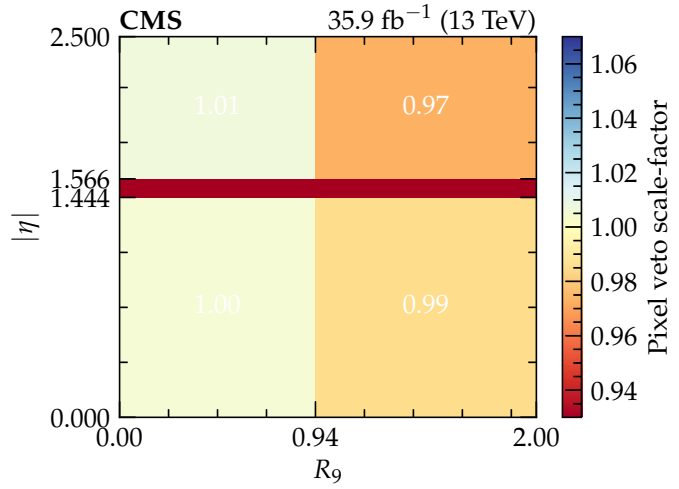


Figure 6.10.: Photon pixel veto scale-factors parameterised by $|\eta|$ and R_9, the ratio of energy deposited in a 3×3 grid of crystals surrounding the seeding crystal with the full supercrystal energy. The R_9 variable is sensitive to unconverted ($R_9 \lesssim 0.94$) and converted ($R_9 \gtrsim 0.94$) photons. The pixel veto removes photon candidates matched to tracks in the pixel detector.

with associated systematic uncertainties from the signal and background models, MC generator and differences between the electron and photon showers.

6.3.3. Jets

The detector response to objects, especially the cluster of particles in jets, is not linear. Therefore, a set of jet energy corrections are measured and applied to map the reconstructed jet energy to the particle-level equivalent (the energy of all generated particles originating from the parton excluding neutrinos). A factorised approach is implemented with each level accounting for different effects [68].

The first step mitigates the contribution of particles from pileup vertices which fall within the cone of the jet. This contribution is determined with the random cone method in zero bias events and in simulations without any hard interactions. Zero bias events are collected by randomly selecting collisions regardless of their content. Since the hard interaction cross-section is small this typically selected events with pileup interactions. The random cone method consists of placing a random set of jet cones in a zero bias event, clustering the particles falling within these cones, and determining the energies of these jets. This energy represents the contribution to a jet from pileup and is parameterised by the event energy density and jet p_{T} and η. A systematic uncertainty is determined by comparing the correction from the random cone method to the generator-level in MC.

The second step involves the correction to particle-level jets. Again, this is determined from QCD dijet simulations by comparing the reconstructed jet p_{T} to particle-level p_{T}, parameterised by the jet p_{T} and η. A systematic uncertainty is determined for this correction as a result of its dependence on the underlying detector calibrations with results from test beam studies.

The third and final step attempts to resolve residual differences between data and MC. These corrections are applied to data since the aim is to replicate the particle-level description which is only available in simulation. This step is split into two stages: the first is an η-dependent correction determined from dijet events with barrel jets used as a reference, and a second a p_{T}-dependent scale difference between data and MC determined from $\mathrm{Z}/\gamma^{*}(\rightarrow \ell\bar{\ell}) + \text{jets}$, $\gamma + \text{jets}$ and QCD multijet events. In these processes the jet energy response is studied by comparing reconstructed jets to particle-level jets on a jet-by-jet basis and also the whole hadronic activity from the correlation of a particle-level jet to the missing transverse energy of the event. These methods do not fully capture the dependence as a result of the correlation between initial (ISR) or final state radiated (FSR) jets. Therefore, an additional correction is determined by probing the η-dependent corrections with a selection to enhance the effect of this correlation. Along with these corrections are associated systematic uncertainties to cover the modelling in MC from an alternative event generator, statistical uncertainties in the data due to trigger prescales, dependence of the corrections over the data taking period, constituent particle energy uncertainties, and the residual difference between the dijet, $\gamma + \text{jets}$ and $\mathrm{Z}/\gamma^{*}(\rightarrow \ell\bar{\ell}) + \text{jets}$ samples after all corrections have been applied.

In addition to the jet energy response difference between data and MC the resolution is corrected for by comparing the balance of jets in a dijet event. The scale determined to correct the MC resolution of the dijet balance is applied to MC jets by smearing their energies with a random variable sampled from a gaussian distribution with a scaled width. Systematic uncertainties are determined by comparing the results to the particle-level imbalance, the effect of removing particles below 10 GeV, non-gaussian tails in the resolution due to rare detector effects and the p_{T}-dependence is propagated through the smearing.

All these corrections, and uncertainties, are propagated through all jet-related quantities such as the missing transverse momentum where the jet energy correction k_i, encoding both scale and resolution corrections, results in a shift in the $p_{\mathrm{T}}^{\mathrm{miss}}$ according to

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} \mapsto \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} + \sum_{i \in \mathrm{jets}} \vec{p}_{\mathrm{T},i}\,(1 - k_i)\ , \tag{6.6}$$

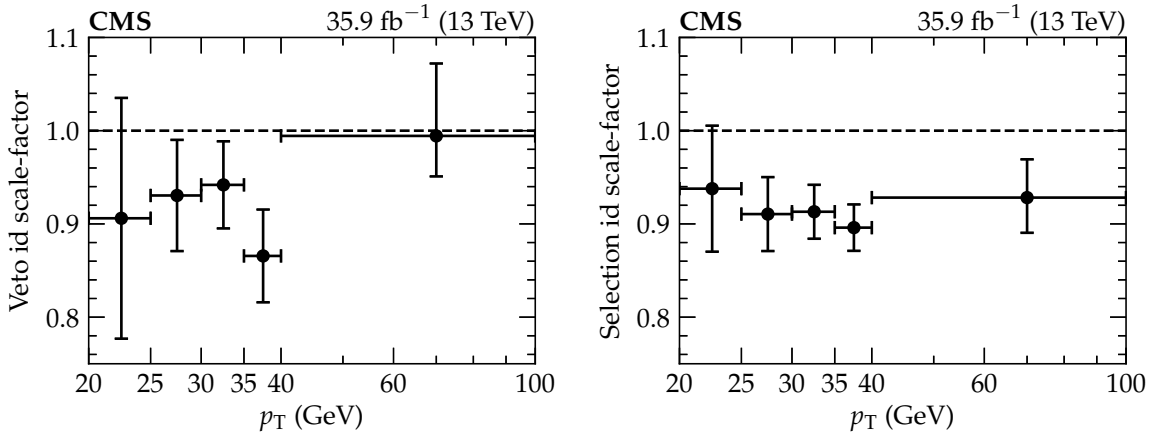
where the sum is over clustered jets, i.e. $p_{\mathrm{T}} > 15\,\mathrm{GeV}$. The corrected $p_{\mathrm{T}}^{\mathrm{miss}}$ quantity is used in subsequent derivatives such as the recoil.

τ_h-tagged jets

The jet energy corrections discussed above are determined under the assumption that the jet originates from a parton. Therefore, the corrections are not applied to τ_h-tagged jets, instead a scale-factor is determined to correct the efficiency of τ_h-lepton identification in MC. The efficiency in data is measured using the tag and probe method in $\mathrm{Z}/\gamma^* \to \tau^+\tau^-$ decays, where one τ-lepton decays into a well reconstructed muon to tag the event and the other into one of the hadronic decay modes to probe the identification. A fit is performed to the invariant mass of the visible particles to extract the signal contribution with templates taken from MC for the signal and most backgrounds. Systematic uncertainties on the identification efficiency are determined from the background normalisation, $p_{\mathrm{T}}^{\mathrm{miss}}$ uncertainties and limited sample sizes [69]. The scale-factors and uncertainties to correct the τ_h identification efficiency in MC is shown in Fig. 6.11.

b-tagged jets

Jets tagged as originating from b-hadrons have the jet energy corrections applied, however, the b-tagging discriminant is not perfectly modelled in MC due to detector simulation limitations and the accuracy of the generator modelling the parton shower and hadronisation. Therefore, a scale-factor is determined for the efficiency of b-tagging jets from QCD multijet, muon-enriched jet, dilepton $t\bar{t}$ and single lepton $t\bar{t}$ samples [65]. The efficiency is parameterised by the jet flavour (grouped in b, c and $udsg$ flavours), p_{T} and η. In simulation, the tagging



(a) Veto identification

(b) Selection identification

Figure 6.11.: τ_h identification scale-factors parameterised by the τ_h p_T. The dashed line is placed at a scale-factor of 1. The veto category is mostly consistent with 1 while the selection category deviates by 5–10% as a result of attempting to reduce misidentifications by probing the complex structure of multiple hadronic τ-lepton decays.

efficiency is calculated by matching jets to generated jets with a particular hadron flavour. Meanwhile, the data efficiency is measured with a pure sample with jets of a certain hadron flavour using a selection which does not bias the jets with respect to the variables in the tagging algorithm. Systematic uncertainties associated to the b-tag scale-factors are determined by varying background normalisations, branching fractions of hadronic decays within the PDG uncertainties, reweighting the distribution of the number of tracks, jet energy scale uncertainties, electron and muon efficiencies and theoretical uncertainties from the event generators. The scale-factors associated with the i-enumerated b-tagged jet, f_i, is measured with a parameteric form with the p_T of the jet given by

$$f_i(p_T) = \begin{cases} p_0 \left(1 + p_1 p_T\right) / \left(1 + p_2 p_T\right), & c \text{ and } b\text{-hadrons} \\ p_0 + p_1 p_T + p_2 p_T^2 + p_3 p_T^3, & udsg\text{-hadrons} \end{cases} \tag{6.7}$$

where p_0, p_1, p_2 and p_3 are parameters measured from a fit to the scale-factors in categories based on the jet flavour, p_T and η. In the context of this analysis, b-tagged jets are vetoed, therefore the event weight becomes

$$w \mapsto w \prod_{i \in \mathrm{b-jets}} \left(1 - f_i\right). \tag{6.8}$$

6.4. Luminosity

The integrated luminosity $\mathcal{L}$ measured for the data-taking period directly scales the weight of a MC event. At CMS the pixels, drift tubes, HF and other detectors perform complimentary measurements of the luminosity by rate counting, R, for visible processes with a cross-section σ_{vis}. The luminosity is determined as

$$\mathcal{L} = \frac{R}{\sigma_{\text{vis}}}\,, \tag{6.9}$$

where the visible cross-section and calibrations of each detector is performed by conducting Van der Meer scans [70] prior to data-taking collisions. The data collected during 2016 corresponds to an integrated luminosity of 35.9 fb^{-1} with uncertainties from the precision of the visible cross-section and calibration measurements along with extrapolations to data-taking collisions resulting in a 2.5% uncertainty [71] on the luminosity measurement.

6.5. Pileup reweighting

The pileup overlay applied to each simulated event is generated by the PYTHIA8.2 package with the number of pileup vertices sampled from a poisson distribution with a tail to higher vertex multiplicities, as shown in Fig. 6.12. In data, for a given instantaneous luminosity, the number of pileup vertices is expected to be poisson distributed. Therefore, a reweighting procedure corrects the MC distribution to reflect that observed in data, without altering the normalisation determined from the product of the process cross-section and the integrated luminosity. The number of interactions in data is estimated from the measured luminosity in each bunch crossing and the average total inelastic cross-section. Furthermore, data measurements are averaged over periods of constant instantaneous luminosity to compare with the mean of poisson distribution sampled in MC, rather than the number of vertices in a particular event. The total inelastic cross-section, measured by CMS, is 69.2 mb, with an uncertainty of 4.6% represented by the bands in Fig. 6.12. These $\pm 1\sigma$ variations are propagated through to the Z invisible width measurement.

6.6. ECAL timing degradation

A gradual degradation in the ECAL timing has caused some L1T ECAL towers to be associated with the incorrect bunch crossing (BX). The L1T logic prevents three successive bunch crossings to be accepted to avoid buffer overflows. Therefore, the deposit in the ECAL is wrongly associated to an early crossing which may lead to a L1T accept, later discarded in the HLT

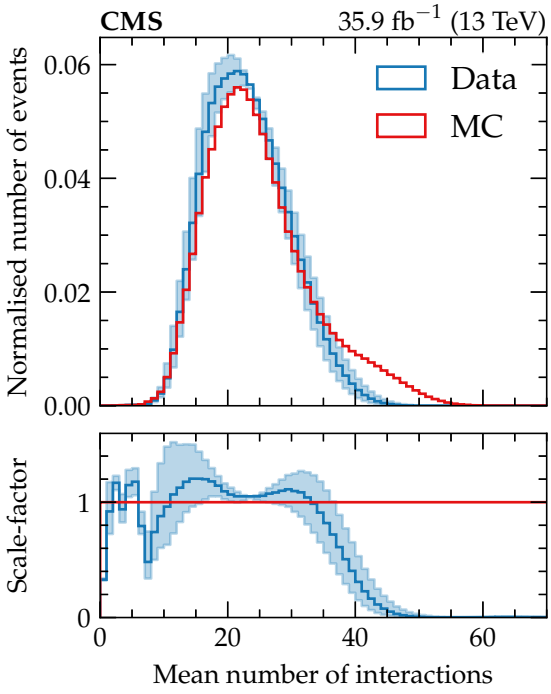


Figure 6.12.: The distribution of the mean number of interactions the events are sampled from, closely following a poisson distribution with a long tail in MC. The uncertainty band is determined by varying the total inelastic cross-section by 4.6%. The average of the mean number of interactions in data is 22.7, under the LHC beam conditions provided during 2016.

where the timing issue is not present. However, the correct crossing is vetoed by the L1T logic resulting in a loss in efficiency in the data collection. This primarily affects events with ECAL deposits in the endcaps where the timing degradation is most prevalent.

The efficiency of this issue is measured by collecting immune events which occur three bunch crossings after a L1T accepted crossing. The immune event is labelled as BX $+ 0$ and the n-enumerated crossing before as BX $- n$ and after as BX $+ n$. Since BX $- 3$ is accepted by the L1T the crossings BX $- 2$ and BX $- 1$ are vetoed, hence they cannot veto BX $+ 0$. Although these events are immune to the issue, they are tagged as events that would have been vetoed if BX $- 1$ or BX $- 2$ contains a L1T object from BX $+ 0$. However, the event is recovered since it is protected by the preceding BX $- 3$. The efficiency is determined from the set of immune events, with and without a problematic L1T object. It is parameterised by the p_{T} and η of the reconstructed jet associated with the problematic L1T object. In MC the efficiency of this issue is determined from a product over the measured efficiencies associated to each object which is applied as a correction to the event weight.

6.7. Theory corrections

The cross-section applied directly to the event weight of a MC process is calculated by theoretical predictions of the highest available order in the strong and electroweak (EW) coupling constants. Associated systematic uncertainties are determined by varying the QCD factorisation and renormalisation scale up and down by a factor of two. This is done for the independent variation of the two scales and the fully correlated variation, with the largest of the six variations taken as the systematic uncertainty. In addition, the uncertainties from the measurements of the sampled NNPDF3.0 parton distribution functions (PDF) [60] are propagated by 100 variations, each randomly sampled from the original distributions of the PDFs.

An alternative method is applied to the signal and dominant V + jets processes. These are generated at NLO in the strong coupling constant, however, instead of scaling the normalisation up to the highest order available, a reweighting of the boson p_{T} distribution is performed for a more accurate prediction. This approach reweights the MC to NNLO QCD and NLO EW with next-to-leading logarithm (NLL) EW corrections [72]. This procedure scales the cross-section parameterised by the boson p_{T}, determined from the generated decay products (with photons within $\Delta R < 0.1$ of a charged lepton summed into their 4-momentum). Systematic uncertainties determined for the QCD corrections include variations in the factorisation and renormalisation scales with the addition of a possible boson p_{T} dependence and encoding the correlation between the Z + jets and W + jets processes.

Another set of systematic uncertainties are determined for the EW corrections to cover higher-order effects and an additional systematic uncertainty on unaccounted contributions from diagrams of mixed order in the strong and EW coupling constants by taking the difference between an additive and factorised combination of the separate QCD and EW corrections. The PDF uncertainties are propagated with the same procedure as for other generated processes. The full set of uncertainties are detailed in Tab. 6.1, with the correlation scheme [72].

6.8. $p_{\mathrm{T}}^{\mathrm{miss}}$ calibration

The missing energy appears in the analysis selection as the recoil variable $\mathcal{U}$, with the criteria $\mathcal{U} > 200\,\mathrm{GeV}$ driven by the trigger thresholds. This parameter acts as a proxy for the boson p_{T} in V + jets processes. The simulation modelling of the recoil as a proxy for the boson p_{T} is confirmed in data by calibration with $\ell\ell$ + jets events. These events have an independent measure of the boson p_{T} through the dilepton p_{T}. Discrepancies in the modelling are typically due to the nonlinear response of the calorimeters to hadronic particles and the modelling of

Label	W + jets	Z + jets	Correlation	Other	Description
$\delta^{(1)}k_{\mathrm{QCD}}$	✓	✓	✓	✓	Factorisation and renormalisation scale variations.
$\delta^{(2)}k_{\mathrm{QCD}}$	✓	✓	✓		Boson p_{T} dependence of the QCD scales.
$\delta^{(3)}k_{\mathrm{QCD}}$	✓				Correlations between QCD effects in Z + jets and W + jets processes, relative to Z + jets.
$\delta^{(1)}k_{\mathrm{EW}}$	✓	✓	✓		Universal unknown Sudakov logarithms beyond NNLO.
$\delta^{(2)}k_{\mathrm{EW}}$	✓	✓			5% of the absolute full NLO EW correction for a conservative estimate of non-Sudakov factors.
$\delta^{(3)}k_{\mathrm{EW}}$	✓	✓			Difference between a naive and rigorous NLL Sudakov approximation.
δk_{mix}	✓	✓	✓		Mixing terms with strong and EW coupling vertices.
δk_{PDF}	✓	✓	✓	✓	100 PDF variations.

Table 6.1.: All systematic uncertainties included from the theory prediction of process cross-sections with a tick for which processes the uncertainty affects split into either W + jets, Z + jets or other processes. A tick in the correlation column signifies the full correlation between the W + jets and Z + jets processes, otherwise the uncertainty is uncorrelated. A brief description of each uncertainty is given.

interactions with the inner detector. A schematic diagram of the event kinematics is shown in Fig. 6.13.

The parallel component $\vec{u}_{\parallel}$ probes imperfect calibrations of the $p_{\mathrm{T}}^{\mathrm{miss}}$ typically from mismeasurement in the hadronic system, whereas the perpendicular component $\vec{u}_{\perp}$ probes the isotropic nature of the energy fluctuations from noise and pileup particles. The magnitude of the $\vec{u}_{\parallel}$ subtracted by $\vec{p}_{\mathrm{T},\ell\ell}$ along with the magnitude of the $\vec{u}_{\perp}$ are expected to be centred on the scale of the $p_{\mathrm{T}}^{\mathrm{miss}}$ and a width corresponding to its resolution. The distribution of these variables is shown in Fig. 6.14.

A maximum likelihood fit is performed to the data, with backgrounds, typically VBS, subtracted by a template determined from MC, and the signal modelled as a Voigtian — the convolution of a Gaussian with a Breit-Wigner which captures the peaking nature of the distribution smeared by the detector resolution. A large tail is seen in the backgrounds within the $u_{\parallel}$ distribution as a result of processes with invisible particles leading to real $p_{\mathrm{T}}^{\mathrm{miss}}$, whereas the background in the $u_{\perp}$ distribution is symmetric. The Voigtian fit is also performed on the signal MC distribution to estimate the scale and resolution in MC. The scale is given by the

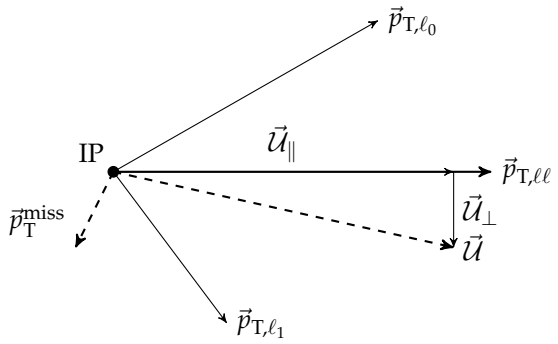


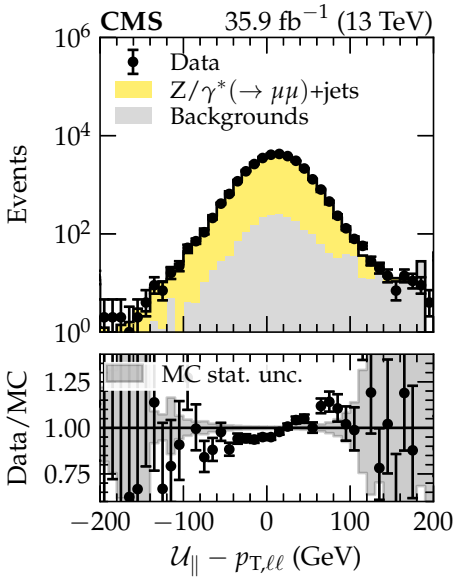
Figure 6.13.: The x-y plane kinematics of a $\mathrm{Z}/\gamma^*(\to \ell\bar{\ell}) + \text{jets}$ event showing the momenta of the leading lepton ℓ_0, subleading lepton ℓ_1, their sum $\vec{p}_{\mathrm{T},\ell\ell}$; the missing transverse momentum $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$; the recoil, defined as $\vec{u} = \vec{p}_{\mathrm{T},\ell\ell} + \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$; and the parallel and perpendicular components of the recoil along the $\vec{p}_{\mathrm{T},\ell\ell}$ axis, $\vec{u}_{\parallel}$ and $\vec{u}_{\perp}$ respectively. The interaction point (IP) is at the centre. The missing energy is spurious as a result of mismeasurement, particles out of acceptance or from pileup vertices, and noise.

mean of the distribution while the width is extracted as

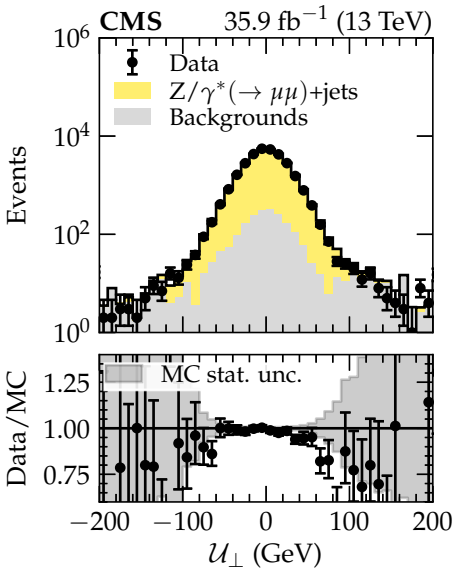
$$\sigma_V = \frac{1}{\sqrt{2\ln 2}} \left(0.5346\gamma + \sqrt{0.2166\gamma^2 + 2\ln(2)\sigma^2} \right) , \tag{6.10}$$

for a Voigtian with a Gaussian centred on u_0 and a width σ and a Breit-Wigner also centred on u_0 and a width γ. When $\gamma \ll \sigma$ the distribution reduces to a Gaussian. These parameters are extracted from a fit in bins of $p_{\mathrm{T},\ell\ell}$ since the detector resolution varies with the energy of particle tracks and deposits. The distribution of $p_{\mathrm{T},\ell\ell}$ is shown in Fig. 6.15 which is divided into 20 bins of equal events in data, rounded to the nearest 5 GeV for convenience. The fit is performed in each $p_{\mathrm{T},\ell\ell}$ bin with 10 GeV bins in the $u_{\parallel} - p_{\mathrm{T},\ell\ell}$ and $u_{\perp}$ dimensions. A prediction for the number of events in each bin is determined from the sum of the Voigtian and the MC background template. The signal normalisation is unconstrained and determined from the fit, along with the u_0, γ and σ_V parameters. Furthermore, the background contribution in each bin is constrained by a gamma distribution[1] centred on the MC prediction. Occasionally the background contribution is negative as a result of NLO MC event generator weights leading to a negative probability density function in the maximum likelihood fit. To avoid the negative contributions, adjacent 10 GeV bins are merged until all background contributions are positive. This proves to be a robust method while maintaining a satisfactory number of bins. Additional fitting issues arise when the γ parameter is near the lower limit at zero. In such cases the γ parameter is fixed at zero and reliable results are recovered. These precautions are also used in the MC fits. Example fits in data and MC are shown in Fig. 6.16. The good agreement in the data fit implies the statistical uncertainty in the background dominates

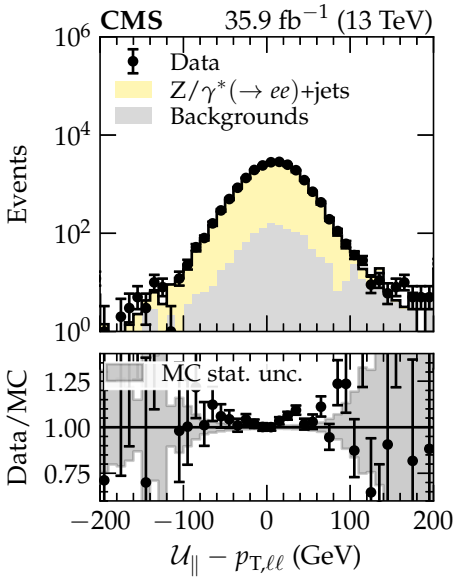
[1] The gamma distribution is the generalised poisson distribution to real numbers, such as for weighted events from simulations.



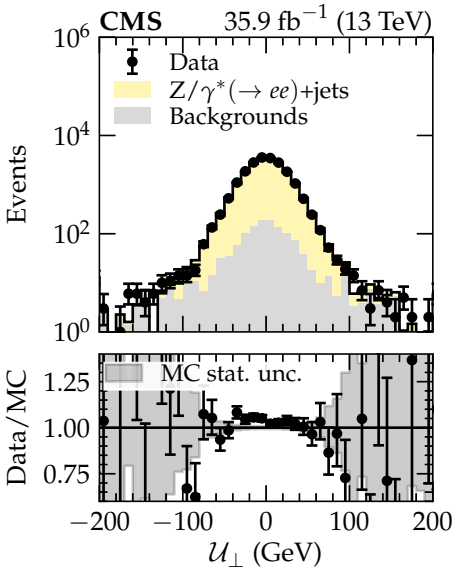
(a) Dimuon parallel component



(b) Dimuon perpendicular component



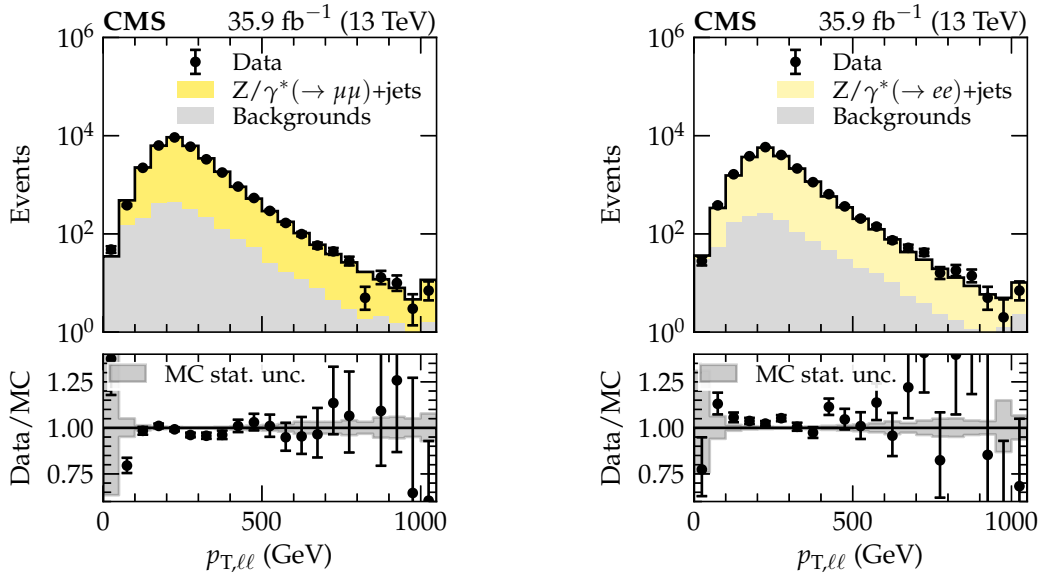
(c) Dielectron parallel component



(d) Dielectron perpendicular component

Figure 6.14.: Distributions of the $\mathcal{U}_{\parallel} - p_{\mathrm{T},\ell\ell}$ and $\mathcal{U}_{\perp}$ for the $\mu\mu$ + jets and ee + jets regions with the $\mathcal{U}$ requirement relaxed to $\mathcal{U} > 100\,\mathrm{GeV}$ to probe their dependence with lower recoil while maintaining a reliable $p_{\mathrm{T}}^{\mathrm{miss}}$ trigger scale factor measurement.

over the systematic uncertainties. Conversely, the systematic uncertainties in the MC fit may dominate. To determine the effect of the systematic uncertainties, the MC fit is performed under the $\pm 1\sigma$ hypotheses for all systematic variations and the difference between the nominal fit parameters is taken as a systematic uncertainty. These are combined in quadrature for the final MC fit uncertainties. Most systematic uncertainties alter the normalisation of these



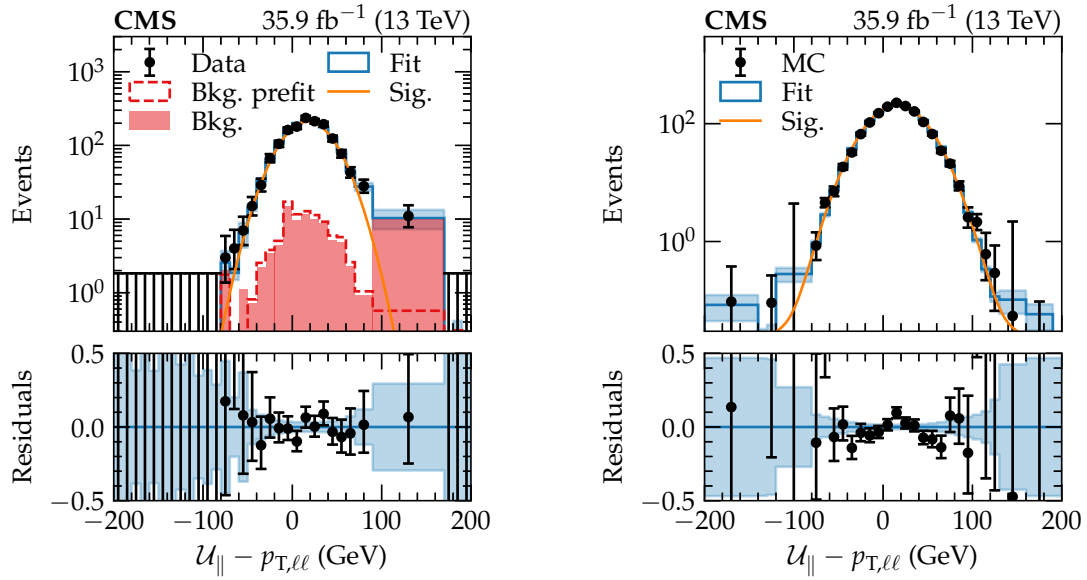
(a) $\mu\mu$ + jets **(b)** ee + jets

Figure 6.15.: Dilepton $p_{T,\ell\ell}$ distributions peaking near 200 GeV as a result of the kinematic selection placed on the recoil and the leading jet.

distributions and hence do not impact the scale and the resolution. However, the lepton and jet energy scale uncertainties have the largest impact.

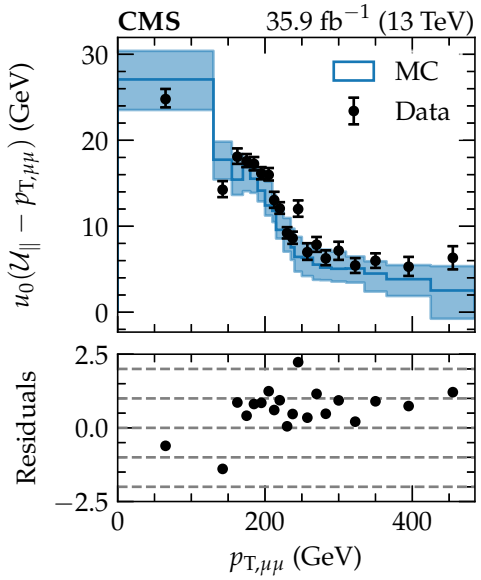
The scale u_0 and resolution σ_V extracted from the fits for each $p_{T,\ell\ell}$ bin is shown in Fig. 6.17 for the $\mu\mu$ + jets region and Fig. 6.18 for the ee + jets region. The scale and resolution in both the $\mu\mu$ + jets and ee + jets region agree within their respective uncertainties. Therefore, no additional corrections are derived for the scale or the resolution since the recoil as a proxy for the boson p_T is well-modelled in MC. A few outliers with residuals beyond ± 1 are seen; however, the number is not significantly beyond expectations given the number of bins probed. Furthermore, the most important parameter, the parallel scale, shows an overestimate in the systematic uncertainties.

Additional studies into the recoil calibration with the number of pileup vertices have been performed elsewhere [73] and show good agreement between data and MC.

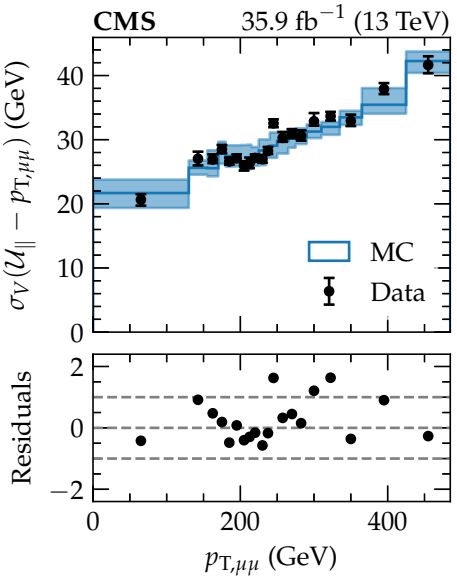


(a) Data fit **(b)** MC fit

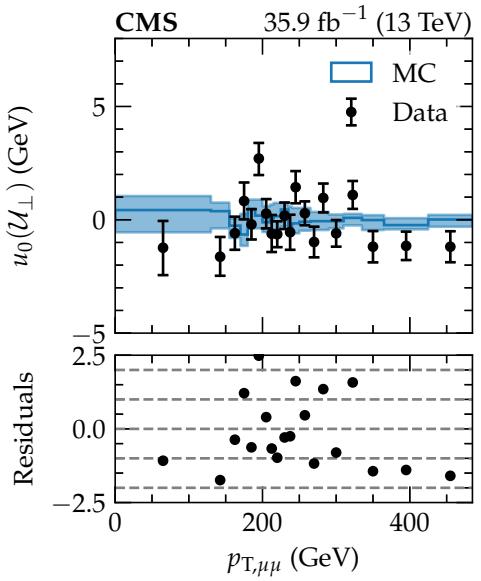
Figure 6.16.: Fit results for the $\mathrm{Z}/\gamma^* \to \mu^+\mu^-$ process in $p_{\mathrm{T},\ell\ell}$ bin 180–190 GeV in data and MC; showing the fit result, background (Bkg.) prefit and postfit, signal Voigtian (Sig.) and the total fit result and it's uncertainty band; for the $\mathcal{U}_{\parallel} - p_{\mathrm{T},\ell\ell}$ distribution. The merging of bins is apparent with the wide bin where the background estimation increases by a factor of 10 as a result of large event weights leading to low effective MC statistics with this wide bin. The residuals represent the difference between the data points and the fit result with good agreement in the data fit and slightly worse in the MC fit as systematic uncertainties are not included to cover the discrepancy.



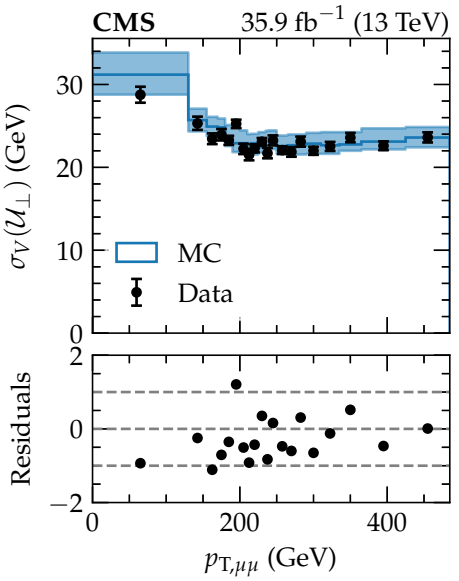
(a) Parallel scale



(b) Parallel resolution

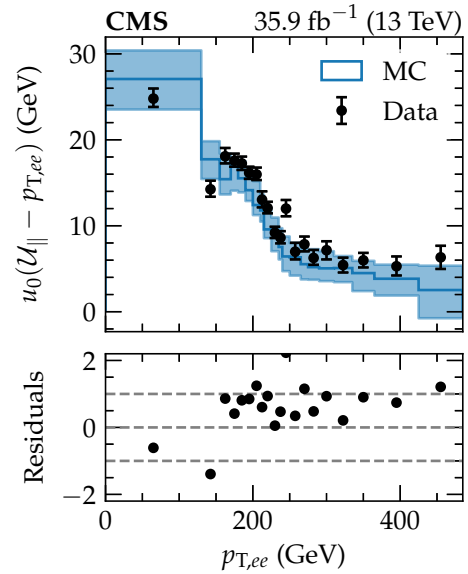


(c) Perpendicular scale

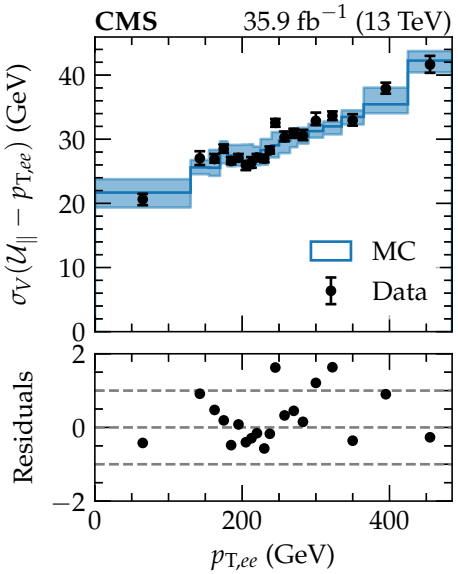


(d) Perpendicular resolution

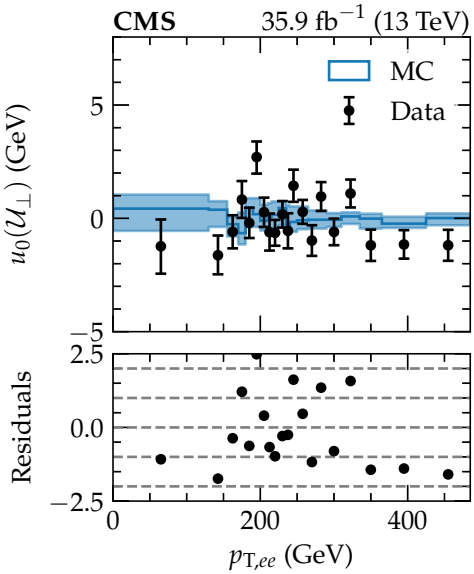
Figure 6.17.: The $p_{\mathrm{T}}^{\mathrm{miss}}$ scale and resolution from the fit to data and MC in the $\mu\mu$ + jets region. The residuals represent the difference between data and MC, scaled by the reciprocal of the uncertainties added in quadrature.



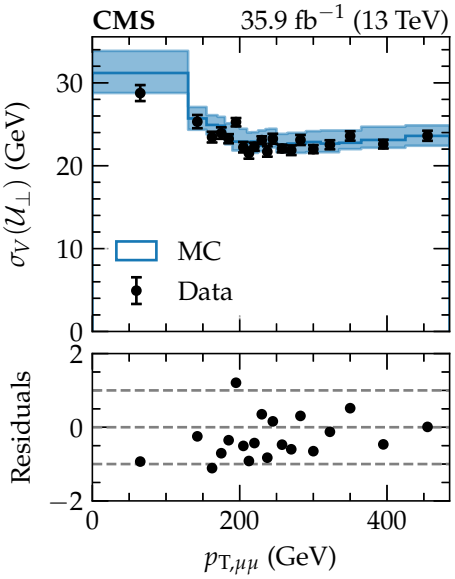
(a) Parallel scale



(b) Parallel resolution



(c) Perpendicular scale



(d) Perpendicular resolution

Figure 6.18.: Similar results to Fig. 6.17 in the ee + jets region.

6.9. Summary

Corrections are applied to simulation to improve the modelling of data, typically determined from data-driven measurements of scale factors to correct the efficiency of a selection in MC to match the data efficiency, along with systematic uncertainties representing limitations in the scale factors.

The first set of corrections fix the emulated trigger efficiency in MC for the single muon and electron trigger paths measured in $Z/\gamma^* \to \ell\bar{\ell}$ events with the tag and probe method. The p_T^{miss} trigger scale factors are determined from events collected by the single muon trigger paths as a reference. Systematic uncertainties estimate the effect of the muon mismatch between the HLT and offline selection, the use of the single muon trigger reference and variations from an independent region. After the trigger corrections are applied, object based selection corrections are measured, typically with the tag and probe method, for the muon identification and isolation; electron identification, isolation and reconstruction; photon identification and track veto; and jet b and τ_h-tag identification. Similarly, the energy and p_T scale corrections and systematic uncertainties are measured for muons, τ_h, electrons, photons and jets, all of which are propagated through the recoil calculation. Global event-based corrections and uncertainties are determined for the luminosity measurement, pileup reweighting, ECAL timing degradation and theory corrections to higher precision in QCD and electroweak vertices including QCD factorisation and renormalisation scale and PDF variations. The use of the recoil variable as a proxy for the boson p_T in $Z(\to \nu\bar{\nu}) + \text{jets}$, $W(\to \ell\nu) + \text{jets}$ and $Z/\gamma^*(\to \ell\bar{\ell}) + \text{jets}$ processes is probed in $Z/\gamma^*(\to \ell\bar{\ell}) + \text{jets}$ events by splitting the recoil into a parallel and perpendicular component to extract the scale and resolution in both data and MC. The results show good agreement, within uncertainties, hence no additional corrections or systematic uncertainties are required.

All corrections are applied to the relevant processes in simulated events. Alternative distributions are provided under the $\pm 1\sigma$ variations for later likelihood fits to extract the $Z(\to \nu\bar{\nu}) + \text{jets}$ and $Z(\to \ell\bar{\ell}) + \text{jets}$ event yields in determining the Z invisible width.

Chapter 7.

Background estimation

7.1. Introduction

The processes used to extract the Z invisible width are $Z \to \nu\bar{\nu}$ and $Z \to \ell\bar{\ell}$. The $Z \to \ell\bar{\ell}$ process is selected by the $\mu\mu$ + jets and ee + jets regions and receives small contributions from background processes, predominantly diboson, as a result of the signal enhancement from the invariant mass selection. Therefore, these backgrounds are estimated from MC after the corrections discussed in Chap. 6 are applied. Furthermore, the distinction between the $Z/\gamma^* \to \ell\bar{\ell}$ and $Z \to \ell\bar{\ell}$ processes requires an estimate of the contribution of the γ^* background as well as the interference between Z and γ^*. With the invariant mass selection, the γ^* and interference contributions are percent-level and hence estimated directly from MC simulations where the Z and γ^* contributions are subsequently turned off in the event generation. Meanwhile, the $Z \to \nu\bar{\nu}$ process is collected in the $p_{\mathrm{T}}^{\mathrm{miss}}$ + jets region which is contaminated by the $W \to \ell\nu$ decay where the lepton is out of acceptance or a misidentified τ_h. This $W \to \ell\nu$ background is estimated by a data-driven control region method. The remaining backgrounds are significantly smaller and estimated from MC, apart from the QCD multijet background which, although small, is estimated with a data-driven method because of large uncertainties associated with the QCD multijet prediction in MC. All these estimation methods are discussed in this chapter.

The estimation methods are performed in orthogonal control regions to the $p_{\mathrm{T}}^{\mathrm{miss}}$ + jets and $\ell\ell$ + jets signal regions. However, decisions on further categorisation within the regions is closely aligned to the signal regions. Although the determination of the Z invisible width requires the overall event yield for the $Z \to \nu\bar{\nu}$ and $Z \to \ell\bar{\ell}$ processes, the $\mathcal{U}$ is divided into bins to exploit the background and systematic uncertainty dependence on this parameter. The chosen bin widths are approximately equivalent to the resolution of the $\mathcal{U}$ with an open-ended final bin, as detailed in Tab. 7.1.

200–220	220–250	250–280	280–310	310–340	340–370	370–400
400–430	430–470	470–510	510–550	550–590	590–640	640–690
690–740	740–790	790–840	840–900	900–960	960–1020	1020–1090
1090–1160	1160–1250	1250–1400	1400–inf			

Table 7.1.: Division of the recoil, in GeV, into 25 bins for the background estimation and Z invisible width extraction. Exploiting the shape of the recoil distribution exposes regions with larger signal to background ratios and various systematic uncertainty contributions.

7.2. $W \to \ell\nu$ prediction

The $W \to \ell\nu$ process contribution in the p_T^{miss} + jets region is estimated from an orthogonal control region defined by the reconstruction and selection of a single lepton, rather than vetoing all leptons. Two control regions are defined: the μ + jets and e + jets, both are used in this estimation method as a source of well-reconstructed leptons and complement each other. These regions are linked to the p_T^{miss} + jets region through the transfer factor method, discussed in the following section. In addition to the two control regions, the τ_h + jets is defined as a validation region to test the use of muon and electron final states to primarily predict the τ-lepton decay of the $W \to \ell\nu$ process.

7.2.1. μ + jets control region

The μ + jets control region provides the primary measurement for the data-driven estimate of $W \to \ell\nu$ in the p_T^{miss} + jets region. It differs from this region by requiring a single well-reconstructed muon in the final state along with a transverse invariant mass between the muon and the p_T^{miss} between 30 and 125 GeV. The kinematic distributions for this region are shown in Fig. 7.1. The most prominent processes include the signal of interest, $W(\to \mu\nu)$ + jets (82.1%), along with backgrounds from: $W \to \ell\nu$ with a τ-lepton decaying into a muon (5.4%); t-quark production (5.3%) and subsequent decay into a muon, jets and neutrinos; diboson production (2.0%); Drell-Yan production of dimuons (2.0%) with one muon out of acceptance or misreconstructed; QCD multijet (1.3%) and other processes (1.9%). The shape of the data and MC distributions are in agreement. Dependence, for instance, with the lead muon η is covered by systematic uncertainties which are larger in the endcaps where the particle flux is greater and the magnetic field is more complex. However, there is a significant discrepancy of 10% between the data and MC normalisation, currently attributed to a normalisation issue with the W + jets production[1]. The analysis is designed to be robust against incorrect normalisations

[1] This issue has been identified as a problematic cross section deviating from the true theory prediction by about 10% across the full boson p_T range for the W + jets process.

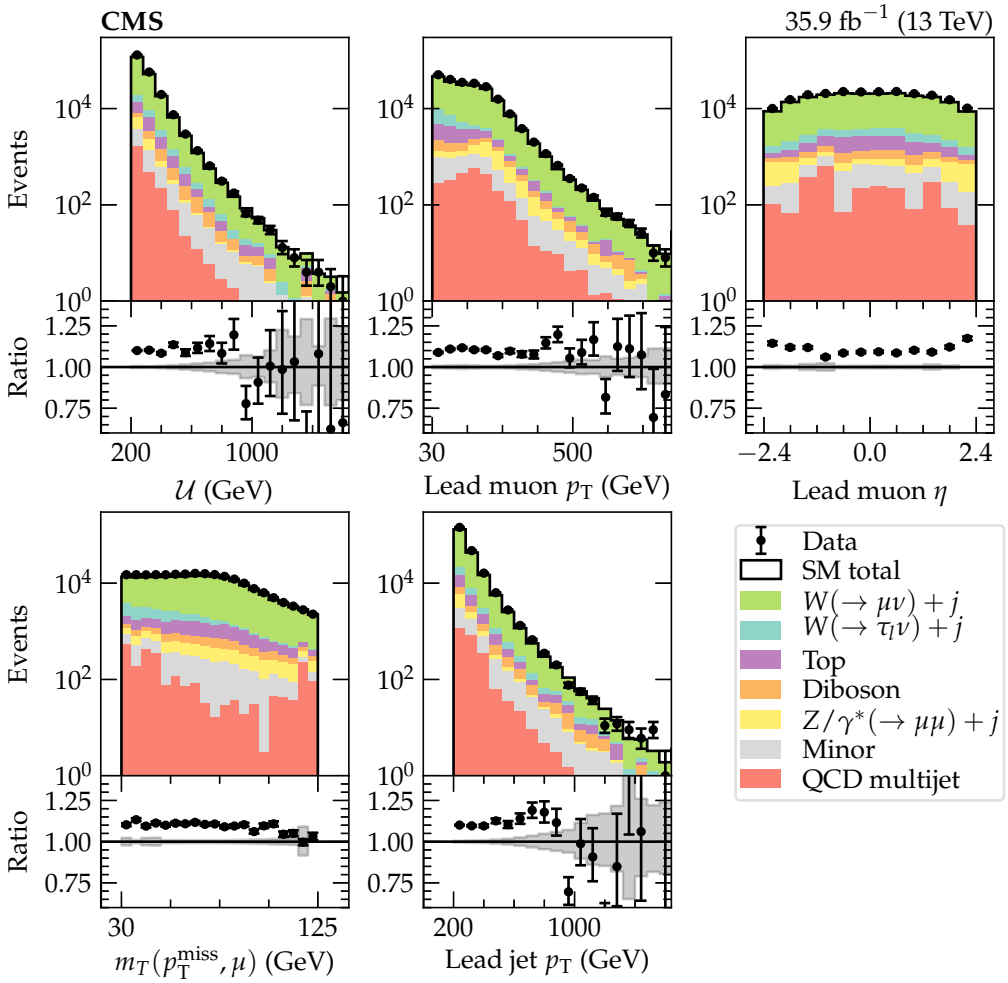


Figure 7.1.: Distributions of kinematic variables within the μ + jets region. The ratio of data to MC is shown in the lower sub-panels along with the statistical uncertainty associated with the MC prediction, without any systematic uncertainties. Processes with the largest cross sections are labelled with a grouping of the smallest contributions into the 'minor' background.

with the V + jets process through data-driven measurements. This is investigated further by inspecting the e + jets control region.

7.2.2. e + jets control region

The selection for the e + jets control region is closely aligned to the μ + jets to provide a complementary measurement of the W($\to \ell\nu$) + jets process. However, the additional selection $p_T^{miss} > 100\,\text{GeV}$ reduces QCD multijet contamination within this region and the statistical power to measure the W($\to \ell\nu$) + jets process. The kinematic distributions of this region are shown in Fig. 7.2 where the same 10% discrepancy between the normalisation in data and MC is observed, although the shapes of the distributions are in good agreement. Therefore,

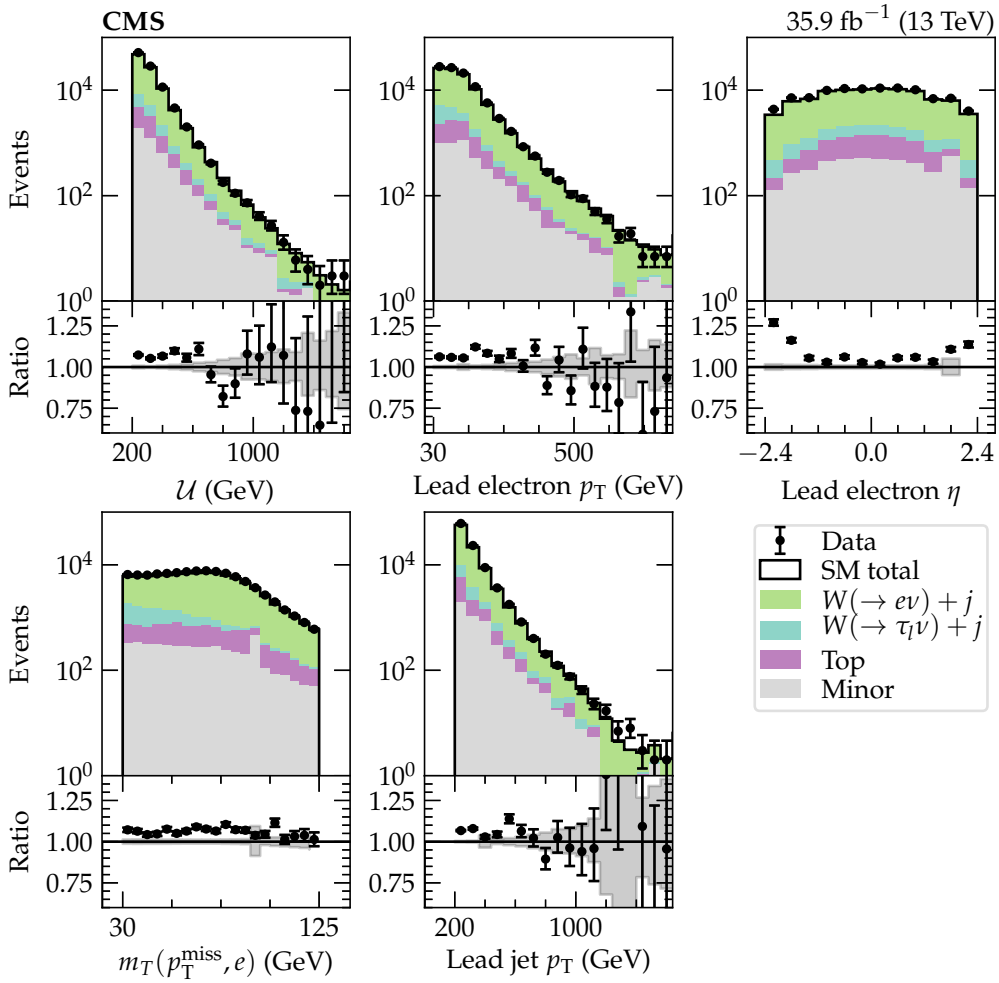


Figure 7.2.: Kinematic variable distributions of the e + jets control region. The ratio of data to MC gives an indication of their agreement within statistical uncertainties on both the data and MC. However, no systematic uncertainties are included. The most prominent processes are the electron (81.1%) and τ-lepton (7.4%) decays of the W boson, with subsequently smaller contributions from t-quark production (6.6%) and other 'minor' backgrounds which include diboson, $Z/\gamma^* \rightarrow e^+e^-$ and QCD multijet processes. The asymmetry in the lead electron η is covered by electron identification uncertainties which are significantly larger in the forward η regions.

the normalisation is not attributed to the muon or electron reconstruction and identification or the trigger since they differ for the μ + jets and e + jets regions. This supports the original hypothesis of a normalisation issue with the W + jets process. Under this hypothesis the normalisation is expected to persist into the p_T^{miss} + jets region, scaled down by the relative contribution of the W + jets process. Therefore, the data-driven estimate of this process is vital for an accurate background prediction. The τ_h + jets region is explored to further support or refute this hypothesis. The different background composition compared to Fig. 7.1 is driven by the additional p_T^{miss} selection in the e + jets region.

7.2.3. τ_h + jets validation region

The τ_h + jets validation region is more aligned with the $p_{\mathrm{T}}^{\mathrm{miss}}$ + jets region than the μ + jets or e + jets control regions since no well-reconstructed lepton exists in the final state, instead a hadronic cluster is tagged as the decay product of a τ-lepton. However, as shown in Fig. 7.3, this region receives a substantial contribution from other hadronic final states with jets mistagged as a τ_h and a lower sample size due to the τ_h-tagging efficiency with a large associated uncertainty. Therefore, for the purpose of measuring the W + jets process, the τ_h + jets region does not

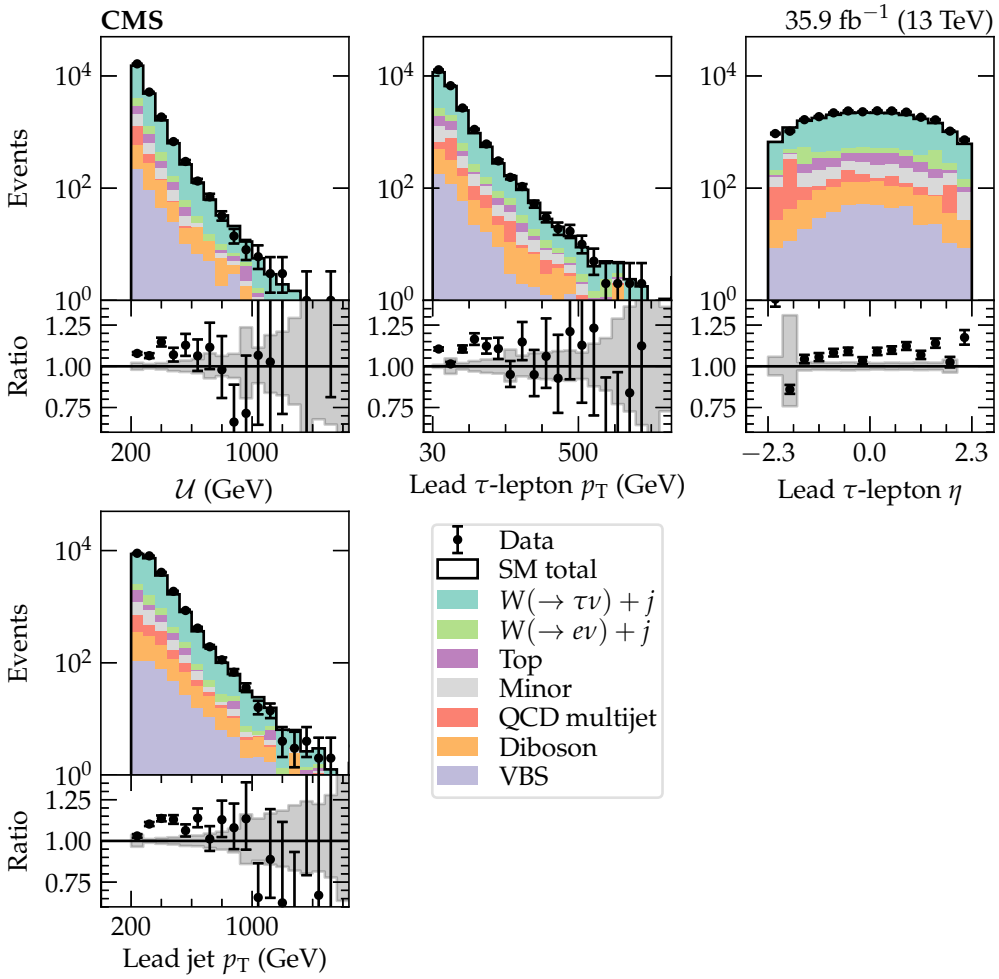


Figure 7.3.: Kinematic distributions for the τ_h + jets validation region. The W + jets process is dominant with the τ-lepton (73.4%) and electron (6.9%) decays, with subsequently smaller backgrounds from t-quark production (5.7%); QCD multijet (3.6%); diboson production (3.2%); vector boson scattering, VBS (1.8%); and other contributions grouped into the 'minor' background.

provide a similar precision to the μ + jets and e + jets regions. However, as a validation region it supports the normalisation offset associated with the W + jets process at 8%, with the shapes of the kinematic distribution in good agreement, within uncertainties, between data and MC.

7.2.4. Systematic uncertainties

The W + jets background estimation is based on a data-driven estimate to improve the MC prediction. The systematic uncertainties associated with the MC predictions, most notably from limited knowledge of selection efficiencies, impact this estimation. The full set of systematic uncertainties affecting the background estimation is as follows.

- All **theory uncertainties** discussed in Sec. 6.7.
- **Veto efficiencies** associated to uncertain corrections in vetoing b-jets, photons and τ_h-leptons. The photons have two associated uncertainties: identification and pixel seed veto efficiencies.
- **Proton beam uncertainties** such as the luminosity measurement (2.5 %) and the pileup profile.
- **Detector-based uncertainties** due to online issues whilst recording data. The most significant uncertainty is due to the ECAL timing issue.
- **Trigger efficiencies** as described in Sec. 6.2.3 and 6.2.2. The $p_{\mathrm{T}}^{\mathrm{miss}}$ trigger efficiency uncertainty is split into three sources: reference trigger, muon multiplicity and independent region. The reference trigger uncertainty is correlated between all regions whereas the muon multiplicity is split into the 0 μ, 1 μ and 2 μ regions and the independent region systematic is uncorrelated between all regions.
- **Muon efficiency uncertainties** associated to veto and selection muon identification and isolation with an uncorrelated statistical and systematic component.
- **Electron efficiency uncertainties** associated to veto and selection electron identification and reconstruction.
- **Jet energy correction uncertainties** determined by varying the scale (JES) and resolution (JER) of jet energies.
- **Unclustered energy uncertainty** encoding the impact of energy deposits, below the clustering threshold, on $p_{\mathrm{T}}^{\mathrm{miss}}$.

7.2.5. Transfer factor method

The transfer factor method is based on applying correction factors, determined from control regions, to improve background predictions in a signal region whilst reducing the systematic uncertainties associated with this estimation. The W + jets prediction in the $p_{\mathrm{T}}^{\mathrm{miss}}$ + jets region,

$N_{\text{pred}}^{p_{\text{T}}^{\text{miss}}+\text{jets}}(\text{W}+\text{jets})$, is determined from the $\ell+\text{jets}$ regions as

$$N_{\text{pred}}^{p_{\text{T}}^{\text{miss}}+\text{jets}}(\text{W}+\text{jets}) = \frac{N_{\text{MC}}^{p_{\text{T}}^{\text{miss}}+\text{jets}}(\text{W}+\text{jets})}{N_{\text{MC}}^{\ell+\text{jets}}(\text{W}+\text{jets})} \left(N_{\text{obs}}^{\ell+\text{jets}} - N_{\text{MC}}^{\ell+\text{jets}}(\text{bkgs.}) \right) , \tag{7.1}$$

where $N_{\text{MC}}^{p_{\text{T}}^{\text{miss}}+\text{jets}}(\text{W}+\text{jets})$ and $N_{\text{MC}}^{\ell+\text{jets}}(\text{W}+\text{jets})$ are the MC predictions for the W + jets process in the $p_{\text{T}}^{\text{miss}}+\text{jets}$ and $\ell+\text{jets}$ regions, respectively; $N_{obs}^{\ell+\text{jets}}$ is the observed number of events in data for the $\ell+\text{jets}$ regions, and $N_{\text{MC}}^{\ell+\text{jets}}(\text{bkgs.})$ is the MC prediction for the backgrounds (processes other than W + jets) in the $\ell+\text{jets}$ region. The transfer factor is given by the ratio

$$t_{\ell+\text{jets}}^{p_{\text{T}}^{\text{miss}}+\text{jets}}(\text{W}+\text{jets}) = \frac{N_{\text{MC}}^{p_{\text{T}}^{\text{miss}}+\text{jets}}(\text{W}+\text{jets})}{N_{\text{MC}}^{\ell+\text{jets}}(\text{W}+\text{jets})} . \tag{7.2}$$

Systematic uncertainties common in both the numerator and denominator typically cancel, reducing the systematic uncertainty associated with the transfer factor, with respect to the individual MC predictions. In practice, the transfer factor method is implemented within the likelihood formalism (App. A) by scaling the W + jets process in the signal and control regions by a common unconstrained parameter r_{W} determined from a simultaneous fit to the data in the $p_{\text{T}}^{\text{miss}}+\text{jets}$ and $\ell+\text{jets}$ region. Here, the likelihood is given by the product over all $\mathcal{U}$ bins with two poisson terms for the $p_{\text{T}}^{\text{miss}}+\text{jets}$ and $\ell+\text{jets}$ region with the expectation values

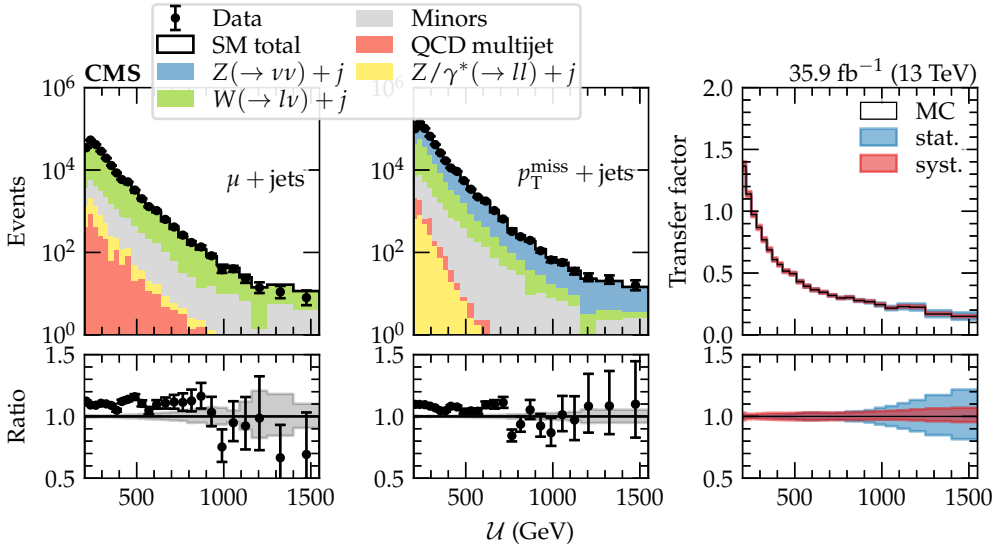
$$r_{\text{Z}} s_{\text{Z}\to\nu\bar{\nu}}^{p_{\text{T}}^{\text{miss}}+\text{jets}}(\boldsymbol{\theta}, \phi^{p_{\text{T}}^{\text{miss}}+\text{jets}}) + r_{\text{W}} s_{\text{W}\to\ell\nu}^{p_{\text{T}}^{\text{miss}}+\text{jets}}(\boldsymbol{\theta}, \phi^{\mu+\text{jets}}) + b^{p_{\text{T}}^{\text{miss}}+\text{jets}}(\boldsymbol{\theta}, \phi^{p_{\text{T}}^{\text{miss}}+\text{jets}}) \tag{7.3}$$

and

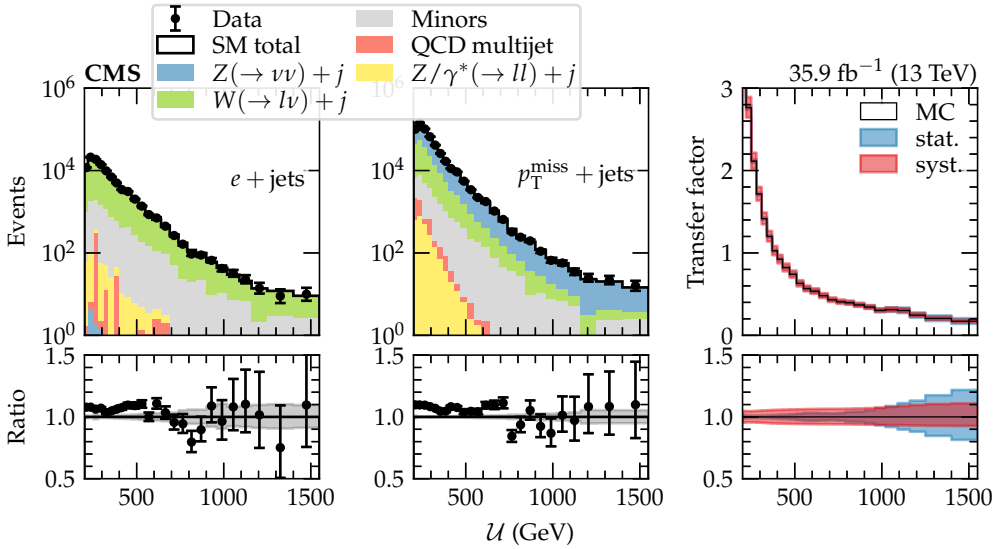
$$r_{\text{W}} s_{\text{W}\to\ell\nu}^{\ell+\text{jets}}(\boldsymbol{\theta}, \phi^{\ell+\text{jets}}) + b^{\ell+\text{jets}}(\boldsymbol{\theta}, \phi^{\ell+\text{jets}}) , \tag{7.4}$$

respectively. The $\text{Z}\to\nu\bar{\nu}$ contribution in the $p_{\text{T}}^{\text{miss}}+\text{jets}$ region, $s_{\text{Z}\to\nu\bar{\nu}}$, is scaled by the parameter of interest r_{Z}; and similarly by r_{W} for the $\text{W}\to\ell\nu$ process $s_{\text{W}\to\ell\nu}$. The remaining backgrounds, b, and all other processes depend on the gaussian constrained, $\boldsymbol{\theta}$, and poisson constrained MC statistical, ϕ, nuisance parameters. Superscripts denote the parameter's associated region. The scaling parameters r_{Z} and r_{W} are correlated across all $\mathcal{U}$ bins, and estimated from a fit to the data. Prior to the fit the MC predictions and the transfer factors are inspected, as shown in Fig. 7.4.

The transfer factor is a steeply falling distribution where leptons from W decays are more likely to be below the p_{T} acceptance, misreconstructed or misidentified with a lower recoil. However, the p_{T} of the lepton increases with the recoil and is significantly harder to miss, resulting in the transfer factor tending to zero. Furthermore, the transfer factor is larger for the



(a) Transfer factor for W + jets from the μ + jets to p_T^{miss} + jets region.



(b) Transfer factor for W + jets from the e + jets to p_T^{miss} + jets region.

Figure 7.4.: Distributions of the $\mathcal{U}$ variable in the μ + jets, e + jets and p_T^{miss} + jets with the transfer factor determined from the W + jets MC estimates. The transfer factor includes the uncertainties associated with the limited number of generated MC events and systematic uncertainties from all sources whereas the kinematic distributions only include the statistical uncertainties. The distributions also include the analysis relevant processes.

e + jets predictor rather than μ + jets because of the tighter selection and lower reconstruction efficiency associated with the electrons.

The systematic uncertainties associated with the μ + jets and e + jets transfer factor are split into the theoretical and experimental uncertainties. The relative systematic uncertainties

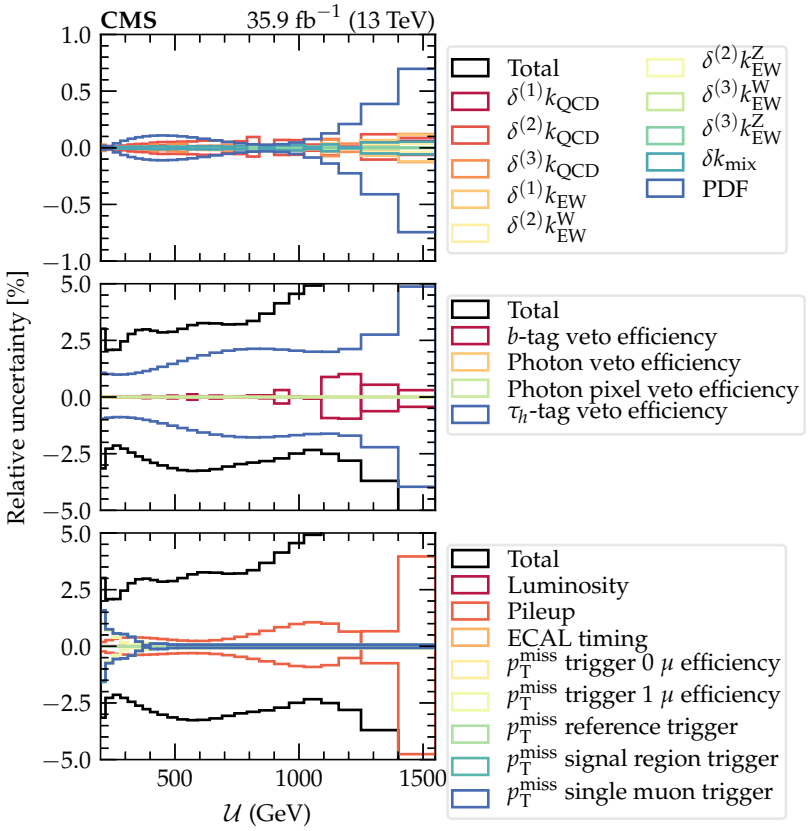


Figure 7.5.: Relative systematic uncertainties (%) on the W + jets transfer factor for the p_T^{miss} + jets region predicted by the μ + jets control region. The uppermost panel includes all the theoretical uncertainties, while the others include experimental uncertainties. The total uncertainty, depicted in each panel, represents all uncertainties summed in quadrature.

associated with the transfer factor $t_{\mu+\text{jets}}^{p_T^{miss}+\text{jets}}(\text{W}+\text{jets})$ and $t_{e+\text{jets}}^{p_T^{miss}+\text{jets}}(\text{W}+\text{jets})$ are shown in Figs. 7.5–7.8, with the uncertainties prior to the transfer factor shown in App. A.3. The theoretical uncertainties primarily cancel as they affect the W + jets process equally in the p_T^{miss} + jets and μ + jets regions. However, the parton distribution function (PDF) uncertainty does not completely cancel as different partons initiate the interaction as a result of kinematic sculpting by requirements on muon p_T and η which are not applied in the p_T^{miss} + jets region. Nevertheless, these contributions are small in comparison to the experimental uncertainties. The dominant systematic uncertainties include the jet energy scale, τ_h-tag veto efficiency and pileup reweighting uncertainty with large contributions from the p_T^{miss} trigger efficiency for $\mathcal{U} < 250\,\text{GeV}$. The alternative templates representing the $\pm1\sigma$ variations in the jet energy scale uncertainty are statistically limited in their precision, whereas they are assumed to be precise within the likelihood fit, leading to non-cancellations in the transfer factor. To partially redeem this issue the alternative templates are smoothed in an attempt to remove statistical fluctuations (App. A). Although this proves to be beneficial in obtaining a good fit to the data the resulting uncertainty is typically dominated by the limited MC sample size.

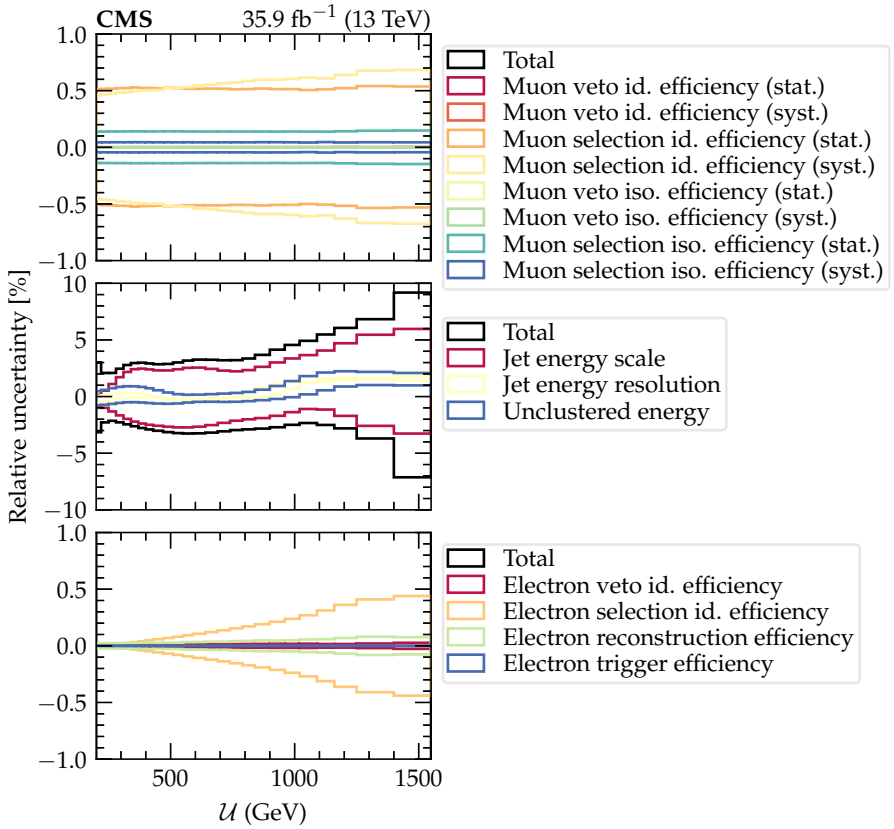


Figure 7.6.: Additional object-based systematic uncertainties to Fig. 7.5.

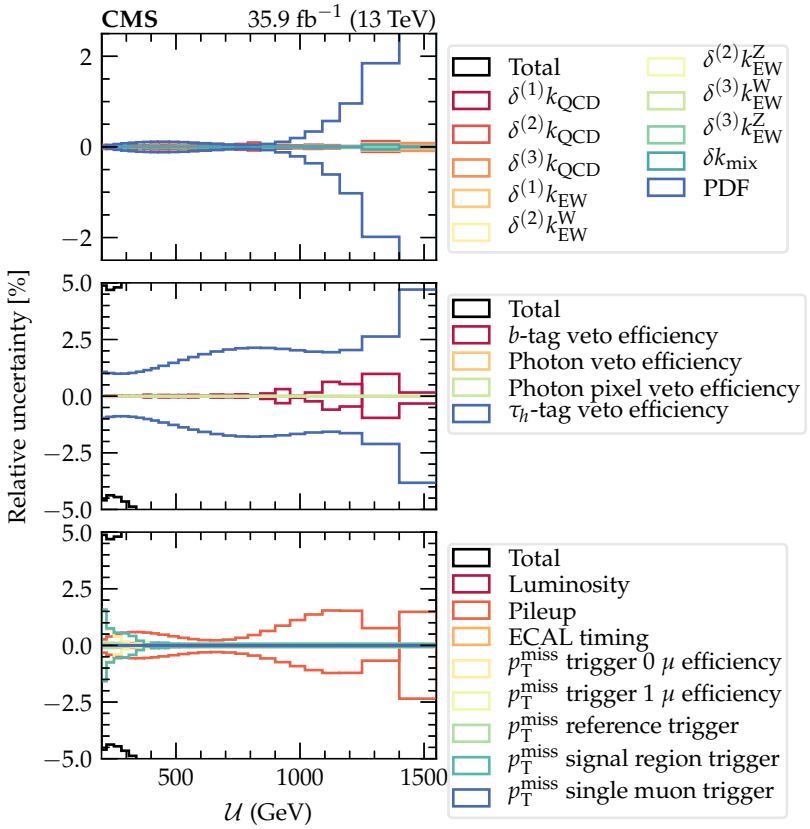


Figure 7.7.: Relative systematic uncertainties (%) on the W + jets transfer factor for the p_T^{miss} + jets region predicted by the e + jets control region. The uppermost panel includes all the theoretical uncertainties, while the others include experimental uncertainties. The total uncertainty, depicted in each panel, represents all uncertainties summed in quadrature.

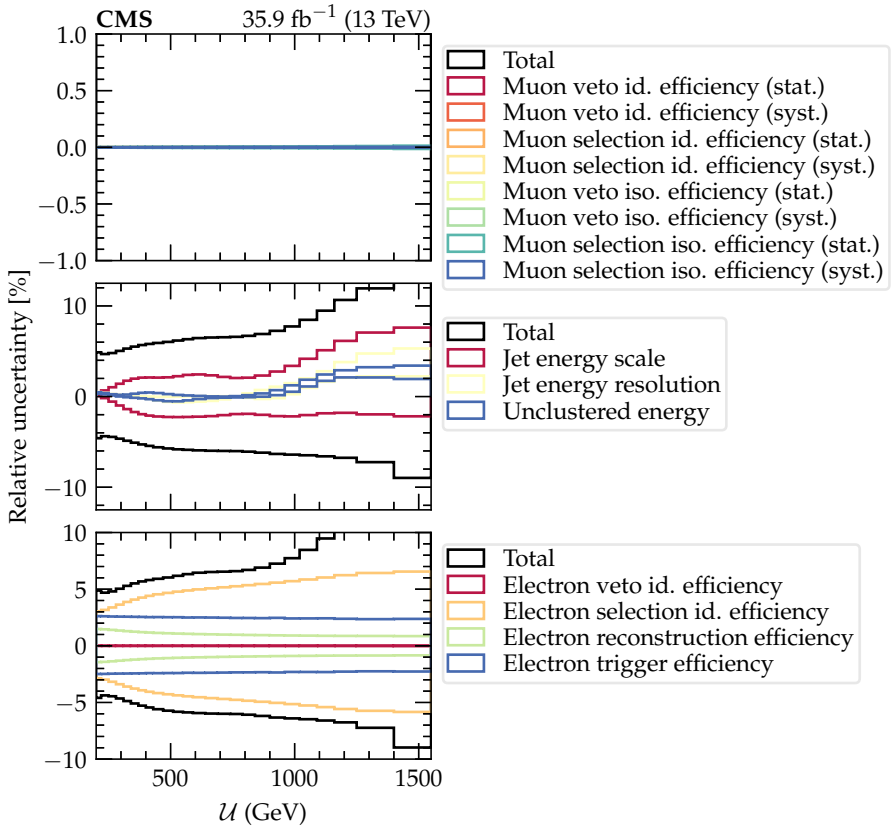


Figure 7.8.: Additional object-based systematic uncertainties to Fig. 7.7.

A set of validation studies are performed to test the transfer factor and the extrapolation involved in predicting the W + jets process in the $p_{\mathrm{T}}^{\mathrm{miss}}$ + jets regions from the W + jets process in the ℓ + jets region.

7.2.6. Validation

Several aspects of the transfer factor approach are tested to determine if additional systematic uncertainties are required to cover mismodelling, undercoverage or bias in the MC prediction. The major background in the $p_{\mathrm{T}}^{\mathrm{miss}}$ + jets region consists of the W + jets process (34.7%), split by decay mode: electron (6.3%), leptonic decay of the τ-lepton (6.9%), muon (9.2%) and τ_h-lepton (12.3%). The data-driven estimate of these processes is performed from the μ + jets and e + jets control regions. The leptonic decay of the τ-lepton is similar to the electron and muon decay modes, with a typically broader $\mathcal{U}$ distribution. However, the τ_h-lepton decay mode differs by being a fully hadronic final state. Therefore, the validation tests are as follows: test the grouping of the muon and electron decay modes, test the extrapolation of the muon and electron decay modes to predict the τ_h-lepton decay mode, and a test of the effects of W polarisation.

The first test verifies the consistency of the μ + jets and e + jets region to validate their combination in measuring the W + jets process. A simultaneous likelihood fit is performed between both regions with expected values given by Eq. 7.4. Two additional unconstrained parameters are applied to the W + jets prediction in the e + jets region to extract the normalisation, r_e, and shape, β_e, difference between both regions. Therefore, the expectation for the W + jets process in the e + jets region is modified to

$$r_{\mathrm{W}} s_{\mathrm{W}\to\ell\nu}^{e+\mathrm{jets}} \mapsto (r_e + \beta_e(\mathcal{U}/\langle\mathcal{U}\rangle - 1))\, r_{\mathrm{W}} s_{\mathrm{W}\to\ell\nu}^{e+\mathrm{jets}}\,, \tag{7.5}$$

where $\langle\mathcal{U}\rangle$ is the weighted average of the $\mathcal{U}$ distribution to scale β_e to a dimensionless parameter, β_e varies the shape of the W + jets process in the e + jets region linearly with respect to the μ + jets region. Perfect consistency between both regions corresponds to $r_e = 1$ and $\beta_e = 0$. The postfit[2] values for the parameters of interest are

$$\begin{aligned} r_e &= 0.95^{+0.05}_{-0.05} \qquad \text{and} \\ \beta_e &= -0.020^{+0.024}_{-0.023}\,. \end{aligned} \tag{7.6}$$

Both are in agreement with a consistent muon and electron channel. The quality of the fit is inspected through the postfit results for the nuisance parameters along with their impact

[2]Postfit refers to the set of best fit values, or parameters determined at these values, estimated from a fit to the data. Conversely, prefit refers to the set of initial values, prior to a fit to the data. The values are typically chosen such that the prefit distributions correspond to the MC predictions.

on the parameter of interest r_e, shown in Fig. 7.9. The postfit nuisance parameters are

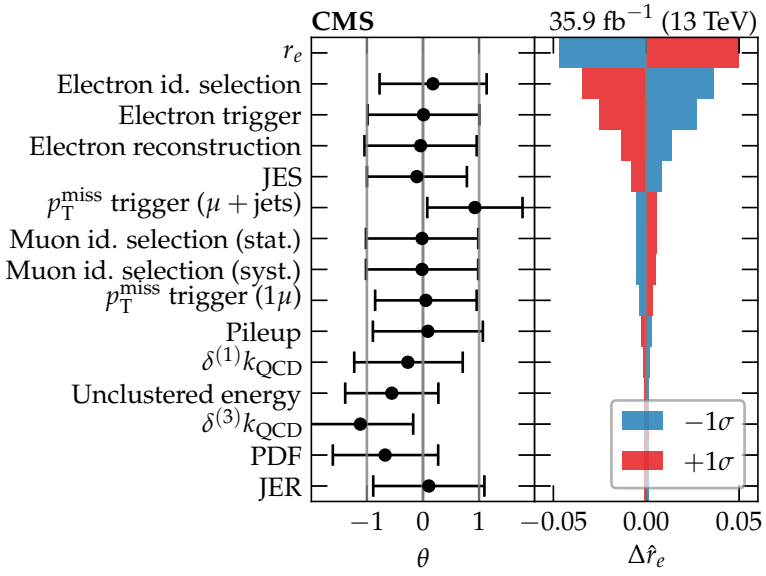


Figure 7.9.: Result of a simultaneous fit between the μ + jets and e + jets region to extract the scale and shape difference between the muon and electron decay modes. The postfit values for the nuisance parameters and their uncertainties are shown as θ as well as the impact of fixing the nuisance parameters at the $+1\sigma$ and -1σ values to determine the effect on the parameter of interest, $\Delta\hat{r}_e$. The set of parameters is restricted to the top 15, in order of impact on the parameter of interest. The θ value for r_e is omitted since it is not a nuisance parameter, instead it is included as a reference for its uncertainty shown in $\Delta\hat{r}_e$.

within a satisfactory limit without significant deviations from the $|\theta| < 1$ region. Therefore, discrepancies between the data and prediction are covered by the unconstrained parameters, with a slight deviation in the low recoil μ + jets region covered by the p_T^{miss} trigger efficiency systematic. The systematic uncertainties with the largest impact on the scale r_e are the electron identification, trigger and reconstruction efficiency uncertainties. The postfit distributions, Fig. 7.10, show the consistency between the postfit prediction and the data. A good agreement between the data and postfit values supports the results of the fit and the consistency of the μ + jets and e + jets. Therefore, the current W + jets prediction formed from a mixture of muon and electron decay modes is kept without the need for additional systematic uncertainties to cover further mismodelling, even though the selections and background compositions differ.

The next validation is a test of the consistency between the μ + jets and e + jets with the τ_h + jets region. This probes the extrapolation performed by the transfer factor method, using the μ + jets and e + jets region to predict the W + jets process in the p_T^{miss} + jets region which mostly consists of τ_h decays of the W boson. This test is performed in a similar way to the previous test, however, the W + jets process in the μ + jets and e + jets regions are scaled by a single unconstrained parameter r_W, whereas the equivalent process in the τ_h + jets region is

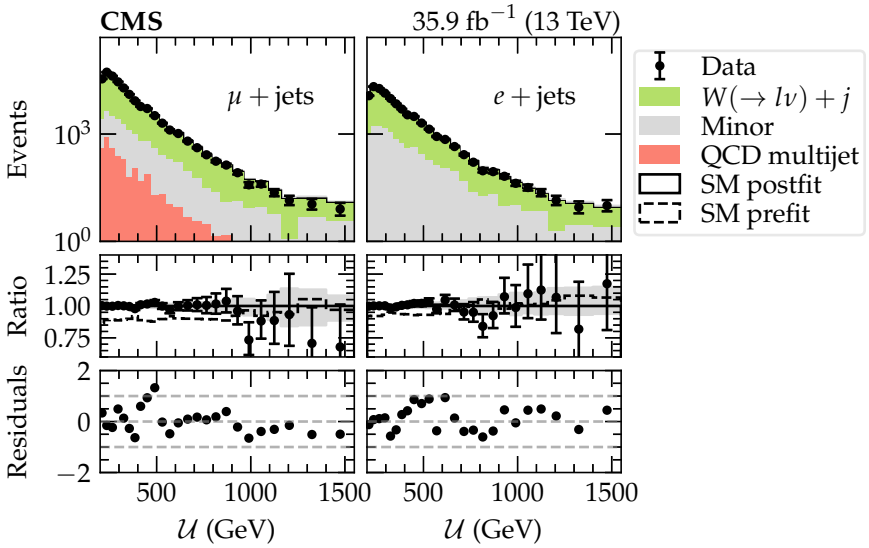


Figure 7.10.: Distributions of the $\mathcal{U}$ in the μ + jets and e + jets with the prefit inputs as well as the postfit results along with its uncertainty. The upper panel shows the results in terms of the event counts with the postfit background composition, while the middle panel shows the ratio to the postfit value and the residuals are defined by the difference between the data and postfit result, normalised by the statistical and systematic uncertainties summed in quadrature.

scaled by

$$(r_\tau + \beta_\tau(\mathcal{U}/\langle\mathcal{U}\rangle - 1))\, r_W s^{\tau_h+\text{jets}}_{W\to\ell\nu} \,. \tag{7.7}$$

The postfit values of these parameters are

$$\begin{aligned} r_\tau &= 1.04^{+0.07}_{-0.06} \qquad \text{and} \\ \beta_\tau &= -0.008^{+0.046}_{-0.047} \,. \end{aligned} \tag{7.8}$$

Again, these values show a consistency in the shape and normalisation of the ℓ + jets and τ_h + jets channels. The nuisance parameters and their impacts on the parameter of interest r_τ is shown in Fig. 7.11. Although the τ_h + jets suffers from a smaller sample size compared to the μ + jets and e + jets regions, the most dominant source of uncertainty comes from the τ_h-tag identification uncertainty, albeit unsurprisingly. The pull[3] on the p_T^{miss} trigger efficiency is consistent with the first test since the same region is used, with an enhanced pull on the PDF uncertainties, although within the satisfactory region. The associated $\mathcal{U}$ distributions from this fit are shown in Fig. 7.12. The μ + jets and e + jets regions are similar to those shown in Fig. 7.10 because the τ_h + jets region does not provide any additional information. Furthermore, the data and postfit result in the τ_h + jets region are in excellent agreement.

[3] A pull on a postfit nuisance parameter refers to the deviation of the postfit value from its prefit.

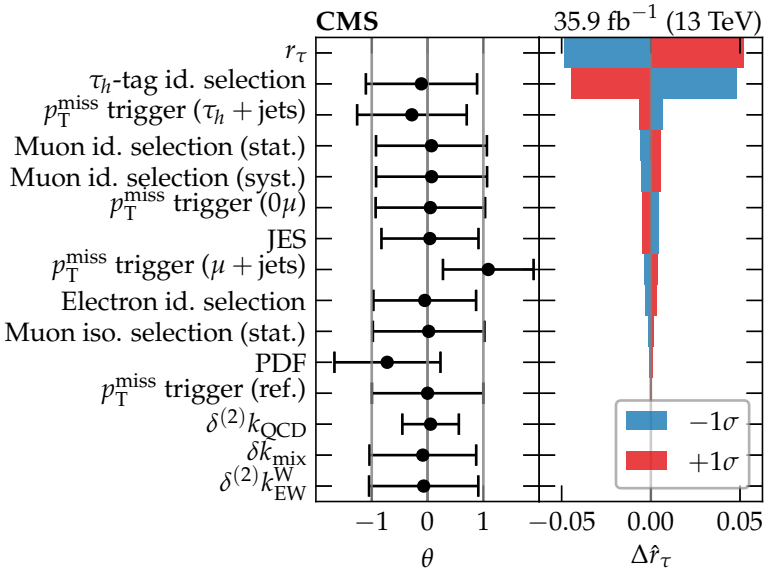


Figure 7.11.: Postfit nuisance parameters and impact on r_τ from a simultaneous fit between the μ + jets, e + jets and τ_h + jets region. The format is aligned with Fig. 7.9.

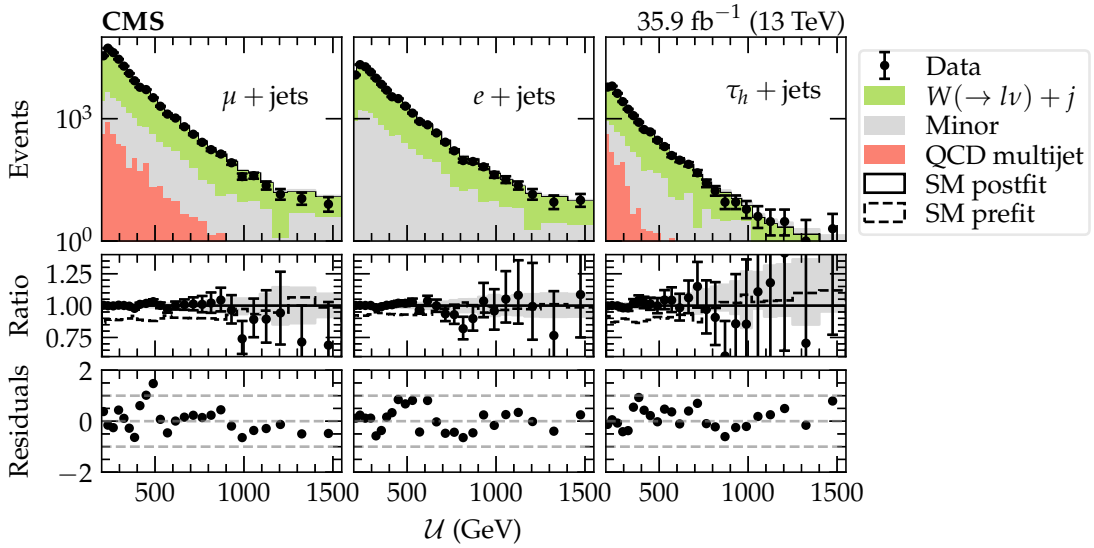


Figure 7.12.: Recoil distributions for the μ + jets, e + jets and τ_h + jets region from a fit to the data to test the extrapolation of a well-reconstructed muon or electron to predict the rate of τ_h decays from the W + jets process. The axis definitions are aligned with Fig. 7.10.

This is partially by design from the unconstrained parameters r_τ and β_τ used to extract the consistency between the various channels.

The final validation tests the extrapolation performed from the different admixture of $W^\pm$ polarisation states between the ℓ + jets and p_T^{miss} + jets regions. The different admixture is a result of the typically softer ℓ^+ p_T compared to ℓ^- p_T to maintain the left-handed chiral state

of the associated neutrino (and right-handed anti-neutrino) [74]. This validation is performed with a similar method, however, the μ + jets is split into two regions: μ^- + jets is used to predict the W + jets process in the μ^+ + jets, hence testing this extrapolation. Since the admixture difference is significantly smaller than the difference between these two regions, any result discrepancies will have a reduced impact on the W + jets prediction. In this fit, both regions are scaled by the unconstrained parameter r_W, with two additional normalisation and shape parameters applied to the μ^+ + jets region as

$$\left(r_{\mu^+} + \beta_{\mu^+}\left(\mathcal{U}/\langle\mathcal{U}\rangle - 1\right)\right) r_W s_{W\to\ell\nu}^{\mu^+\text{+jets}} \,. \tag{7.9}$$

The postfit values for the parameters from a fit to the data are

$$\begin{aligned} r_{\mu^+} &= 0.976^{+0.013}_{-0.013} \qquad \text{and} \\ \beta_{\mu^+} &= 0.007^{+0.025}_{-0.024} \,. \end{aligned} \tag{7.10}$$

The uncertainty on the normalisation parameter is noticeably smaller than previous validation tests as a result of the closer alignment between both regions leading to further cancellations with the transfer factor method. The null hypothesis of consistent shape and normalisation between the μ^+ + jets and μ^- + jets is within the 95% confidence level[4] with an excellent shape agreement and a normalisation near the boundary, although within the consistent limits. The postfit nuisance parameters and their impacts on the unconstrained parameter r_{μ^+} are shown in Fig. 7.13. The pulls are consistent with previous results including the μ + jets region, without any significant deviations from the prefit values. No significant constraints are estimated for the nuisance parameters, other than a slight constraint on the $\delta^{(2)}k_{QCD}$, also seen in Fig. 7.9. This nuisance parameter is associated to the higher-order QCD corrections discussed in Sec. 6.7. In particular, this systematic uncertainty is included to account for unknown shape dependencies from the choice of the factorisation and renormalisation scale. However, this introduces a partial double-counting of the uncertainty on $\delta^{(1)}k_{QCD}$, the normalisation effect of the choice of factorisation and renormalisation scales [72]. Moreover, the postfit results are in excellent agreement with the data, as shown in Fig. 7.14.

The validation tests show a consistency between the three lepton-based regions, with the set of systematic uncertainties covering discrepancies between these regions. Furthermore, the offset between data and the MC prediction is completely consistent between all three regions and corrected by the transfer factor approach. Therefore, no additional systematic uncertainties are derived or placed on the W + jets transfer factor method. This method is implemented within the final result as a simultaneous fit between the signal and control regions.

[4] A null hypothesis is typically excluded if it falls outside of the 95% confidence interval.

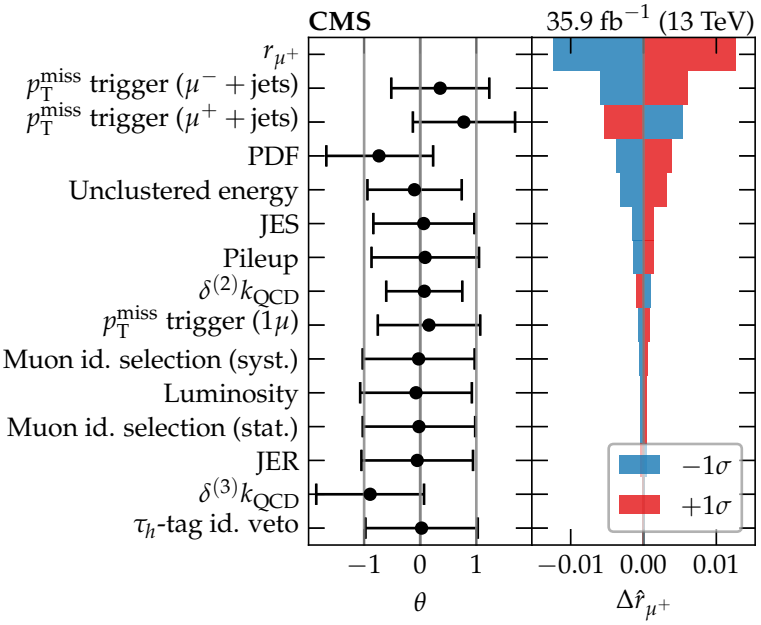


Figure 7.13.: Postfit nuisance parameters from a simultaneous fit to the $\mu^+ +$ jets and $\mu^- +$ jets regions to extract the shape and normalisation difference. The impacts on the r_{μ^+} parameter are shown. The format is aligned with Fig. 7.9. The enhanced postfit uncertainty on the jet energy resolution is a result of its asymmetric uncertainties evident from its $\Delta\hat{r}_{\mu^+}$ impact.

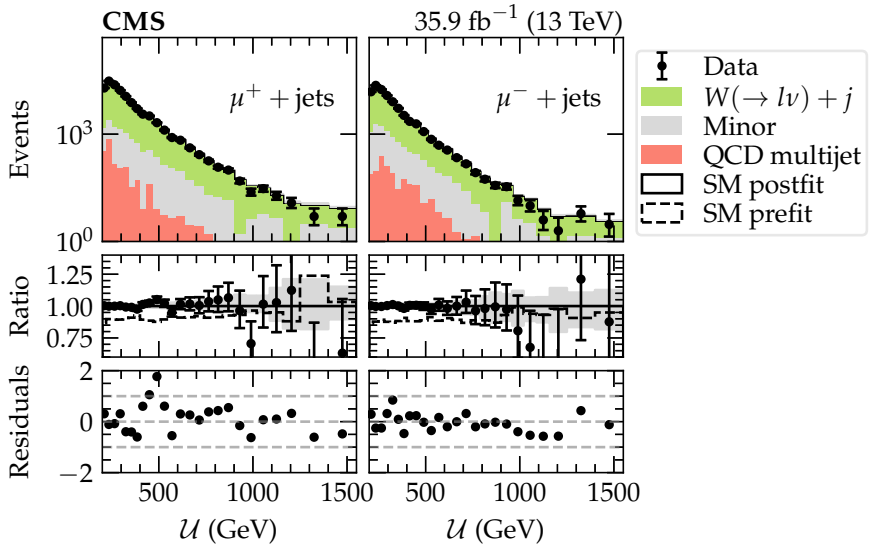


Figure 7.14.: Postfit $\mathcal{U}$ distributions in the simultaneous fit between the $\mu^+ +$ jets and $\mu^- +$ jets regions. The axis definitions are aligned with Fig. 7.10.

7.3. QCD multijet estimation

The QCD multijet contamination within the $p_{\mathrm{T}}^{\mathrm{miss}}$ + jets signal region is less than 1%, with a sharply falling $\mathcal{U}$ distribution compared to the signal, leading to approximately a 0.1% contribution in the tail of the $\mathcal{U}$ distribution. Nevertheless, a data driven estimate controls possibly large systematic uncertainties associated with the QCD multijet simulation.

The estimate is performed from a QCD multijet enriched sideband aligned with the $p_{\mathrm{T}}^{\mathrm{miss}}$ + jets region, with the $\Delta\phi_{\mathrm{min}}$ selection inverted to

$$\Delta\phi_{\mathrm{min}} < 0.5\,. \tag{7.11}$$

Equivalent sideband regions are defined, by inverting the $\Delta\phi_{\mathrm{min}}$ selection, for the μ + jets and e + jets for an independent W + jets estimation within this sideband, performed with the transfer factor method. QCD multijet events contaminate the $p_{\mathrm{T}}^{\mathrm{miss}}$ + jets region through severe jet mismeasurement, leading to a jet below threshold, or otherwise misreconstructed, leading to significant missing energy with no overlap between a jet and the $p_{\mathrm{T}}^{\mathrm{miss}}$ in the ϕ-plane. By inverting the $\Delta\phi_{\mathrm{min}}$ selection, the severely mismeasured jet is required to be reconstructed and above threshold. This jet is selected from the nearest jet to the missing energy in the transverse plane (known as the $p_{\mathrm{T}}^{\mathrm{miss}}$-aligned jet), as shown in Fig. 7.15 along with other kinematic distributions. The sideband region predominantly consists of the QCD multijet process (75.2%) with contamination from $\mathrm{W}(\to \ell\nu)$ + jets (10.9%) and the signal $\mathrm{Z}(\to \nu\bar{\nu})$ + jets (9.5%), followed by a small contribution from t-quark and minor background processes (4.3%).

A disagreement of 10–20% between data and MC is not covered by scaling the $\mathrm{W}(\to \ell\nu)$ + jets process up by the 10% determined from Sec. 7.2. Instead, this is attributed to mismodelling of the QCD multijet process. This disagreement is parameterised by a linear fit to the p_{T} distribution of the $p_{\mathrm{T}}^{\mathrm{miss}}$-aligned jet, extrapolated to the $p_{\mathrm{T}}^{\mathrm{miss}}$ + jets region at 0–40 GeV, within each $\mathcal{U}$ bin. A 2-dimensional fit is performed with the likelihood formalism (App. A). The final six $\mathcal{U}$ bins are combined to avoid categories of low statistical power. The linear fit is performed across 10 $p_{\mathrm{T}}^{\mathrm{miss}}$-aligned jet p_{T} bins from 40–1040 GeV. The W + jets prediction is performed by the transfer factor method, discussed in Sec. 7.2, inclusive in the $p_{\mathrm{T}}^{\mathrm{miss}}$-aligned jet p_{T}. Furthermore, the $\mathrm{Z}(\to \nu\bar{\nu})$ + jets signal contamination within this region is allowed to vary to determine the correlation between the data-driven QCD multijet estimate and the $\mathrm{Z}(\to \nu\bar{\nu})$ + jets signal. The postfit distribuition is shown in Fig. 7.16. The parameterisation model introduced in the fit significantly improves the agreement between data and the prediction, with the extrapolation to the $p_{\mathrm{T}}^{\mathrm{miss}}$ + jets region shown in Fig. 7.17. To a first order approximation, within uncertainties, the scale factors follow a linear trend with the $\mathcal{U}$. Therefore, a linear fit determines the scaling factor to the higher $\mathcal{U}$ bins, with higher order effects determined from a quadratic fit taken as an uncertainty on the scale factors, ranging from a 5–100% uncertainty on the correction.

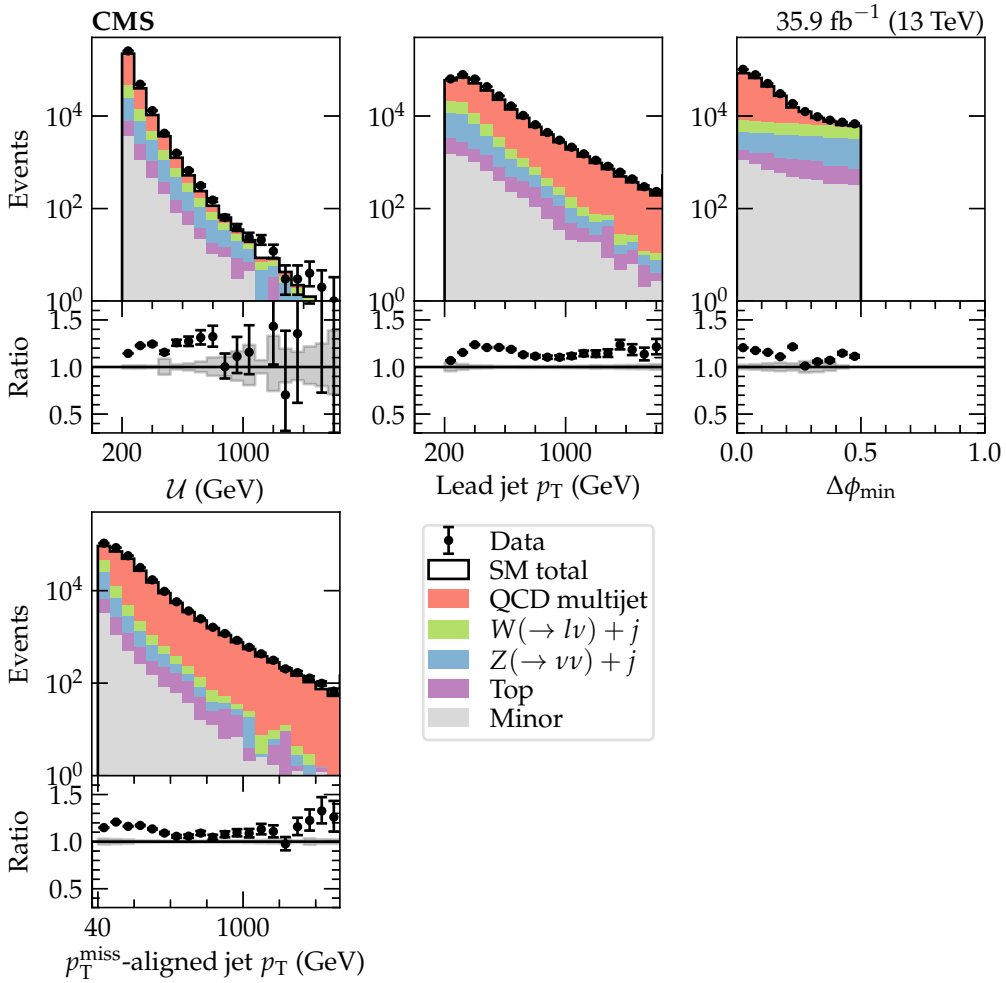


Figure 7.15.: Kinematic distributions within the QCD-enriched sideband. The ratio is taken with respect to the SM total and includes the statistical uncertainty associated with the limited sample size.

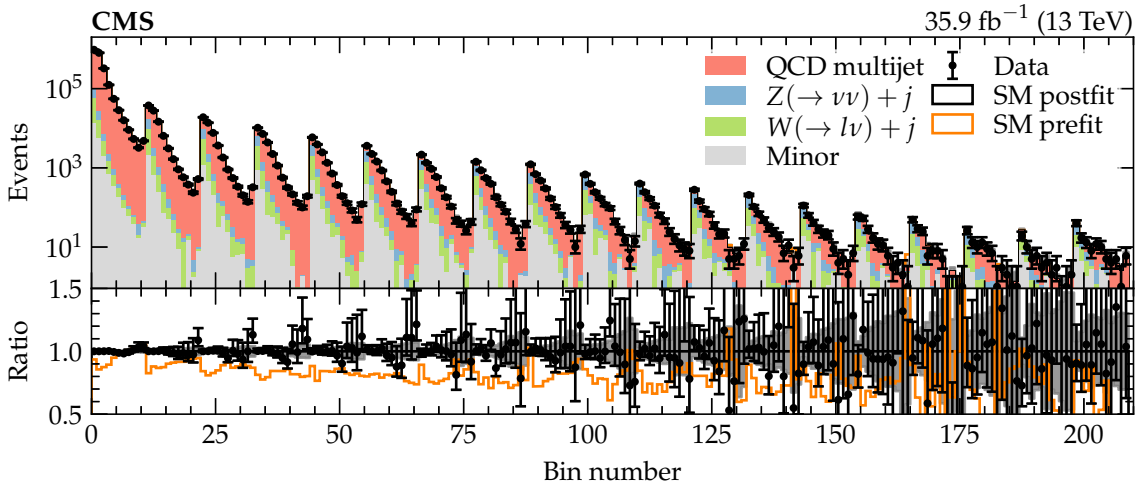


Figure 7.16.: Postfit results of a fit to data in the QCD multijet enriched sideband allowing the normalisation on the $Z(\to \nu\bar{\nu}) + \text{jets}$ and $W(\to \ell\nu) + \text{jets}$ processes to vary and a linear shape variation on the QCD multijet process. The 2-dimensional binning is unfolded into 1-dimension labelled by the bin number, with 0–9 corresponding to the p_T^{miss}-aligned distribution of the first $\mathcal{U}$ bin, 10–19 to the second $\mathcal{U}$ bin, and so on for 19 $\mathcal{U}$ bins in total. The ratio is taken with respect to the SM postfit result.

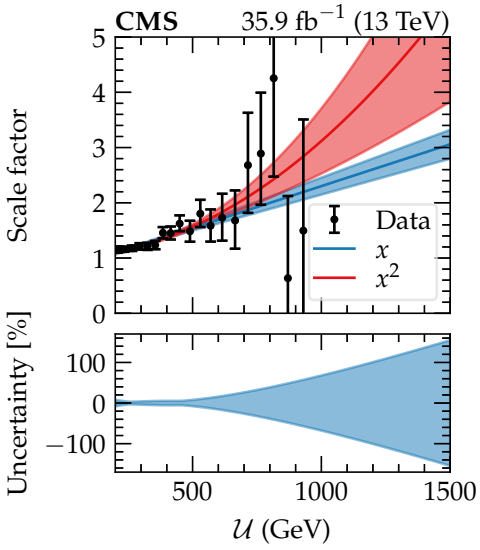


Figure 7.17.: The QCD multijet scaling factor determined from a linear extrapolation from the QCD multijet enriched sideband to the $p_T^{miss} + \text{jets}$ region. The data points include the statistical and systematic uncertainties determined from the fit. Additional linear (x) and quadratic (x^2) fits are performed with the maximum difference taken as an uncertainty band, shown in the lower panel.

These corrections are applied as scaling parameters to the QCD multijet process within the $p_{\mathrm{T}}^{\mathrm{miss}}$ + jets region, with the uncertainty as alternative templates, decorrelated across the $\mathcal{U}$ bins to allow the shape of the QCD multijet prediction to vary, as well as the normalisation. Furthermore, the predictions are found to have a 20% anti-correlation with the $\mathrm{Z}(\rightarrow \nu\bar{\nu})$ + jets normalisation, and hence this correlation is included within the final fit.

7.4. γ^* contribution

To determine the contributions to the $\mathrm{Z}/\gamma^* \rightarrow \ell\bar{\ell}$ process from $\mathrm{Z} \rightarrow \ell\bar{\ell}$ and $\gamma^* \rightarrow \ell\bar{\ell}$, the processes are generated at NLO accuracy through the matrix element and parton shower, without the detector simulation to avoid the expensive computation. Regardless, the sample size of the generated $\mathrm{Z} \rightarrow \ell\bar{\ell}$ and $\gamma^* \rightarrow \ell\bar{\ell}$ processes are limited and the statistical uncertainty, along with QCD scale and PDF variations, are propagated through the ratio of cross sections of both processes to the $\mathrm{Z}/\gamma^* \rightarrow \ell\bar{\ell}$ interaction. The generated processes do not include any of the corrections discussed in Chap. 6, including the higher-order QCD and EW corrections since these will mostly cancel in the ratio of cross sections, with the small remainder estimated by the uncertainties. To avoid inconsistencies between various event generators, the 3 processes are generated by MADGRAPH5_AMC@NLO v2.2.2. The invariant mass distribution of the generator-level dilepton is shown in Fig. 7.18. The inclusive ratio of cross sections, with respect to the $\mathrm{Z}/\gamma^* \rightarrow \ell\bar{\ell}$ process, within the fiducial region $71 < m_{\ell\ell} < 111\,\mathrm{GeV}$, defined by the analysis selection is used to determine the contributions from $\gamma^* \rightarrow \ell\bar{\ell}$ and the interference with the $\mathrm{Z} \rightarrow \ell\bar{\ell}$ process. The $\gamma^* \rightarrow \ell\bar{\ell}$ component is treated as a background while the interference is treated as the $\mathrm{Z} \rightarrow \ell\bar{\ell}$ signal scaled by the square-root of any signal strength modifiers, to account for the $\mathrm{Z} \rightarrow \ell\bar{\ell}$ amplitude appearing in the interference cross section term. Therefore, within the likelihood to extract the Z invisible width, the contribution from the $\mathrm{Z}/\gamma^* \rightarrow \ell\bar{\ell}$ process in the $\ell\ell$ + jets region is

$$(k_{\mathrm{Z}} r_{\mathrm{Z}} + (1 - k_{\mathrm{Z}} - k_{\gamma^*})\sqrt{r_{\mathrm{Z}}} + k_{\gamma^*})\, s^{\ell\ell+\mathrm{jets}}_{\mathrm{Z}/\gamma^* \rightarrow \ell\bar{\ell}}\,, \tag{7.12}$$

where r_{Z} is the unconstrained parameter scaling the $\mathrm{Z} \rightarrow \ell\bar{\ell}$ process, k_{Z} is the constrained parameter measured from the ratio of $\mathrm{Z} \rightarrow \ell\bar{\ell}$ to $\mathrm{Z}/\gamma^* \rightarrow \ell\bar{\ell}$ cross sections, and k_{γ^*} is similarly defined as the ratio of $\gamma^* \rightarrow \ell\bar{\ell}$ to $\mathrm{Z}/\gamma^* \rightarrow \ell\bar{\ell}$. An uncertainty associated with the extrapolation from the generator-level selection to the analysis region is performed by reweighting the distribution of partons initiating the interaction to the final analysis region. This captures effects caused by the analysis selection such as the lepton η restriction and the high $\mathcal{U}$ requirements. The values are determined after the reweighting procedure and the difference between the nominal result and the reweighted result is taken as a systematic uncertainty. The measured values are summarised in Tab. 7.2.

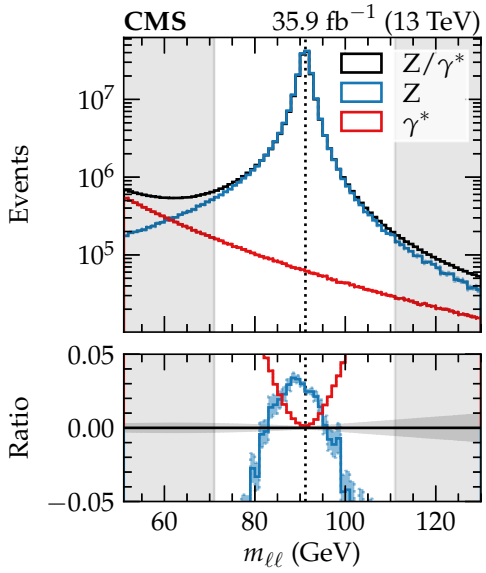


Figure 7.18.: Invariant mass distributions of the generated dilepton from the $Z/\gamma^* \to \ell\bar{\ell}$, $Z \to \ell\bar{\ell}$ and $\gamma^* \to \ell\bar{\ell}$ processes. The lower panel shows the ratio with respect to the $Z/\gamma^* \to \ell\bar{\ell}$ process, with the $Z \to \ell\bar{\ell}$ ratio subtracted by one to place it on the same scale as the $\gamma^* \to \ell\bar{\ell}$ ratio. The uncertainty bands represent the statistical uncertainty of the MC samples. The region of interest, $71 < m_{\ell\ell} < 111\,\text{GeV}$, is bounded by the grey regions.

	k_Z	k^*_γ
Parameter value	1.0165	0.0150
MC Statistical uncertainty [%]	±0.1	±0.1
PDF uncertainty [%]	±2.3	±2.4
μ_F and μ_R uncertainty [%]	-	-
Reweighting uncertainty [%]	±1.4	±13.5

Table 7.2.: The measured values for the k_Z and k^*_γ parameters with their associated systematic uncertainties. The factorisation and renormalisation scales have a negligible impact on these parameters.

7.5. Remaining backgrounds

The remaining backgrounds in the p_T^{miss} + jets and $\ell\ell$ + jets regions are estimated directly from MC with all corresponding sources of systematic uncertainties included in the final fit to extract the Z invisible width.

7.6. Summary

The $Z(\rightarrow \nu\bar{\nu}) + \text{jets}$ signal collected in the $p_{\text{T}}^{\text{miss}} + \text{jets}$ region is largely contaminated by the $W(\rightarrow \ell\nu) + \text{jets}$ processes. This process is estimated through a data-driven prediction from the $\ell + \text{jets}$ control regions through the transfer factor approach. The consistency of the normalisation and shape of the $\mathcal{U}$ distribution in the $\mu + \text{jets}$, $e + \text{jets}$ and $\tau_h + \text{jets}$ regions; as well as between the $\mu^+ + \text{jets}$ and $\mu^- + \text{jets}$; is used to test the extrapolation from the control regions to the $p_{\text{T}}^{\text{miss}} + \text{jets}$ region performed in the transfer factor approach. The validation test results show a consistent set of control regions.

Although the QCD multijet process is a minor background within the $p_{\text{T}}^{\text{miss}} + \text{jets}$ region, the large cross section and difficulty in modelling the process has led to a data-driven estimate of this background. A MC-to-data correction is extrapolated from the QCD multijet sidebands through a linear fit to the $p_{\text{T}}^{\text{miss}}$-aligned jet p_{T} distributions. A further linear parameterisation determines the correction factor across the $\mathcal{U}$ distribution, with an alternative quadratic fit to incorporate higher order effects along the $\mathcal{U}$ distribution through a systematic uncertainty.

To determine the Z invisible width, the rate of the $Z \rightarrow \ell\bar{\ell}$ process is required. However, the $\ell\ell + \text{jets}$ consists of the inseparable processes: $Z \rightarrow \ell\bar{\ell}$, $\gamma^* \rightarrow \ell\bar{\ell}$ and their interference. Instead, these are estimated through a generator-level study where the Z and γ^* contributions are successively suppressed. Systematic uncertainties in this estimation include the theoretical PDF and QCD scale uncertainties, as well as the difference with the estimation under a reweighting of the partons initiating the interaction to the analysis reconstructed-level regions.

The estimation of these processes are used within the final fit to extract the rate of $Z(\rightarrow \nu\bar{\nu}) + \text{jets}$ and $Z(\rightarrow \ell\bar{\ell}) + \text{jets}$ events to determine the Z invisible width.

Chapter 8.

Invisible width measurement

8.1. Introduction

The key ingredients into extracting the Z invisible width from Eq. 1.1 includes the $Z(\to \nu\bar{\nu}) + \text{jets}$ and $Z(\to \ell\bar{\ell}) + \text{jets}$ processes. These are estimated from the $p_{\text{T}}^{\text{miss}} + \text{jets}$ and $\ell\ell + \text{jets}$ region with background subtraction through data driven methods or MC estimation. The comparison of the predictions to data for these regions are presented in this chapter, acting as the input for the likelihood fit. The details of the likelihood fit are discussed along with the results of the fit in terms of the nuisance parameters, their impact on the parameter of interest itself, the Z invisible width.

8.2. Signal regions

The SM prediction for the $Z(\to \nu\bar{\nu}) + \text{jets}$ and $Z(\to \ell\bar{\ell}) + \text{jets}$ processes from MC simulations predict the rate of events by estimating the efficiency and acceptance of the selection applied, with data-driven corrections where relevant. The normalisation on these processes is estimated from a likelihood fit to that data, exploiting the shape of the $\mathcal{U}$ distribution. The shapes of each process is taken from MC, with the associated uncertainties affecting the normalisation and morphing the shapes, within their gaussian constraints.

8.2.1. $\ell\ell + \text{jets}$ region

The selection applied to the $\ell\ell + \text{jets}$ is designed for similar kinematics between the signal processes. This is primarily achieved through the $\mathcal{U}$ parameter which is well modelled in MC, as shown in Sec. 6.8. Furthermore, the selections differ by kinematic requirements on the final state leptons, including their p_{T} and invariant mass. These kinematic distributions are shown for the $\mu\mu + \text{jets}$ and $ee + \text{jets}$ regions in Fig. 8.1 and 8.2, respectively. Both regions are

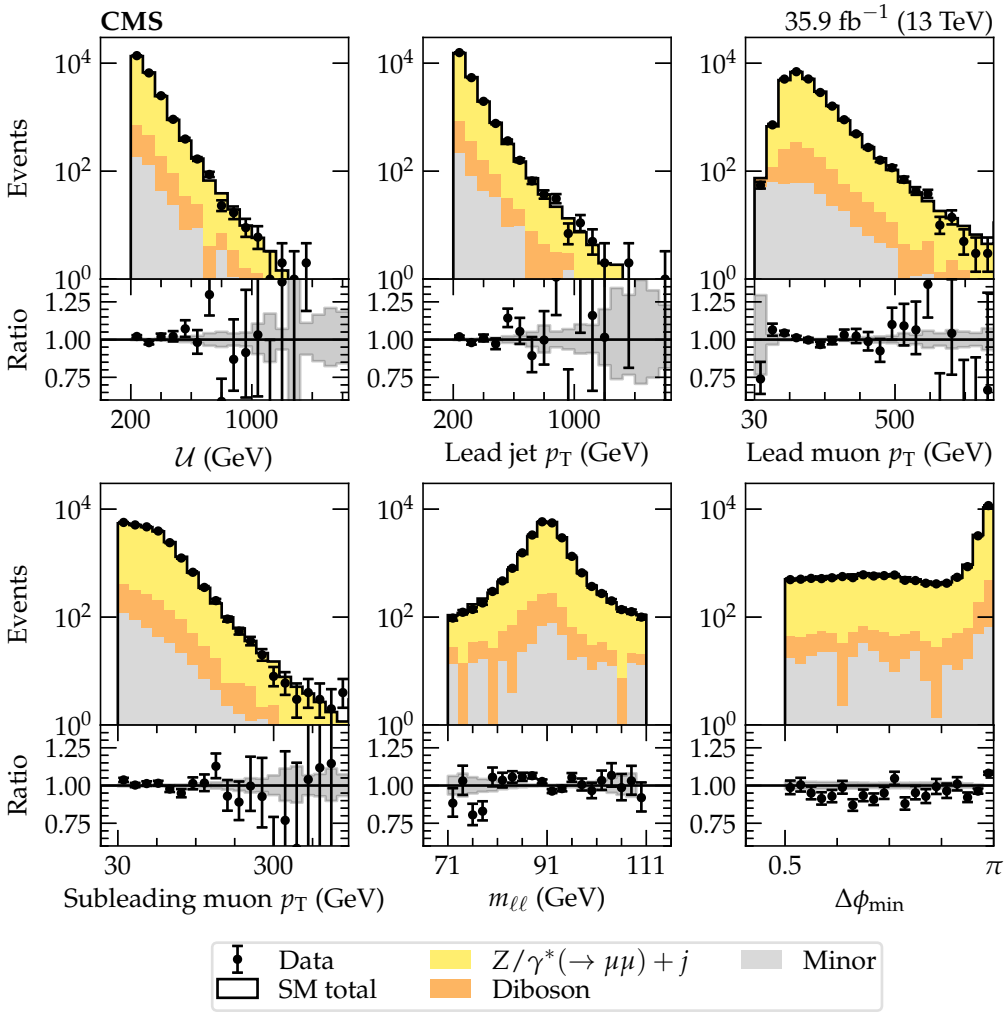


Figure 8.1.: Kinematic distributions within the $\mu\mu$ + jets region. The ratio is taken with respect to the SM total prediction and includes the MC statistical uncertainties, whereas systematic uncertainties have been omitted.

enriched in the $Z/\gamma^*(\to \ell\bar{\ell}) +$ jets process (94%), with the largest background from diboson production (4–5%) and a smaller contribution from other minor backgrounds (1–2%). These backgrounds are determined from NLO MC, without any additional data-driven correction. The systematic uncertainties associated with these processes are included in the background subtraction performed within the likelihood fit. The data and MC distributions are within agreement, both for their shape and normalisations. The requirement placed on the recoil and lead jet p_T result in the sculpting of the lead and subleading muon distributions, while the invariant mass window is centred on the Z mass.

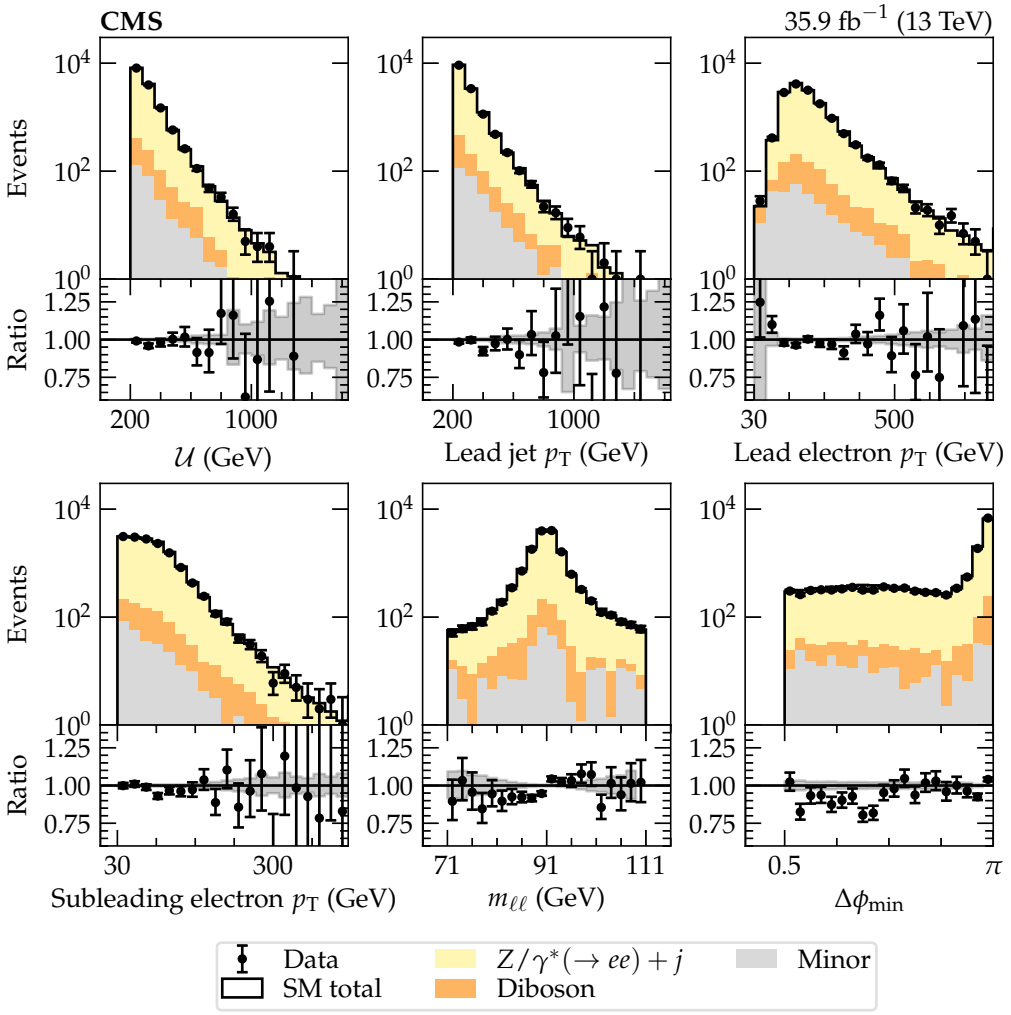


Figure 8.2.: Kinematic distributions within the ee + jets region, equivalent to those shown in Fig. 8.1.

8.2.2. p_T^{miss} + jets region

The p_T^{miss} + jets region consists primarily of the $Z(\rightarrow \nu\bar{\nu})$ + jets process (59.1%) with a significant background from $W(\rightarrow \ell\nu)$ + jets (34.6%) and minor contributions from t-quark production (1.7%), QCD multijet (0.7%) and other processes (3.9%). The $W(\rightarrow \ell\nu)$ + jets process is modified by the data-driven estimate through the transfer factor approach. Similarly, the QCD multijet contribution includes a data-driven estimate from an extrapolation from the QCD multijet enriched sidebands. The other backgrounds are estimated from NLO MC predictions. The kinematic distributions for the p_T^{miss} + jets region, without the data-driven estimates, are shown in Fig. 8.3. The data and SM prediction distributions disagree, however, the large $W(\rightarrow \ell\nu)$ + jets contribution has a 10% correction determined from the data-driven transfer factors. The correction covers the discrepancy seen between data and the SM prediction.

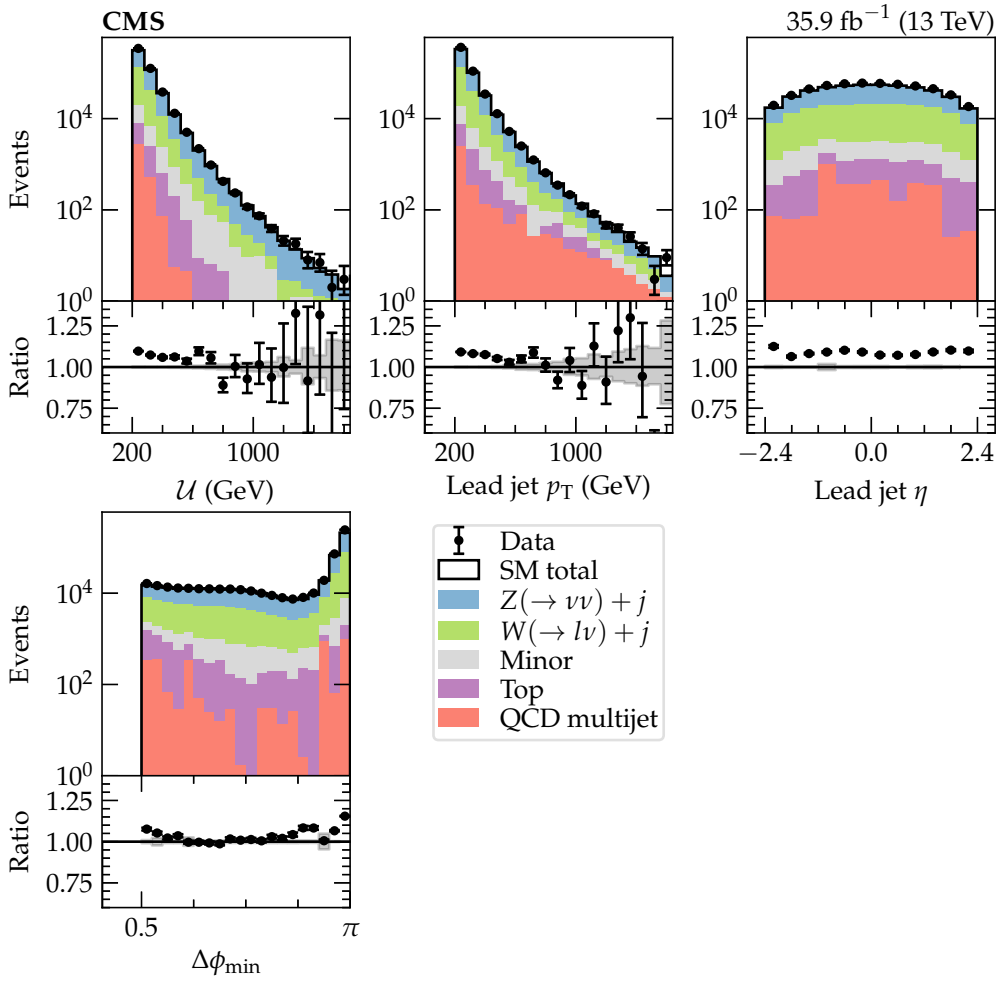


Figure 8.3.: Kinematic distributions within the p_T^{miss} + jets region, similar to those shown in Fig. 8.1.

8.3. Likelihood fit

The Z invisible width is estimated from a simultaneous fit to the data between the p_T^{miss} + jets, $\ell\ell$ + jets and ℓ + jets ($\ell = e, \mu$) regions under the likelihood formalism (App. A). The W + jets estimation through the transfer factor approach is implemented by a global unconstrained parameter which scales the W + jets process in the p_T^{miss} + jets and ℓ + jets regions. Such a parameter is constrained by the data in the W + jets enriched control regions and applied to the same process in the p_T^{miss} + jets region, improving the accuracy of the prediction along with correcting normalisation issues. To extract the invisible width, the $Z(\rightarrow \ell\bar{\ell})$ + jets process is scaled by a parameter r_Z while $Z(\rightarrow \nu\bar{\nu})$ + jets is scaled by $r_Z r_{inv}$. The observed number of events for a particular process may be expressed as the product of the cross section, branching fraction, analysis efficiency and acceptance. Setting $r_{inv} = r_Z = 1$ corresponds to the SM expected number of events for the process. Therefore, taking the ratio of Eq. 1.1 with its SM

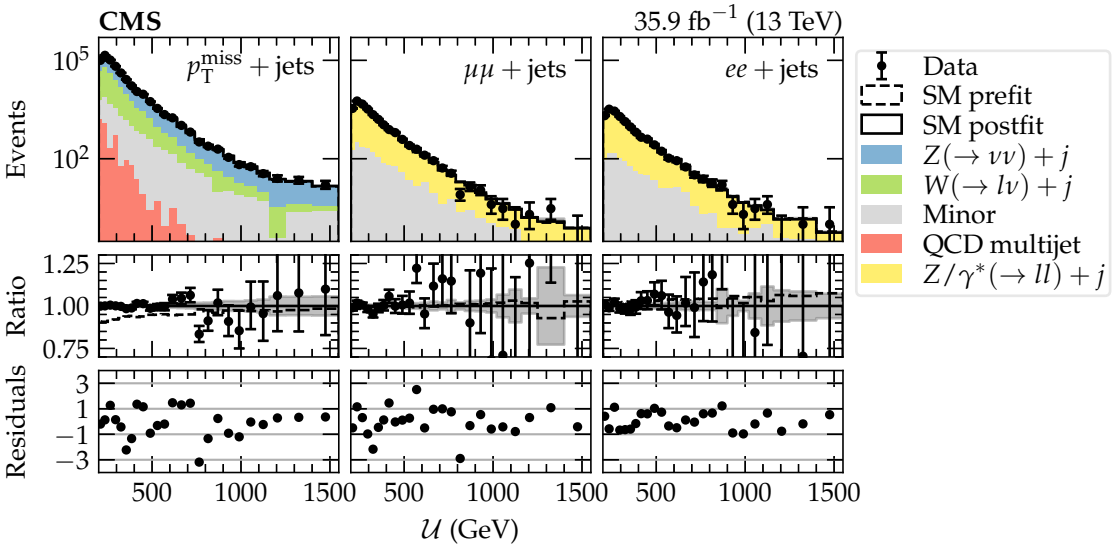


Figure 8.4.: $\mathcal{U}$ distributions for the signal regions to determine the rate of $Z(\to \nu\bar{\nu})$ + jets and $Z(\to \ell\bar{\ell})$ + jets. The ratio is taken with respect to the SM postfit and includes the prefit result to show the effect of including the scaling factors. The residuals display the difference between the data and the SM postfit results, normalised by their uncertainties summed in quadrature.

expectation gives

$$r_{\text{inv}} = \frac{\Gamma(\text{Z} \to \text{inv})}{\Gamma_{\text{SM}}(\text{Z} \to \text{inv})}\,, \tag{8.1}$$

where Γ_{SM} corresponds to the SM expected width. Systematic uncertainties, discussed in Sec. 7.2.4 and included for the p_T^{miss} + jets and $\ell\ell$ + jets regions, are encoded as nuisance parameters allowing the shape and normalisation of the various processes to vary according to a single nuisance parameter per systematic fully correlated between $\mathcal{U}$ bins.

The postfit distributions from the simultaneous fit are shown in Fig. 8.4 A significant improvement with the postfit results in the p_T^{miss} + jets region compared to the prefit as a result of the data-driven scaling parameter of 10% determined for the $W(\to \ell\nu)$ + jets process. The QCD multijet process is corrected by up to a factor of 3 in the tail of the $\mathcal{U}$ distribution. However, this process remains to be a small background. There is no significant difference between the postfit and prefit results in the $\ell\ell$ + jets regions. Furthermore, one data point is beyond the 3σ band of the postfit result, however, since there are 75 bins here, along with another 50 in the ℓ + jets regions, this is not a concern and falls within standard expectations. The associated postfit nuisance parameters and their impact on the r_{inv} parameter is shown in Fig. 8.5. The postfit nuisance parameters are all within the $|\theta| < 1$ level, with no significant constraint other than the $\delta^{(2)}k_{\text{QCD}}$ parameter which is associated with a possible shape dependence with the $\mathcal{U}$ distribution from the choice of factorisation and renormalisation scale. This parameter has a

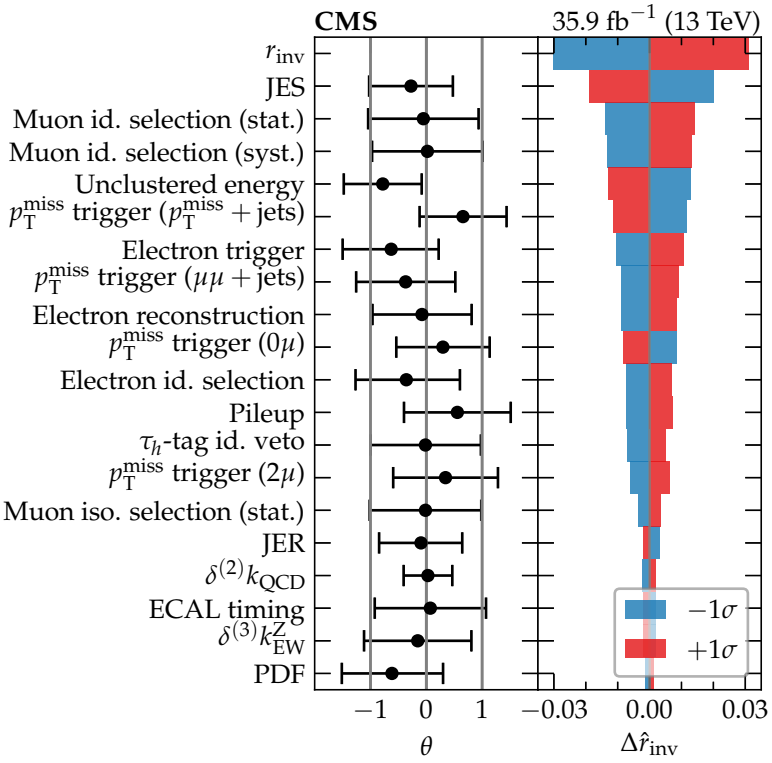


Figure 8.5.: Nuisance parameter (θ) pulls, constraints and impacts on r_{inv} ($\Delta\hat{r}_{\mathrm{inv}}$). Only the 20 most impacting parameters are displayed.

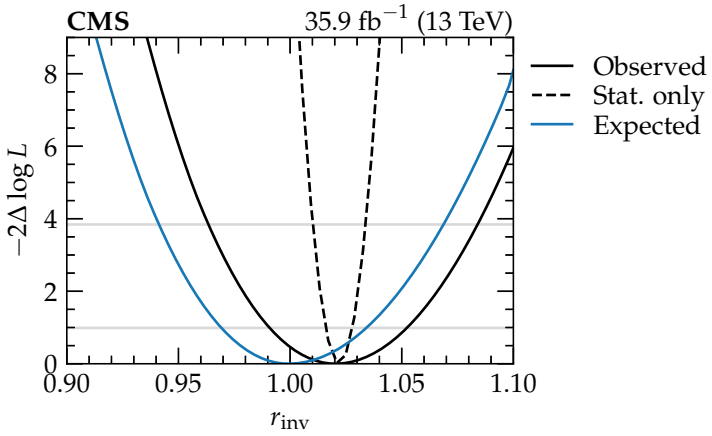


Figure 8.6.: Likelihood profile across the r_{inv} dimension for the fit to data (observed) and an alternative fit where the data is set to the prediction in the poisson terms of the likelihood (expected). This is typically known as the Asimov dataset and gives the expected results with all the parameters set to their input values. In addition, a fit to the data with the systematic uncertainties fixed at their postfit values (stat. only) is included to compare the effect of statistical and systematic uncertainties. The measurement of the r_{inv} parameter is systematically limited.

partial double-counting with the $\delta^{(1)}k_{\text{QCD}}$ parameter which probes the normalisation variation from the choice of these scales. The largest impacts on the r_{inv} parameter come from the JES uncertainties, unclustered energy variations, muon identification selection, electron and $p_{\text{T}}^{\text{miss}}$ trigger. The JES and unclustered energy uncertainties dominate from the lack of cancellation between the various regions as a result of the limited MC sample size (App. A.3) used to determined the JES alternative templates. The muon and electron efficiencies are dominant since they significantly affect the ℓ + jets and $\ell\ell$ + jets regions but not the $p_{\text{T}}^{\text{miss}}$ + jets region. Lastly, the $p_{\text{T}}^{\text{miss}}$ trigger uncertainties have a component uncorrelated between the various regions, hence leads to a non-cancellation and a large impact on r_{inv}.

The multi-dimensional minimisation of the likelihood may be profiled onto a single dimension by straddling the minimum for various fixed values of the dimension of interest. The value of the minimum is subtracted from subsequent likelihood profile values to give $-2\Delta \log L$, where the confidence interval in terms of the number of standard deviations n_{sig} is given by the intersection points of $-2\Delta \log L = n_{\text{sig}}^2$. The profile for the r_{inv} is shown in Fig. 8.6. It is also instructive to perform the scan in 2-dimensions with a modified fit where the $\text{Z}(\to \nu\bar{\nu})$ + jets and $\text{Z}(\to \ell\bar{\ell})$ + jets processes are scaled by separate parameters, r_{inv} and $r_{\ell\ell}$ respectively. This 2-dimensional scan is shown in Fig. 8.7, showing that the measurement is consistent with the SM prediction.

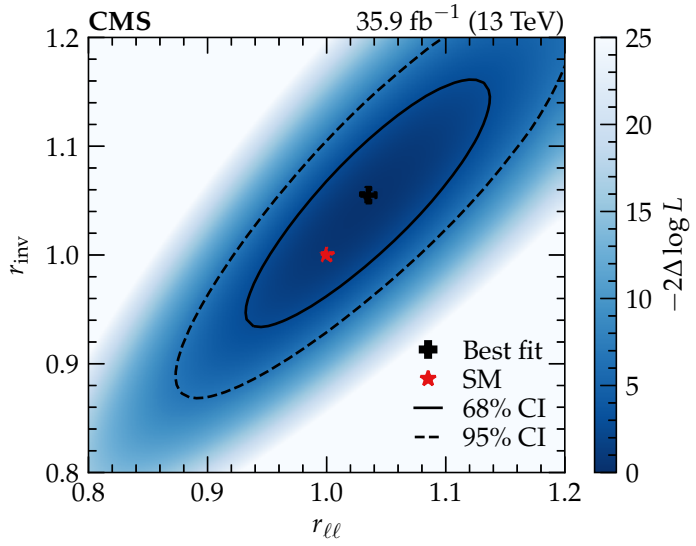


Figure 8.7.: Two-dimensional likelihood profile in the r_{inv}–$r_{\ell\ell}$ plane for a fit with uncorrelated scaling parameters on the $\mathrm{Z}(\to \nu\bar{\nu}) + \text{jets}$ and $\mathrm{Z}(\to \ell\bar{\ell}) + \text{jets}$ processes. The SM prediction is within the 68% confidence interval (CI) with the best fit point slightly higher in both parameters. The nominal fit with the single parameter of interest r_{inv} can be interpreted as the gradient of the line through the origin to any point on the plane. Therefore, the uncertainty associated with this parameter is less than the uncertainty of the separate parameters.

The best fit value of the parameter of interest and its associated uncertainty is measured as

$$r_{\mathrm{inv}} = 1.021 \pm 0.006\ (\text{stat.})\ ^{+0.031}_{-0.030}\ (\text{syst.})\ . \tag{8.2}$$

As with the likelihood profiles, this is consistent with the SM value of one, and is systematically dominated. This is translated into an invisible width, by scaling the SM prediction [9],

$$\Gamma_{\mathrm{inv}} = 512 \pm 3\ (\text{stat.})\ ^{+16}_{-15}\ (\text{syst.})\ \text{MeV}\ , \tag{8.3}$$

with the summary of the direct measurement of the invisible width shown in Fig. 8.8. The measurement presented is comparable to the LEP combined direct measurement.

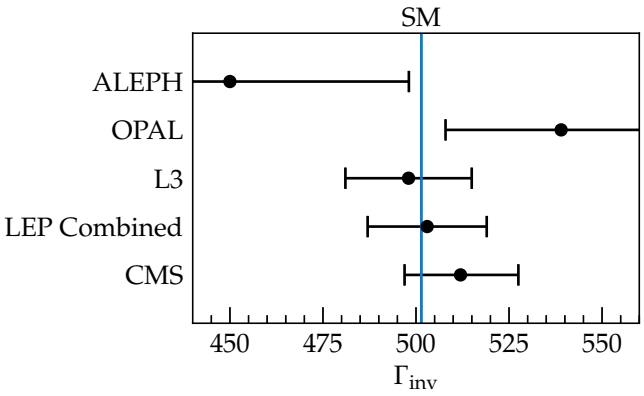


Figure 8.8.: Measurement of the Z invisible width by the LEP experiments and the result with the CMS experiment presented here. The SM prediction is included with a negligible uncertainty.

Chapter 9.

Conclusions

A measurement of the Z invisible width is performed with $35.9\,\text{fb}^{-1}$ of data collected by the CMS experiment during the 2016 proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$. The invisible width is measured from the $p_{\text{T}}^{\text{miss}}$ + jets signature left in the detector as the Z boson decays into neutrinos which pass through undetected. These are compared to the muon and electron (collectively ℓ) reference decays of the Z boson by collecting events with the $\ell\ell$ + jets signature. The selection placed on these events is similar to the $p_{\text{T}}^{\text{miss}}$ + jets region to restrict to similar kinematics. In particular, the $\mathcal{U}$ is used in place of the missing transverse energy as a proxy for the boson p_{T} in V + jets events. The modelling of this variable, as a proxy for the boson p_{T}, is validated in data by measuring its scale and resolution in events with a fully reconstructed dilepton for the boson p_{T}. This parameter is further used in the ℓ + jets control regions, where a data-driven estimate of the large $\text{W}(\to \ell\nu)$ + jets background in the $p_{\text{T}}^{\text{miss}}$ + jets region is performed through the transfer factor approach. Other backgrounds are significantly smaller and estimated from MC, apart from the QCD multijet process. A QCD multijet enriched sideband is collected by inverting the selection on $\Delta\phi_{\text{min}}$. Within this region, a correction factor for the QCD process is determined by extrapolation, with conservatively large uncertainties on the small background. In addition to these background estimations, the $\text{Z}/\gamma^* \to \ell\bar{\ell}$ process is divided into the $\text{Z} \to \ell\bar{\ell}$ and $\gamma^* \to \ell\bar{\ell}$ constituents to determine the contributions from each subprocess and their interference.

The invisible width is extracted with a simultaneous likelihood fit between the $p_{\text{T}}^{\text{miss}}$ + jets, ℓ + jets and $\ell\ell$ + jets regions. The parameter of interest is extracted as a scaling parameter on the $\text{Z}(\to \nu\bar{\nu})$ + jets process, relative to $\text{Z}(\to \ell\bar{\ell})$ + jets. This parameter is translated into an invisible width measurement of

$$\Gamma_{\text{inv}} = 512 \pm 3\ (\text{stat.})\ ^{+16}_{-15}\ (\text{syst.})\ \text{MeV}\,, \tag{9.1}$$

consistent with the SM and comparable to the LEP combined measurement. The measurement presented is systematically dominated, with the limiting sources from the jet energy scale, muon identification and trigger efficiency uncertainties.

The limiting systematic uncertainties imply that this measurements will not show significant gains from a larger dataset from the continual running of the LHC at increased luminosities. However, similar precision-based measurements are crucial tests of the SM, especially with the lack of BSM physics at the LHC. These include furthering our understanding of the Higgs boson since its discovery at the LHC, as well as probing the electroweak sector with precision measurements.

The results shown are currently in the review process within the CMS collaboration. Most aspects of the analysis are in place, missing only the lepton momentum scale uncertainties. These uncertainties are expected to have a 1–2 % affect on the final result, uncorrelated with all other uncertainties. In addition to these, a few techniques are under investigation to improve the limited MC sample size affecting the jet energy scale uncertainties with the hope of improving the r_{inv} measurement to about 2.7 %.

Appendix A.

The likelihood formalism

A.1. Definition

Throughout the analysis, parameter estimation is performed under the likelihood formalism, whereby the maximum likelihood estimator for a set of m parameters $\boldsymbol{\theta}$, $L(\boldsymbol{\theta})$, is maximised

$$\frac{\partial \ln L(\boldsymbol{\theta})}{\partial \theta_j} = 0 \,, \tag{A.1}$$

where j enumerates the set $\boldsymbol{\theta}$ from 1 to m. For a set of n statistically independent measurements $\boldsymbol{x}$, each sampled from a probability density function $f(x_i; \boldsymbol{\theta})$ given the set of parameters, the likelihood function is

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i; \boldsymbol{\theta}) \,. \tag{A.2}$$

The set of parameters which maximises this functions is denoted $\hat{\boldsymbol{\theta}}$.

For large sample size, typical with CMS data, the maximisation procedure become computationally demanding. Instead, the data is binned and the measurements $\boldsymbol{x}$ represent the event count in the bins. Typically the event count in a particular bin is considered to be a random variable, hence the probability density function is a poisson distribution P for the binned data, given a prediction $\boldsymbol{\nu}$,

$$f(x_i; \boldsymbol{\theta}) = P(x_i; \nu_i(\boldsymbol{\theta})) \,. \tag{A.3}$$

The poisson distribution is replaced by a multinomial distribution if the event count is considered as fixed. The predictions $\boldsymbol{\nu}$ are determined from a functional form or estimated from MC simulations.

The set of parameters may contain unconstrained parameters such as a signal strength multiplier as well as constrained parameters, known as nuisances parameters, which encode

systematic uncertainties. For the estimation of a signal strength multiplier the parameters are divided into unconstrained $\boldsymbol{\mu}$, gaussian constrained $\boldsymbol{\theta}$ and poisson constrained $\boldsymbol{\phi}$ characterising limited MC sample sizes [75]. With these parameters the likelihood function is given by

$$L(\boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{\phi}) = \prod_{i=1}^{n} P(x_i; \nu_i(\boldsymbol{\mu}, \boldsymbol{\theta}, \phi_i)) P(n_i; \gamma_i(\phi_i)) \prod_{j=1}^{m} N(\theta_j) \,, \tag{A.4}$$

where $N(\theta_j)$ is a gaussian distribution with a zero mean and unit width, $P(n_i; \gamma_i(\phi_i))$ is a poisson distribution extended to real numbers for weighted MC events with an effective number of entries n_i (the ratio of the sum of weighted events to its variance) with a ϕ_i dependent prediction γ_i. This prediction is given by

$$\gamma_i(\phi_i) = n_i \left(1 + \frac{1}{\sqrt{n_i}}\right)^{\phi_i} \tag{A.5}$$

to allow ϕ_i to taken any real number value without resulting in a negative prediction. The factor $1 + 1/\sqrt{n_i}$ is chosen to yield convenient values for ϕ_i, typically between -1 and $+1$. The poisson constraints are applied to each bin on the overall MC prediction, summing all signal and background contributions. For a single signal and single background process, with contributions s_i and b_i, the prediction ν_i takes the form

$$\nu_i(\mu, \boldsymbol{\theta}, \phi_i) = \frac{\gamma_i(\phi_i)}{n_i} \left(\mu s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta})\right) \tag{A.6}$$

The $\boldsymbol{\theta}$ dependence of $s_i(\boldsymbol{\theta})$ and $b_i(\boldsymbol{\theta})$ is determined from alternative MC templates generated at $\theta_k = -1, 0, +1$, with all other values interpolated or extrapolated [76]. To encode the shape dependence of these nuisance parameters along a particular dimension the set of bins $i = 1$ to n are split into N categories, each with M bins along the 'shape' dimension. The effect of the nuisance parameters is factorised into the inclusive normalisation r_{norm} and a shape morphing term r_{morph}:

$$s_{i,j}(\boldsymbol{\theta}) = s_{i,j} \prod_{k=1}^{m} r_{\text{norm},i,k}(\theta_k) \sum_{j=1}^{m} (1 + r_{\text{morph},i,j,k}(\theta_k)) \,, \tag{A.7}$$

where the signal contribution is subscripted by the categorisation i and shape dimension bin j with MC prediction for the signal $s_{i,j}$, and a similarly defined relationship for the background $b_{i,j}(\boldsymbol{\theta})$. The normalisation term is determined from the MC templates generated at $\theta_k = -1, 0, +1$: $s^-_{i,j,k}$, $s_{i,j}$ and $s^+_{i,j,k}$; as

$$r_{\text{norm},i,k}(\theta_k) = \left(\frac{1}{\sum_{j=1}^{M} s_{i,j}} \sum_{j=1}^{M} \begin{cases} s^+_{i,j,k} & \text{if } \theta_k > 0 \\ s^-_{i,j,k} & \text{otherwise} \end{cases} \right)^{|\theta_k|} \,. \tag{A.8}$$

Meanwhile, the shape morphing term extrapolates linearly for $|\theta_k| > 1$ with a smooth polynomial below this range:

$$r_{\text{morph},i,j,k} = \frac{\theta_k}{2\hat{s}_{i,j}} \left(\Delta_{i,j,k} + \Sigma_{i,j,k} p(\theta_k)\right) \ , \tag{A.9}$$

where the hat denotes parameters normalised by their sum across the shape dimension, and the other parameters are defined as

$$\Delta_{i,j,k} = \hat{s}^{+}_{i,j,k} - \hat{s}^{-}_{i,j,k} \ , \tag{A.10}$$

$$\Sigma_{i,j,k} = (\hat{s}^{+}_{i,j,k} - \hat{s}_{i,j}) + (\hat{s}^{-}_{i,j,k} - \hat{s}_{i,j}) \ , \tag{A.11}$$

$$p(\theta_k) = \begin{cases} \frac{1}{8}\theta_k \left(\theta_k^2(3\theta_k^2 - 10) + 15\right) & \text{if } |\theta_k| < 1 \\ \theta_k / |\theta_k| & \text{otherwise .} \end{cases} \tag{A.12}$$

This procedure is known as vertical template morphing[76] which uses normalised parameters to avoid double-counting the normalisation effects, with a linear dependence on θ_k for symmetric uncertainties and a smooth, differentiable polynomial to alter asymmetric uncertainties. A similar dependence is defined for all other processes.

A.2. Limited sample size of alternative templates

The accuracy of the vertical template morphing procedure is highly dependent on the quality of the alternative MC templates $s^{+}_{i,j,k}$ and $s^{-}_{i,j,k}$ which may suffer from the limited sample size in MC. Although the statistical size of the nominal template $s_{i,j}$ is incorporated in the likelihood, the alternative MC templates are assumed to be precisely known. This issue typically exists for background processes with a small cross section compared to a larger process, hence should not have a substantial impact on the parameters of interest. However, an alternative template with a small sample size which results in an abnormally large systematic effect in a particular bin will lead to a significant constraint on the associated nuisance parameter. If this systematic uncertainty is correlated with one of the larger processes then the impact of this systematic uncertainty may be severely underestimated on the parameter of interest.

To avoid issues with the limited sample size of alternative templates, additional constrained nuisance parameters are introduced in each bin for each systematic uncertainty, to allow the fit to effectively smooth out fluctuations due to the limited sample size. However, this would introduce a significant number of parameters to fit, beyond the computational power available. Instead the alternative templates are smoothed prior to the fitting procedure, as required. The procedure involves regularising large statistical variations in the relative $\pm 1\sigma$ variation templates. Bins centres are interpolated with a smoothing spline, where each point is inversely

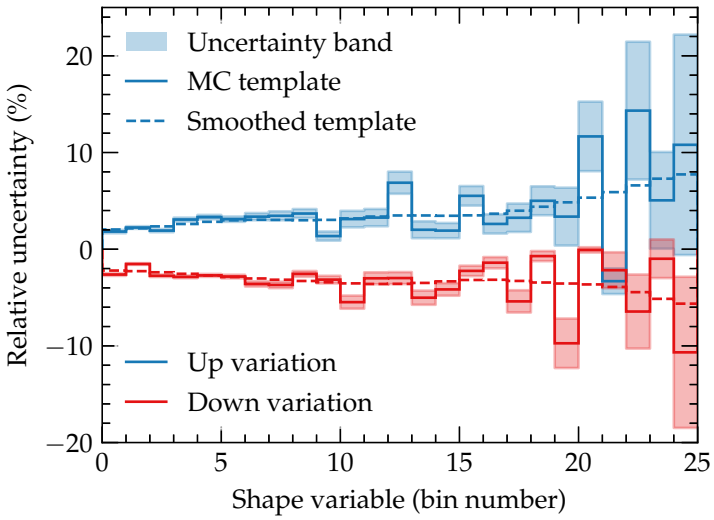


Figure A.1.: Smoothing of a statistically limited MC alternative template by a spline followed by a gaussian smear. The uncertainty band represents the uncorrelated uncertainty between the up and down variation as a representative uncertainty. The resulting smoothed template has significantly reduced the variation between adjacent bins, especially in the final bins where the statistical uncertainty is the greatest.

weighted by it's uncertainty, with a subsequent gaussian smearing to prevent large changes between adjacent bins. The normalisation effect of the variations is fixed during this smearing procedure. An example of the smoothing procedure is shown in Fig. A.1. This method proves to reduce the constraint on systematic uncertainties with a limited sample size, leading to an overall increase in the uncertainty on the parameter of interest. The smoothing procedure is only used where it is required, ensuring the method does not lead to a significant loss in information.

A.3. Alternative templates

The nuisance parameters affect the normalisation and shape of the $\mathcal{U}$ distribution by interpolating between and extrapolating from alternative templates on the MC, shown in Figs. A.2 and A.3 for the V + jets processes. As detailed in Sec. A.2, these templates are smoothed in situations where the MC sample size is too small. Some of these templates cancel in the transfer factor for r_{inv} measurement, such as the theory uncertainties. However, the JES, JER and unclustered energy uncertainties are evidently limited by the MC sample size. Although smoothing aids at determined the 'true' template, it cannot reproduce an underlying correlation such as for the JES uncertainty between various processes. Therefore, the 2–5 % JES uncertainty on the $Z(\to \nu\bar{\nu})$ + jets process in the $p_{\text{T}}^{\text{miss}}$ + jets region may be consistent with all other processes, however, the limited MC sample size limits our knowledge of this. This

results in non-cancellation of uncertainties, such as JES, which are expected to cancel in ratios of the various V + jets processes.

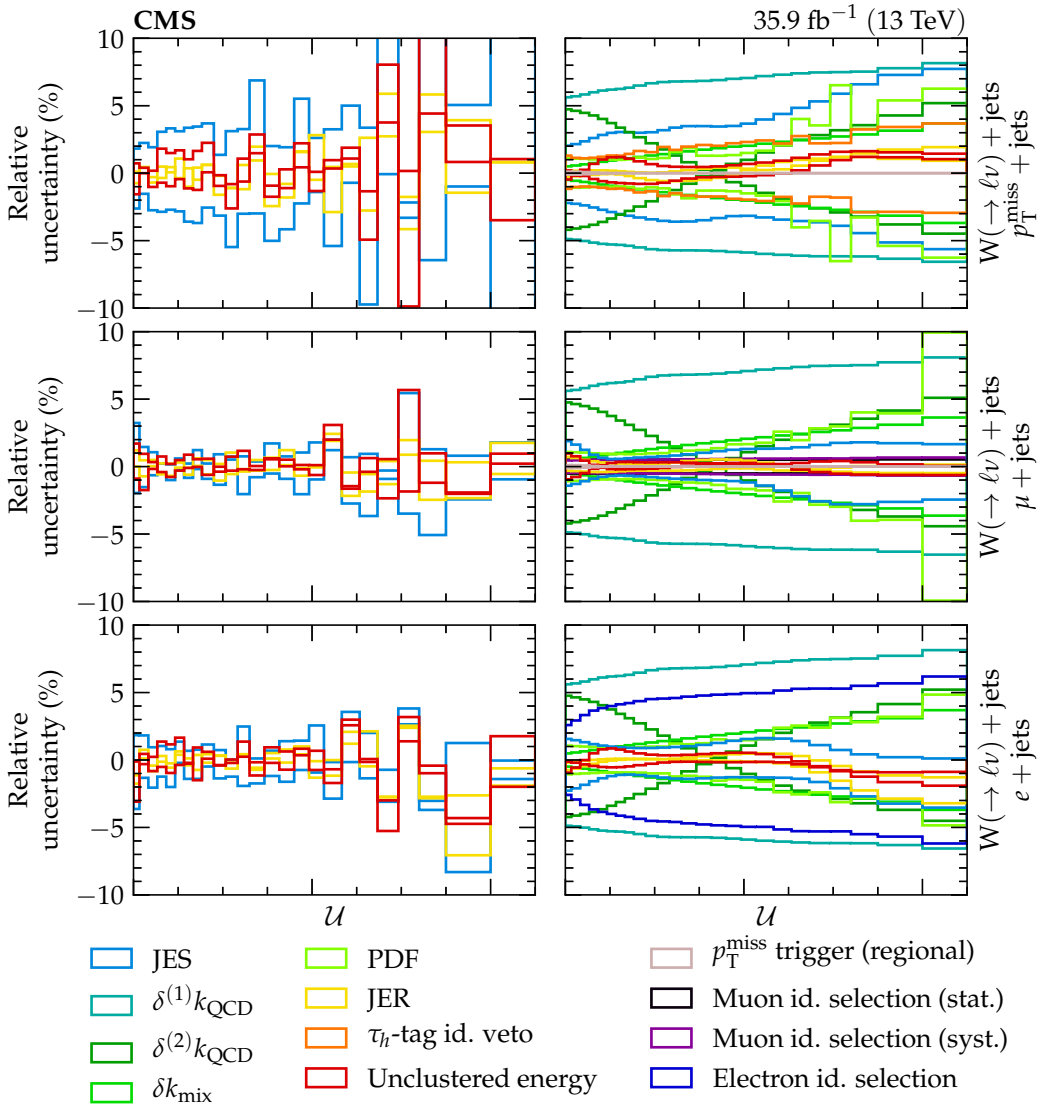


Figure A.2.: The 12 most significant alternative templates on the W + jets process. The left column shows the template prior to smoothing while the right column shows the same template after smoothing. Templates which only appear in the right column are not smoothed.

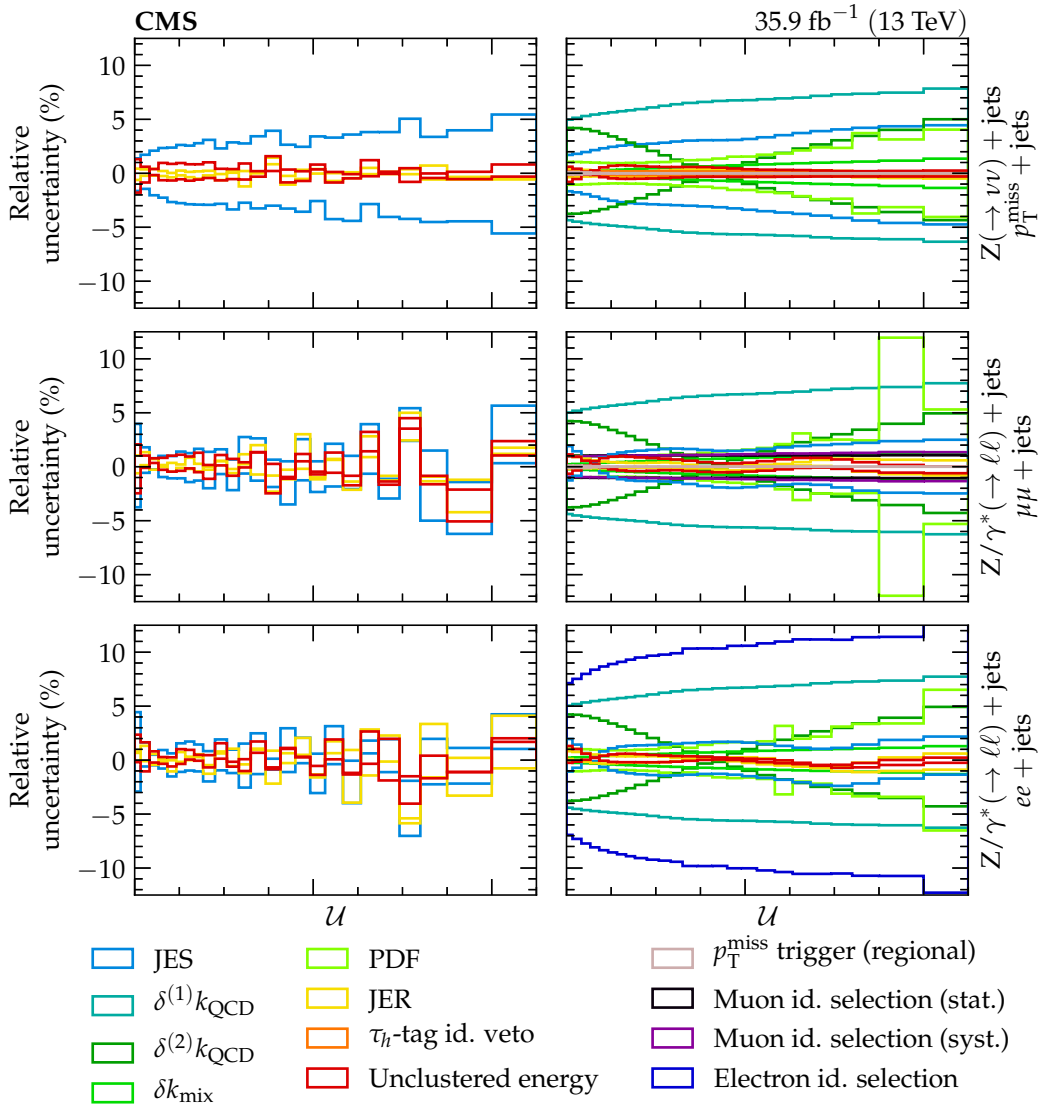


Figure A.3.: The 12 most significant alternative templates on the V + jets processes within the signal region. The left column shows the template prior to smoothing while the right column shows the same template after smoothing. Templates which only appear in the right column are not smoothed.

Bibliography

[1] CMS Collaboration, "Search for natural and split supersymmetry in proton-proton collisions at $\sqrt{s} = 13$ TeV in final states with jets and missing transverse momentum", *JHEP* **05** (2018) 025, doi:10.1007/JHEP05(2018)025, arXiv:1802.02110.

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

[3] E. Corbelli and P. Salucci, "The Extended Rotation Curve and the Dark Matter Halo of M33", *Mon. Not. Roy. Astron. Soc.* **311** (2000) 441–447, doi:10.1046/j.1365-8711.2000.03075.x, arXiv:astro-ph/9909252.

[4] Super-Kamiokande Collaboration, "Measurement of the flux and zenith angle distribution of upward through going muons by Super-Kamiokande", *Phys. Rev. Lett.* **82** (1999) 2644–2648, doi:10.1103/PhysRevLett.82.2644, arXiv:hep-ex/9812014.

[5] SNO Collaboration, "Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory", *Phys. Rev. Lett.* **89** (2002) 011301, doi:10.1103/PhysRevLett.89.011301, arXiv:nucl-ex/0204008.

[6] KamLAND Collaboration, "First results from KamLAND: Evidence for reactor anti-neutrino disappearance", *Phys. Rev. Lett.* **90** (2003) 021802, doi:10.1103/PhysRevLett.90.021802, arXiv:hep-ex/0212021.

[7] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. 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** (Aug, 2012) 1–29.

[8] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett. B* **716** (Jul, 2012) 30–61.

[9] Particle Data Group Collaboration, "Review of Particle Physics", *Phys. Rev. D* **98** (Aug, 2018) 030001, doi:10.1103/PhysRevD.98.030001.

[10] L3 Collaboration, "Determination of the number of light neutrino species from single

photon production at LEP", *Phys. Lett.* **B431** (1998) 199–208, doi:10.1016/S0370-2693(98)00519-X.

[11] OPAL Collaboration, "Measurement of single photon production in e+ e- collisions near the Z0 resonance", *Z. Phys.* **C65** (1995) 47–66, doi:10.1007/BF01571303.

[12] ALEPH Collaboration, "A Direct measurement of the invisible width of the Z from single photon counting", *Phys. Lett.* **B313** (1993) 520–534, doi:10.1016/0370-2693(93)90027-F.

[13] M. Carena, A. de Gouvea, A. Freitas et al., "Invisible Z boson decays at e+ e- colliders", *Phys. Rev.* **D68** (2003) 113007, doi:10.1103/PhysRevD.68.113007, arXiv:hep-ph/0308053.

[14] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys. Rev. Lett.* **13** (Aug, 1964) 321–323, doi:10.1103/PhysRevLett.13.321.

[15] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (Oct, 1964) 508–509, doi:10.1103/PhysRevLett.13.508.

[16] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global Conservation Laws and Massless Particles", *Phys. Rev. Lett.* **13** (Nov, 1964) 585–587, doi:10.1103/PhysRevLett.13.585.

[17] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett.* **B716** (2012) 1–29, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.

[18] CMS Collaboration, "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC", *Phys. Lett.* **B716** (2012) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

[19] E. Noether, "Invariant variation problems", *Transport Theory and Statistical Physics* **1** (1971), no. 3, 186–207, doi:10.1080/00411457108231446, arXiv:https://doi.org/10.1080/00411457108231446.

[20] M. E. Peskin and D. V. Schroeder, "An Introduction to quantum field theory". Addison-Wesley, Reading, USA, 1995.

[21] J. L. Troutman, "The Euler-Lagrange Equations", pp. 142–190. Springer US, New York, NY, 1983. doi:10.1007/978-1-4684-0158-5_7.

[22] P. A. M. Dirac, "The Quantum Theory of the Electron", *Proceedings of the Royal Society of London Series A* **117** (February, 1928) 610–624, doi:10.1098/rspa.1928.0023.

[23] N. Cabibbo, "Unitary Symmetry and Leptonic Decays", *Phys. Rev. Lett.* **10** (Jun, 1963)

531–533, doi:10.1103/PhysRevLett.10.531.

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

[25] B. Pontecorvo, "Inverse beta processes and nonconservation of lepton charge", *Sov. Phys. JETP* **7** (1958) 172–173.

[26] Z. Maki, M. Nakagawa, and S. Sakata, "Remarks on the unified model of elementary particles", *Prog. Theor. Phys.* **28** (1962) 870–880, doi:10.1143/PTP.28.870.

[27] R. K. Ellis, W. J. Stirling, and B. R. Webber, "QCD and collider physics", *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.* **8** (1996) 1–435.

[28] V. V. Sudakov, "Vertex parts at very high-energies in quantum electrodynamics", *Sov. Phys. JETP* **3** (1956) 65–71.

[29] G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language", *Nucl. Phys.* **B126** (1977) 298–318, doi:10.1016/0550-3213(77)90384-4.

[30] Y. L. Dokshitzer, "Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics.", *Sov. Phys. JETP* **46** (1977) 641–653.

[31] V. N. Gribov and L. N. Lipatov, "Deep inelastic e p scattering in perturbation theory", *Sov. J. Nucl. Phys.* **15** (1972) 438–450.

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

[33] G. P. Salam, "Towards Jetography", *Eur. Phys. J.* **C67** (2010) 637–686, doi:10.1140/epjc/s10052-010-1314-6, arXiv:0906.1833.

[34] O. S. Brüning, P. Collier, P. Lebrun et al., "LHC Design Report". CERN Yellow Reports: Monographs. CERN, Geneva, 2004.

[35] CMS Collaboration, G. L. Bayatian, S. Chatrchyan, G. Hmayakyan et al., "CMS Physics: Technical Design Report Volume 1: Detector Performance and Software". Technical Design Report CMS. CERN, Geneva, 2006.

[36] T. Sakuma and T. McCauley, "Detector and Event Visualization with SketchUp at the CMS Experiment", *J. Phys. Conf. Ser.* **513** (2014) 022032, doi:10.1088/1742-6596/513/2/022032, arXiv:1311.4942.

[37] L. Borrello, A. Messineo, E. Focardi et al., "Sensor Design for the CMS Silicon Strip Tracker", Technical Report CMS-NOTE-2003-020, CERN, Geneva, (Aug, 2003).

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

[39] J. Mans, J. Anderson, B. Dahmes et al., "CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter", Technical Report CERN-LHCC-2012-015. CMS-TDR-10, (Sep, 2012).

[40] G. Bayatian et al., "Design, performance and calibration of the CMS forward calorimeter wedges", *Eur. Phys. J.* **C53** (2008) 139–166, doi:10.1140/epjc/s10052-007-0459-4.

[41] CMS Collaboration, G. L. Bayatyan, N. Grigorian, V. G. Khachatrian et al., "CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems". Technical Design Report CMS.

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

[43] CMS Collaboration, P. Sphicas, "CMS: The TriDAS project. Technical design report, Vol. 2: Data acquisition and high-level trigger". Technical Design Report CMS.

[44] I. Bird, P. Buncic, F. Carminati et al., "Update of the Computing Models of the WLCG and the LHC Experiments", Technical Report CERN-LHCC-2014-014. LCG-TDR-002, (Apr, 2014).

[45] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST* **9** (May, 2014) P10009. 80 p, doi:10.1088/1748-0221/9/10/P10009.

[46] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems", *Transactions of the ASME–Journal of Basic Engineering* **82** (1960), no. Series D, 35–45.

[47] T. Speer, K. Prokofiev, R. Frühwirth et al., "Vertex Fitting in the CMS Tracker", Technical Report CMS-NOTE-2006-032, CERN, Geneva, (Feb, 2006).

[48] R. Frühwirth, W. Waltenberger, and P. Vanlaer, "Adaptive Vertex Fitting", Technical Report CMS-NOTE-2007-008, CERN, Geneva, (Mar, 2007).

[49] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", *JINST* **12** (2017), no. 10, P10003, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.

[50] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm", *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.

[51] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet User Manual", *Eur. Phys. J.* **C72** (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.

[52] R. Frederix, S. Frixione, V. Hirschi et al., "The automation of next-to-leading order electroweak calculations", *JHEP* **07** (2018) 185, doi:10.1007/JHEP07(2018)185, arXiv:1804.10017.

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

[54] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements", *Eur. Phys. J.* **C76** (2016), no. 3, 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.

[55] R. Frederix and S. Frixione, "Merging meets matching in MC@NLO", *JHEP* **12** (2012) 061, doi:10.1007/JHEP12(2012)061, arXiv:1209.6215.

[56] P. Artoisenet, R. Frederix, O. Mattelaer et al., "Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations", *JHEP* **03** (2013) 015, doi:10.1007/JHEP03(2013)015, arXiv:1212.3460.

[57] S. Alioli, P. Nason, C. Oleari et al., "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP* **06** (2010) 043, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.

[58] S. Alioli, P. Nason, C. Oleari et al., "NLO single-top production matched with shower in POWHEG: s- and t-channel contributions", *JHEP* **09** (2009) 111, doi:10.1007/JHEP02(2010)011,10.1088/1126-6708/2009/09/111, arXiv:0907.4076.

[59] E. Re, "Single-top Wt-channel production matched with parton showers using the POWHEG method", *Eur. Phys. J.* **C71** (2011) 1547, doi:10.1140/epjc/s10052-011-1547-z, arXiv:1009.2450.

[60] NNPDF Collaboration, "Parton distributions for the LHC Run II", *JHEP* **04** (2015) 040, doi:10.1007/JHEP04(2015)040, arXiv:1410.8849.

[61] S. Agostinelli, J. Allison, K. Amako et al., "Geant4—a simulation toolkit", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** (2003), no. 3, 250 – 303, doi:https://doi.org/10.1016/S0168-9002(03)01368-8.

[62] CMS Collaboration, "Muon Identification and Isolation efficiency on full 2016 dataset". https://cds.cern.ch/record/2257968, Mar, 2017.

[63] CMS Collaboration, "Electron and photon performance in CMS with the full 2016 data sample.". https://cds.cern.ch/record/2255497, Mar, 2017.

[64] CMS Collaboration, "Reconstruction and identification of τ lepton decays to hadrons and ν_τ at CMS. Reconstruction and identification of tau lepton decays to hadrons and tau neutrino at CMS", *JINST* **11** (Oct, 2015) P01019. 72 p, doi:10.1088/1748-0221/11/01/P01019.

[65] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV", *JINST* **13** (2018), no. 05, P05011, doi:10.1088/1748-0221/13/05/P05011, arXiv:1712.07158.

[66] CMS Collaboration, "Filtering out noise in HCAL barrel and endcaps". https://cds.cern.ch/record/2221745, Sep, 2016.

[67] CMS Collaboration, "Muon HLT performance on 2016 data". https://cds.cern.ch/record/2297529, Dec, 2017.

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

[69] CMS Collaboration, "Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV", *JINST* **13** (2018), no. 10, P10005, doi:10.1088/1748-0221/13/10/P10005, arXiv:1809.02816.

[70] S. van der Meer, "Calibration of the effective beam height in the ISR", Technical Report CERN-ISR-PO-68-31. ISR-PO-68-31, CERN, Geneva, (1968).

[71] CMS Collaboration, "CMS Luminosity Measurements for the 2016 Data Taking Period", cms-pas-lum-17-001, (2017).

[72] J. M. Lindert et al., "Precise predictions for $V+$ jets dark matter backgrounds", *Eur. Phys. J.* **C77** (2017), no. 12, 829, doi:10.1140/epjc/s10052-017-5389-1, arXiv:1705.04664.

[73] CMS Collaboration, "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector", *JINST* **14** (2019), no. 07, P07004, doi:10.1088/1748-0221/14/07/P07004, arXiv:1903.06078.

[74] Z. Bern et al., "Left-Handed W Bosons at the LHC", *Phys. Rev.* **D84** (2011) 034008, doi:10.1103/PhysRevD.84.034008, arXiv:1103.5445.

[75] R. J. Barlow and C. Beeston, "Fitting using finite Monte Carlo samples", *Comput. Phys. Commun.* **77** (1993) 219–228, doi:10.1016/0010-4655(93)90005-W.

[76] J. S. Conway, "Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra", in *Proceedings, PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, CERN,Geneva, Switzerland 17-20 January 2011*,

pp. 115–120. 2011. arXiv:1103.0354. doi:10.5170/CERN-2011-006.115.