Observation and Measurement of the Electroweak $W^{\pm}W^{\pm}jj$ Process in Proton-Proton Collisions at $\sqrt{s}$ = 13 TeV with the ATLAS Detector

Giulia Gonella

Fakultät für Mathematik und Physik
Albert-Ludwigs-Universität Freiburg

Observation and Measurement of the Electroweak $W^{\pm}W^{\pm}jj$ Process in Proton-Proton Collisions at $\sqrt{s}$ = 13 TeV with the ATLAS Detector

Dissertation

zur Erlangung des Doktorgrades der

Fakultät für Mathematik und Physik der

Albert-Ludwigs-Universität

Freiburg im Breisgau

vorgelegt von

Giulia Gonella

August, 2019

Giulia Gonella: *Observation and Measurement of the Electroweak $W^{\pm}W^{\pm}jj$ Process in Proton-Proton Collisions at $\sqrt{s}$ = 13 TeV with the ATLAS Detector*

Datum der mündlichen Prüfung: 25/10/2019

Dekan: Prof. Dr. W. Soergel
Betreuer der Arbeit: Prof. Dr. K. Jakobs, Dr. C. Weiser
Koreferent: Prof. Dr. B. Heinemann
Prüfer: Prof. Dr. S. Dittmaier, Prof. Dr. M. Schumacher

This was really the way my whole road experience began,
and the things that were to come are too fantastic not to tell.

Jack Kerouac, On the Road

Abstract

Measurements of the production of W boson pairs through vector boson scattering are a vital test for the electroweak sector of the Standard Model (SM) and the mechanism of the electroweak symmetry breaking. They could detect small deviations from the predictions through which new physics could manifest itself. In the SM the couplings of the gauge bosons among themselves and with the Higgs boson are precisely balanced such that divergences in the WW scattering amplitude are avoided, and small deviations from the predictions could lead to significant effects.

This thesis presents the first measurement with the ATLAS detector at the LHC of the electroweak production of two W bosons with the same electric charge, in the signature with two charged leptons, missing transverse energy and two jets. The process is observed for the first time at ATLAS with a significance of 6.5σ using data recorded in proton-proton collisions at $\sqrt{s}$ = 13 TeV corresponding to an integrated luminosity of 36.1 fb^{-1}, and its cross section in a defined fiducial phase space is measured to be $\sigma^{\text{fid}}_{\text{meas}} = 2.89^{+0.51}_{-0.48}(\text{stat})\ ^{+0.29}_{-0.28}(\text{syst})$ fb.

The requirement of two same-charge W bosons leads on one hand to a better suppression of background processes, among them the WW production induced by the strong interaction. On the other hand it is experimentally more challenging due to the contributions from opposite-charge dilepton production, where the charge of one electron is wrongly reconstructed. This thesis details the estimation of this background, whose contribution in the signal region kinematic phase space is determined with a relative uncertainty of 15%.

A precise estimation of this background requires a good knowledge of the probability of reconstructing an electron with the wrong charge sign. A measurement of this quantity is presented in this thesis, and the probabilities vary from less than 1% to about 10% for increasing transverse momentum and pseudorapidity of the electrons. Correction factors are provided for the modelling of this effect in simulation, which deviate from unity by up to 30% for wrongly reconstructed electrons, and are determined with relative systematic uncertainties between 2% and 20%.

The contained amount of data limits the precision and sensitivity of the measurement. However, vector boson scattering is one of the most favourable ways to assess clues on the theoretical extension of the SM in a model independent way, and larger datasets at higher energies will make the inner mechanism of the electroweak symmetry breaking and possible new physics more accessible.

Kurzzusammenfassung

Messungen der Produktion von W-Bosonpaaren durch Vektorbosonstreuuung stellen einen unerlässlichen Test des elektroschwachen Sektors des Standard Models (SM) sowie des Mechanismus der elektroschwachen Symmetriebrechung dar. Sie könnten kleine Abweichungen von den Vorhersagen aufdecken, durch die sich neue Physik manifestieren könnte. Im SM werden Divergenzen in der WW-Streuamplitude durch eine feine Abstimmung der Kopplungen der Eichbosonen untereinander sowie mit dem Higgsboson vermieden. Kleine Abweichungen von der Vorhersage könnten deutliche Auswirkungen haben.

Diese Dissertation präsentiert die erste Messung der elektroschwachen Produktion zweier W-Bosonen gleicher elektrischer Ladung mit dem ATLAS Detektor am LHC. Die Messung basiert auf der Signatur mit zwei geladenen Leptonen, fehlender transversaler Energie und zwei Jets. Mithilfe eines Datensatzes von Proton-Proton-Kollisionen, der bei einer Schwerpunktsenergie von $\sqrt{s}$ = 13 GeV aufgenommen wurde und einer integrierten Luminosität von 36,1 fb^{-1} entspricht, wird der Signalprozess mit einer Signifikanz von 6,5σ nachgewiesen. Der Wirkungsquerschnitt in einem vordefinierten, den experimentellen Gegebenheiten angepassten Phasenraum wird zu $\sigma^{\text{fid}}_{\text{meas}} = 2.89^{+0.51}_{-0.48}(\text{stat})\ ^{+0.29}_{-0.28}(\text{syst})$ fb gemessen.

Die Forderung zweier gleichgeladener W-Bosonen unterdrückt einerseits Untergrundprozesse wie eine durch die starke Wechselwirkung induzierte WW Produktion. Andererseits ist sie aufgrund von Beiträgen durch die Produktion ungleichgeladener Leptonen, bei der die Ladung eines Elektrons fehlerhaft rekonstruiert wird, experimentell herausfordernd. Diese Arbeit führt die Abschätzung dieses Untergrundes aus, dessen Beitrag in der Signalregion mit einer relativen Unsicherheit von 15% bestimmt wird.

Eine präzise Abschätzung dieses Untergrunds erfordert eine gute Kenntnis darüber wie wahrscheinlich ein Elektron mit falschem Ladungsvorzeichen rekonstruiert wird. Dies wird in der vorliegenden Arbeit untersucht, wobei die gemessenen Wahrscheinlichkeiten von unter 1% bis zu etwa 10% variieren und mit steigendem Transversalimpuls der Elektronen sowie mit ihrer Pseudorapidität zunehmen. Korrekturfaktoren für die Modellierung der Ladungsrekonstruktion in der Simulation werden zur Verfügung gestellt. Diese weichen für fehlerhaft rekonstruierte Elektronen um bis zu 30% von eins ab und werden mit relativen systematischen Unsicherheiten von 2% bis 30% bestimmt.

Die Größe des genutzten Datensatzes limitiert die Präzision und die Sensitivität der Messung. Allerdings stellt Vektorbosonstreuung eine der vorteilhaftesten Wege dar Hinweise auf theoretische Erweiterungen des Standardmodells in einer modellunabhängigen Art aufzuspüren, und größere Datensätze bei höheren Energien werden den Kern der elektroschwachen Symmetriebrechung und mögliche neue Physik besser zugänglich machen.

Contents

1 Introduction

Over the last decades experiments have put the descriptive and predictive power of the Standard Model (SM), the theory of fundamental interactions among elementary particles, to numerous stringent tests. The successful structure of the SM describes the electromagnetic, weak and strong fundamental forces as mediated by gauge vector bosons. These emerge to preserve gauge symmetries which are imposed in the model. The electroweak gauge vector fields Z, W^+, W^- and γ are predicted to self-interact due to the non-Abelian nature of their symmetry group structure $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$. The Z and $W^\pm$ bosons acquire mass via the mechanism of spontaneous symmetry breaking, which is incorporated in the SM by introducing scalar fields with a specific potential term. The consequence is a physical Higgs boson.

The fortune of the SM is spoilt by phenomena like the baryon-antibaryon asymmetry in the Universe or the presence of dark matter, which do not find explanations within its framework. Any deviation from the SM predictions might indicate the direction for its completion. This is the foundation of the Large Hadron Collider at CERN, conceived to push the limit of our current understanding of fundamental interactions and to reveal their potential unexpected behaviour. By analysing data from proton-proton collisions collected by detectors like ATLAS [1], the SM predictions can be challenged, and to this end the electroweak sector is one of the key fields from the start on. The luminosity delivered at centre-of-mass energies never achieved in previous experiments, in particular during Run 2 when 139 fb^{-1} of data has been collected at $\sqrt{s}$ = 13 TeV, launched a new era in particle physics for probing the inner electroweak anatomy and its spontaneous breaking mechanism.

A favoured way to do this is through detailed measurements of so-called Vector Boson Scattering (VBS) processes, which refers to interactions where partons from colliding protons emit gauge bosons interacting with each other and create a forward-backward, high-mass two-jets system. The emitted bosons interact not only via triple and quartic couplings making VBS processes a test for multi-boson interactions, but also through the Higgs boson. In particular, the scattering of longitudinally polarised W bosons plays a central role in testing the Higgs boson's action in the SM. In fact, the W bosons scattering cross section would grow with the centre-of-mass energy if diagrams involving the Higgs boson, with couplings as predicted by the SM, would not restore unitarity by cancelling the divergences.

Final states like those of VBS can also be produced in other electroweak processes which are not separable from the genuine VBS diagrams without breaking the gauge invariance. Processes that involve the strong interaction at leading order can lead to the same final state, and their impact in analyses can only be reduced through dedicated selection criteria that exploit the different preferred kinematic topologies for the electroweak processes and those with the strong contributions.

The quantity of diagrams involved in VBS processes brings along difficulties for the theoretical calculations, making precise predictions complex to achieve. The VBS processes are challenging for experiments as well, due to the low cross-sections compared to other SM processes and the complex final states which require a precise understanding of background process. In this scenario, the study of the scattering of two W bosons with the same electric charge is favoured, because the contributions of diagrams that include strong vertices at leading order are suppressed with respect to the electroweak-induced ones, and the number of diagrams involved is limited by the

same-charge requirement. Furthermore, other background processes that produce opposite-charge lepton signatures, like $t\bar{t}$, W^+W^- or $Z \to \tau\tau$, are suppressed more efficiently.
Indeed, the $pp \to W^\pm W^\pm jj \to \ell^\pm \nu \ell^\pm \nu jj$ was the first VBS process whose evidence of 3.6σ over background-only expectation was reported by the ATLAS experiment in proton-proton collisions at a centre-of-mass energy $\sqrt{s}$ = 8 TeV with 20.8 fb^{-1} of data [2]. The CMS Collaboration presented a 2.0σ excess in a similar 8 TeV study [3]. Both ATLAS and CMS [4] Collaborations observed the process in Run 2 at $\sqrt{s}$ = 13 TeV, and the ATLAS measurement is documented in this thesis and has been published in Ref. [5]. The selection is optimised in order to increase the expected significance of the signal process over the background, and the measurement of its fiducial cross section is performed with a maximum likelihood fit methodology.
One of the most arduous backgrounds to control in same-charge signatures arises from the charge misidentification of electrons. The charge misidentification probability is generally small, but processes producing opposite-charge leptons have a much higher cross section than the signal, and can therefore yield a large amount of events passing the same-charge requirement if one electron is wrongly reconstructed.
In order to estimate the contribution of this background, it is necessary to precisely measure the probability of reconstructing the wrong electric charge of electrons and to apply corrections to the simulation to adequately model this effect. The measurement of the probabilities of charge misidentification is presented in this thesis and has been published in Ref. [6]. The results of this measurement have been provided to the whole ATLAS Collaboration.
The theoretical framework of the SM is outlined in Chapter 2, where VBS processes are also discussed. The event simulation techniques are presented in Chapter 3. A description of the ATLAS detector is given in Chapter 4, and Chapter 5 is dedicated to object reconstruction techniques. In Chapter 6 the measurement of charge misidentification probabilities is detailed and the optimisation of the event selection for the $W^\pm W^\pm jj$ analysis is described in Chapter 7. Chapter 8 is dedicated to the estimation of background processes and the analysis results are presented in Chapter 9. Finally, a summary is given in Chapter 10.

2 Theoretical Framework

The Standard Model (SM) of particle interactions is a unified framework able to describe and provide predictions on three out of four fundamental forces that act among elementary particles, the electromagnetic, the weak and the strong interaction, not accounting for gravity. The construction of the SM has been an extraordinary success of the interplay of theory and experiment, which played complementary roles in formulating the model that emerged in the shape that we know today. The former constructed a solid theoretical basis respecting the needed requirements for a descriptive and predictive quantum field theory, and was constantly challenged by the new experimental discoveries over the last decades. The result is a renormalisable, non-Abelian gauge theory of quantum fields, based on the symmetry groups $SU(3)_C \times SU(2)_L \times U(1)_Y$. The SM can be constructed using a few guiding principles. The most important are its particles content, its external and internal symmetries, such as the Lorentz and gauge invariance, the first rigid and exact, the second local and spontaneously broken, and the requirement of renormalisability strictly linked to the dimension of the allowed operators. The SM particle constituents are described in Section 2.1, while Section 2.2 offers a brief overview of the theoretical structure of the electroweak (EW) and Quantum Chromodynamics (QCD) sectors. In Section 2.3 the mechanism of electroweak symmetry breaking is discussed in its application to the SM. The discussion is mainly based on [7–10], while a useful review is offered in [11]. A phenomenological overview on the Vector Boson Scattering (VBS) processes is given in Section 2.4, with a focus on the scattering of W bosons.

2.1 Particle Content

The SM elementary particles consist of twelve fermions and twelve vector bosons, having half-integer and integer spin respectively, and one scalar boson. Fermions are organised in three generations, each one including two quarks and two leptons, and are organised based on the symmetry group of the *Weak Isospin*. This does not conserve parity symmetry, therefore, fermions in each generation occupy states of different chiralities, where left-handed states form a doublet and right-handed states a singlet.

The underlying symmetries imply conservation of the quantum numbers, or *charges*, related to them. Quantum numbers characterise the particles, together with their spin and mass. The three fermion generations have identical quantum numbers but different masses. The third component of the weak isospin I_3 is related to the two other quantum numbers of *hypercharge* Y and *electromagnetic charge* Q, deriving from the underlying symmetry groups $SU(2)_L \times U(1)_Y$, with the Gell-Mann Nishijima relation

$$Q = I_3 + Y/2. \tag{2.1}$$

The four fermions in each family are distinguished by their charges under strong and electromagnetic interactions. Quarks are the only fermions charged under the strong interaction. They are organised in triplets of the $SU(3)_C$ group carrying red, green or blue charge, while leptons live in colour neutral singlets. Left-handed quarks are organised in up-type, up u, charm c and top t with I_3 = +1/2, and down-type down d, strange s and bottom b with I_3 = -1/2. They carry electric charge +2/3 and -1/3

Table 2.1: Quantum numbers of leptons and quarks in the Standard Model. The indices L and R refer to the left-handed chirality of the doublets and the right-handed chirality of the singlets.

	Generation			Quantum Numbers		
	1st	2nd	3rd	I_3	Y	Q
Leptons	$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L$	$\begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L$	$\begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L$	+1/2 -1/2	-1 -1	0 -1
	e_R	μ_R	τ_R	0	-2	-1
Quarks	$\begin{pmatrix} u \\ d \end{pmatrix}_L$	$\begin{pmatrix} c \\ s \end{pmatrix}_L$	$\begin{pmatrix} t \\ b \end{pmatrix}_L$	+1/2 -1/2	1/3 1/3	2/3 -1/3
	u_R	c_R	t_R	0	4/3	2/3
	d_R	s_R	b_R	0	-2/3	-1/3

Table 2.2: Properties of gauge bosons in the Standard Model.

Boson	Interaction	Mass [GeV]	Symmetry gauge	Charge
γ	EM	0	$\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$	Q
$W^\pm$	weak	80.4 [12]		I_3
Z		91.2 [12]		
g	strong	0	$\mathrm{SU}(3)_C$	colour

respectively. Left-handed lepton doublets are composed by the electron e, the muon μ and the tau τ which carry electromagnetic charge -1 and occupy the down I_3 = -1/2 state. Up-type leptons are the neutrinos, which have neither electric nor colour charge, hence they participate in the SM interactions only in their left-handed component. The quantum numbers of quarks and leptons are listed in Table 2.1.

Forces between fermions are mediated by spin-1 particles, gauge bosons, which derive from the generators of the symmetry groups. The photon γ mediates the electromagnetic interaction, the eight gluons g the strong, while the Z and $W^\pm$ bosons the weak interaction. The Z and the charged W bosons correspond to a linear combination of the $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ group generators, and acquire mass thanks to the spontaneous breaking of the electroweak symmetry. This mechanism is related to the only scalar elementary particle in the SM known up to now, the Higgs boson. Gauge bosons properties are summarised in Table 2.2.

2.2 Symmetries of the Standard Model Lagrangian

The SM is a quantum field theory, i.e. the class of theories that emerged from the combination of special relativity and quantum mechanics. This means that quantum-mechanical single-particle wave functions are turned into fields intended as operators obeying canonical commutation relations. The SM makes use of the Lagrangian formalism, where monomials in the values of the fields and their derivatives are summed up to form a Lagrangian $\mathcal{L}$ that describes the full dynamics of the fields.

The electroweak sector The free propagation of a spin-1/2 particle field ψ is described by the Dirac Lagrangian[1]

$$\mathcal{L}_{\text{Dirac}} = \bar{\psi}\left(i\gamma^{\mu}\partial_{\mu} - m\right)\psi, \tag{2.2}$$

where γ^{μ} (μ= 0,...,3) are the four Dirac matrices and ψ is a four-component field. The Lagrangian is invariant under Lorentz transformations. Besides global transformations that relate physically inequivalent states, other symmetries are required, which relate equivalent field configurations whose redundancy should be factored out. These are gauge symmetries. The Lagrangian is invariant under global U(1) transformation, but in order to respect the invariance under local gauge transformation of the fields

$$\psi(x) \to e^{i\alpha(x)}\psi(x), \tag{2.3}$$

the derivative must be replaced with its *covariant* formulation

$$D_{\mu} \equiv \partial_{\mu} + ieA_{\mu}. \tag{2.4}$$

The newly introduced A_{μ} is a gauge vector field, that transforms as

$$A_{\mu} \to A_{\mu} - \frac{1}{e}\partial_{\mu}\alpha(x). \tag{2.5}$$

A kinetic term can be built in terms of gauge invariant quantities, such as the electromagnetic field tensor $F_{\mu\nu} = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu}$. The Lagrangian is hence written as

$$\mathcal{L}_{\text{QED}} = \bar{\psi}\left(i\partial\!\!\!/ - m\right)\psi - \frac{1}{4}\left(F_{\mu\nu}\right)^2 - \bar{\psi}e\gamma^{\mu}A_{\mu}\psi, \tag{2.6}$$

with $\partial\!\!\!/ = \partial_{\mu}\gamma^{\mu}$. The Eq. (2.6) is the Lagrangian of Quantum Electrodynamic (QED) that describes the interaction between the electron with electric charge e and the photon, i.e. the gauge vector boson A_{μ} and accounts for nearly all observed phenomena for macroscopic scales down to 10^{-13} cm [7]. The quantity $\bar{\psi}e\gamma^{\mu}\psi = J^{\mu}$ is the conserved Noether current associated to the U(1) symmetry and accounts for the interaction between fermions with electric charge e and the boson A_{μ}.

Gauge invariance is hence related to charge conservation, besides being the basis of the QED Lagrangian itself. Furthermore, it motivates the absence of the photon mass, since the $\mathcal{L}_{\text{QED}}$ invariance would fail by inserting a mass term for the boson.

The great success of QED as a gauge theory motivates the extension of the model to describe other interactions, like the weak force. This force was proved by experiment to maximally violate parity symmetry, such as the measurement of the asymmetry in the angular distribution of the decay products in β^{-} decays, from Chien-Shiung Wu [14]. Therefore the Dirac spinors need to be split into left- and right-handed components

$$\chi_L = \begin{pmatrix} \nu_{\text{L}} \\ \ell_L \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} u_{\text{L}} \\ d_{\text{L}} \end{pmatrix}, \quad \chi_R = \ell_{\text{R}} \quad \text{and} \quad u_{\text{R}}, d_{\text{R}} \tag{2.7}$$

[1] Natural units are used throughout this thesis, which implies $\hbar$ = 1, c = 1. The respective values in SI unites are $\hbar$ = 1.055·10^{-34} Js and c = 2.998·10^{8} m/s [13].

each one being an irreducible representation of the Lorentz group, and never interacting with each other. The Fermi four-fermion interaction theory that described the weak interaction by an effective coupling G_F and the product of two currents, included indeed left-handed fields only. The prove of the non-renormalisability of the Fermi theory pushed the theoretical effort towards a formulation of a gauge theory of weak interaction. Indeed, renormalisability is the real rationale behind the formulation of any theory that wants to provide predictions for physical observables.
The two currents can be written in terms of

$$J^{+}_{\mu} = 2\bar{\chi}_L\gamma_\mu T^a\chi_L \quad \text{and} \quad J^{-}_{\mu} = 2\bar{\chi}_L\gamma_\mu T^b\chi_L, \tag{2.8}$$

where $T^{a,b}$ are two matrices accounting for the interaction between the left-handed doublets. They are the two generators of a symmetry group forming a Lie algebra, obeying to the commutation relations $[T^a, T^b] = i\epsilon^{abc}T_c$, with ϵ^{abc} being the antisymmetric tensor. A third current related to J^c_μ can hence be interpreted as one of the three conserved currents of an SU(2) gauge symmetry, as T^a = 1/2τ^a, with τ^a, (a = 1,2,3) the Pauli matrices, generators of SU(2). It is a neutral weak current that was not accounted for in the Fermi theory. It can not be identified with the electromagnetic current, since this relates electrically charged fields and connects not only left-handed but also right-handed states.
Following the flow of QED, a model based on gauge symmetry groups can be built. In order to obtain a chiral theory, the gauge invariance for the fields is required under SU(2)$_L$ and U(1)$_Y$ transformations [15–17]

$$\chi_L \to e^{i\alpha(x)T_a + i\beta(x)Y}\chi_L \qquad \chi_R \to e^{i\beta(x)Y}\chi_R. \tag{2.9}$$

To maintain the invariance after the fields transformation, the derivative gets modified in the covariant formulation as

$$D_\mu = \partial_\mu - igW^a_\mu T^a - ig'\frac{Y}{2}B_\mu. \tag{2.10}$$

The terms W^a_μ, with a = (1,2,3) are the gauge vector fields associated to SU(2)$_L$ and B_μ the one associated to the U(1)$_Y$ group, where Y and g' are respectively generator and new coupling constant of the introduced hypercharge symmetry group. The vector A_μ from the electromagnetic interaction is an appropriate linear combination of the neutral W^3_μ and B_μ, obtained by fixing the values for Y so that the electron couples to the photon with the correct charge e, yielding to the Gell-Mann Nishijima relation of Eq. (2.1).
The Lagrangian can be written for the fields χ_L and χ_R

$$\mathcal{L}_{\text{EW}} = i\bar{\chi}_L\gamma^\mu D_\mu\chi_L + i\bar{\chi}_R\gamma^\mu D_\mu\chi_R - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}W^a_{\mu\nu}W^{\mu\nu}_a \tag{2.11}$$

includes the electromagnetic interaction in a unified description with the weak one, and completes formally the gauge boson-fermion interactions in the fermionic electroweak sector. The Lagrangian involving the $B_{\mu\nu}$ and $W^a_{\mu\nu}$ terms is the one of the gauge sector

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}W^a_{\mu\nu}W^{\mu\nu}_a, \tag{2.12}$$

where

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \quad \text{and} \quad W^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + g\epsilon^a{}_{bc} W^b_\mu W^c_\nu . \tag{2.13}$$

It generates triple and quartic bosonic couplings for W^a_μ. The self-interactions among gauge bosons of $\mathrm{SU}(2)_L$ are a direct consequence of the non-vanishing commutator between the generators of the group, which is defined non-Abelian. The interaction for the physic fields is obtained by defining

$$\begin{cases} W^\pm_\mu = \frac{1}{\sqrt{2}}\left(W^1_\mu \mp i W^2_\mu\right) \\ Z_\mu = \cos\theta_\mathrm{W} W^3_\mu - \sin\theta_\mathrm{W} B_\mu \\ A_\mu = \sin\theta_\mathrm{W} W^3_\mu + \cos\theta_\mathrm{W} B_\mu . \end{cases} \tag{2.14}$$

The last two equations are the rotations of the neutral fields and θ_W is the magnitude of the *weak angle* responsible for the rotation. It can be written in terms of the coupling constants as

$$\sin\theta_\mathrm{W} = \frac{g'}{\sqrt{g^2 + g'^2}} \qquad \cos\theta_\mathrm{W} = \frac{g}{\sqrt{g^2 + g'^2}},$$

and can be determined at leading-order by the masses of the W and Z bosons. The mass terms in the Lagrangian as it is defined in Eq. (2.11) should be set to zero to maintain gauge invariance, despite the experimental evidence of massive fermions and gauge bosons. The relation with θ_W and the mechanism to introduce the needed mass terms will be clarified in Section 2.3.

Quantum Chromodynamics In order to find a categorisation for the hadrons appearing at colliders, a model was proposed by Gell-Mann and Zweig that explained hadrons and mesons as composed by elementary components, the quarks. Their multiplets structure would reproduce the ones of hadrons, and they would be regulated by the SU(3) symmetry, the *flavour* symmetry. However, the wave function of baryons did not reflect the correct symmetry required to obey the Fermi-Dirac statistic, making the SU(3)-flavour description not satisfactory. The problem could be cured by an additional quantum number with a totally antisymmetric function. This finds experimental evidences in the high-energy annihilation of electrons and positrons, where the ratio of the production cross section of hadrons compared to muons agrees with the experiments only assuming the existence of three charges for each quark.

The most natural theoretical candidate is a non-Abelian gauge theory with gauge group SU(3) coupled to quarks in the fundamental representation, with related conserved charge named *colour* [18–20]. This derives from the usual requirement of gauge invariance of the Lagrangian for the quark fields, belonging to the fundamental representation of SU(3), under the transformation

$$q_i \to e^{i\alpha(x)\frac{\lambda_a}{2}} q_i . \tag{2.15}$$

The generators of the group are eight 3×3 traceless matrices λ^a, which do not commute with each other $\left[\frac{\lambda^a}{2}, \frac{\lambda^b}{2}\right] = f^{abc}\frac{\lambda^c}{2}$. The needed covariant derivative implies as usual new vector fields, G_μ.

They are eight gluons that transforms as

$$G^a_\mu \to G^a_\mu - \frac{1}{\alpha_s}\partial_\mu \alpha_a - f_{abc}\alpha_b G^c_\mu \qquad a = 1, \ldots, 8 \tag{2.16}$$

The Lagrangian has the form

$$\mathcal{L}_{\mathrm{QCD}} = -\frac{1}{4}F^a_{\mu\nu}F^{\mu\nu a} + \sum_f \bar{q}^i_f \left(i\gamma^\mu (D_\mu)_{ij} - m_f \delta_{ij}\right) q^j_f \tag{2.17}$$

where f is the flavour index, i, j the quark colour index and a the gluon colour index. The term

$$F^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + \alpha_s f^{abc} G^b_\mu G^c_\nu \tag{2.18}$$

contains the non-Abelian term $\alpha_s f^{abc} G^b_\mu G^c_\nu$ that makes the difference between QCD and QED and describes the self-interaction of three or four gluons, which is given by the fact that they have the colour charge and derives exclusively by the gauge invariance of the Lagrangian.

The Lagrangian includes kinetic terms for the gluons and the quarks, and terms for the quark-quark, gluon-quark and three- and four-gluon interactions.

Divergences that arise when considering higher order process, ultraviolet divergences, are reabsorbed by renormalisation methods. They imply a definition of an energy scale on which the parameters of the theory depend, the renormalisation scale μ_R, and an energy Q at which the *effective* coupling can be measured. In particular the behaviour of the coupling constant with the energy determines the conditions of validity of the perturbation theory. Due to the gluon self-interactions, the effective coupling of QCD $\alpha_s = \alpha_s(Q^2)$ increases at low energies, which implies that the perturbative regime does not hold for lower energies, and quarks and gluons are confined inside hadrons. The energy scale considered to separate the confinement state to the *asymptotic freedom* that characterises the perturbative regime is ~ 1 GeV, and the usual scale for α_s is considered the Z boson mass, for which $\alpha_s(m_Z)$ = 0.12 [12].

2.3 Spontaneous Breaking of the Electroweak Symmetry

The finite range of the weak force and experimental evidences required the need for terms in the Lagrangian accounting for masses of the up-to-now massless fields. However, the vector bosons $W^\pm$ and Z and the fermions can not acquire mass by adding the term to the electroweak Lagrangian of Eq. (2.11) without breaking the gauge invariance. For the same reason the mass terms for quarks in Eq. (2.17) have to be dropped in order to maintain the gauge invariance. The masses for bosons and fermions emerge by including the Brout-Englert-Higgs mechanism into the SM framework [21–24]. A potential is added, which maintains the invariance of the Lagrangian with respect to certain global symmetries, which are *broken* in the sense that the solutions to the equations, i.e. the vacuum states, are not symmetric anymore under those symmetries. For each global symmetry spontaneously broken the theory will contain a massless boson which will couple to the gauge bosons of the symmetry to provide them with an additional degree of freedom. The additional degree of freedom are the longitudinal polarisation states. Which and how many bosons will acquire mass depends on the expression of the potential and the representation of the gauge group based on which the scalar of the theory transforms. In the case of the non-Abelian electroweak model, three vector bosons have to acquire mass, while one has to remain massless, the photon, i.e. the electric charge has to leave the vacuum invariant. A potential is added

$$V(\Phi) = -\mu^2|\Phi|^2 + \lambda\left(|\Phi|^2\right)^2 \tag{2.19}$$

that presents two parameters μ and λ. The latter has to be $\lambda > 0$ to preserve the stability of the vacuum, while μ defines the shape of the potential. By choosing $\mu^2 < 0$ the potential has a set of degenerate minima as ground states, which spontaneously break the initial symmetry. The potential depends on a complex doublet

$$\Phi(x) = \begin{pmatrix} \Phi^+ \\ \Phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} \Phi_1 + i\Phi_2 \\ v + H + i\Phi_3 \end{pmatrix}. \tag{2.20}$$

By minimising the potential, the value of the field in the ground state is chosen as $|\Phi_0| = \sqrt{-\mu^2/2\lambda}$, with v vacuum expectation value that defines the set of minima of the potential. An expansion around the chosen ground state gives

$$\Phi = \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}, \tag{2.21}$$

where $H(x)$ is the physical field of the *Higgs boson*. The Lagrangian

$$\mathcal{L}_{\text{Higgs}} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + (D^\mu\Phi)^\dagger(D_\mu\Phi) - V(\Phi) \tag{2.22}$$

contains the couplings between Higgs and gauge bosons in its second term, while the potential presents triple and quartic self-couplings of the Higgs boson, and the Higgs boson mass term $m_H^2 = 2|\mu|^2 = 2\lambda v^2$. From the interaction term $(D^\mu\Phi)^\dagger(D_\mu\Phi)$ and using Eq. (2.14) the masses for the $W^\pm$ and Z bosons emerge

$$m_W^2 = \frac{1}{4}g^2v^2 \qquad m_Z^2 = \frac{1}{4}(g^2 + g'^2)v^2, \tag{2.23}$$

from which by exploiting the relation $G_\text{F}/\sqrt{2} = g^2/(2\sqrt{2}m_W)^2$ it is possible to obtain the vacuum expectation value $v \approx 246$ GeV. The relation between the θ_w and the masses of the bosons finally emerges from $m_W^2/m_Z^2 = g^2/(g^2 + g'^2) = \cos^2\theta_\text{w}$.

Fermion masses Mass terms for fermions would violate gauge invariance, therefore are generated as well through the coupling to the Higgs field. The interaction term which is added has the form of a *Yukawa coupling* $\mathcal{L}_{\text{Yuk}} = -\lambda_\ell\bar{\chi}_L\Phi\chi_R$ with Φ a spinor with vacuum expectation as Eq. (2.21) and $\bar{\chi}_L$ and χ_R the left-handed doublet and right-handed singlet from Eq. (2.7). This coupling term preserves charge, isospin, hypercharge, baryonic number and the right dimension of the Lagrangian, besides being Lorentz and Gauge invariant.

In case of charged leptons

$$\mathcal{L}_{\text{Yuk, lep}} = -\frac{\lambda_\ell v}{\sqrt{2}}\left(\bar{\ell}_L \ell_R + \bar{\ell}_R \ell_L\right) - \frac{\lambda_\ell}{\sqrt{2}}\left(\bar{\ell}_L \ell_R + \bar{\ell}_R \ell_L\right) H, \tag{2.24}$$

where the factors $-\frac{\lambda_\ell v}{\sqrt{2}} \equiv m_\ell$ are identified with the mass terms with ℓ = electron, muon, tau. The absence of the right-handed neutrino implies a vanishing mass term for the left-handed neutrino. Similarly, the Yukawa interaction can be written for quarks considering interaction eigenstates in left-handed doublet $Q_L'^f$ and right-handed singlets $u_R'^f$ and $d_R'^f$

$$\mathcal{L}_{\text{Yuk, q}} = -\frac{\lambda_d v}{\sqrt{2}} \bar{d}_L'^f d_R'^f - \frac{\lambda_u v}{\sqrt{2}} \bar{u}_L'^f u_R'^f + \text{h.c.}, \tag{2.25}$$

where f stands for a specific flavour. When more than one generation is introduced an additional coupling term can mix them. However, it is possible to diagonalise the Higgs couplings by choosing a new basis for the quark fields, where the two bases are related by the unitary transformation $u_L'^f = U_u^{fg} u_L^g$ and $d_L'^f = U_d^{fg} d_L^g$. The new basis is the mass basis, since it diagonalises the mass matrix, and u_L^f and d_L^f are the mass eigenstates. The terms in the Lagrangian which are sensitive to the mixing are the charged currents. The one for the W boson is

$$J_W^{\mu +} = \frac{1}{\sqrt{2}} \bar{u}_L^f \gamma^\mu \left(U_u^\dagger U_d\right)_{fg} d_L^g = \frac{1}{\sqrt{2}} \bar{u}_L^f \gamma^\mu V_{fg} d_L^g, \tag{2.26}$$

where V_{fg} is the unitary Cabibbo-Kobayashi-Maskawa matrix that allows weak-interaction transitions between the three quark generations. Its elements are fundamental parameters of the SM that need to be determined experimentally.

2.4 Vector Boson Scattering

The scattering of two gauge bosons radiated from initial state quarks is a key process for the understanding of electroweak symmetry breaking, which carries many theoretical and experimental challenges. In particular the scattering of charge-like W bosons offers a powerful tool, being highly sensitive to the mechanism of the breaking of the electroweak symmetry. In the following a phenomenological overview of the WW scattering process is offered, highlighting the importance of the same-charge W scattering case.

2.4.1 Scattering of *W* Bosons

The interactions among weak gauge bosons that appear in the SM Lagrangian of the gauge sector in Eq. (2.12) involve up to four bosons. They derive from the non-Abelian nature of the gauge theory, due to the weak charge that vector bosons carry.

In particular, the Lagrangian related to the quartic couplings predicts the existence of $WWWW$, $WWZZ$, $WWZ\gamma$ and $WW\gamma\gamma$ vertices.

Quartic couplings appear in hadron-initiated processes in the scattering of two vector bosons

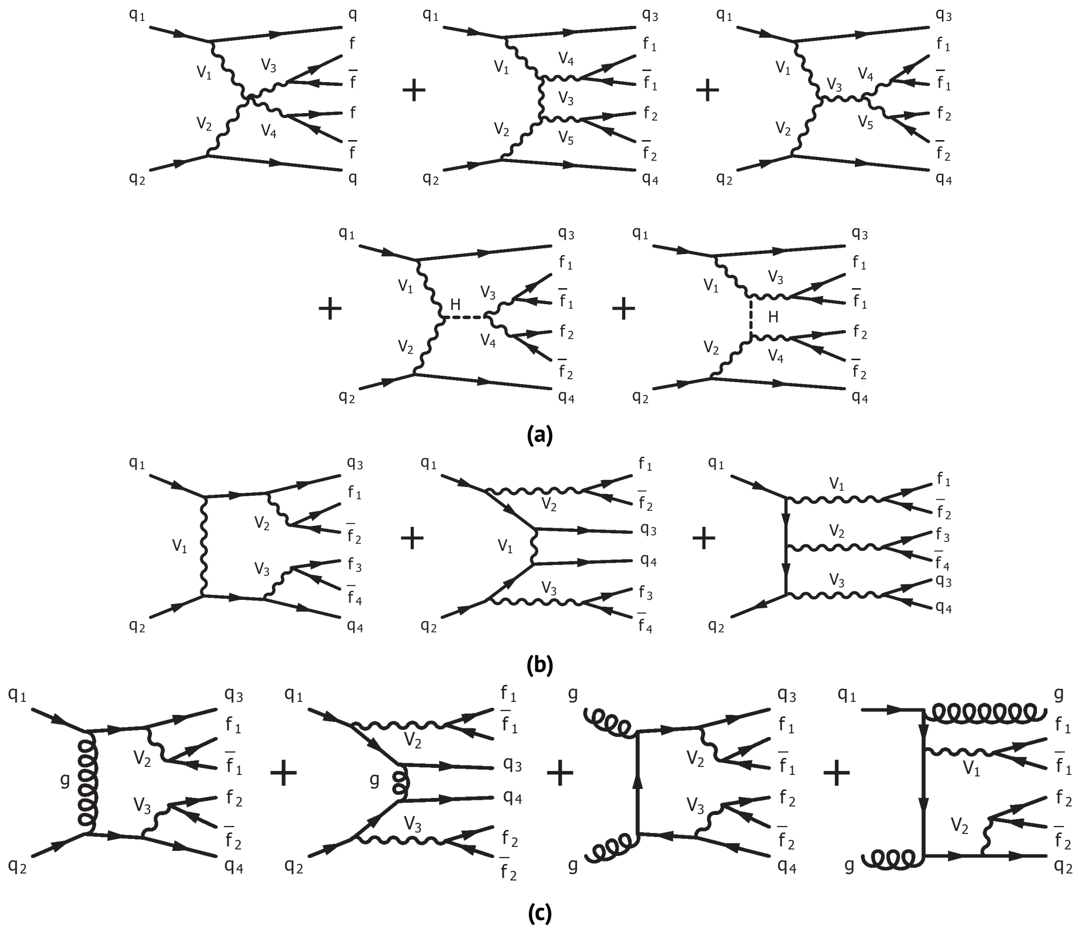


Figure 2.1: Feynman diagrams corresponding to the pure VBS $VVjj$ production processes (a), representative examples from the EW-induced production without the VBS topology (b) and the QCD-induced processes (c).

emitted by two initial state quarks, i.e. in VBS processes. A complete description of these processes is complex, due to irreducible contributions that can not be neglected to preserve gauge invariance. Besides the diagrams containing the quartic couplings, other two classes of contributions to the VBS definition need to be considered. As depicted in Fig. 2.1a, the s- and t-channel exchange of Z/γ^* are present, together with the s- and t-channel exchange of the Higgs boson. These contributions offer a fundamental way to access information on the electroweak symmetry breaking and on the role of the Higgs boson within the SM. The Higgs boson is in fact responsible for the unitarity preservation of the VBS cross section at all energies. Among the scattering processes of different gauge bosons, the one involving longitudinally polarised W boson is deeply connected to the mechanism of electroweak symmetry breaking. The polarisation of a longitudinally polarised massive W boson at rest is described by the polarisation vector $\epsilon_\mathrm{L} = (0, 0, 0, 1)$. Lorentz-transforming it into a generic reference frame leads to $\epsilon = 1/M\,(p_z, 0, 0, E)$, which depends on the energy. This becomes dangerous for the unitarity preservation in case of $E \gg M$. Indeed, the unitarity condition is equivalent to the requirement of having all probabilities of transition from an initial to a final state summing up to one [25], at all perburbative orders, and this translates into having the tree-level amplitude of any $2 \to 2$ process asymptotically flat with energy [26]. If only the diagram involving

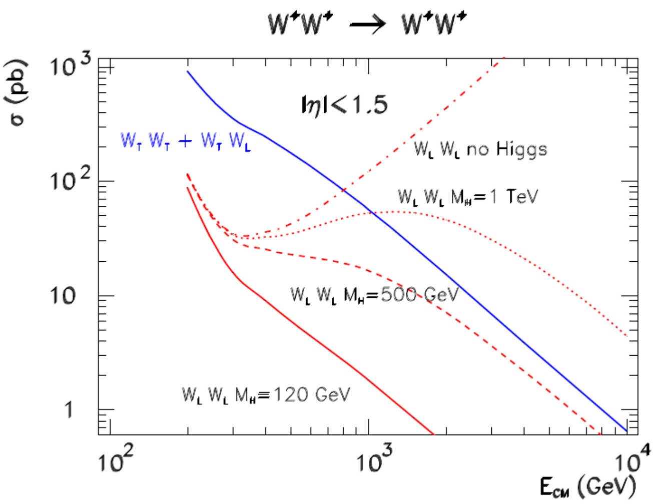


Figure 2.2: Total W^+W^+ scattering cross section as a function of the centre-of-mass energy for different polarisations and Higgs boson masses. Calculations are performed with MADRAPH5 and a selection on the scattering angle corresponding to η = ±1.5 with respect to the direction of the incoming W boson is applied [25].

the quartic coupling was present, the matrix element would behave as $\mathcal{M} \sim \epsilon_L\epsilon_L\epsilon_L\epsilon_L \sim s^2$. Adding the diagrams involving triple gauge couplings modifies the dependency on the energy, which becomes linear, despite the cross section still growing unphysically as $\mathcal{M}_{\text{gauge}} = -g^2 \frac{s}{4M_W^2} + \mathcal{O}(s^0)$. The second cancellation which suppresses the dependency on the energy is due to the Higgs boson. Higgs-mediated diagrams contribute with the term $\mathcal{M}_H = g_{HWW}^2 \frac{s}{M_W^4} + \mathcal{O}(s^0)$ that will cancel exactly the divergences if $g_{HWW} = gM_W$, i.e. if the Higgs boson couples to the W boson proportionally to M_W^2. This means that the presence of the Higgs boson cures the unitarity violation of the WW interaction which happens in the no-Higgs scenario, as shown in Fig. 2.2. Moreover, this implies that VBS could indicate new physics in case of larger cross sections, even if this was related to delayed unitarity cancellations, i.e. in case the theory completing the SM at higher energies would restore unitarity at experimentally unreachable scales.

Contributions and interference The same final state of two quarks and the decay products of the two bosons can be achieved at hadron colliders not only via the interaction of the two W bosons, but also via processes where the diagrams can not be interpreted as production times decay of the bosons [27]. The class of events where the two W bosons do not interact directly is not separable from the purely VBS diagrams. All contributing processes are addressed as electroweak processes, where the order of the electroweak coupling constant α_{EW} is six including all decay products' vertices. This class of processes includes the actual VBS mechanism of Fig. 2.1a, but it includes also s-channel contributions from non-resonant vector bosons or triple bosons production. Diagrams contributing to this class but not VBS, are shown in Fig. 2.1b. This class together with the VBS production is addressed as electroweak production.
A further contribution comes from another class of diagrams, referred to as QCD or strong, where the order of the strong coupling constant α_s is two and of α_{EW} is four. Examples of QCD diagrams are given in Fig. 2.1c.

The scattering of two same-charge W bosons, is the simplest VBS process to compute due to the low number of Feynman diagrams involved, since the s-channel diagrams are suppressed by the same-charge requirement. Moreover it has the largest electroweak-to-strong cross section ratio. The square amplitude of the $W^{\pm}W^{\pm}jj$ process can be expressed as

$$|\mathcal{M}|^2 = |\mathcal{M}_1 + \mathcal{M}_2|^2 = |\mathcal{M}_1|^2 + |\mathcal{M}_2|^2 + 2 \times \mathcal{R}e(\mathcal{M}_1^* \cdot \mathcal{M}_2)$$

where $\mathcal{M}_1$ and $\mathcal{M}_2$ are the amplitudes for the QCD and electroweak productions respectively, and an interference term is present, which is of $\mathcal{O}(\alpha_s, \alpha_{\text{EW}}^5)$. From a theoretical point of view the three contributions $\mathcal{O}(\alpha_{\text{EW}}^6)$, $\mathcal{O}(\alpha_s^2, \alpha_{\text{EW}}^4)$ and $\mathcal{O}(\alpha_s, \alpha_{\text{EW}}^5)$ are separable in a gauge invariant way, due to the different orders in the coupling constants [28], but an experimental measurement will include the interference part. Furthermore, it will not be able to identify the pure VBS contribution to the electroweak-induced set of diagrams. This can be enhanced experimentally by dedicated selections that exploit the different topology of the processes where the two W bosons interact via VBS.

Higher order corrections There are four contributions that form the next-to-leading order (NLO) corrections to the $pp \to \ell^{\pm}\nu\ell^{\pm}\nu jj$ process, $\mathcal{O}(\alpha_{\text{EW}}^7)$, $\mathcal{O}(\alpha_s, \alpha_{\text{EW}}^6)$, $\mathcal{O}(\alpha_s^2, \alpha_{\text{EW}}^5)$ and $\mathcal{O}(\alpha_s^3, \alpha_{\text{EW}}^4)$. The two orders $\mathcal{O}(\alpha_s, \alpha_{\text{EW}}^6)$ and $\mathcal{O}(\alpha_s^2, \alpha_{\text{EW}}^5)$ receive both QCD and EW corrections, with the consequence that at NLO it is not possible to distinguish EW from QCD production [29]. The electroweak corrections to the $pp \to \ell^+\nu\ell^+\nu jj$ process are found to be very large, reaching $\sim -16\%$ [30], because of virtual corrections that correspond to the insertion of massive vector particles in the scattering process.
The QCD corrections were calculated [31–34] in the VBS approximation [35, 36], which neglects certain contributions, such as the s-channel triboson production. Studies performed on physical observables such as m_{jj} or Δy_{jj} have shown [37] that these are good approximations at leading order (LO) in the kinematic region of VBS processes at the percent level, but can not be neglected at higher orders.
A complete calculation for the process $pp \to \mu^+\nu_\mu e^+\nu_e jj$ has been recently performed [28], which does not rely on the VBS approximation.

2.4.2 Effective Field Theory Parametrisation

The VBS processes are a probe for new physics, whose effects can be investigated in the high-energy behaviour of the scattering cross section, or in the bosonic quartic couplings. These can present modifications with respect to the SM expectations, that manifest themselves into values different from the SM predictions, and can be quantified with parameters that modify the couplings, i.e. the anomalous quartic gauge couplings (aQGCs). Another instrument used for the investigation of new physics is the gauge-invariant Effective Field Theory (EFT) for VBS [29]. In this case the SM is considered as a low-energy limit of an unknown UV-complete theory, with a well-separated energy scale Λ corresponding to the scale of new physics. In this approach, the amplitudes can be expanded in powers of E/Λ, where E is the scale at which processes are measured. The method is commonly used to investigate the Higgs boson couplings, for which a state-of-the-art review is offered in Ref. [38], and is a powerful tool for bosonic couplings as well, being a well-defined model-independent field theory. In this approach new gauge-invariant operators are ordered by their canonical dimension, so that the new Lagrangian is constructed as

$$\mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \frac{1}{\Lambda^2}\mathcal{L}_{\text{dim-6}} + \frac{1}{\Lambda^4}\mathcal{L}_{\text{dim-8}} + \mathcal{O}\left(\frac{1}{\Lambda^4}\right).$$

This can be restricted considering at first place the dimension-6 part, where

$$\mathcal{L}_{\text{dim-6}} = \sum_i C_i O_i, \tag{2.27}$$

with $\{O_i\}$ being a gauge-invariant set of operators that form a complete basis and C_i Wilson coefficients, that should be extracted simultaneously from a measurement.
In total, 59 terms plus their Hermitian conjugates form a complete basis for dimension-6 terms, and are associated to 2499 independent parameters [39]. Among them, VBS processes receive contributions from 14 operators, once constraints on CP conservation and flavour symmetry are imposed. Operators of dimension-8 can introduce decorrelation of effects between triple and quartic gauge couplings [40], but due to the lack of a complete basis up to now they are currently not considered in present interpretations.

3 Phenomenology and Simulation of Proton-Proton Collisions

At hadron colliders such as the Large Hadron Collider (LHC), in one inelastic highly-energetic hadron collision hundreds of particles are typically produced, spanning their momenta in a wide orders of magnitude range. This chapter is devoted to the phenomenological description of proton-proton scattering giving an overview of the processes involved, and of the tools available to compute or parametrise them. Starting from the master formula for a cross section computation, its main constituents are elaborated in Section 3.1. An overview of further physics phenomena happening during the scattering and their simulation techniques is presented in Section 3.2. The chapter mainly follows Refs. [12, 41, 42].

3.1 Cross Section Formulation

The SM described in Chapter 2 provides the theoretical framework to compute the probability of the transition from a certain initial state to a final one, i.e. the *cross section* of a process. For a generic process $ab \to n$ the total cross section, which is the result of the integration of the *differential cross section* $d\sigma/d\Omega$ over the full solid angle Ω, is expressed as

$$\sigma_{ab\to n} = \int \left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \mathrm{d}\Omega \right) = \int \frac{1}{F} |\mathcal{M}_{ab\to n}|^2 \mathrm{d}\Phi. \tag{3.1}$$

The *matrix element* $\mathcal{M}$ is the scattering amplitude derived from the Lagrangian density typically using perturbative expansion, and averaged over initial-state spin and color degrees of freedom. The term F is the particle flux, proportional to the square of the centre-of-mass energy, and the phase space factor $d\Phi$ is the element of the kinematic space which can be integrated over all possible particle momenta. In case of scattering of composite particles such as protons, the well-known perturbative approach can not be used to describe the initial state, because this is characterised by low-momentum QCD processes. The dynamics of the proton's components, quarks and gluons, generally referred to as *partons*, happens at length scales of $\mathcal{O}(1\ \mathrm{fm})$, while the one of the hard process occurs at length scales which can be three orders of magnitude lower [44, 45]. Nevertheless, the whole process can be factorised in soft and hard dynamics, according to the factorisation theorem [46], and the hadronic cross section is obtained by converting the partonic prediction into the hadronic one, so that the cross section of a two-proton scattering $p_\mathrm{A} p_\mathrm{B} \to X$ is expressed as

$$\sigma_{p_\mathrm{A} p_\mathrm{B}\to X} = \sum_{a,b} \int \mathrm{d}x_a \mathrm{d}x_b f_a(x_a, \mu_\mathrm{F}^2) f_b(x_b, \mu_\mathrm{F}^2) \cdot \hat{\sigma}_{ab\to X}(x_a x_b s, \mu_\mathrm{R}^2, \mu_\mathrm{F}^2). \tag{3.2}$$

Parton Distribution Functions The weighting functions in Eq. (3.2) are the parton distribution functions (PDFs) $f_{a/b}(x_{a/b}, \mu_\mathrm{F}^2)$, which represent the probability density to find a parton a/b inside the proton which carries a fraction $x_{a/b}$ of the proton's longitudinal momentum. Hence, x indicates

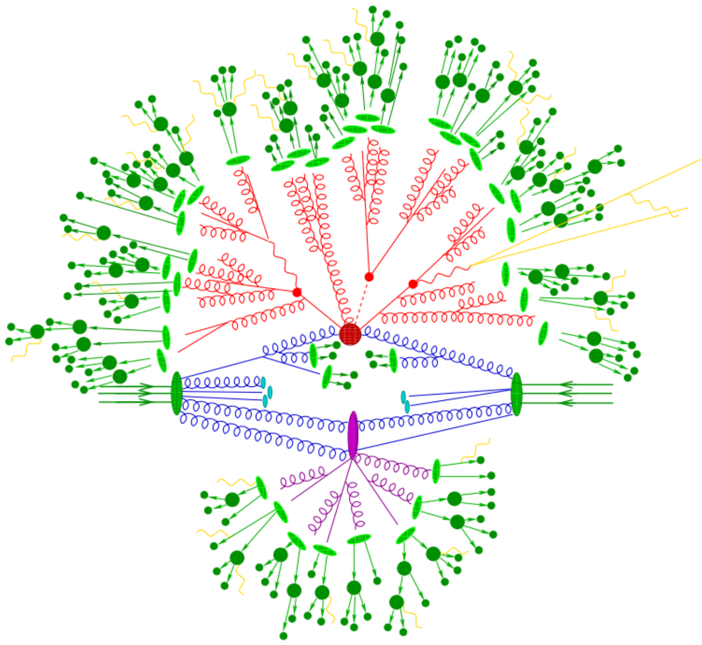
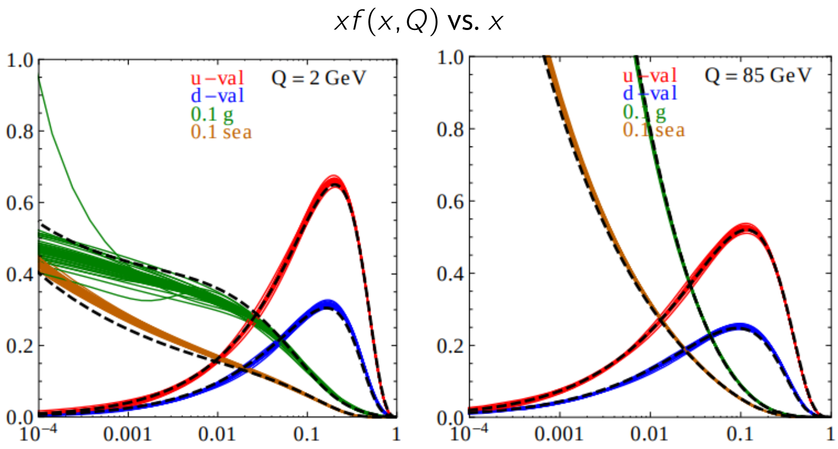


Figure 3.1: Schematic representation of hadronic collision. The red circle in the center is the hard scattering computed with the matrix element, where the additional radiated gluons and quarks are simulated by the parton shower. The purple circle is a secondary hard scattering of additional partons. The light green ovals are the hadrons, created during the hadronisation in simulation, while the dark green ones are their decays. The yellow lines represents the soft photons radiation [43].

the fraction of longitudinal momentum of the proton carried by the parton. The further dependency on the *factorisation scale* μ_F absorbs divergences from collinear emissions of initial state partons, and is the cut-off that separates the long- and short-distance physics.

The full x dependency of the PDFs can not be determined from perturbation theory, but is instead obtained from fits of theoretical predictions to experimental data, like deep inelastic electron-proton scattering measurements. PDFs can be parametrised with a generic functional form and measured on a particular process and dataset at a chosen energy scale μ_0, which is usually the momentum transferred in the hard interaction. The parametrisation will evolve for different values of μ_F. This evolution is described by the DGLAP equations [48–50], which allow to convert a PDF measured at a certain scale to the scale of interest. In Fig 3.2 the dependency on x for certain PDFs at two different energy scales is shown. There the different behaviour between *valence* and *sea* quarks is visible. The former determine the quantum numbers of the hadrons, and have higher probability to carry a large x value compared to the sea quarks that are the quark-antiquark pairs produced and annihilated as virtual particles by the strong interaction inside the hadrons. Depending on the choice of the dataset, the reference value for μ_F and the perturbative order, many strategies can exist to retrieve the PDF distributions [51, 52].

Perturbative sum of hard matrix element The cross section of the subprocess $ab \to X$ can be computed within the theoretical framework of the SM perturbatively. The validity of the perturbative approach is ensured by the principle of asymptotic freedom [50] (see Section 2.2). Therefore Eq. (3.2)



(a) **(b)**

Figure 3.2: Parton distribution functions from the CT10NNLO analysis shown for $\mu_F \equiv Q = 2$ GeV (a) and $\mu_F \equiv Q = 85$ GeV (b) as a function of x. The dashed curve is the central fit, and the colored bands show the uncertainty. The four curves shown are $xu_{\mathrm{valence}} = x(u - \bar{u})$, $xd_{\mathrm{valence}} = x(d - \bar{d})$, $0.10xg$ and $0.10xq_{\mathrm{sea}}$ [47].

can be written as

$$\begin{aligned}\sigma_{p_A p_B \to X} &= \sum_{a,b} \int \mathrm{d}x_a \mathrm{d}x_b f_a(x_a, \mu_\mathrm{F}^2) f_b(x_b, \mu_\mathrm{F}^2) \cdot \hat{\sigma}_{ab \to X}(x_a x_b s, \mu_\mathrm{R}^2, \mu_\mathrm{F}^2) = \\ &= \alpha_\mathrm{s}^k \sum_{n=0}^{\infty} \alpha_\mathrm{s}^n \sum_{a,b} \int \mathrm{d}x_a \mathrm{d}x_b f_a(x_a, \mu_\mathrm{F}^2) f_b(x_b, \mu_\mathrm{F}^2) \cdot c^{(n)}(x_a x_b s, \mu_\mathrm{R}^2, \mu_\mathrm{F}^2),\end{aligned} \tag{3.3}$$

where the partonic cross section $\hat{\sigma}_{ab \to X}$ has been rewritten as a perturbative sum over α_s of hard matrix elements, the coefficients $c^{(n)}$ carrying the dependency on the kinematics. The leading power k differs for different processes, while calculations are classified by the power of the coupling constant inside the sum. If only the first term of the expansion is included (n = 0) the calculation is said to be LO, while it is said to be NLO if terms are included up to n = 1, next-to-next-to-leading order (NNLO) for an expansion up to n = 2 and so on. The singularities that arise from virtual loop contributions and real emissions are absorbed by the coupling constant at each order causing the dependency of α_s on the renormalisation scale choice $\alpha_\mathrm{s} \equiv \alpha_\mathrm{s}(\mu_\mathrm{R})$ introduced in Chapter 2. From a formal point of view the cross section computed at all orders of the perturbation theory is independent from the choice of the renormalisation and factorisation scales (see Section 2.2). However the truncation of the series introduces a dependency on both μ_R and μ_F at each order. This dependency gives hints on the perturbative uncertainty, which is reduced going to higher orders. This is estimated by evaluating the effects of the variation of the scales.

3.2 Event Simulation

The kinematic information on the four-momentum of the stable particles in the final state is obtained with simulation tools, that take into account further physics processes that happen besides the

hard scattering, such as soft emission, hadronisation, further partons interactions, as well as the interaction of the particles with the detector which is usually performed separately with different tools. These tools are computer simulations that exploit Monte Carlo techniques (MC), able to implement the probabilistic description of particles generation and interaction.
The differences between the various generators on the market, such as PYTHIA [53], Herwig [54], SHERPA [55], lay in the diverse implementation of the modelling options, treatment of the soft emissions and its interface with the matrix element generation, and some other aspects that are briefly described in the following.

Soft emissions and hadronisation In increasing precision calculations, the sequence of emissions of quarks and gluons subsequent to the hard scattering need to be accounted for. However, in case of emissions which have small opening angles with respect to the emitting parton (collinear) and for low p_{T} ones (soft), shown in red in Fig. 3.1, logarithmic divergences appear. For sufficiently inclusive observables these divergences cancel against virtual emissions order by order in perturbation theory [56, 57], but for more exclusive quantities only partial cancellation occurs. In these case logarithmic terms appear, which can spoil the convergence of the series, and fixed-order calculations need to be substituted by resummed calculations. Depending on the order of the logarithm in the resummed terms these predictions are named leading-logarithmic (LL) or next-to-leading-logarithmic (NLL), and they gain in relevance as they affect small-p_{T} areas of observables, where most of the events are, or cases of vetoes on low-p_{T} jets. The use of resummation techniques is limited to high-accuracy computation of single observables. For the event generation specific programs called Parton Shower (PS) [43] are dedicated to the treatment of gluon emission and quark splitting. They are based on the sequential evolution from the hard scattering to lower and lower momentum scales up to the non-perturbative limit ($\sim$1 GeV). They make use of the probability for a parton to split, parametrised by a *splitting function* (DGLAP splitting), which depends on the fraction of momentum transferred, and the no-splitting probability given by the *Sudakov form factor* [58].
As the energy decreases with the PS evolution, the QCD coupling constant increases. At the point when perturbation theory breaks down *hadronisation* occurs, meaning that colorless bound states are formed, which will approximately carry the direction of the original parton. The simulation of these processes, shown in light-green in Fig. 3.1, is performed in different ways for different programs. PYTHIA is performing a *Lund string fragmentation* [59], while Herwig and SHERPA a *cluster fragmentation* [60]. The difference mainly lays in the modelling of baryon production, but all of them are approximations of the process with parameters to be determined using experimental data.
The final stage of the event generation is the simulation of the decay chain of unstable hadrons ($c\tau \leq 10$ mm) produced during the hadronisation, such as b- and c-hadrons or τ leptons, the dark green circles in Fig. 3.1. The parameters of the theoretical models are often tuned on data, and the framework commonly used in ATLAS are EVTGEN [61] and TAUOLA [62] for τ lepton decays.

Underlying event The parton-parton interaction constitutes the hard scattering and produces the majority of the outgoing particles, but other interactions can take place changing the topology of the colliding system and consequently the particle multiplicity in the final state. The interactions of

the partons not involved in the hard scattering or produced after it are referred to as underlying events, shown in purple in Fig. 3.1. Together with this, the beam remnants left after the protons interaction can contribute to the observed activity. All these processes need to be considered in the simulation procedures, which is done with a phenomenological approach, meaning that the predictions from simulation are tuned with experimental data.
Beside these events, also pile-up need to be accounted for in simulation. It refers to additional interactions happening in the bunches, which are the compressed collection of protons made to collide. Such additional interactions can derive both from protons from the same bunch of the hard-scattering process, in-time pile-up, and from protons in other bunches, out-of-time pile-up. To account for pile-up, the signal expected from the hard interaction is superimposed with an inclusive simulated sample of proton-proton collision. The average number of interactions per bunch crossing $\langle\mu\rangle$ is reweighed and translated into a modified weight for the simulation, in the so-called *pile-up reweighting* procedure. This is only needed to adjust the simulation to the pile-up profile of the data.

Detector interaction In order to rely on simulated events and be able to compare them to data in physics analyses, the detector response is to be accounted for. To do so the interaction of the created particles with the detector material is simulated as well [63]. This is done with a special toolkit called GEANT4 [64], which in ATLAS is integrated into the Athena software framework [65]. This simulates the interaction of the particles created in a collision with the detector material, making it possible to infer the detected signals. Simulated events and real data are finally reconstructed with the same reconstruction framework, that will be the object Chapter 5, and elaborates the input information into the physics objects used in the analyses.

4 The Large Hadron Collider and the ATLAS Experiment

The experimental facilities described here belong to the CERN (*Conseil Européen pour la Recherche Nucléaire*) complex, located on the Franco-Swiss border near Geneva. CERN hosts the proton-proton LHC [66] and its detectors.

There are seven experiments located at the LHC ring. The ATLAS (A Toroidal LHC ApparatuS) [1] and CMS (Compact Muon Solenoid) [67] experiments cover a wide range of topics from precision measurements of the SM to searches for new resonances or phenomena. The ALICE (A Large Ion Collider Experiment) experiment [68] collects mainly data from heavy-ion collisions to investigate strongly interacting matter at high energy density, while LHCb (Large Hadron Collider beauty) [69] focuses on measuring properties of hadrons that contain a b quark. The three experiments MoEDAL, TOTEM and LHCf [70–72] perform more specific research, from magnetic monopoles to cross section measurements or calibration studies. An overview of the LHC is given in Section 4.1, while Section 4.2 describes the ATLAS detector and its main components.

4.1 The Large Hadron Collider

The LHC is a hadron accelerator and particle-particle collider that is situated in a nearly circular 26.7 km long underground tunnel originally built for the Large Electron-Positron Collider (LEP) [73]. It operates with proton or heavy ion lead beams counter-rotating in two separate vacuum pipes. These are made to collide in four interaction points, where four detectors are located: the multi-purpose ATLAS and CMS detectors, as well as the ALICE and LHCb ones. The four sections of the collider where these experiments sit, together with other sections dedicated to the injection, extraction, acceleration and cleaning of the beams, are straight segments, while approximately 22 km of the total length consists of curved sections. The LHC was built with the aim of searching for the Higgs boson, revealing physics beyond the SM and precisely measuring parameters in unexplored phase spaces [75]. It was designed to reach centre-of-mass energies of up to 14 TeV for protons and 5.4 TeV for ions per nucleon. This is made possible by the use of hadrons, which lose only a small fraction of their energy via synchrotron radiation due to their larger mass, compared to electrons or positrons, previously used by LEP.

To reach the nominal beam energy, before being injected in the LHC, bunches of protons are accelerated in several steps through interconnected linear and circular accelerators, as shown in Fig. 4.1. Protons are extracted from hydrogen gas by ionisation, and are transferred to the first accelerator, the linear LINAC2, where they are brought up to 50 MeV. After this, protons are injected into the Proton Synchrotron Booster which accelerates them to an energy of 1.4 GeV, and subsequently into the Proton Synchrotron. Here a further boost is applied to increase the beam energy to 25 GeV. The next step is the Super Proton Synchrotron, a 6.9 km circular accelerator which brings the protons' energy to 450 GeV before entering the LHC ring, where they are accelerated to the final beam energy.

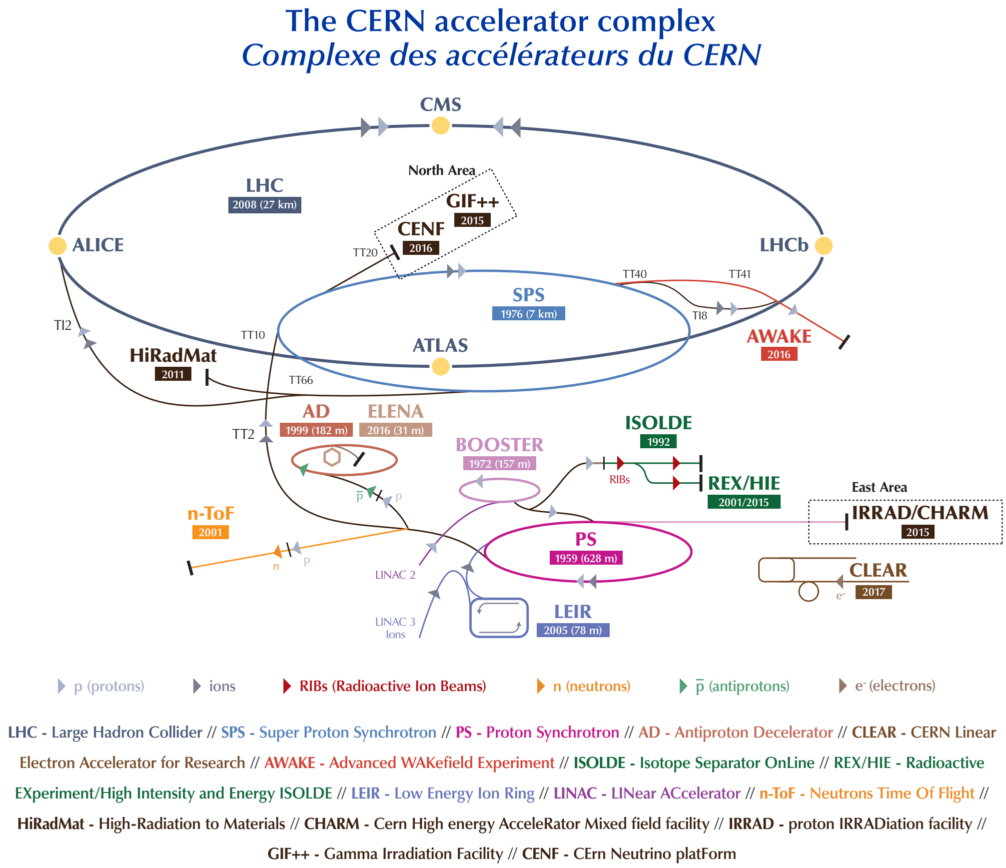
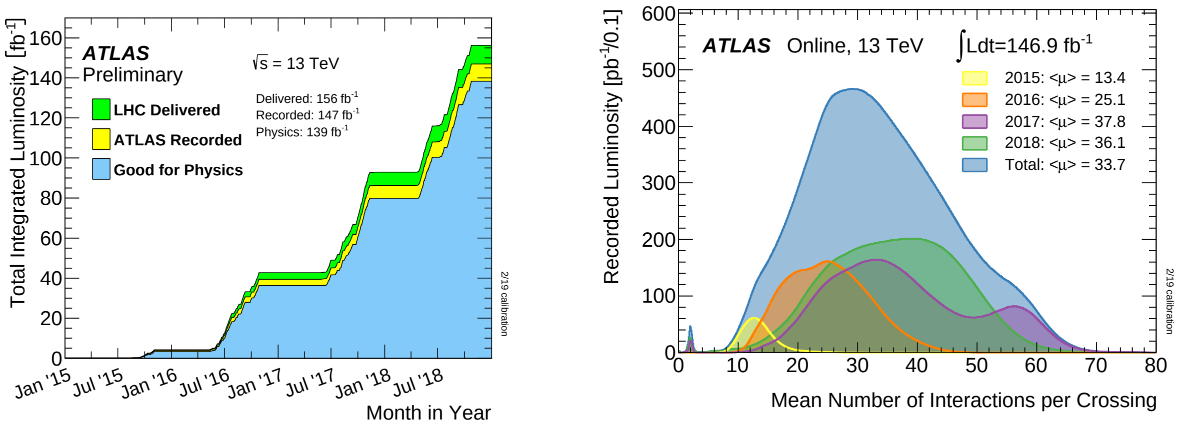


Figure 4.1: The CERN accelerator infrastructure [74].

Despite being originally designed to operate at a maximum energy of 7 TeV per beam, the standard proton-proton operating energy throughout Run 2 was 6.5 TeV, for a centre-of-mass energy of $\sqrt{s}$ = 13 TeV. To accelerate the beams within the LHC, eight 400 MHz radio frequency (RF) cavities per beam are placed in cryomodules to keep them working in a superconducting state.

To keep such energetic proton beams in the desired orbit, superconducting magnets are needed. The bending of the beams is obtained by the use of 1232 15 m-long niobium-titanium (NbTi) dipole magnets providing a magnetic field of up to 8.3 T. To reach this value the magnets need to be operated in a superconductivity state, which is provided with liquid helium via a cooling system, keeping an operational temperature of 1.9 K. The dipoles are equipped with sextupole, octupole and decapole magnets, which correct for small imperfections in the magnetic field at their extremities. To focus the protons in the bunches into a small transverse size, 392 quadrupole magnets are used. They have alternating polarity, and the four magnetic poles are arranged symmetrically around the beam pipe to squeeze the beam either vertically or horizontally, important especially at the interaction point to increase the probability of collisions.

Beside the energy, the other key feature that characterises the performance of a collider is related to the rate of collisions at the interaction point. This is parametrised by the *instantaneous luminosity* L, measured in particles per area per time, that directly relates the event rate of a physics process



(a) **(b)**

Figure 4.2: (a) Cumulative integrated luminosity versus time delivered to ATLAS (green), recorded by ATLAS (yellow) and certified to be of good quality data (blue) in 2015-2018 during stable beam proton-proton collisions at $\sqrt{s}$ = 13 TeV [76]. (b) Luminosity-weighted distribution of the mean number of interactions per crossing for proton-proton collisions at $\sqrt{s}$ = 13 TeV for the years 2015-2018 [76].

to its cross section

$$\frac{dN_{pp\to X}(t)}{dt} = L(t)\cdot\sigma_{pp\to X},$$

where $\sigma_{pp\to X}$ is the cross section for the $pp \to X$ process. The luminosity depends on the beam parameters. For two Gaussian distributed proton bunches of transverse widths σ_x and σ_y, containing $N_{p,1}$ and $N_{p,2}$ protons each, and colliding at frequency f the luminosity can be parametrised according to

$$L = f\cdot\frac{N_{p,1}\cdot N_{p,2}}{4\pi\sigma_x\cdot\sigma_y}.$$

This can be further expressed in terms of the normalised transverse emittance ϵ_n and the amplitude function at the interaction point β^* as

$$L = \frac{f\cdot N_{p,1}\cdot N_{p,2}\cdot\gamma}{4\pi\cdot\epsilon_n\cdot\beta^*}F.$$

The term γ is the relativistic gamma factor and F accounts for a geometrical correction due to the crossing angle under which the beams are brought to collision. Given the nominal values of the parameters, such as a frequency of 11245 Hz, the number of particles per bunch of approximately 10^{11} protons/bunch and $\sigma_x \approx \sigma_y \approx 16\cdot10^{-4}$ cm, the design value for the LHC is an instantaneous luminosity of $L = 10^{34}$ cm^{-2}s^{-1}.

The *integrated luminosity* is obtained by integrating the luminosity over a defined time period. For the full Run 2 period from 2015 to 2018, 158 fb^{-1} have been delivered by the LHC, of which 147 fb^{-1} have been recorded by ATLAS. Of these, 139 fb^{-1} have been considered to be of good quality, i.e. the reconstructed physics objects reflect specific quality criteria, and have been used for physics analyses. The total integrated luminosity and data quality are shown in Fig 4.2a. To determine the luminosity, counting rates are measured by specific luminosity detectors (see 4.2.5) and the van-der-Meer beam-separation method [77] is used to measure the beam profile to determine the

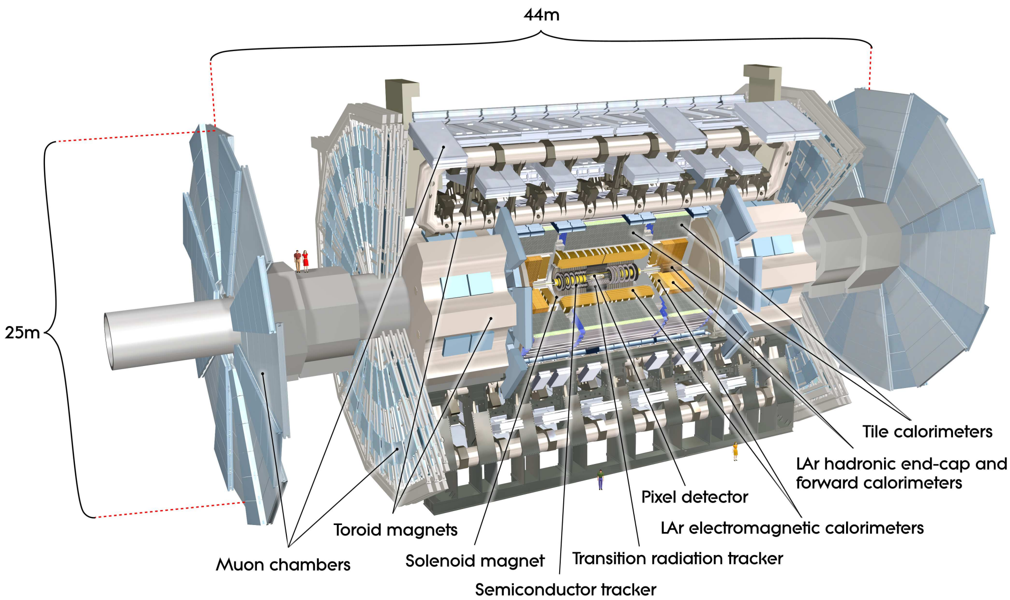


Figure 4.3: Cut-away view of the ATLAS detector [1].

absolute luminosity calibration.
The additional interactions of protons from the same bunch or from other bunch-crossings, the pile-up, increase strongly with higher instantaneous luminosity values. It is typically characterised by the mean number of interactions per bunch crossing, which is depicted in Fig. 4.2b for the years 2015-2018.

4.2 The ATLAS Detector

The geometry of the ATLAS detector is cylindrical, with a forward-backward symmetry with respect to the interaction point. The central region of the detector is usually addressed as barrel, the forward regions as endcaps, and a schematic view is presented in Fig. 4.3. It covers a solid angle of almost 4π and the magnets structure includes a thin superconducting solenoid around the Inner Detector (ID) and three large superconducting toroids, one for the barrel and two for the endcaps. The ID, immersed in the magnetic field of $B = 2$ T, was designed to track charged particles and measure their kinematic properties via their curvature. Electromagnetic (EM) and hadronic calorimeters measure the energy and direction of electromagnetically and hadronically interacting particles and are surrounded by a muon spectrometer (MS). The bending planes in the ID and in the MS are perpendicular to one another. The coordinate system used by the detector is described in the following sections, together with its sub-detectors.

4.2.1 Coordinate System

The right-handed coordinate system used by ATLAS is centred on the nominal interaction point, the x-axis points to the center of LHC ring, the y-axis points towards the surface and the z-axis coincides approximately with the beam axis. A set of spherical coordinates (r, ϕ, θ) is also used, where the azimuthal angle ϕ ranges from $-\pi$ to π and the polar angle θ is measured from the positive z-axis. To describe the polar position of particles, the *rapidity* is defined

$$y = \frac{1}{2}\ln\left[\frac{E+p_z}{E-p_z}\right],$$

where E is the total energy and p_z is the z-component of the momentum. This is conveniently used since differences in rapidity are Lorentz invariant under boosts along the beam axis, and the rest frames of the parton-parton collisions have different boosts with respect to the laboratory frame due to the different fractions of proton momentum carried by the partons. For massless particles or for negligible masses in the highly relativistic case, the rapidity is approximated by the *pseudorapity*

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

The particle production is approximately constant per (pseudo)rapidity unit, which is one additional reason behind the choice of these two variables to describe the polar position of particles.
Distances between particles are usually measured in the (ϕ, η) parameter space and are usually expressed in terms of

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}.$$

The transverse component of the momentum p_{T} and energy E_{T} are commonly used, and are defined in the x-y plane as

$$p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2}$$

and

$$E_{\mathrm{T}} = \sqrt{p_{\mathrm{T}}^2 + m^2}.$$

4.2.2 Inner Detector

The ATLAS ID [78] provides information for the pattern recognition of charged particles above a certain p_{T} threshold (nominally 0.5 GeV) within a range of $|\eta| < 2.5$ and full coverage in the ϕ angle. It is designed for an excellent momentum resolution and aims at the measurement of both primary and secondary vertices. It is contained in a magnetic field of $B = 2$ T provided by the superconducting solenoid. The field is parallel to the beam axis and bends charged particles in the ϕ direction. The magnetic field is constant in the radial direction while it decreases along z outside the solenoid region, going away from the interaction point. The ID consists of three subdetectors: the pixel detector, the Semi-Conductor Tracker (SCT) and the Transition Radiation Tracker (TRT). The layout of the ID system is shown in Fig. 4.4.

Pixel Detector

The semiconductor pixel detector [80] is the innermost layer of the ID that provides very high-granularity and high-precision measurements as close as possible to the interaction point and in the full range of $|\eta| < 2.5$. Ranging up to 150 mm in the radial direction, it was originally located at

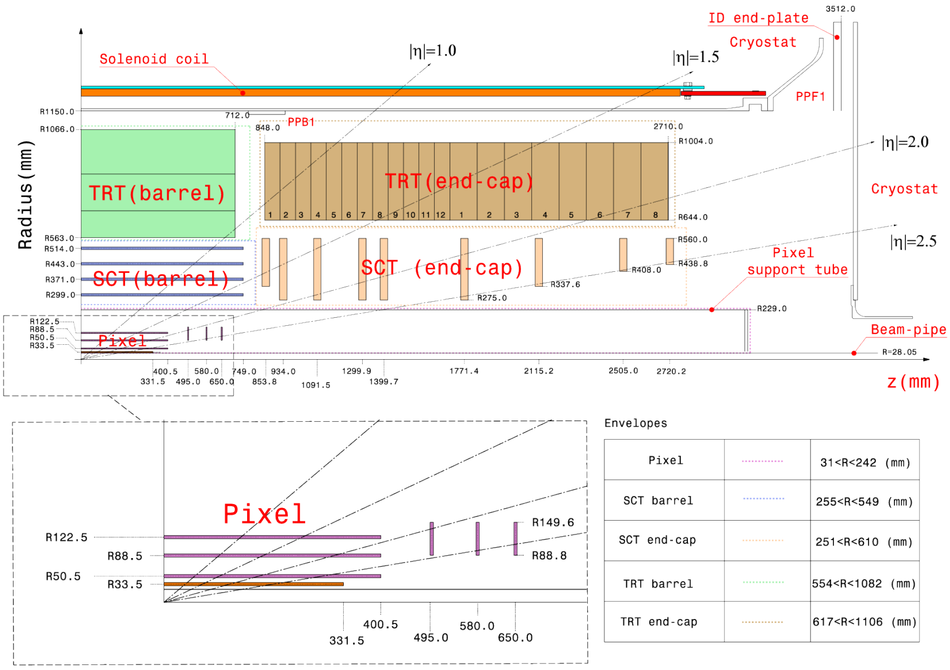


Figure 4.4: The ATLAS inner detector in R-z view. The whole inner detector is shown in the top panel, while a zoom of the pixel detector region is given in the bottom pannel [79].

50.5 mm to the beam axis. This distance was reduced with the insertion of a new pixel layer at 33.25 mm, the insertable B-layer (IBL) [81], which improved the spatial resolution and enhanced the tracking accuracy [82]. The pixel detector consists of a central region where four layers of silicon pixel modules are mounted around the beam axis, and two endcap regions with three disks each perpendicular to the beam axis, for a total of 1744 modules. In the barrel pixels have a size of 50 μm in the R-ϕ plane and 400 μm along z. Typically the pixel detector provides four measurement points, with an intrinsic resolution of 10 μm in azimuthal direction and 115 μm in axial (radial) direction of the barrel (endcap). The pixel detector determines the ability of the ID to reconstruct the interaction point with high resolution as well as vertices of particles' decays.

Semiconductor Tracker

The SCT surrounds the pixel detector, it consists of strips and has coarser granularity, as the track density decreases at larger radii. It is mounted in a distance of $R = 299$ mm to 514 mm from the beam axis, covering a region up to $|\eta| < 2.5$ and consists of four cylindrical barrel layers and nine planar endcap disks on each side. Typically four precision measurements per charged particle are detected by silicon microstrip sensors in the barrel. The barrel modules are two-sided, with two adjacent sensors mounted on each side. Both in the barrel and in the endcap the two modules are tilted with a stereo angle of $\sim$40 mrad with respect to each other to break the degeneracy along the z-direction: this ensures a good position resolution for charged particles. The intrinsic accuracy per module is 17 μm in R-ϕ and 580 μm in z, and it is possible to distinguish charged particles separated by at least $\sim$200 μm. The SCT contributes significantly to the measurement of momenta, impact parameters and vertex positions.

Transition Radiation Tracker

The outermost part of the ID is the TRT, which provides coverage up to $|\eta| = 2.0$ in the remaining region that goes up to 1082 mm from the beam axis. It consists of 4 mm-diameter straw tubes interleaved with polypropylene as radiator material, which are placed parallel to the beam axis in the barrel region, and are radial in the endcap regions. The tubes are filled with a Xe-based gas mixture, which provides electron identification capability through the detection of photons emitted by transition radiation created by the radiator material in between the straws. Each electronic channel provides a drift-time measurement giving R-ϕ information, and particle identification is also provided, as transition radiation is more likely for electrons than other charged particles, which can therefore be discriminated via different energy thresholds. The single hit resolution is $\sim$120 μm in the barrel and 130 μm in the endcaps, which is lower compared to the silicon detectors. However, for each track typically more than 30 hits are recorded, which significantly improves the spatial resolution of the track reconstruction if the information is combined with the other detectors.

Material distribution

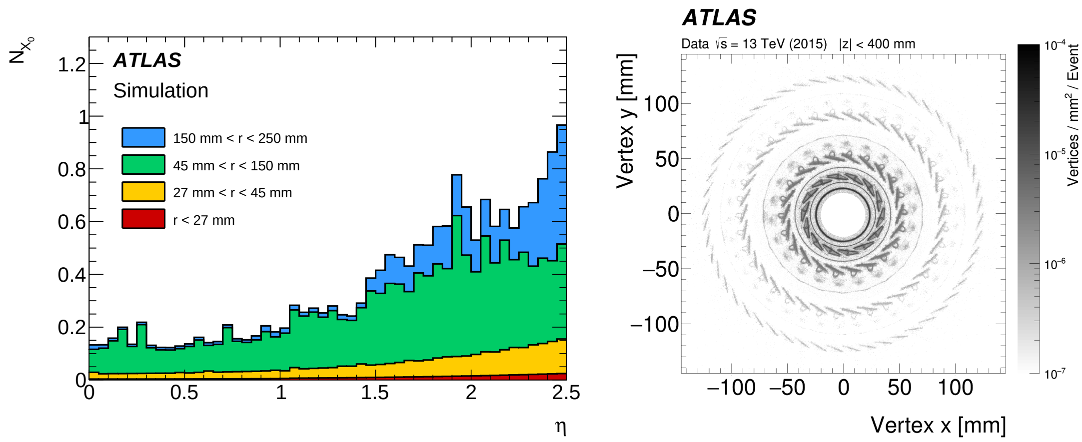
The knowledge of the amount and type of material inside the ID is directly linked to the performance in the track and object reconstruction and calibration. It is measured indirectly via the reconstruction of both photon conversion vertices and vertices from hadronic interactions [79]. The former are sensitive to the electromagnetic interactions of particles inside the material, such as the electron-positron production of photons. This method takes advantage of the excellent theoretical understanding of electromagnetic interactions, but the radial resolution is limited by the collinearity of the electron-positron pair. The reconstruction of hadronic vertices is sensitive to nuclear interactions. It suffers from greater difficulties in the modelling, but gains in resolution thanks to the large opening angles between the daughter particles. A complementary third approach is to measure nuclear interaction rates of charged hadrons through hadronic interaction[1]. The amount of material is usually evaluated in terms of the interaction length λ_0 and the radiation length X_0, which are the mean path length after which the flux of relativistic primary hadrons is reduced to a fraction 1/e, and the distance after which the energy of a high-energy electron ($E \gg 2m_e$) is reduced to $1/e \cdot E$.

Fig. 4.5a presents the amount of material in units of X_0 as a function of $|\eta|$, separately for four regions in the radial direction. In Fig. 4.5b the distribution of vertices from hadronic interactions is shown for the pixel detector in the transverse plane.

4.2.3 Calorimeter System

The ATLAS calorimeters provide accurate measurements of energy and position of electrons, photons and hadrons, as well as information on their discrimination. This is obtained by the characterisation

[1] This is done by considering the collection of *tracklets*, i.e. tracks with hits in the pixel detector only, and measuring the fraction of them which can be matched to combined tracks. A charged hadron that undergoes hadronic interaction in the material between the pixel detector and the SCT typically leaves hits in the pixel only, hence this efficiency is related to the amount of material crossed by the particle.



(a) **(b)**

Figure 4.5: (a) Number of radiation lengths as a function of η in its positive range indicating the amount of material associated to electromagnetic interactions. The regions are categorised for different radial distances, which correspond to the beam pipe (red), the IBL (yellow), the pixel barrel (green) and the pixel service regions (blue) [79]. (b) Distribution of hadronic-interaction vertices for data in a x-y view for the pixel detector [79].

of the properties of the electromagnetic and hadronic showers induced by the particles respectively in the electromagnetic and hadronic calorimeters. In the η region matched to the inner detector, the high granularity is suited for electron and photon precision measurements, while the rest of the system has coarser granularity which is sufficient for jet reconstruction and measurements of missing energy from weakly interacting particles. The calorimeter system covers the range $|\eta| < 4.9$. A liquid Argon (LAr) technique [83] is used for all electromagnetic (EM) calorimetry covering the interval $|\eta| < 3.2$ and for hadronic calorimetry from $|\eta| = 1.4$ up to the acceptance limit. At large radii radiation levels are lower, therefore a less expensive iron-scintillator hadronic tile calorimeter is used [84]. The total thickness in terms of interaction lengths of the full calorimeter system is 11 λ_0 at $\eta = 0$, which contains the hadronic showers. An overview of the calorimeter system is shown in Fig. 4.6.

Electromagnetic calorimeter

The EM calorimeter is a sampling calorimeter with active layers of liquid Argon alternated with absorber lead plates. The electrodes to measure the energy deposited in the liquid argon and the lead absorbers are built with an accordion-shaped geometry to provide complete ϕ symmetry without azimuthal cracks. The barrel part of the EM calorimeter covers $|\eta| < 1.475$, while the two endcaps components cover the range $1.375 < |\eta| < 3.2$.
The barrel has a thickness of 22 radiation lengths and consists of 16 modules. In Fig. 4.7a a single module of the barrel is shown. The barrel is longitudinally divided into three layers that have increasing granularity to obtain an elaborated analysis of the shower shape. The first layer is finely segmented into strips, having a granularity of $\Delta\eta \times \Delta\phi = 0.0031 \times 0.098$. This allows for the distinction of close-by particles, e.g. two photons from a $\pi_0 \to \gamma\gamma$ decay. The second

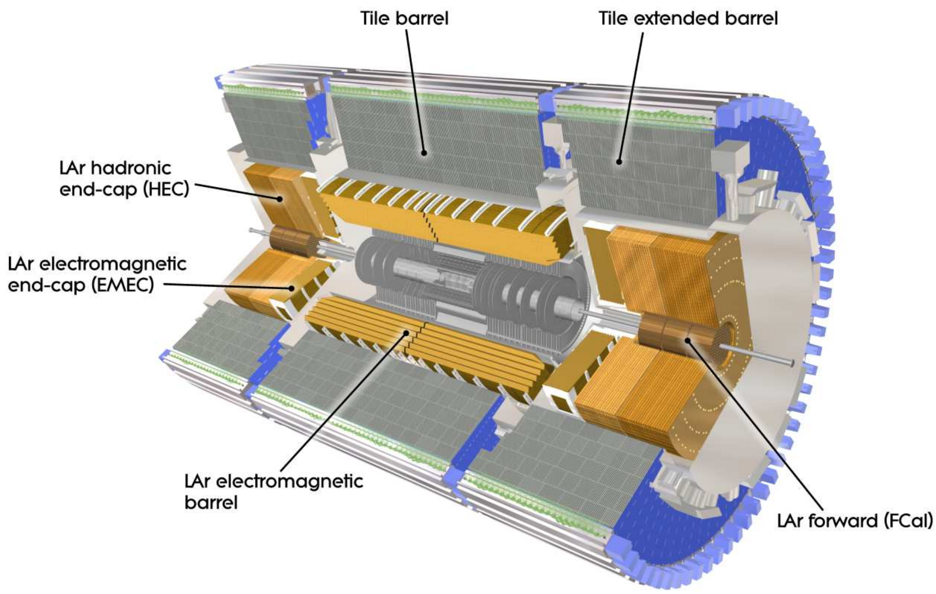


Figure 4.6: Cut-away view of the ATLAS calorimeter system [1].

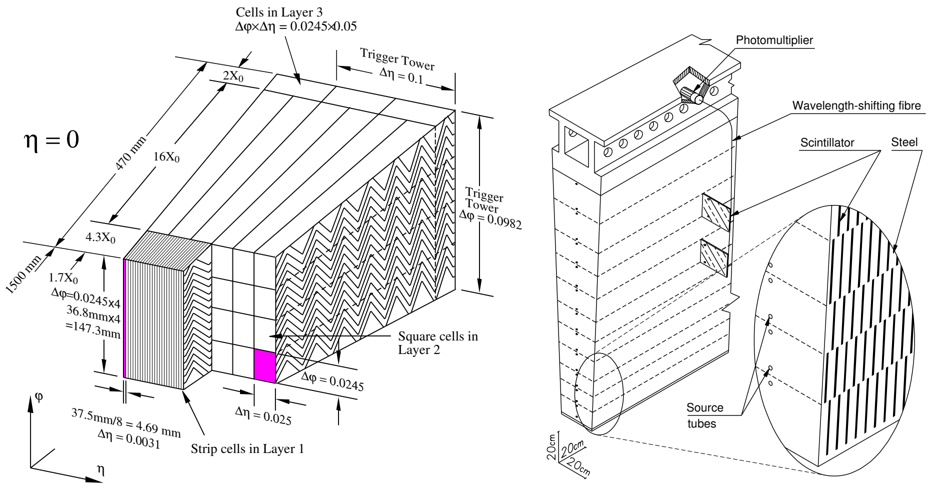
layer aims at energy measurement and has a coarser granularity of 0.025×0.025 in $\eta \times \phi$. The granularity of this layer defines the azimuthal size of the so called *towers*, $N_\eta \times N_\phi$ = 200×256 elements spanning longitudinally the three layers of the EM calorimeter. Combined clusters of 4×4 $\Delta\eta \times \Delta\phi$ = 0.025×0.025 elements are the so called *trigger towers*, which provide inputs to the first level trigger for a fast decision. The third layer of the EM calorimeter has twice coarser granularity in η and the same ϕ segmentation and is intended to correct for the overlap of the energy deposition in the following hadronic calorimeter. In front of the EM barrel calorimeter a presampler layer is placed. It is a 11 mm thick LAr layer which provides shower sampling and corrects for energy loss in front of the electromagnetic calorimeter.

Each EM calorimeter endcap consists of two coaxial wheels, with a $\sim$3 mm spacing between them at $|\eta|$ = 2.5. The inner part is divided into three layers, while the outer one has two, with a coarser transverse granularity. The presampler is present in the endcaps as well, with a lower thickness of 5 mm up to $|\eta| < 1.8$.

The resolution in energy that can be achieved by the EM calorimeter is parametrised as

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c, \tag{4.1}$$

where a is the sampling term, b the noise term and c the constant term, all depending on η. The term a mainly contributes at low energy and gets worse with increasing η starting from values around $10\%/\sqrt{E[\text{GeV}]}$, due to the increase in the amount of material in front of the calorimeter. The noise term is dominated by pile-up noise at high $|\eta|$ and has a target value of around 170 MeV. For increasing energies the relative energy resolution tends asymptotically to the constant term c, originated from the calibration of the calorimeter, which has a design value of 0.7% [83].



(a) **(b)**

Figure 4.7: (a) Barrel module of the electromagnetic calorimeter where the different layers and granularities are visible [1]. (b) Tile hadronic calorimeter module, where the mechanical assembly and the optical readout are depicted [1].

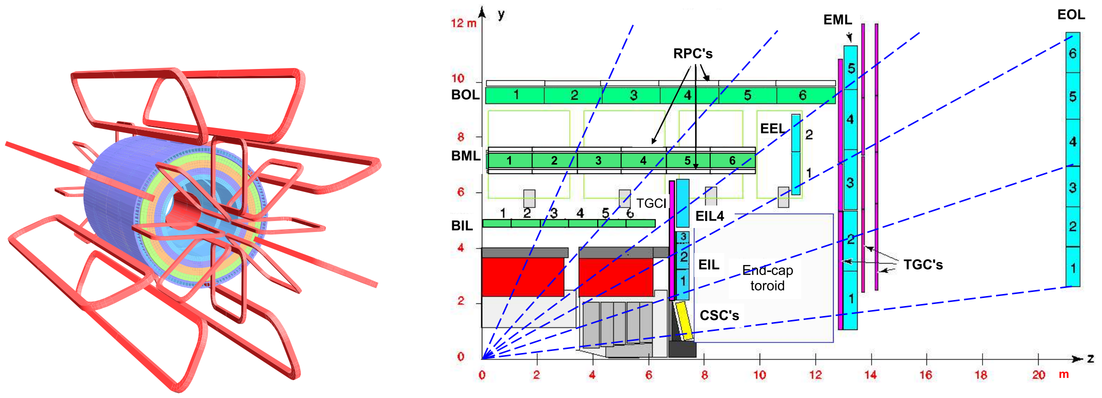
Hadronic calorimeter

In the barrel the tile hadronic calorimeter is a sampling calorimeter where steel is used as absorber and scintillating plastic tiles as active material, and is placed outside the envelope of the EM calorimeter. Together with the endcap and forward calorimeters it provides energy and momentum measurements used for jet reconstruction, and it suppresses contributions of particles different from muons reaching the MS. The main features required from the physics performance relates to good energy resolution, uniformity in η and ϕ directions and reduction of cracks and dead material. The barrel of the tile calorimeter covers the region $|\eta| < 1.0$ and it has two extended barrels covering the range $0.8 < |\eta| < 1.7$. A 68 cm gap between the two hosts the readout cables from the ID and the EM calorimeter. Like the EM calorimeter, also the tile barrels are longitudinally divided into three layers, the first two having a granularity of $\Delta\eta \times \Delta\phi$ = 0.1×0.1 and the last one having $\Delta\eta \times \Delta\phi$ = 0.2×0.1. A single module structure is shown in Fig. 4.7b. Radially the tile calorimeter extends up to 4.25 m, the greater dimensions are due to the larger particle penetration depths of hadrons compared to electrons. Azimuthally, the barrel components are divided into 64 wedge-shaped modules that cover a sector of $\Delta\phi$ = 0.1.

The hadronic calorimeters in the endcaps and in the forward area use a LAr technology. Each hadronic endcap calorimeter consists of two independent wheels placed behind the endcaps of the EM calorimeter, covering a range of $1.5 < |\eta| < 3.2$. The segmentation of the modules changes at $|\eta|$ = 2.5, being $\Delta\eta \times \Delta\phi$ = 0.1×0.1 for $|\eta| < 2.5$ and twice as coarse for larger $|\eta|$ values. The relative energy resolution of the hadronic tile and endcap calorimeters is [85]:

$$\frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E[\mathrm{GeV}]}} \oplus 3\%.$$

The forward calorimeter [86] is integrated in the endcap cryostats and consists of three modules



(a) (b)

Figure 4.8: (a) Geometry and position of the barrel and endcaps toroid magnets [1]. (b) Section of the muon system in a bending plane containing the beam axis. The dashed lines indicate trajectories of infinite-momentum muons traversing three muon stations [1].

in each endcap. The first one is optimised for electromagnetic measurements and is made of copper, while the other two are made of tungsten and mainly measure hadronic interactions. It has a coverage of 3.1 $|\eta| < 4.9$. Its front face is placed at 4.7 m to the interaction point, which is a relatively short distance and makes dealing with radiation challenging. On the other hand this allows for a more uniform coverage. The granularity of the forward calorimeter is $\Delta\eta \times \Delta\phi = 0.2 \times 0.2$ and the relative energy resolution is [85]:

$$\frac{\sigma_E}{E} = \frac{100\%}{\sqrt{E[\mathrm{GeV}]}} \oplus 10\%.$$

4.2.4 Muon Spectrometer

Muons need a further detector, as they can traverse the calorimeter system unstopped, only depositing in it a small amount of energy. The MS [87] provides muon momentum measurements from 3 GeV up to 1 TeV and features superconducting toroid magnets which provide the magnetic field to deflect muon tracks in the R-z plane. The system is instrumented with separate trigger and high-precision tracking chambers and it occupies 90% of the total volume of the ATLAS detector. The total track measurement is obtained with the complementary information provided by the inner detector. The MS has been designed for efficient muon identification and high-precision momentum measurements. The toroid system consists of three large air-core toroids, depicted in Fig. 4.8a where the magnets windings and positions are shown. A large barrel toroid provides a magnetic field in the range $|\eta| < 1.4$, has an outer diameter of 20.1 m and is made of 8 coils assembled radially and symmetrically around the beam axis. In the region $1.6 < |\eta| < 2.7$ two smaller endcap magnets are positioned, while the interval $1.4 < |\eta| < 1.6$ is the so called transition region, where the magnetic deflection is provided by a combination of barrel and endcap field, but with a lower bending power. The two endcap toroids are aligned to it in the endcaps, but rotated in azimuth by an angle of 22.5° to optimise the performance in the transition region. The bending power is

higher in the endcaps where it reaches 6 Tm, while typical values in the barrel are around 3 Tm. All the magnets are helium-cooled in order to reach the superconducting state.
The chambers are arranged in the barrel region in three cylindrical layers around the beam axis, while in the transition and the endcap regions they lay in three layers perpendicularly to it. A section of the system is shown in Fig. 4.8b. For most of the $|\eta|$ range the Monitored Drift Tubes (MDT) aim at measuring precisely the track coordinates in the principal bending direction of the field, having an average resolution of 80 μm. In the innermost plane, over the range $2.0 < |\eta| < 2.7$ and close to the interaction point, Cathode Strip Chambers (CSC) are positioned. These are multiwire proportional chambers with cathodes segmented into strips and with high granularity. Other types of chambers, the Resistive Plate Chambers (RPC) in the barrel and Thin Gap Chambers (TGC) in the endcaps, cover $|\eta| < 2.4$ and feed the trigger system with data. The main purposes are to provide bunch crossing identification, well defined p_{T} thresholds and measure a coordinate in the direction orthogonal to the one determined by the precision chambers. The momentum resolution of the spectrometer is $\frac{\sigma_{p_{\mathrm{T}}}}{p_{\mathrm{T}}}$ = 2-4% in the range p_{T} = 10-200 GeV and $\frac{\sigma_{p_{\mathrm{T}}}}{p_{\mathrm{T}}}$ = 10% at p_{T} = 1 TeV [87].

4.2.5 Forward Detectors

Four smaller detectors are positioned in the forward region, starting from ± 17 m from the interaction point up to ± 240 m. The LUCID (LUminosity measurement using Cerenkov Integrating Detector) [88] and ALFA (Absolute Luminosity For ATLAS) [89] detectors have the main aim of measuring the luminosity delivered to ATLAS. The ZDC (Zero-Degree Calorimeter) [90] plays a key role in determining the centrality of heavy-ion collisions, being able to measure neutral particles at pseudorapidity values up to $|\eta| \geq 8.2$. The AFP (ATLAS Forward Proton) [91] detector aims at detecting diffractive events, i.e. where one or two protons do not dissociate in the collision and are scattered at very small angles.

4.2.6 Trigger System

The high centre-of-mass energy and luminosity at the LHC cause an elevated event rate. This needs to be decreased in order for the data to be stored: the ATLAS trigger system [92] reduces it from the designed bunch-crossing rate of 40 MHz to about 1000 Hz on average. Finally, events, which have a size of the order of 1 MB, are written to mass storage.
The ATLAS trigger system is divided into two levels [93]. The first one, Level-1 (L1), is a hardware-based system and performs the first, fast reduction using a reduced granularity from the calorimeter (L1Calo) and muon (L1Muon) subdetectors. The calorimeter triggers use a sliding-window algorithm based on analog signals from the hadronic and EM calorimeter cells called *trigger towers*. Once digitised, the signal is converted into E_{T} values which are transmitted to different processors responsible for the identification of $e^{\pm}$, γ or $\tau^{\pm}$ candidates on one side and of jets and E_{T} objects on the other side. The former objects are identified through the use of predefined thresholds in the energy deposit in 2×2 clusters of towers. The jets and E_{T} processors use the same technique in clusters around a trigger tower core identified as local maximum. Muons are first identified by the muon trigger chambers RPC and TGC via a simple tracking algorithm and the information is

combined in the muon central processor.
The geometric and kinematic information from the L1Calo and L1Muon are combined in the L1Topo trigger. This computes refined global event quantities, such as invariant masses or angular variables, to be sent to the central trigger processor (CTP). This performs the final L1 decision, based on hit coincidences in different subdetector layers within certain windows, and implements *menus* made of combinations of trigger selections. For each event the L1 trigger defines the (η, ϕ) coordinates of regions where its selection processes have found an event with interesting features, a *region of interest* (RoI). The time for the L1 to make a decision on the event is of about 2.5 μs.
If an event is accepted by L1 it is re-processed at the High Level Trigger (HLT) stage. This is a software-based trigger, obtained by merging and optimising two of the levels previously used in Run 1, the second and Event Filter levels. The RoI formed at L1 are sent to the HLT, where sophisticated selection algorithms use the full detector granularity in the RoI or in the full event. Most of the trigger reconstruction algorithms are optimised to minimise differences between the HLT and the offline analysis selections. The event rate is reduced from the L1 output of $\mathcal{O}$(100) kHz to approximately 1 kHz within a processing time of roughly 200 ms. A small fraction of the bandwidth is dedicated to the so called *prescaled triggers*, when only a fraction of events passing the HLT are selected and stored, in order to decrease the final output rate.

4.2.7 Data Acquisition

All the information from the detector needs to be digitised and stored, in order to be available in case of a positive trigger decision, when they are transferred to the Data Acquisition System (DAQ). The data are stored temporarily in local buffers by the readout system and the ones in the RoI are required by the HLT. The selected events are then reconstructed. Events passing the full chain of the trigger are stored at CERN in the so called Tier-0 facility, replicated and distributed through the LHC computing grid in Tier-1 and Tier-2 locations, computing facilities all over the world.

5 Object Reconstruction and Particle Identification

Particles that pass through the detector are reconstructed starting from the responses of the subcomponents, which are combined by dedicated algorithms into high-level physics *objects*. These are identified and their four-momenta and kinematic properties are retrieved. This chapter is dedicated to the techniques used in the ATLAS collaboration to perform the reconstruction and evaluate the related efficiencies and uncertainties, starting from the construction of the tracks and interaction vertices in Section 5.1. Particular attention is given to the objects of interest for the studies presented in this thesis. Electrons are detailed in Section 5.2 and muons in Section 5.3. Section 5.4 offers an overview on the reconstruction of the spray of hadrons named jets, while Section 5.5 focuses on the missing transverse momentum.

5.1 Track and Vertex Reconstruction

Track reconstruction The tracking system records the hits and drift circles of the charged particles traversing the ID.

The track reconstruction [94] starts from the creation of clusters from hits in the pixel and SCT that are adjacent and have a deposited charge above a certain threshold. Three-dimensional *space points* are then created, which represent the points where the charged particle traversed the active material of the ID. After the clusterisation, track seeds are formed from combinations of space points in the four pixel layers and the first SCT layer. Track candidates are formed afterwards, through the use of a combinatorial Kalman filter [95] which incorporates other space points compatible with the preliminary trajectory, so that track candidates are extended till the fourth layer of the SCT. Multiple track candidates per seed are created in case of more than one compatible space point on the same layer. This makes an ambiguity solving step necessary, where candidates are ranked based on a score. The score is computed based on the intrinsic resolution and multiplicity in different subdetectors of the clusters in the candidate. Also the χ^2 of the track fit is taken into account together with the logarithm of the track momentum, which promotes energetic tracks. Tracks which fail any of the criteria are rejected, while the others are ranked. The ambiguity algorithm is aided by an artificial neural network [96] trained to identify merged clusters and used to predict cluster positions. A final high-resolution fit with the full information is performed on candidates fulfilling the requirements and fitted tracks are added to the final collection. The surviving tracks are extended into the TRT through algorithms that search for new hits and provide candidates for the silicon-seeded tracks. These extensions are fitted and evaluated via methods similar to the first ambiguity solving step and the final collection is created.

In cases where track seeds from silicon have a low score due to ambiguous hits, their seeds are suppressed at the ambiguity solving stage. In other cases hits in the innermost layers can be missing, e.g. for converted photons or long-lived particles ($\tau > 3{\times}10^{11}$ s). Also, a substantial energy loss at large radii can lead the extension to the TRT in wrong directions, which can happen mostly for electrons. For all these cases a different algorithm is used, which starts from a global pattern recognition in the TRT where segments are formed and followed back into the silicon detectors [97].

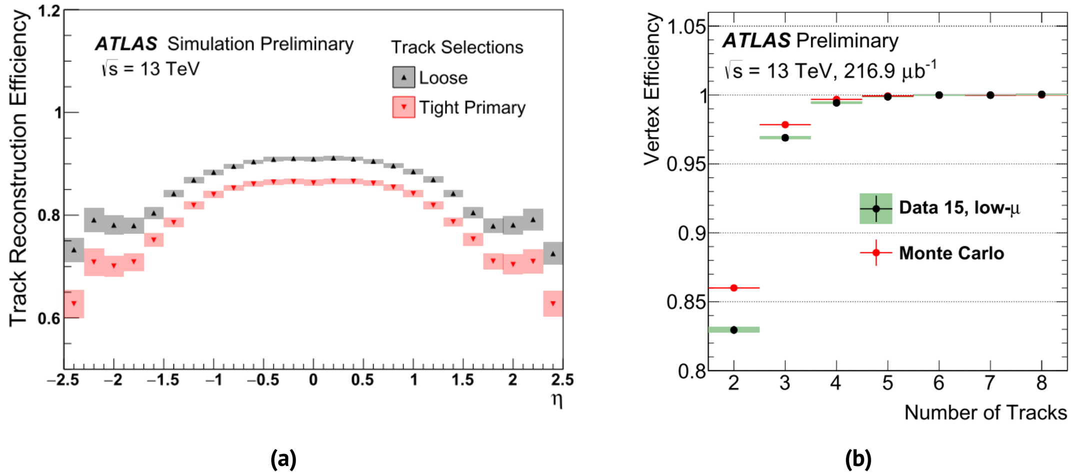


Figure 5.1: (a) Track reconstruction efficiency evaluated on simulated minimum bias charged pions as a function of truth η for Loose and Tight Primary selections. The different selections are based on the track p_T ($>$ 400 MeV in both cases), and number of hits in pixel and SCT detectors, required to be higher for *Tight Primary* tracks. Extension in the TRT is not required. The bands indicate the systematic uncertainties [98]. (b) Efficiency of vertex reconstruction as a function of the number of tracks in a low-μ data sample, compared to simulation [101].

The track reconstruction efficiency is measured as a function of (p_T, η) in simulations [98], by associating each cluster to the truth particle, i.e. a particle in simulation, which has the largest energy deposition in simulation, and considering the ratio of the number of reconstructed tracks matched and the number of generated particles. Fig. 5.1a shows the efficiency as a function of η for two sets of tracks, depending on different quality criteria based on momentum, pseudorapidity and number of hits, measured on charged pions from minimum bias events[1]. In general in the regions $|\eta| > 1$ the efficiency is lower due to the higher amount of material, while above $p_T \geq 5$ GeV it reaches a plateau around $\sim$90% for the looser selection and $\sim$85% for the tighter one.
Two quantities are defined, that give an estimation of the quality of the reconstructed track. One is the *transverse impact parameter significance* d_0/σ_{d_0}, defined as the ratio between the distance from the proton beam to the track's point of closest approach to the beam in the transverse plane d_0 and its significance. The other is the *longitudinal impact parameter* z_0, which is the distance of the track from the primary vertex measured along the beam axis at the point of closest approach in the R-ϕ plane.

Vertex reconstruction The interaction vertex is reconstructed by [102] defining a set of tracks that satisfy certain selection criteria, and identifying a seed position. By an iterative fit the best vertex position is estimated where less-compatible tracks are down weighted and the vertex position recomputed. When this is fixed, incompatible tracks are removed and used for the reconstruction of other vertices. The vertex considered to be the one of the hard scattering is defined as *primary*. It requires at least two tracks associated to it and is the vertex with the largest $\sum p_T^2$ of the associated

[1] The PYTHIA8 event generator [99] was tuned based on data from inclusive-inelastic scattering. Data from minimum bias and underlying events are modelled with differente diffractive descriptions, with the contribution mixed according to the cross section. This is a standard procedure, which gives insights into the description of the strong interaction at low energy scales [100].

tracks [101]. The vertex reconstruction efficiency is determined in data as the fraction of events with a reconstructed vertex among events with at least two reconstructed tracks, and it is shown in Fig. 5.1b. The dataset used contains events with a reduced number of interactions per bunch-crossing (low-μ events), while the simulated sample uses the minimum-bias tune of PYTHIA 8.185 [103]. Tracks have a minimum kinematic selection, requiring $p_{\mathrm{T}} > 400$ MeV for $|\eta| < 2.5$, and requirements on silicon hits are also placed.
The uncertainty of the vertex position is also estimated in simulation, proportional to the separation between two vertices reconstructed out of two sub-sets of tracks of approximately the same $\sum p_{\mathrm{T}}$ and corrected with scale factors derived with data [104]. Finally, further algorithms search for secondary vertices and photon conversions.

5.2 Electrons

Electrons are reconstructed starting from energy deposits inside the EM calorimeter, which are matched to tracks in the inner detector, aiming at a high efficiency. The techniques for reconstruction and identification are explained in the following, together with some details on their isolation, energy calibration, trigger information and an overview of the efficiency measurement. Particular attention is given to the electron reconstruction as this will be of importance for the studies described in Chapter 6.

Reconstruction In the first step of the reconstruction procedure a sliding-window algorithm [105] searches for energy deposits in the EM calorimeter in windows of size 3×5 middle layer cells of $\Delta\eta \times \Delta\phi$ = 0.025×0.025, inside longitudinal towers, defined in see Section 4.2.3. The total transverse energy of the towers contained in the window must be above the threshold of 2.5 GeV, taking into account the energy of the cells in all longitudinal calorimeter layers. In this case a *precluster* is formed and the position of its *seed* computed as the weighed barycentre of all cells within a 3×3 window around the central tower. Here a removal of duplicates is performed, and in case two preclusters are within 2×2 cells the more energetic one is kept. Once the clusters are formed, the cells are assigned to each of them, depending on their distance from the seed position, different for each calorimeter layer. In the last step clusters of different $N_\eta^{\mathrm{cluster}} \times N_\phi^{\mathrm{cluster}}$ sizes are built, based on the hypothesised particle types - electron, converted or unconverted photon - and position in the calorimeter.
The track reconstruction begins with pattern recognition procedure, where a track seed is formed out of three hits in the silicon detectors. The energy loss due to interactions with detector material is evaluated under the assumption of a pion track, and attempts are made to build a track with at least seven hits. If the track falls within a region of $\Delta R < 0.3$ from the barycentre of the cluster, the pattern recognition is performed again under the electron assumption in terms of energy loss. Candidates are then fit with the ATLAS Global χ^2 Track Fitter [106] under either the pion or electron hypothesis, according to the recognition algorithm used previously.
At this stage a track-cluster matching is performed. Tracks are extrapolated to the calorimeter middle layer and the position is compared to the cluster barycentre. In the matching the energy loss

due to bremsstrahlung emission is taken into account and the number of hits in the silicon detector has to be larger than or equal to four. Tracks loosely matched in this way are refit with a Gaussian Sum Filter [107], optimised for electrons taking into account non-linear bremsstrahlung effects.

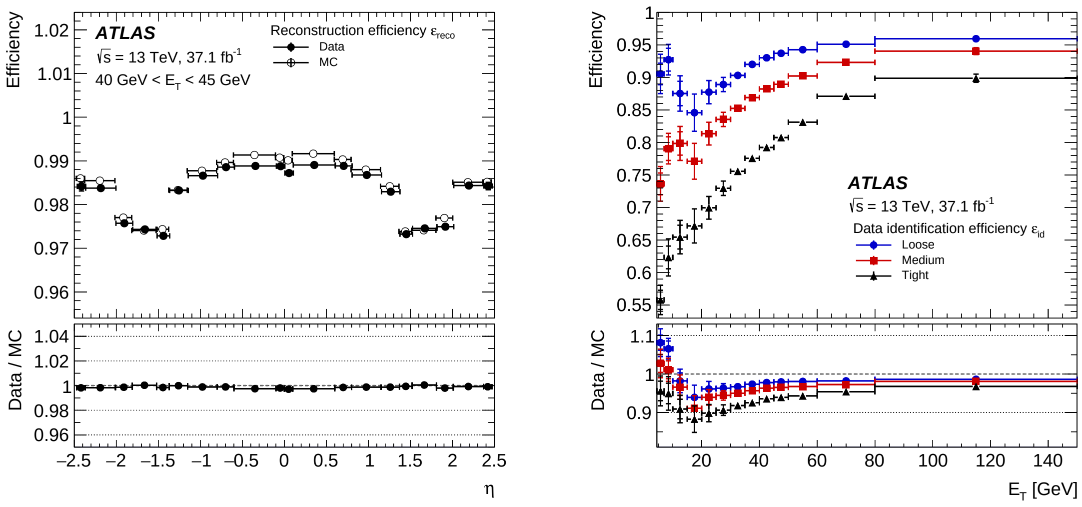
If more than one track satisfies the matching conditions, one is chosen as *primary* by an algorithm that uses the cluster-track distance, the number of hits in the pixel detector and the presence of hits in the first layer of the SCT. In case of absence of tracks associated to a cluster, the object is considered as unconverted photon. If the association is done with two tracks of opposite curvature, but collinear at the production vertex and compatible with electrons in the TRT, then the cluster is classified as a converted photon. This classification can also happen for a cluster matched to single tracks in order to increase the efficiency of conversion detection, but the track must have no hits in the silicon detectors in this case. Clusters for electrons are reformed with a different tower granularity and the energy is calibrated, as will be described later in this section.

Efficiencies for the reconstruction, shown in Fig. 5.2a as a function of η, are between 97%-99% for electrons with $E_\mathrm{T} > 15$ GeV. The four momentum is reconstructed out of η and ϕ extracted from the track, and the energy retrieved from the calibrated cluster. Tracks are required to be originated from the primary interaction vertex, to reduce the background from conversions.

Identification Identification algorithms use variables related to track measurements, shapes of showers in the calorimeter and track-cluster matching quantities to distinguish signal from background electrons. The simulated samples used for the optimisation of such algorithms are $Z \to ee$ and $J/\Psi \to ee$, which need to be corrected for inaccuracies in the description of the detector and the shower shapes modelling performed with GEANT4 [64].

The baseline algorithm is based on a likelihood method that uses the signal and background probability density functions of variables which are chosen as discriminating. A likelihood discriminant is built and requirements are placed on it to distinguish between signal-like or background-like objects. Furthermore, electrons are required to pass simple selection criteria on the number of hits on the track.

Based on the different selections on the discriminant, three inclusive operating points are defined, *Loose*, *Medium* and *Tight*, in increasing background rejection power. The identification operating points are optimised in several bins of $|\eta|$ and E_T, due to the dependency of the variables on both the material distribution and the energy. The signal efficiency for the Tight identification over the E_T range goes from 78% to 90%, while the highest efficiency is measured for the Loose operating point and is around 95%. Lower signal efficiencies for Tight electrons have the benefit of a higher background rejection. The Loose identification efficiency increases to 85% at E_T = 20 GeV and to 95% at E_T = 100 GeV [6]. Two methods, namely the Z_mass and the Z_iso method, are performed on $Z \to ee$ events to obtain the identification efficiencies. They differ for the background treatment and are treated as systematic variation on a single measurement. Also $J/\Psi \to ee$ events are used to retrieve efficiencies for low E_T electrons, i.e. 7 GeV $< E_\mathrm{T} <$ 20 GeV. Identification efficiencies are shown in Fig. 5.2b for the three operating points.



(a) **(b)**

Figure 5.2: (a) Electron reconstruction efficiencies relative to reconstructed clusters measured in $Z \to ee$ events in simulation (open points) and data (closed points) collected in 2015+2016 as a function of η. The inner uncertainties are statistical and the total show the statistical and systematic component [6]. (b) Electron identification efficiencies measured in $Z \to ee$ events as a function of E_{T} for the Loose, Medium and Tight operating points. The vertical uncertainty bars represent the statistical (inner bars) and total (outer bars) uncertainties [6].

Isolation The signature in the detector is similar for *prompt* electrons, originated from heavy resonance decays such as Higgs, W or Z bosons, and for *background* electrons, which can be originated from non-prompt decays of c- and b-hadrons, electrons from converted photons or even light hadrons misidentified as electrons. Identification criteria used for rejecting such background objects are enforced by requirements on the detector activity around a lepton candidate, referred to as *isolation*. Two quantities are defined: the track-based $p_{\mathrm{T}}^{\mathrm{varcone}}$ and the calorimeter-based $E_{\mathrm{T}}^{\mathrm{cone}}$.

The former is defined as the scalar sum of the transverse momenta of the tracks associated to the primary vertex in a cone around the lepton. Tracks are considered if $p_{\mathrm{T}} > 1$GeV, and the p_{T} of the tracks associated to the electron is excluded from the sum, as well as the associated ones from e.g. converted bremsstrahlung photons. The size of the variable cone for electrons is $\Delta R = \min(0.2, 10\ \mathrm{GeV}/p_{\mathrm{T}})$, and the variable is referred as $p_{\mathrm{T}}^{\mathrm{varcone0.2}}$. This is sensitive to charged particles only, but additional interactions from pp collisions can be suppressed due to the requirement on the tracks to be originated from the primary vertex.

The $E_{\mathrm{T}}^{\mathrm{cone}}$ is the sum of the transverse energy of the *topological clusters* in a cone around the electron candidate. Topological clusters are built from neighbouring cells in the calorimeter that have energy deposit above a certain threshold, which is estimated from noise from electronic and pile-up simulations [108]. From the transverse energy of the clusters in the cone, an energy deposit contained in a rectangular smaller cluster of $\Delta\eta \times \Delta\phi = 0.125 \times 0.175$ is subtracted to remove the energy deposited by the electron, and pile-up effects are corrected for on an event-by-event basis. Usually the size of the cone is $\Delta R = 0.2$ and the variable is called $E_{\mathrm{T}}^{\mathrm{cone0.2}}$. To determine the isolation criteria *relative* isolation variables are used, where the quantities defined above are divided by the p_{T} of the lepton candidate.

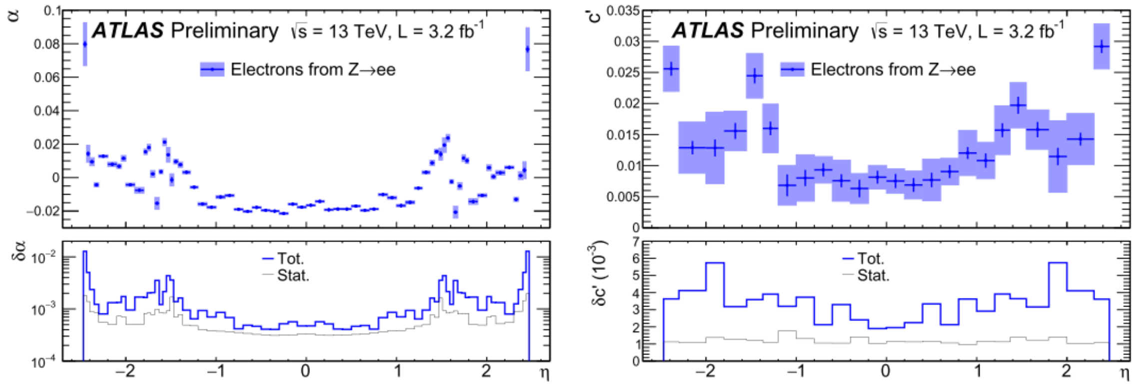


Figure 5.3: Electron calibration energy scale factor α and additional constant term for the energy resolution, measured in $Z \to ee$ events as a function of η. The bands on the top plots are the total uncertainties, while in the bottom plots total and statistical uncertainties are depicted [109].

Isolation requirements are placed on the $p_{\mathrm{T}}^{\mathrm{varcone0.2}}$ and $E_{\mathrm{T}}^{\mathrm{cone0.2}}$. Different operating points are defined for the isolation, depending on the target of the selection. A first class targets a given isolation efficiency measured on $Z \to ee$ events, either constant or as a function of E_{T}. This is the case for the Gradient isolation operating point with 90(99%) efficiency at 25(60) GeV [6]. A second class has a fixed threshold on the isolation variable. This is the case of the FixedCutTight isolation.

Energy calibration This is performed in two steps, the first based on simulation through multivariate techniques, and a second *in situ* calibration [109]. The first step is performed with a single-particle sample without pile-up, and covers the region $0 < |\eta| < 2.5$. The technique is the same as was used in Run 1 [110] with some updates related to the reconstruction and the insertion of the IBL. The multivariate analysis aims at finding the calibration constants to correct the electromagnetic cluster for the energy loss in the material upstream of the calorimeter, the energy deposits in neighbouring cells and the leakage beyond the liquid Argon calorimeter. To do so simulations need to be corrected to reach an accurate modelling of the detector geometry and material distribution, which is measured in data using the ratio of the energy in different calorimeter layers. The response-calibration is applied to cluster energies both in data and simulation. To account for residual disagreement between data and simulation an in-situ procedure is performed. The miscalibration of the energy scale is parametrised by the different response between data and simulations $E_i^{\mathrm{data}} = E_i^{\mathrm{MC}}(1 + \alpha_i)$, for a given η bin i. The difference in the energy resolution of data and simulation, parametrised by Eq. (4.1), is quantified by an additional constant term for each η bin. The two terms, shown in Fig. 5.3, are derived using $Z \to ee$ events and applied both to electrons and photons. The constant term lays between 1%-3% depending on the pseudorapidity region, the energy scale for electrons from $Z \to ee$ decays is measured with a precision better than one per mil in the barrel and a few per mil in the endcaps, while the constant term is measured with a precision between two and five per mil. The energy corrections derived on simulation are applied to data while the resolution in simulation is corrected for the different resolution observed in data.

Trigger Electrons are reconstructed and identified both at the L1 trigger level and at the HLT. Regions in the calorimeter made of 4×4 trigger towers for a total size of $\Delta\eta \times \Delta\phi = 0.4{\times}0.4$ (see Section 4.2.3) are used to calculate the energy, and the choice is made based on thresholds which differ depending on the η region. A veto can also be set to certain values on the leakage into the hadronic calorimeter, and requirements on the isolation can be put at this stage as well, on the transverse energy contained in the ring-area around the core of the cluster. Regions of interest are then identified, which are $\Delta\eta \times \Delta\phi$ = 0.4×0.4 areas where the L1 selection requirements are satisfied (see Section 4.2.6), and the second trigger step can be performed.
At the HLT candidates are reconstructed and potential background electrons are already rejected in order to decrease the rate. Fast algorithms related to the EM calorimeter build clusters starting from the region of interest identified at L1. Requirements are imposed on quantities like the energy deposited in the hadronic calorimeter, the shower shape, the energy difference between the first two largest energy deposits in the clusters and the transverse energy of the cluster itself. Tracks are matched to clusters before a second HLT step, where additional requirements are placed. Given an offline identification algorithm and an isolation operating point, the trigger efficiency is defined as the number of events selected by the trigger in a sample of events where reconstructed electrons are identified by the offline algorithm.

Efficiency measurement The electron efficiency is given by the product of different efficiency measurements for the reconstruction, identification, isolation and trigger. These are performed separately, but each with respect to the previous step. For all the measurements a tag-and-probe method is used, performed in two-dimensional bins of both E_{T} and η. Samples of $Z \to ee$ and $J/\Psi \to ee$ are used, where a strict selection is applied to one of the electrons (tag) together with some requirements on the di-electron system, so that the other electron of the pair (probe) remains as unbiased as possible. The efficiency is then measured on the probe electron, after the residual background has been accounted for. Since efficiencies in data not always match the ones in simulation, simulated samples will need to be corrected for these differences. For this reason efficiencies are measured in data and simulation and their ratio, the scale factor, is the final quantity used as a multiplicative factor to correct simulated electrons. The use of the ratio accounts also for the differences of efficiencies for electrons from different processes, like $Z \to ee$ or $J/\Psi \to ee$, because these residual effects are expected to cancel out in the ratio, making scale factors universal. The combination of the measurements itself is done using these ratios. Measurements are performed in two-dimensional bins of (E_{T}, η) due to the dependency of the rates on these two variables and scale factor values are usually close to unity. In Fig 5.4 the combination is shown for the product of reconstruction, identification and isolation efficiencies as a function of E_{T}.
The uncertainties related to the measurement are estimated by varying some parameters of the measurement, such as the isolation criteria for the tag and the probe electron, the fit model for signal and background or the signal invariant mass range.
Scale factors are determined with uncertainties around a few percent for E_{T} around 5 GeV and reaching less than 1% for E_{T} values higher than 100 GeV.
In order to address the issue of the *misidentification* of an electron's charge, in Run 2 new studies

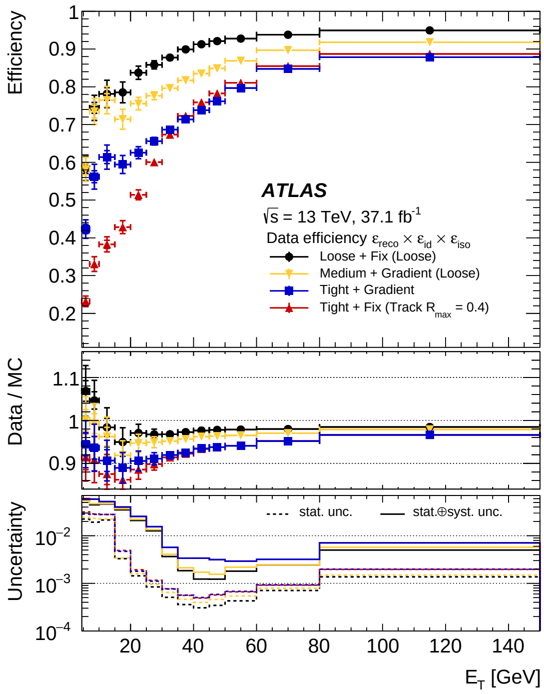


Figure 5.4: Product of reconstruction, identification and isolation efficiencies for data in $Z \to ee$ events as a function of E_T for different operating points. The inner uncertainties are statistical while the total include the statistical and systematic components. The lower panel shows the ratio between data and simulation results, and the relative statistical and total uncertainties [6].

have been performed and for the first time centrally produced charge misidentification scale factors have been delivered. Chapter 6 is dedicated to these studies and measurements.

5.3 Muons

Muons leave a small fraction of their energy in the ID or in the calorimeters. They are reconstructed by combining information mainly from the ID and the MS, and then identified with high efficiency. A detailed description of the techniques can be found in [111], in the following an overview is presented.

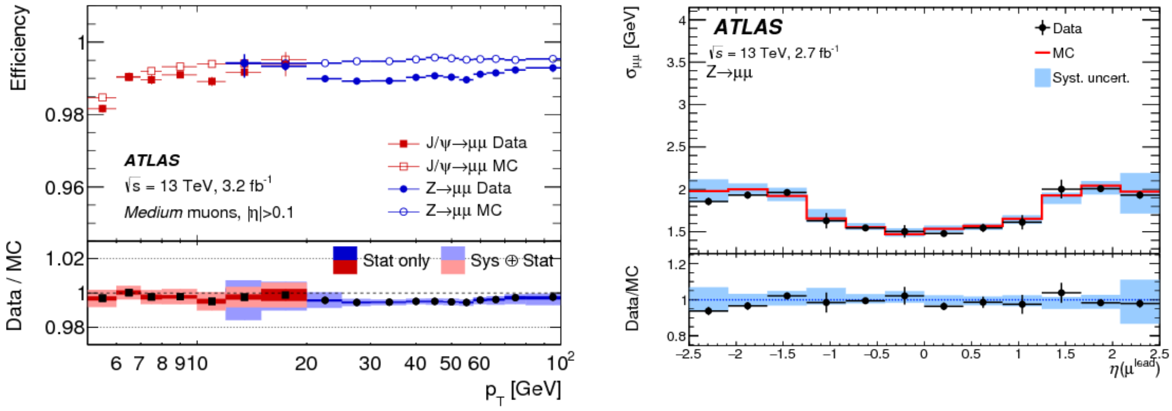
Reconstruction The information from the ID and the MS is reconstructed independently and then combined to form the muon track. The technique used to combine them defines different muon types.

The reconstruction in the ID is the same as described in Section 5.1. In the MS, track segments are first formed starting from a search of hit patterns in each chamber. In the MDT, aligned hits in the bending plane are searched for by applying a Hough transformation method [112]. Hits in the RPC and TGC are used to measure the coordinates orthogonal to the bending plane. Track candidates are then built by fitting information from different layers, using as seed the middle one and selecting segments based on hit multiplicity and fit quality. At least two segments matched by using relative position and angles are required to build a candidate, with the exception of the transition region

between the endcaps and the barrel, where only a single high-quality segment can be used. The same segment can be shared by more tracks, but an overlap removal procedure selects later the best assignment. However, all tracks which share segments in the two innermost detector layers are kept, if they have segments in all the three layers and do not share hits in the outermost one. This ensures high efficiency in case of close-by muons reconstruction.
Depending on the subdetectors used in the reconstruction, four types of muons are defined. The combined (CB) muons are reconstructed independently in the ID and MS and the combined track is formed with a global fit that uses hit information from both subdetectors. They are defined only in the $|\eta| < 2.5$ range. To extend the acceptance to the region not covered by the ID, in the range $2.5 < |\eta| < 2.7$ extrapolated muons (ME) are defined. The reconstruction is based only on the information from the MS with a loose requirement on compatibility with an origin in the interaction point. For muons with low p_{T} or ending up in regions with reduced MS acceptance, the segmented-tagged (ST) algorithm is used. A track in the ID is extrapolated to the MS and classified as muon if associated with at least one segment in the MDT or CSC. In order to recover tracks in the areas of the MS which are poorly instrumented, also the information from the calorimeter is included: if a track in the ID can be matched to an energy deposit consistent with a minimum-ionising particle in the calorimeter, it is identified as calorimeter-tagged (CT) muon.

Identification and isolation This is performed to suppress backgrounds but selecting at the same time muons with high efficiency while ensuring a robust momentum measurement. Background comes mainly from in-flight decays of pions and kaons in the ID, where tracks are usually distinguished by a "kink" in the reconstructed track. Quality requirements are applied on variables that offer good discrimination between signal and background, as verified in simulated $t\bar{t}$ events. For CB muon tracks selections are applied for the *q/p significance* [2] and $\rho\prime$[3]. Also the normalised χ^2 of the combined track fit is considered in the identification criteria, while requirements on the number of hits in the ID and MS are used to ensure a robust momentum measurement.
Among the four identification working points provided, two are considered in this analysis: the *Loose* one for the muon pre-selection, and the *Medium* one. All muon types are used for the Loose selection, which is designed to maximise the reconstruction efficiency while providing good-quality tracks. The four classes are inclusive, so all the CB and ME muons satisfying the Medium requirement are included in the Loose selection. The Medium identification criteria can be applied to minimise the systematic uncertainties associated with the reconstruction and calibration, and only CB and ME are used in this case. To select Medium muons, requirements on the number of hits in the MS chambers are applied, as well as a loose selection on the compatibility between the ID and MS measurement through the q/p significance variable.
Isolation is defined as for electrons in Section 5.2 using $p_{\mathrm{T}}^{\mathrm{varcone}}$ and $E_{\mathrm{T}}^{\mathrm{cone}}$ variables. The calorimeter-based variable $E_{\mathrm{T}}^{\mathrm{cone}}$ uses the same ΔR = 0.2 as for electrons, and is named $E_{\mathrm{T}}^{\mathrm{cone20}}$, while the track-based one has a ΔR = min(0.3, 10 GeV/p_{T}^{μ}), and is named $p_{\mathrm{T}}^{\mathrm{varcone30}}$. Seven isolation working

[2] Absolute difference between the ratio charge/momentum of the muon measured in the ID and in the MS, divided by the sum in quadrature of the respective uncertainties
[3] Absolute value of the difference between the p_{T} measured in the ID and MS, divided by the p_{T} of the combined track.



(a) **(b)**

Figure 5.5: (a) Reconstruction efficiency for Medium muons as a function of p_T in the region $0.1 < |\eta| < 2.5$ measured in $Z \to \mu\mu$ and $J/\Psi \to \mu\mu$ data and simulated events. The error bars indicate the statistical uncertainty. The bottom pad shows the ratio of the efficiencies in data and simulation with statistical and systematic uncertainties [111]. (b) Dimuon invariant mass resolution for $Z \to \mu\mu$ events for data and corrected simulations as a function of the p_T of the leading muon. The ratio of the fitted resolution in data and simulation is shown in the lower pad. Both statistical and systematic uncertainties are displayed [111].

points are available. The Gradient operating point is selected by requirements both on $p_T^{\text{varcone30}}/p_T^{\mu}$ and $E_T^{\text{cone30}}/p_T^{\mu}$ and chosen to have $\geq 90(99)\%$ efficiency at 25(60) GeV.

Efficiency measurement The reconstruction efficiency [111] is measured with a tag-and-probe method on $Z \to \mu\mu$ and $J/\Psi \to \mu\mu$ events, and is close to 99% for $|\eta| < 2.5$ and in the p_T range $5 < p_T < 100$ GeV, as shown in Fig. 5.5a for the Medium selection. The isolation efficiency depends on the selection and the muon p_T, and varies from 93% to 100%. The simulation usually describes those values well, but to account for small differences in muon reconstruction, identification, isolation and trigger efficiencies, in analyses scale factors are applied to each muon, which are of the order of a few percent away from unity. Furthermore a set of corrections is applied to the simulated muon momentum as well, to account for momentum scale and resolution differences between data and simulation. In particular the momentum scale correction accounts for the material traversed by the muons before the MS and the related energy loss and it is usually around a few percent.

The resolution is obtained by fitting the width of the invariant mass peaks and is displayed in Fig. 5.5b for $Z \to \mu\mu$ events. The relative muon p_T resolution ranges from values smaller than or equal to 2%, up to 2.9%, and is larger for larger $|\eta|$ values.

5.4 Jets

The quarks and gluons produced in the collisions can undergo soft and collinear showering and then hadronise, appearing in the detector as collimated showers of energetic hadrons, and photons from their decays. They are commonly denoted as jets, and are observed in the ATLAS calorimeter as energy deposits [113].

Jets are first reconstructed out of energy information in the calorimeter but then need to be associated to tracks in the ID and calibrated, as detailed in [114]. Neighbouring cells in the calorimeter with energy deposit above a certain threshold - estimated from noise from electronic and pile-up simulations - are considered for building so called topological clusters. These are treated as massless particles assumed to originate from the geometrical centre of the detector, and are the inputs for reconstruction algorithms. The one most commonly used in ATLAS is the anti-k_t [115] one, a jet-clustering algorithm that generalises the existing sequential recombination jet algorithms [116,117]. The radius parameter ΔR, which defines the extension of the jet, is chosen as ΔR = 0.4 throughout this thesis.

Jets are reconstructed at the electromagnetic scale, which corresponds to the energy deposited by particles that interact electromagnetically. The construction is therefore initially done assuming energy depositions in the calorimeter by particles produced in electromagnetic showers, then jets need to be calibrated by scaling their energy to the truth jet energy scale (JES), that follows a local calibration step.

Using the same anti-k_t algorithm, truth jets are built out of stable interacting particles from simulation (with lifetime $> 10^{-10}$ s, and excluding muons and neutrinos), and are geometrically matched to the reconstructed calorimeter jets within a distance ΔR = 0.3. Truth jets are used in the calibration procedure that combines methods on both simulation and data, and is performed in several steps. It corrects the jet direction to point to the primary vertex rather than the centre of the detector. It also accounts for pile-up by removing the contribution of particles of infinitesimal momentum from the p_{T} of each jet, and then deriving a residual correction. Afterwards a simulation-based calibration is performed to correct both energy and direction of the jet to the one matched at truth level. A further step aims at reducing the dependency of the JES on the particle composition and their energy distribution inside the jet, which affects the longitudinal and transverse features of the jet. A residual in situ calibration is performed on data, to account for differences in jet response with respect to simulations. This is done by balancing the jet p_{T} against well-measured reference objects, photons, Z bosons or other jets, and the corrections obtained are applied to data jets.

Roughly 80 JES uncertainty terms are related to the final calibration, most of them accounting for assumptions on the event selection, simulation, sample statistics and propagated uncertainties or uncertainties related to the use of other objects. Also pile-up uncertainties, differences in jet response in simulation and data and simulated jet composition play a role. Uncertainties are around 4.5% at 20 GeV, while they are between 1-2% for 200 GeV $< p_{\mathrm{T}} <$ 2 TeV for central jets. They are fairly constant with respect to η and there are dedicated uncertainties introduced for $2.0 < |\eta| < 2.6$ to account for details in the calorimeter energy reconstruction. In Fig. 5.6 the JES uncertainties for 2015 data is depicted for jets of $\eta = 0$.

The resolution is measured using a multi-step process and is parametrised as function of three terms: the noise, the stochastic effect from the sampling nature of the calorimeters, and the constant term.

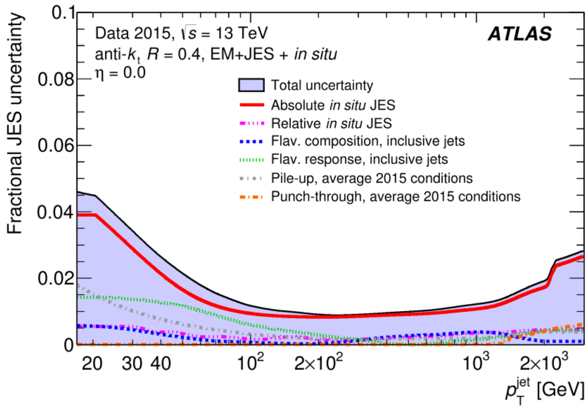


Figure 5.6: Combined uncertainty in the JES of fully calibrated jets as a function of the jet p_T at $\eta = 0$. Uncertainties include pile-up and the ones propagated from Z/γ^*+jets (absolute *in situ*) and η-intercalibration (relative *in situ*). The flavour composition assumes the quark and a gluon composition taken from PYTHIA dijet simulation [114].

***b*-jet tagging** The long lifetime, high mass, and decay multiplicity of b-hadrons are exploited by algorithms in order to identify jets containing b hadrons. Several have been developed [118], and the most discriminating variables resulting from these are combined in artificial neural networks. The one used in Run 2 is the MV2c10 algorithm [119], it makes use of boosted decision trees (BDTs) and exploits the improved tracking software. To measure the efficiency for a b-hadron to be tagged by the algorithm, two methods are used [119], a tag-and-probe one and a combinatorial likelihood approach. For both efficiencies are measured in data and simulation to compute scale factors which are close to one with uncertainties ranging from 2% to 12% depending on the jet p_T. Different b-tagging operating points are defined, based on the output of the b-tagging algorithm. They correspond to 60%, 70%, 77% and 85% b-tagging efficiencies in simulated $t\bar{t}$ events. Two sets are provided depending on whether the stated efficiency is averaged over the p_T spectrum but varies with jet p_T, or the efficiency is constant over p_T but the cut value changes. The scale factors are applied in simulated events per jet in analyses to correct event rates after a b-tagging requirement.

5.5 Missing Transverse Momentum

Particles not interacting inside the detector, which within the Standard Model are neutrinos only, are indirectly reconstructed by exploiting momentum conservation in the transverse plane [120]. The longitudinal momenta of the interacting partons are unknown, but their transverse momenta are usually negligible compared to momentum transfers in hard scattering interactions. Hence, the negative transverse momentum imbalance of the final state particles corresponds to the total transverse momentum carried by undetected particles. This quantity is typically referred to as the missing transverse momentum E_T^{miss}, whose x and y components are [120]

$$E_{x(y)}^{\text{miss}} = E_{x(y)}^{\text{miss},e} + E_{x(y)}^{\text{miss},\gamma} + E_{x(y)}^{\text{miss},\tau} + E_{x(y)}^{\text{miss,jets}} + E_{x(y)}^{\text{miss},\mu} + E_{x(y)}^{\text{miss,soft}}.$$

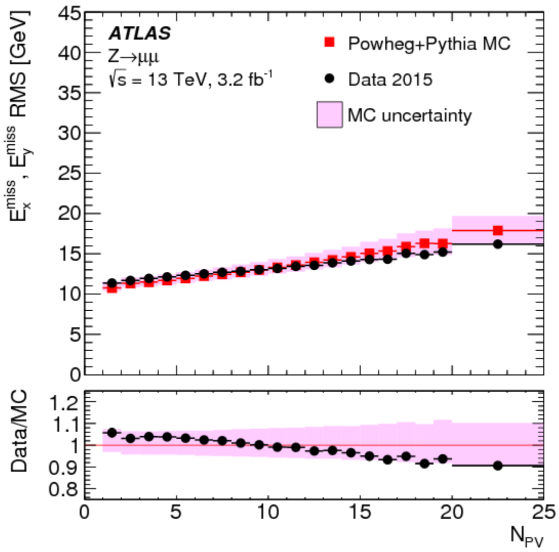


Figure 5.7: The root mean square width of the $E_{x(y)}^{\text{miss}}$ distributions in bins of the number of primary vertices in an inclusive sample of $Z \to \mu\mu$ events for data and Monte Carlo (MC) simulation. The shaded bands represent the combined statistical and systematic uncertainties on the resolution measurement [120].

The first five contributions in the sum are the so-called *hard terms*, i.e. the $x(y)$ components of the negative vectorially summed transverse momenta of the respective calibrated objects: electrons, photons, taus, jets and muons. The last term comprises the tracks in the event not associated to the reconstructed hard objects but still originating from the hard-scatter primary vertex, or it uses energy clusters in the calorimeter not associated with hard objects.
The $\mathbf{E}_{\text{T}}^{\text{miss}}$ vector is constructed out of $E_{x(y)}^{\text{miss}}$, hence its modular value is

$$E_{\text{T}}^{\text{miss}} = \sqrt{\left(E_x^{\text{miss}}\right)^2 + \left(E_y^{\text{miss}}\right)^2},$$

and its azimuthal angle ϕ^{miss} is given by

$$\phi^{\text{miss}} = \arctan\left(E_x^{\text{miss}}/E_y^{\text{miss}}\right).$$

The performance and features of the reconstruction of the $E_{\text{T}}^{\text{miss}}$ are evaluated in $W \to \ell\nu$ events, where genuine $E_{\text{T}}^{\text{miss}}$ is expected, and in $Z \to \mu\mu$ events, where it is not [120]. The modelling of the different components is evaluated in data and simulation, and the response and resolution are measured and studied as a function of the pile-up activity. The resolution is measured in terms of the root mean square width of the $E_{x(y)}^{\text{miss}}$ distribution, which is shown in Fig. 5.7 for $Z \to \mu\mu$ events. It ranges from 10–15 GeV, with a 10% difference between data and simulation in the region of higher pile-up activity.

6 Measurement of Charge Misidentification Probabilities

The electric charge of electrons[1], which are reconstructed as detailed in Ch. 5, is determined from the curvature of the primary tracks in the ID associated to them.
This chapter is dedicated to studies and measurements related to the electron charge misidentification. After the description in Section 6.1 of the data and simulated-events samples used, the categorisation of electrons performed in simulation is explained in Section 6.2, while the tracking and kinematic features of electrons with misidentified charge are presented in Section 6.3. These offer a base of concepts for Section 6.4, where a measurement of electron charge misidentification probabilities and corrections to their modelling in simulation are presented.

6.1 Dataset and Simulated Samples

The data used for the measurement were collected by the ATLAS detector in 2016, from pp collisions at $\sqrt{s}$ = 13 TeV. The data quality is ensured by requiring that detectors were operating nominally, and the integrated luminosity available amounts to 33.9 fb^{-1}.
A $Z \to ee$ sample is generated with POWHEG-BOX v2 [121–124], and interfaced to PYTHIA 8.186 [125] for the parton shower. The PDF set used for the matrix element is the CT10 [126], while non-perturbative effects are modelled with the AZNLO tune [127] and the CTEQ6LI PDF set.
To simulate the pp collisions from pile-up PYTHIA 8.186 is used with the MSTW2008LO PDF [128], and simulated events are reweighted to match the data distribution of the average number of interactions per proton bunch crossing.
As discussed in Section 3.2 the interactions of the particles with the ATLAS detector are simulated and the reconstruction procedures used for data are used to reconstruct the particles. Details on electron reconstruction are given in Section 5.2, where the description of the identification and isolation operating point used in the following can be found as well. For each electron both the *true* quantities at event generation and simulation stage, and the *reconstructed* quantities at detector level are studied. Reconstructed particles are matched to *true* particles, to be used as a reference to assess the reconstruction. The choice of the matched *true* particle is based on variables that quantify how likely the true particle caused the detector responses that led to the reconstruction of that specific particle.[2] After truth matching, other algorithms [130] classify the true objects. This is generally done based on the parent of each truth-matched particle, and in most cases this is sufficient to univocally identify signal or background objects for a specific process. However special care needs to be taken to classify electrons, especially when performing charge misidentification studies and measurements.

[1] In this context the name *electron* will refer either to electron or positron unless specified.
[2] The main quantity to evaluate this likelihood is the *truth matching probability*. It is the ratio of the number of hits in the ID caused by the true electron to the number of all ID hits of the electron's reconstructed track, weighted based on the average number of expected measurements in each subdetector [94,129].

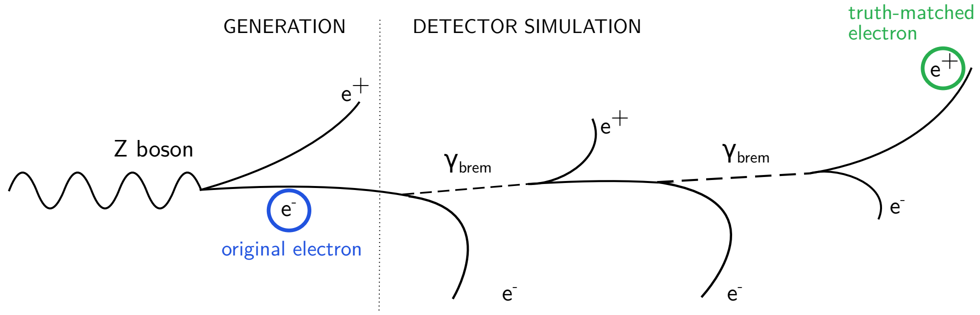


Figure 6.1: Schematic representation of the chain of bremsstrahlung emission and pair production processes of an electron generated from a Z boson. One of the two electrons which stems from the Z boson in the generation step emits a bremsstrahlung photon in the detector simulation phase, and an electron−positron pair is produced. The process is iterated due to further interactions with the detector material, and the reconstruction algorithm reconstructs the electron from pair production, which has a different charge with respect to the generated one. The *true* electron matched to the reconstructed one in these cases is usually the simulated electron from pair production (green circle), while the *original* one that has to be used for charge comparison is the generated one (blue circle).

6.2 Electron Classification in Simulation

The classification becomes crucial for cases of an electron's interaction with detector material resulting in the emission of a bremsstrahlung photon that converts into an electron−positron pair. This is studied in the following for $Z \to ee$ decays, but could be extended to W or Higgs boson decays, or other physics processes where electrons can be produced at generation-level. For the studies only electrons generated from the Z decay are considered, i.e. not deriving from background processes such as hadron decays or pair production of final or initial state radiation photons. In cases where the generated electron interacts with the material of the ID and produces a photon with subsequent conversion into an electron−positron pair, the reconstructed electron can be one of the two electrons from the photon conversion. In this case, based on the electron reconstruction procedure, the energy deposition in the electromagnetic calorimeter of one of the electrons from conversion can be matched to a track in the ID coming from the original electron. In these cases the truth-matched electron is the one originating from the photon conversion. However, the generated electron determines the true charge.
In Fig. 6.1 this truth matching procedure is schematised for an electron generated from the Z boson, that produces a bremsstrahlung and conversion chain during the simulation phase. The true electron matched to the reconstructed one is one of the electrons from the photon conversion produced in the GEANT simulation (encircled in green), while the generated electron from the Z boson (encircled in blue) is the one to be used as reference to asses the charge assignment. This will be referred to in the following as *original*.
The details on the selection procedure for the original electron using the truth information are given in Appendix A.
Based on these considerations, the electrons classified as *prompt* in a $Z \to ee$ sample are:

- signal electrons generated directly from the Z boson, labelled as *single-track* in the following;
- electrons from conversion of bremsstrahlung photons, for which it is possible to trace back

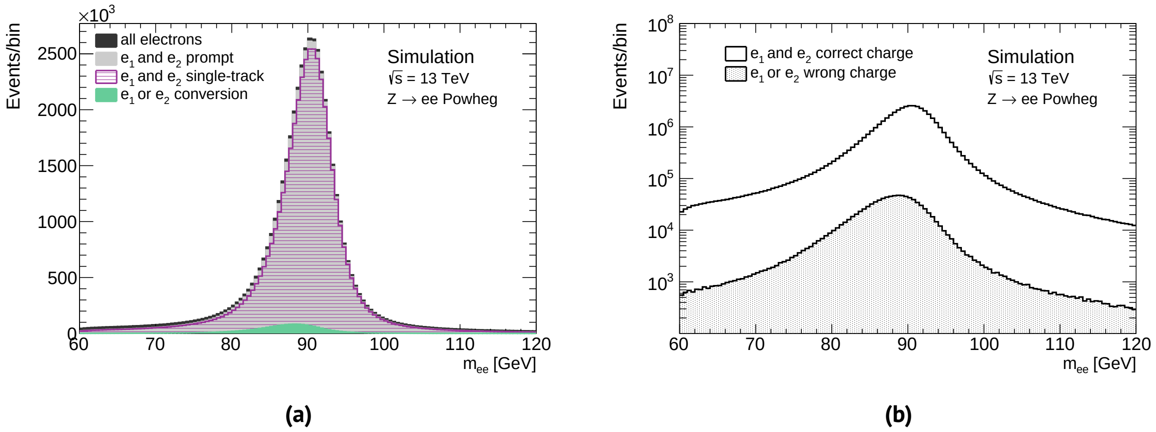


Figure 6.2: (a) Reconstructed invariant mass of the two electrons in a $Z \to ee$ simulated sample. The contribution from signal electrons from Z (single-track) or electrons from conversion of bremsstrahlung photons that stem from signal electrons (conversion) are shown superimposed. They form together the class of prompt electrons for which the charge misidentification measurement is meaningful, and they are the largest fraction of events in a simulated $Z \to ee$ sample. (b) Reconstructed invariant mass of prompt electrons in a simulated $Z \to ee$ sample split into events where both electrons are reconstructed with the correct charge and events where one of the two electrons or both are reconstructed with the wrong charge.

the *original* signal electron generated from the Z boson at generation-level. They are labelled as *conversion* in the following.

A $Z \to ee$ sample mostly produces prompt electrons, even without a specific kinematic selection. This can be seen in Fig. 6.2a, where the two components of the prompt definition, i.e. single-track and conversion electrons, are shown as well. The main contribution to the prompt electrons comes from the single-track ones, whilst a smaller fraction comes from conversions.
The fraction of *prompt* electrons in a $Z \to ee$ sample with wrongly identified charge is expected to be small. This is visible in Fig. 6.2b where the events with prompt electrons are split into events where the charge assignment is correct for both electrons and where one of the two or both are misreconstructed. For this, the reconstructed electron charges are compared to those of the respective *original* electrons, as previously defined.
In case of single-track electrons the misidentification happens in less than 1% of the cases, which is visible in Fig. 6.3a and 6.3b where all single-track electrons are shown together with the wrong track ones. The conversion case is examined in Fig. 6.3c and 6.3d. The wrong-charge assignment reaches up to 40-50% of all conversion electrons, since the probability of reconstructing the electron or the positron is the same. It is notable, that the wrong charge electrons populate large η ranges for both families, and their p_T distribution is shifted towards higher values. These dependencies will be further analysed in Section 6.3.
Photon conversion is thus the main contribution to the wrong charge assignment for prompt electrons in general, accounting for up to 80% of all cases of misidentification assignments. This is visible in Fig. 6.4, where the transverse momentum and the pseudorapidity distributions for wrong charge events are shown and the contributions of the single-track or conversion electrons are presented superimposed.

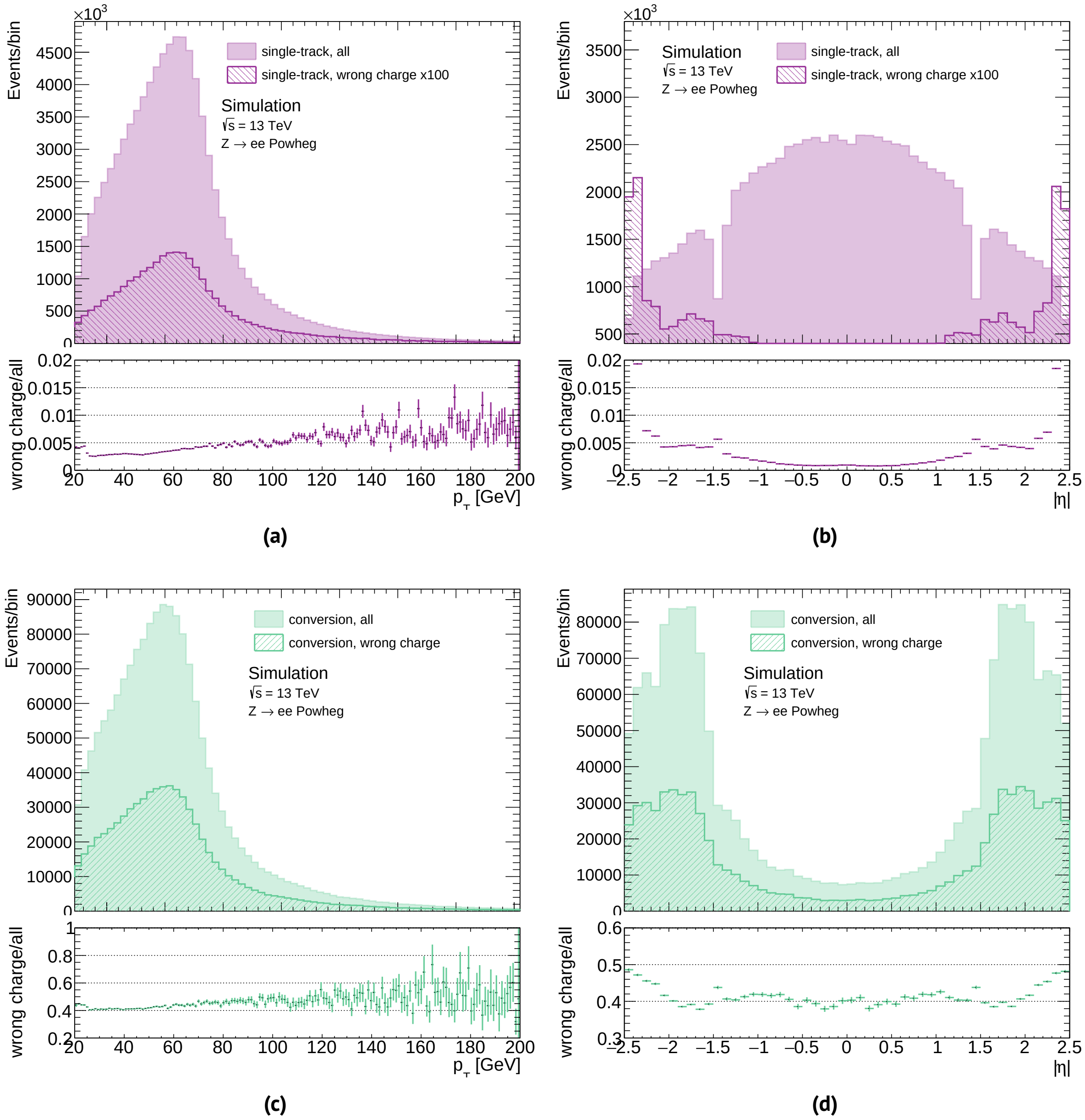
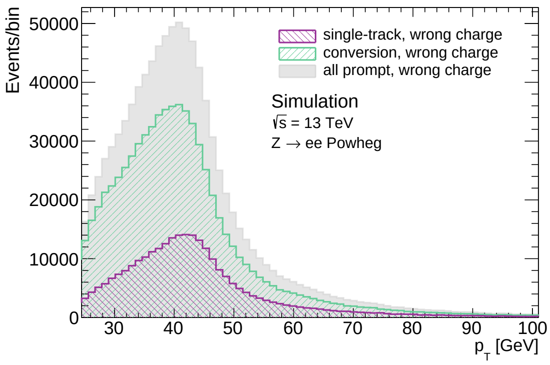
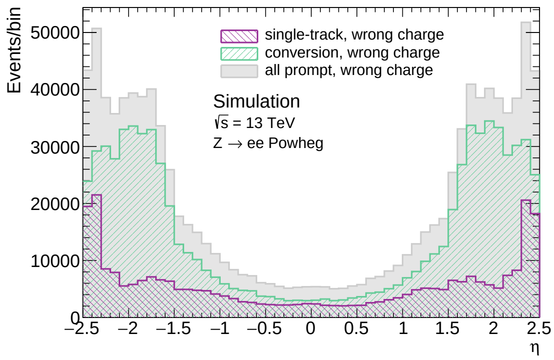


Figure 6.3: Reconstructed transverse momentum (a,c) and pseudorapidity (b,d) distributions for all single-track electrons (top) and electrons from conversion of bremsstrahlung photon (bottom), superimposed to the wrong charge cases. For the single-track cases the wrong charge distribution is scaled by a factor of 100 to be visible. The fraction of wrong charge electrons is shown for each case in the lower panel. Both electron candidates from $Z \to ee$ simulated process are used.



(a)



(b)

Figure 6.4: Reconstructed transverse momentum (a) and pseudorapidity (b) distributions for wrong charge electrons from simulated $Z \to ee$ events. The impact of the two sources is depicted superimposing the distributions of electrons from conversion processes and single-track from Z boson decays.

6.3 Properties of Charge Misidentification

It is useful to categorise the causes of charge misidentification into two main families:

- the matching of the *wrong track* to the electron candidate;
- the misreconstruction of the curvature of the *correct track*.

In this context the "correct" track refers to the one produced by the primary electron, as defined in Section 5.2.
The conversion of bremsstrahlung photons discussed in the previous section belongs to the first family, since at least three particles are present in close proximity, with the chance of picking a wrong one with the wrong charge. Early interactions of electrons in the beam pipe or the innermost pixel layer can result not only in one bremsstrahlung emission, but multiple ones. This is shown in a single-electron event display in Fig. 6.5. Beside the photon conversion, tracks with insufficient number of pixel hits can be lost and other tracks in the ID can be matched to energy deposits in the electromagnetic calorimeter. These cases belong to the wrong track-matching category as well.
The second family of charge misidentification sources refers to the wrong reconstruction of the curvature of the correct track. This is the cause of misidentification for the single-track electrons, i.e. for signal electrons with unambiguous track assignment. The misreconstruction of the curvature is primarily relevant for high p_{T} electrons due to their nearly straight tracks.
Properties of the different categories are studied in the $Z \to ee$ POWHEG sample, where unbiased electrons are obtained with a tag-and-probe method, which has been introduced in Section 5.2. It imposes very tight selection criteria on the tag electron which is used to select the events, in order to extract properties of the probe, which is thus not affected by selection biases. The Tight identification operating point is used for tag electrons, while probe electrons are selected with the Loose operating point. The counted tracks are the ones associated to the electron's cluster, while the kinematic properties are the ones of the reconstructed electrons.
Electrons with correct track assignment and correct charge are most likely to have one track candidate, as it is visible in Fig. 6.6a. Electrons with a misassigned track tend to have large track candidate multiplicity, in particular they have a most probable value of three tracks, regardless of whether the charge is correctly or wrongly reconstructed. In general wrong-track electrons with correct or incorrect charge reconstruction have very similar properties because the charge reconstruction is in effect random for them.
From Fig. 6.6b, where the pseudorapidity of the reconstructed electrons is shown, it is visible how the electrons with wrong charge assignment follow the material distribution inside the inner detector, described in Section 4.2.2. Indeed most of the electrons with the charge wrongly assigned are in the range $1.5 < |\eta| < 2.2$, corresponding to a region in the inner detector with relative large amount of material, as shown in Fig. 4.5. The effect of the charge misidentification related to the shape distortion of the track is highly visible in the η distribution, especially with the increase of passive material at larger $|\eta|$.
For electrons that are reconstructed with the correct track and charge assignment, the energy deposited in the cluster of the electromagnetic calorimeter mainly corresponds to the electron's

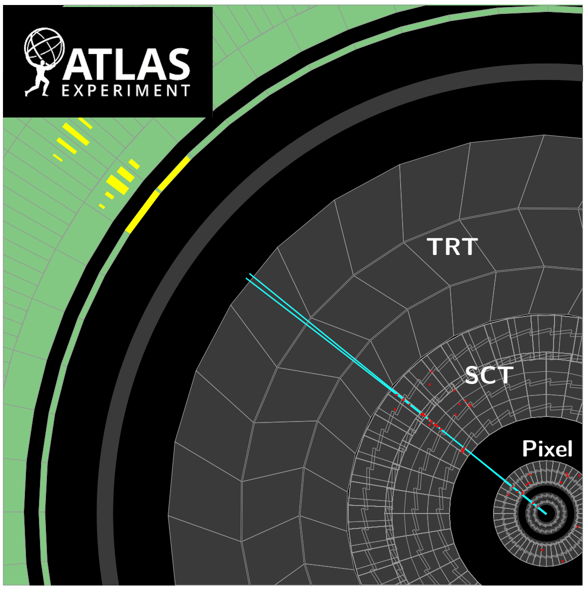


Figure 6.5: Event display of a simulated event, taken from a sample of single-electron events generated with flat η and p_{T} distribution in the 3.5 GeV $< p_{\mathrm{T}} <$ 100 GeV range, followed by a linear ramp down to 300 GeV and a flat distribution again up to 3 TeV. The electron has been associated with a wrong track, and hence electric charge. The reconstruction in the detector is depicted in the transverse view, with red hits visible in the pixel detector with the IBL included and in the SCT. The TRT hits are present but not shown. The electromagnetic calorimeter is represented in green. Two tracks (cyan, both approximately $E_{\mathrm{T}} =$ 20 GeV and of opposite charge) have been reconstructed from hits in the ID and have been associated to the energy cluster in the electromagnetic calorimeter (yellow) corresponding to E_{T}=40 GeV. In reality, neither track originates from the primary electron, since an early interaction in the beam pipe or innermost pixel layer has resulted in the creation of multiple electrons and photons producing hits used to reconstruct the tracks. Adapted from [6].

energy, except some late radiation. This is visible in Fig. 6.6c, where the distribution of the ratio between the cluster energy and the track momentum peaks at $E/p = 1$ for such electrons. For correct track electrons which have the wrong charge the largest probability is at E/p = 0, showing that the momentum of the track is much larger compared to the energy deposited in the cluster. This indicates that electrons with misreconstructed charge belonging to this type are mainly straight tracks with poorly measured momenta, since the relative momentum resolution scales with $1/p_{\mathrm{T}}$. The distribution for conversion electrons is broader, and in general there is a significant fraction of events with $E/p \neq 1$, due to a wrong track being matched to the candidate electron or a wrong track curvature being measured, that results in an erroneous measurement of the electron's momentum. The small peak visible in the distribution for wrong track and wrong charge electrons is also related to the imperfection in the truth classification mentioned above, which brings contamination from electrons of the correct-track family.

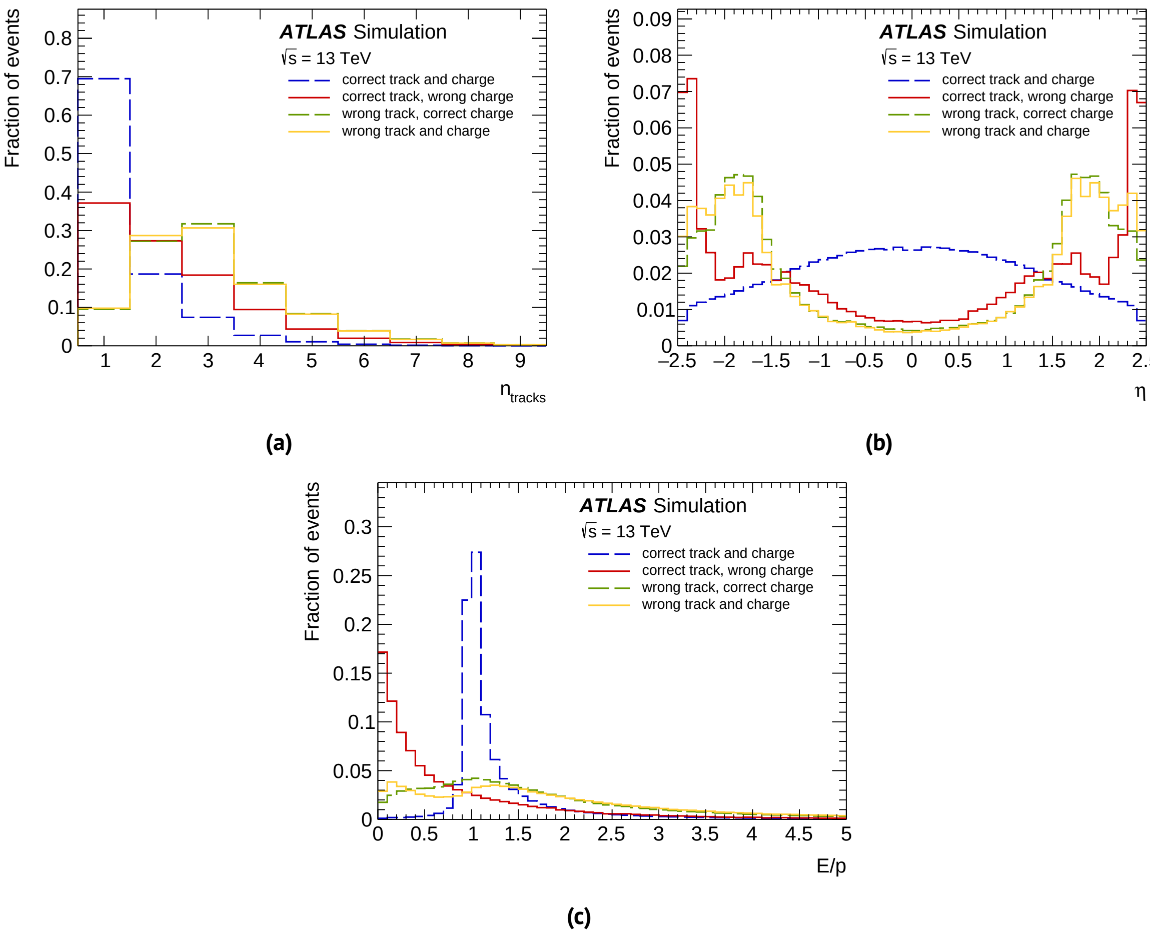


Figure 6.6: Kinematic and tracking properties of electrons from a simulated $Z \rightarrow ee$ sample. The multiplicity of track candidates associated to an electron's cluster (a), the pseudorapidity of the reconstructed electron (b), the energy-to-momentum ratio where the energy is the cluster one and the momentum is the track momentum (c) are shown. The distributions are shown for reconstructed electrons where the track is correctly reconstructed and the charge correctly assigned (blue), the track is correctly reconstructed but the charge assignment is wrong (red), the wrong tracks is picked but the charge is still correctly assigned (green), the wrong track is picked and the charge is wrongly assigned (yellow), all with respect to the primary electron. The distributions are normalised to the same area [6].

6.4 Modelling Correction Factors

The charge misidentification is mainly related to the interaction of the electrons with the material in the detector, whose description can be mismodelled in simulation. Therefore, correction factors are applied to simulation to be used for measurements sensitive to charge misidentification. These corrections are the scale factors and are built with the charge misidentification probabilities, measured in $Z \rightarrow ee$ simulated events and data, where the measurement can not distinguish among the classifications discussed in the previous section. Probabilities and correction factors are measured for five operating points. In particular the Tight and Medium identification operating points are combined with FixedCutTight and Gradient isolation ones. A further selection is applied in one case, which refers to a dedicated BDT [6] variable aiming at reducing the charge misidentification probability. The BDT is trained on variables related to the electron's track, electromagnetic shower

Table 6.1: Electron selection for one operating point used in the measurement of charge misidentification probabilities.

Electron selection	
Identification	Tight
Isolation	FixedCutTight
Trigger	Single Electron trigger
$p_T^{e_0,e_1}$	> 25 GeV
Cluster $\|\eta\|$	$<2.5, \notin [1.37, 1.52]$
$\|d_0/\sigma_{d_0}\|$	< 5
$\|z_0 \times \sin\theta\|$	< 0.5mm

Table 6.2: Binning and kinematic ranges for the measurement of charge misidentification probabilities.

Nominal measurement	
Cluster $\|\eta\|$ bin borders	[0.0, 0.60, 1.10, 1.37] and [1.52, 1.70, 2.30, 2.50]
p_T bin borders	[25, 60, 90, 130, 1000] GeV
Nominal Z central mass window	$75 < m_{ee} < 105$ GeV
Nominal sidebands	$55 < m_{ee} < 70$ GeV and $110 < m_{ee} < 125$ GeV

and combined tracking and calorimeter quantities, and is optimised aiming at the best rejection of wrong-charge electrons given an efficiency loss of 3% for correctly-reconstructed electrons. In the following, the procedure and results for the Tight identification and FixedCutTight isolation selection will be shown, and a comparison among three operating points will be given at the end.

6.4.1 Probabilities Measurement

The samples used for the measurement are described in Section 6.1. The kinematic requirements for the electrons are listed in Table 6.1 for the specific operating point presented in this chapter. The *single-lepton trigger* [93] requires each event to have at least one lepton passing the trigger thresholds, and ensure that at least one of the two signal leptons is matched to the trigger object.

In dielectron data events it is not possible to determine which of the two electrons underwent charge misidentification, hence charge misidentification probabilities are measured simultaneously in a *likelihood fit*. The use of this method in this case is preferable compared to a tag-and-probe procedure. In order for this method to be used for charge misidentification probabilities extraction, an assumption should be made that the probe electron is the one affected by charge misidentification due to the looser identification and isolation requirements. The use of a likelihood method instead, does not imply any assumption on which of the two electrons underwent charge misidentification, and requires to have the same identification and isolation criteria on the two electrons.

The construction of the likelihood function starts from the Ansatz that in the absence of background, given the probabilities ε_i and ε_j for a wrong charge reconstruction of each of the two electrons in a bin-pair i and j, the number of expected same-charge events $\mathrm{n}_{\mathrm{SC}}^{\mathrm{exp}}$ is given by

$$\mathrm{n}_{\mathrm{SC}}^{\mathrm{exp}} = \mathrm{n}p = \mathrm{n}\left[(1-\varepsilon_i)\,\varepsilon_j + \left(1-\varepsilon_j\right)\varepsilon_i\right]. \tag{6.1}$$

The total number of events n, counted in this case in a Z mass window in the selected bin−pair, is multiplied by the probability p to have exactly one of the two electrons reconstructed with the wrong charge, $p = (1-\varepsilon_i)\,\varepsilon_j + \left(1-\varepsilon_j\right)\varepsilon_i$. The probability to observe within the same mass range and in the same bin−pair a number n_{SC} of same-charge events can be expressed using binomial counting as

$$P(\mathrm{n}_{\mathrm{SC}}) = \binom{\mathrm{n}}{\mathrm{n}_{\mathrm{SC}}} p^{\mathrm{n}_{\mathrm{SC}}}\,(1-p)^{\mathrm{n}-\mathrm{n}_{\mathrm{SC}}}\,. \tag{6.2}$$

This can be approximated by a Poissonian distribution for large n and small p

$$P\left(\mathrm{n}_{\mathrm{SC}}|\varepsilon_i,\varepsilon_j\right) = \frac{(\mathrm{n}_{\mathrm{SC}}^{\mathrm{exp}})^{\mathrm{n}_{\mathrm{SC}}}\,e^{-\mathrm{n}_{\mathrm{SC}}^{\mathrm{exp}}}}{\mathrm{n}_{\mathrm{SC}}!}, \tag{6.3}$$

where n_{SC} is the measured number of same-charge events and the dependency on the charge misidentification probabilities becomes explicit in the expression of $\mathrm{n}_{\mathrm{SC}}^{\mathrm{exp}}$ in Eq. (6.1). The single-bin−pair likelihood function can now be defined and corresponds to

$$L(\varepsilon_i,\varepsilon_j) \equiv P\left(\mathrm{n}_{\mathrm{SC}}|\varepsilon_i,\varepsilon_j\right). \tag{6.4}$$

The charge misidentification probabilities are obtained by minimising the sum of the negative logarithm of the likelihood function over all bins

$$-\ln L = -\sum_{i,j} \ln L(\varepsilon_i,\varepsilon_j). \tag{6.5}$$

In this way, all the combinations (1,j) and (i,1) are used to determine the charge misidentification probability in bin 1, and similarly for the other bins.

The likelihood fit is performed double-differentially in p_{T} and absolute cluster pseudorapidity of the electrons. The function is built by counting events in a fixed range of 15 GeV around the Z boson mass, where a fraction of events with the same-charge is expected due to wrong charge assignment during the charge reconstruction. Electrons are counted in 24 bins of (p_{T}, $|\eta|$) to reflect the dependency of the charge misidentification probability on these two quantities, as seen in Section 6.3. The transition region of the calorimeter $1.37 < |\eta| < 1.52$ is excluded from the measurement[3]. The range and binning are summarised in Table 6.2.

The background contribution in data mainly comes from hadronic jets, Dalitz decays[4] of neutral pions and semi-leptonic decays of heavy-flavour hadrons. The number of same-charge events coming from these processes that enter the m_{ee} window used for the event counting is estimated

[3] The probabilities measured in the $|\eta|$ bin $1.10 < |\eta| < 1.37$ are also provided for electrons in the bin $1.37 < |\eta| < 1.52$. This because being the charge mainly determined with the information from the ID, the transition region of the electromagnetic calorimeter does not have an impact on the trend of the charge misidentification probability.

[4] Dalitz decays refer generally to all four body processes involving a massive particle decaying into one massless boson and two fermions. In this case it refers mainly to the process $\pi^0 \to \gamma e^+ e^-$.

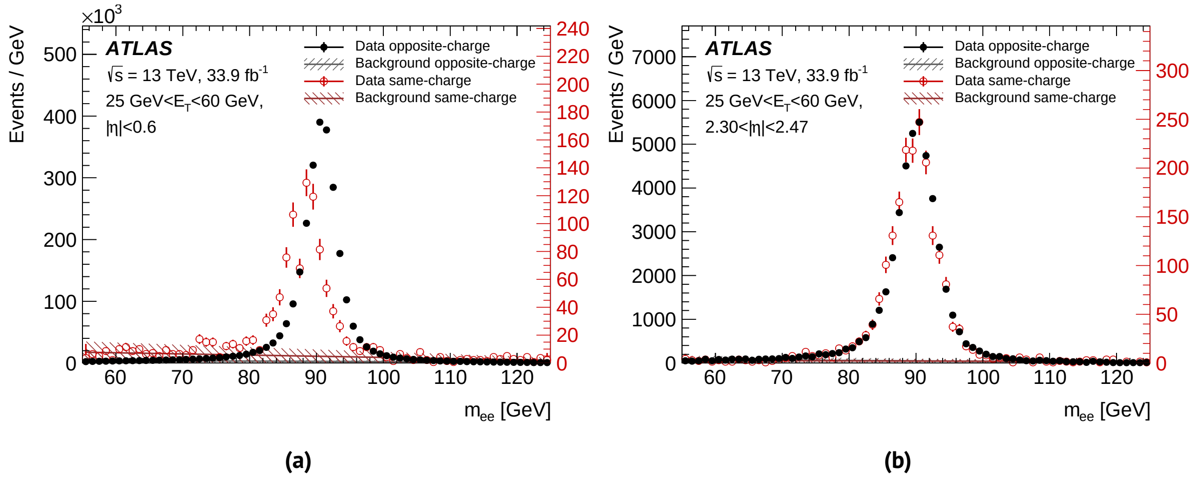


Figure 6.7: Distribution of the invariant mass of electron pairs with opposite charge (closed black circles) and same charge (open red circles) for the same p_T bin 25 GeV $< p_T <$ 60 GeV and for $0.0 < |\eta| < 0.6$ (a) and $2.30 < |\eta| < 2.47$ (b). The estimated number of events coming from background processes is depicted as a red (black) line for same-(opposite-)charge events. The hatched bands show the uncertainties assigned. Electrons are selected with the requirements detailed in Table 6.1 and both distributions are scaled to the same unit area [6].

with a *sideband subtraction method*. Low- and a high-m_{ee} sidebands are defined as detailed in Table 6.2 and the number of same-charge events $\mathrm{n}_{\mathrm{SC}}^{\mathrm{bkg,low/high}}$ is measured in each of them. The number of same-charge events in the central Z mass window coming from background processes $\mathrm{n}_{\mathrm{SC}}^{\mathrm{bkg}}$ is extrapolated assuming a linear background distribution

$$\mathrm{n}_{\mathrm{SC}}^{\mathrm{bkg}} = \frac{1}{2}\left(\mathrm{n}_{\mathrm{SC}}^{\mathrm{bkg,low}} + \mathrm{n}_{\mathrm{SC}}^{\mathrm{bkg,high}}\right).$$

This is taken into account in the likelihood function in the number of expected events $\mathrm{n}_{\mathrm{SC}}^{\mathrm{exp}}$. This number consists therefore of two contributions, i.e. the number of expected events from charge misidentification of electrons from $Z \to ee$ decays and the number of same-charge events coming from the background. Hence, the expression of $\mathrm{n}_{\mathrm{SC}}^{\mathrm{exp}}$ in the likelihood function is

$$\mathrm{n}_{\mathrm{SC}}^{\mathrm{exp}} = \mathrm{n}\left[(1-\varepsilon_i)\,\varepsilon_j + \left(1-\varepsilon_j\right)\varepsilon_i\right] + \mathrm{n}_{\mathrm{SC}}^{\mathrm{bkg}}. \tag{6.6}$$

The background contribution in opposite-charge events is much smaller than both the contribution from genuine electrons and the background yield in same-charge events, and is therefore neglected. This is shown in Fig. 6.7, where the dielectron invariant mass distribution m_{ee} is depicted for same-charge and opposite-charge data in the same p_T range for two $|\eta|$ bins separately. The opposite-charge background is shown but not accounted for in the measurement, and is obtained by applying the sideband subtraction method to all events, and subtracting the same-charge contribution. The same-charge background contribution used in the subtraction is included in the figure but very small as well, and is counted in the low- and high-m_{ee} sidebands and extrapolated linearly in the central mass region.

The same-charge background events considered for the subtraction in the central window do not

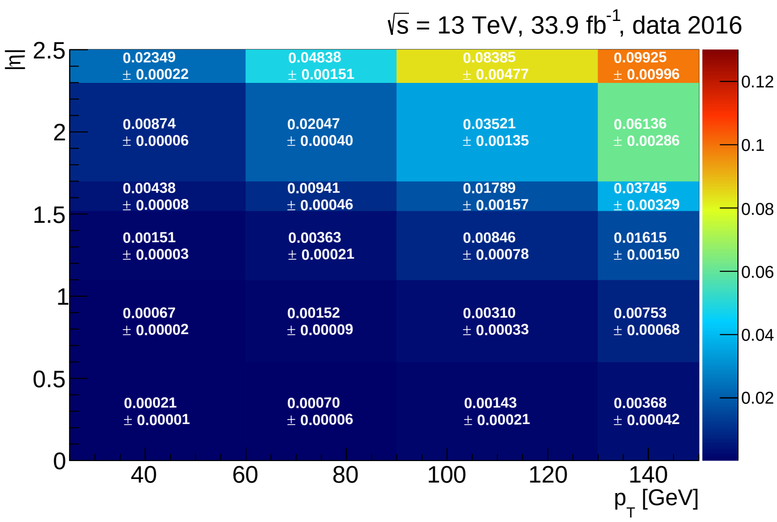
contribute to the probabilities measurement, as also visible from Eq. (6.6), since they are counted regardless of their same-charge being genuine or deriving from charge misidentification.
The resulting probabilities are shown in Fig. 6.8 as a two-dimensional (p_{T}, $|\eta|$) map, and as a function of p_{T} and $|\eta|$ separately for all bins.
The charge misidentification probabilities are shown for different operating points in Fig. 6.9 as a function of E_{T} and $|\eta|$. The Tight and Medium identification operating points are compared, combined with the FixedCutTight isolation operating point in both cases. One operating point includes a selection on the BDT variable. From the charge misidentification probabilities shown in the figure, the BDT requirement is proven to be effective in suppressing the charge misidentification.

6.4.2 Scale Factors Measurement

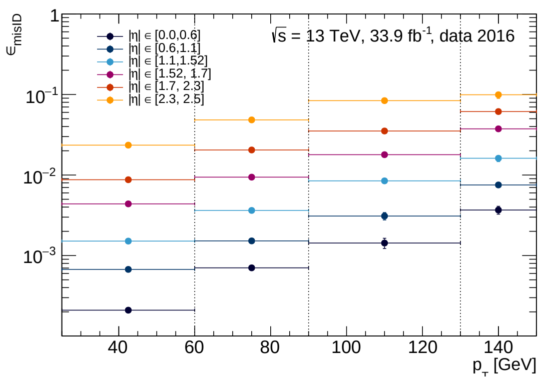
The charge misidentification probabilities measured both in data and simulation with the likelihood method typically agree within 10%. However differences can become as large as 30% for some kinematic regions. This is visible in Figs. 6.10 and 6.11, where they are shown for all the p_{T} and $|\eta|$ bins, together with their ratio.
In order to correct the efficiencies in the simulation, and hence make the probability of misidentification match the one observed in data, scale factors are extracted. These are computed differently depending on whether an electron has its charge wrongly reconstructed or not. The scale factors for wrong-charge and right-charge electrons are given by

$$\mathrm{SF}_{\mathrm{wrong}} = \frac{\varepsilon_{\mathrm{data}}}{\varepsilon_{\mathrm{MC}}} \qquad \mathrm{SF}_{\mathrm{right}} = \frac{1-\varepsilon_{\mathrm{data}}}{1-\ \varepsilon_{\mathrm{MC}}}, \tag{6.7}$$

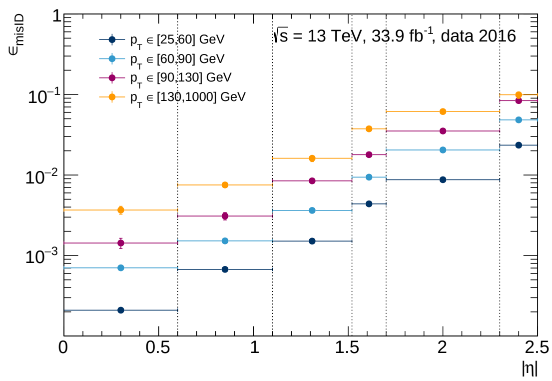
and the ones for wrongly reconstructed electrons are shown in Fig. 6.12, together with the respective statistical uncertainties. The scale factors for wrongly reconstructed electrons reach a deviation from unity of up to 30%, while the corrections to right charge electrons never exceed 0.5% and are shown in Fig. C.5 in Appendix C. By weighting both correct- and wrong-charge electrons the normalisation of the sample remains invariant.



(a)



(b)



(c)

Figure 6.8: Charge misidentification probabilities as a two-dimensional map in (p_T, $|\eta|$) (a) and as a function of the electron p_T (b) and $|\eta|$ (c). Electrons are selected with the Tight identification and FixedCutTight isolation operating point. Only statistical uncertainties are shown.

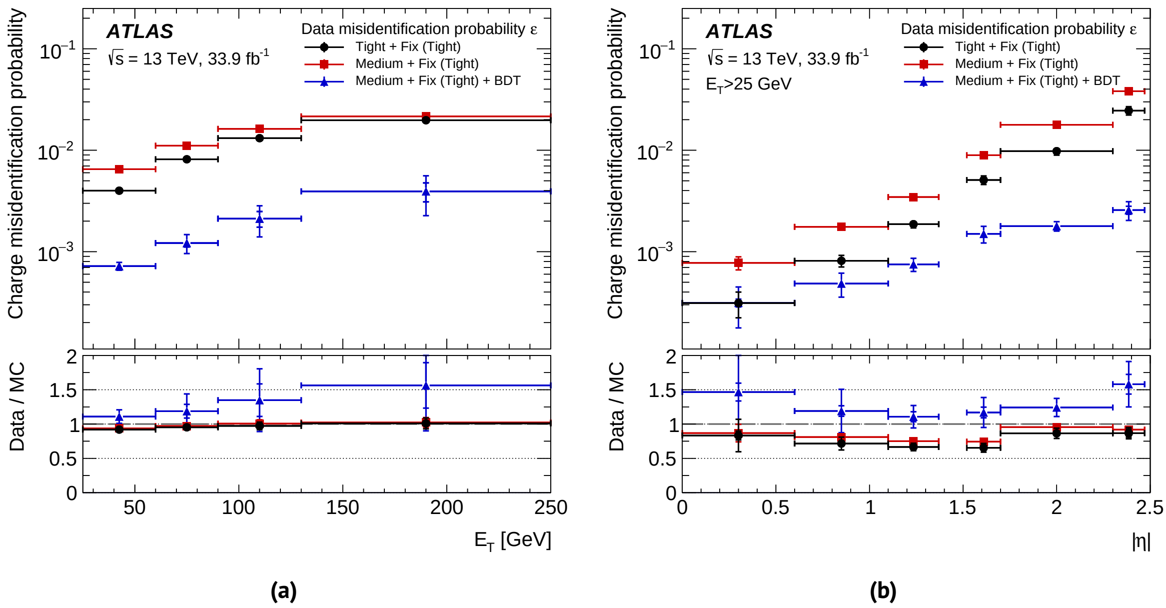


Figure 6.9: Charge misidentification probabilities as a function of E_T (a) and $|\eta|$ (b) for different operating points. Two identification operating points, Tight and Medium are compared, while an additional requirement is imposed for the third point, on a BDT dedicated to the charge misidentification suppression. The FixedCutTight isolation is the same in all three cases. The inner uncertainties are statistical while the total include statistical and sytematic uncertainties [6].

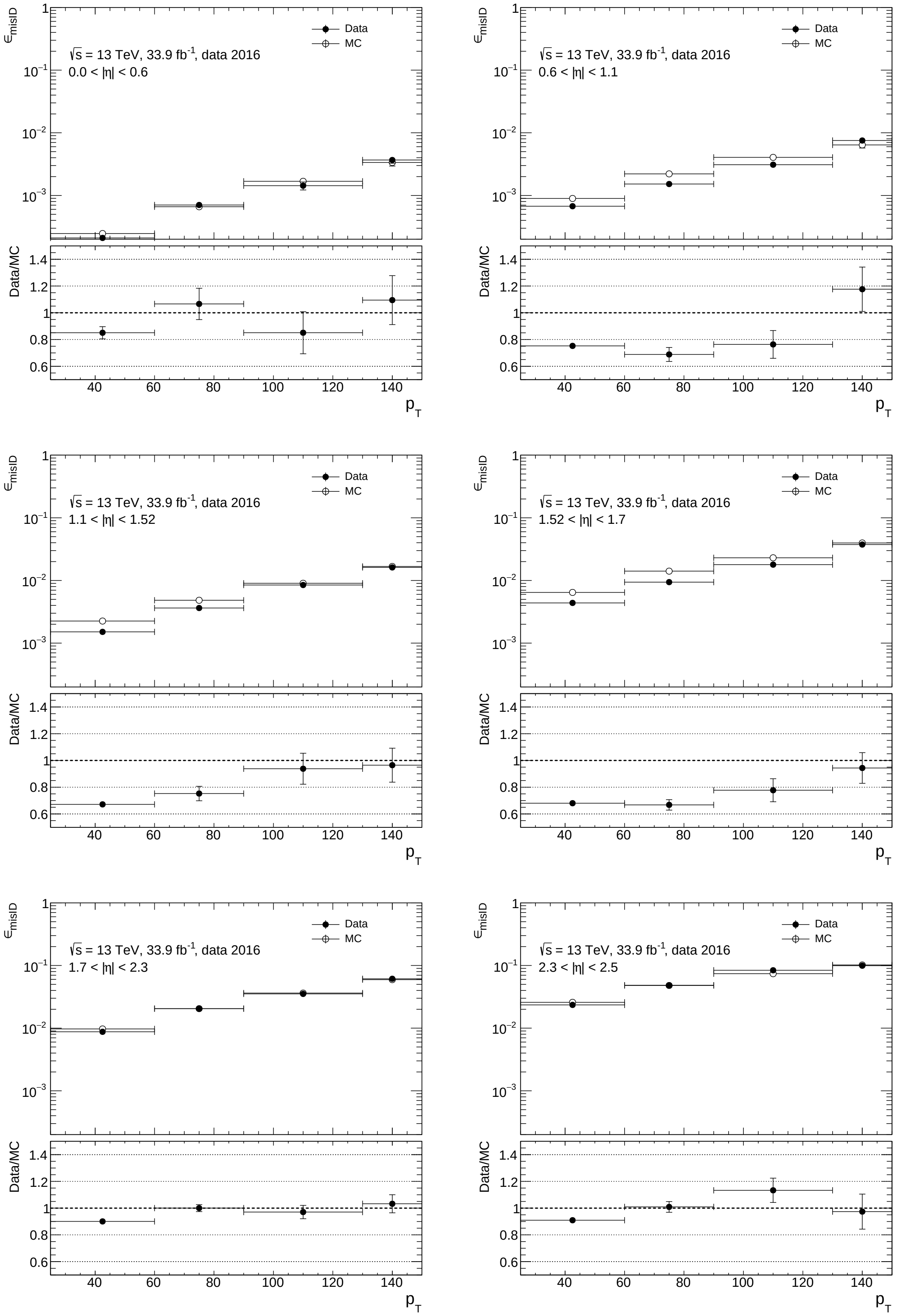


Figure 6.10: Charge misidentification probabilities for data and simulation as a function of p_{T} for the six $|\eta|$ bins (top panels). The ratio $\varepsilon_{\mathrm{data}}/\varepsilon_{\mathrm{MC}}$ is shown (bottom panels). Electrons are selected with the Tight identification and FixedCutTight isolation. Only statistical uncertainties are shown.

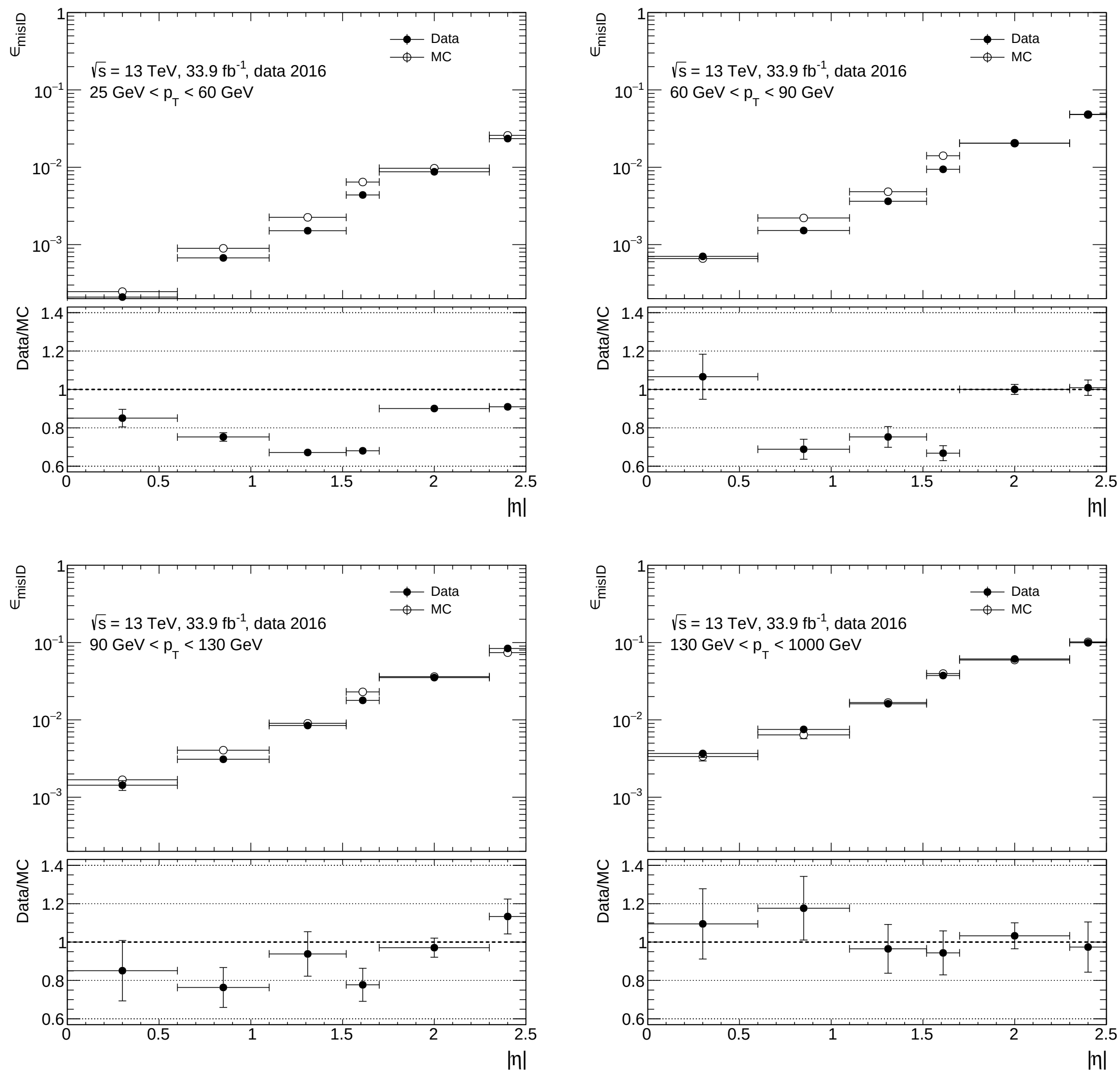


Figure 6.11: Charge misidentification probabilities for data and simulation as a function of $|\eta|$ for the four p_{T} bins (top panels). The ratio $\varepsilon_{\text{data}}/\varepsilon_{\text{MC}}$ is shown (bottom panels). Electrons are selected with the Tight identification and FixedCutTight isolation. Only statistical uncertainties are shown.

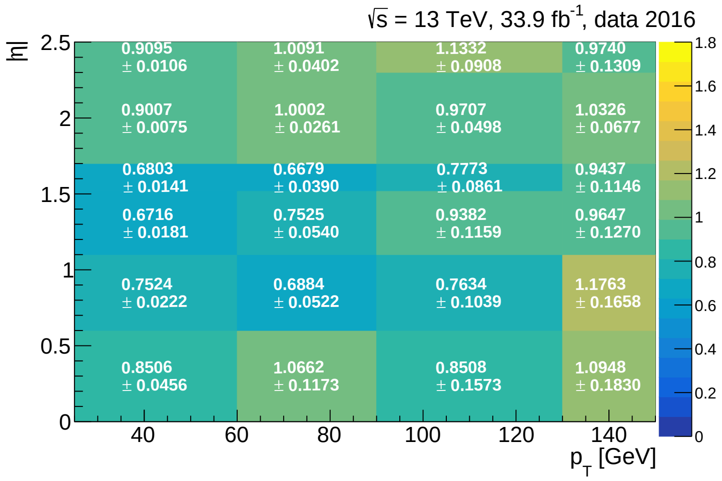


Figure 6.12: Charge misidentification scale factors for wrongly reconstructed electrons for electrons selected with the Tight identification and FixedCutTight isolation operating point. Only statistical uncertainties are shown.

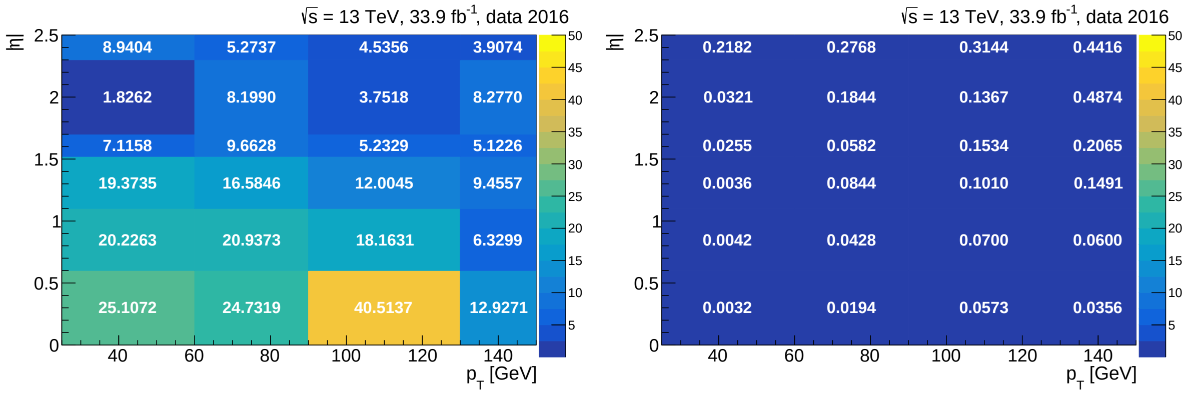
6.4.3 Systematic Uncertainties

The systematic uncertainties on the scale factors are derived from four variations in the procedure of the probabilities measurement. Two variations concern both the measurement in data and simulation, whilst two affect either simulation only or data only. The differences between the varied scale factors and the nominal one are symmetrised and considered as systematic variations which are combined in quadrature to obtain the total systematic uncertainty. For the two cases where the variation affects data and simulation, the probabilities are varied coherently and only the effect on the scale factors is studied, i.e. effects that are well modelled are allowed to cancel in the ratio.

The systematic uncertainties associated with the choice of the Z boson mass window used to count same- and opposite-charge events are evaluated both for data and simulation. This is done by increasing and decreasing the range, measuring two sets of probabilities in the 70 GeV $< m_{ee} <$ 110 GeV and 80 GeV $< m_{ee} <$ 100 GeV windows, and computing two sets of scale factors. The sidebands maintain the definition in Table 6.2. The maximum variation between the nominal scale factor in each bin and the one obtained with a different m_{ee} window is assigned as systematic uncertainty. This uncertainty targets the dependency of the scale factors on the dielectron invariant mass, and it accounts for the shift of the same-charge distribution towards lower m_{ee} values, which can bias the probabilities measurement.

A further source of systematic uncertainty is the trigger selection on the electrons, which impact probabilities both in data and simulation. Differences can exist in scale factors for electrons that were selected by the trigger and those that were not. The nominal measurement requires both of them to pass the trigger selection. This systematic uncertainty source is estimated by obtaining a second set of scale factors with data and simulation probabilities where only one electron is required to fire the trigger.

A *closure* test is performed, using the true electron charge. The charge misidentification probabilities



(a) **(b)**

Figure 6.13: Relative total systematic uncertainties in percent on scale factors for wrongly (a) and correctly (b) reconstructed electrons, selected with the Tight identification and FixedCutTight isolation operating point.

in simulation can be obtained as the ratio of the number of wrong charge electrons to all electrons in a certain (p_{T}, $|\eta|$) bin i

$$\varepsilon_i^{\text{true}} = \frac{(\text{wrong charge electrons})_i}{(\text{all prompt electrons})_i}, \tag{6.8}$$

where *all prompt electrons* refers to all the electrons that satisfy the prompt criteria from Section 6.2. A set of scale factors is obtained using the nominal data probabilities and the ones from simulation obtained as in Eq. (6.8). The difference between the nominal set of scale factors and the new one is considered as the third source of systematic uncertainty.

The full size of the estimated background is assigned as a systematic uncertainty to account for the shortcomings of the simple sideband subtraction.

The four uncertainties are added in quadrature to obtain the total systematic uncertainties for both electrons with correctly and wrongly reconstructed charge. The relative values are shown for the two cases in Fig. 6.13, while Fig. 6.14 shows the four different contributions individually in absolute values for wrongly reconstructed electrons. The largest contribution to the uncertainty comes from the trigger selection which can result in a 20% variation of the nominal value for certain bins, in particular for low $|\eta|$ values and high p_{T}. The variation of the m_{ee} window is the smallest contribution to the total uncertainty. Systematic uncertainties for correctly reconstructed electrons are below one percent, while for misreconstructed electrons span a wide range that can reach 15% or more in certain bins.

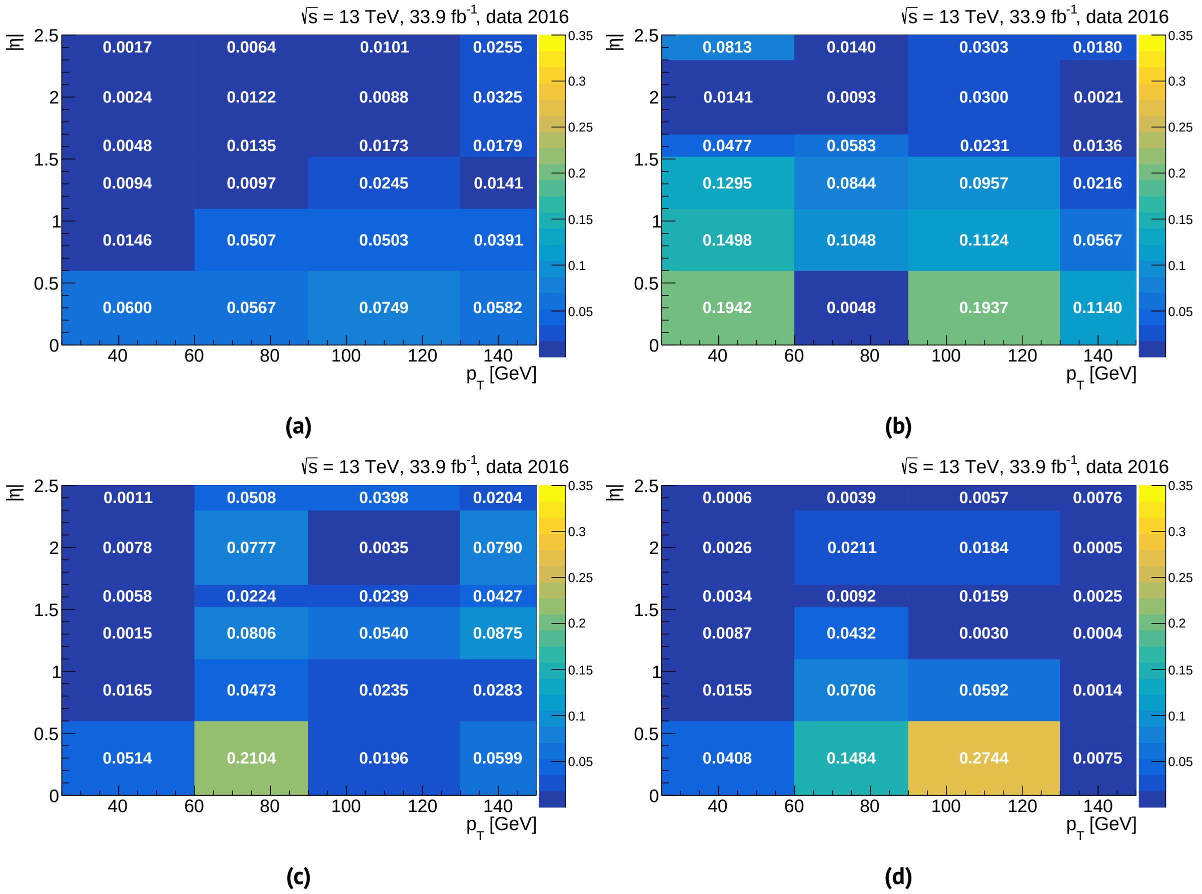


Figure 6.14: Systematic uncertainties in absolute values for scale factors for wrongly reconstructed electrons. The uncertainties are shown individually for the Z mass window variation (a), the trigger selection (b), the truth closure (c) and the background subtraction (d). Electrons are selected with the Tight and FixedCutTight isolation operating point.

6.4.4 Results For the $W^{\pm}W^{\pm}jj$ Analysis Selection

The selection criteria applied by the $W^{\pm}W^{\pm}jj$ analysis, that will be the object of Ch. 7, 8 and 9, slightly differs from the operating points for which the scale factors are provided. Electrons in the analysis are selected with the provided Tight and Gradient isolation operating point, but the two the main differences lay in the p_{T} threshold, which is lowered to 20 GeV due to optimisation and background studies, and an additional requirement is imposed to solve the ambiguity between the electron-photon identification (see Section 7.3). The set of dedicated scale factors is derived on $Z \to ee$ data and simulation events with the samples listed in Section 6.1. The measurement is performed in five variable-sized bins of p_{T} with lower edges [20, 27, 60, 90, 130] GeV up to 1000 GeV, and six $|\eta|$ bins with edges [0.0, 0.6, 1.1, 1.37] and [1.52, 1.7, 2.3, 2.5]. The calorimeter transition region $1.37 < |\eta| < 1.52$ is excluded from the measurement. Corrections are in general similar to the ones for the operating point presented in this chapter and shown in Fig. 6.12, and the ones for wrongly-reconstructed electrons vary from unity by 20%-30%, whilst the corrections for correctly-reconstructed electrons are at per mil level. The scale factors and their absolute systematic uncertainties are shown in Fig. 6.15 for wrongly-reconstructed electrons, while results

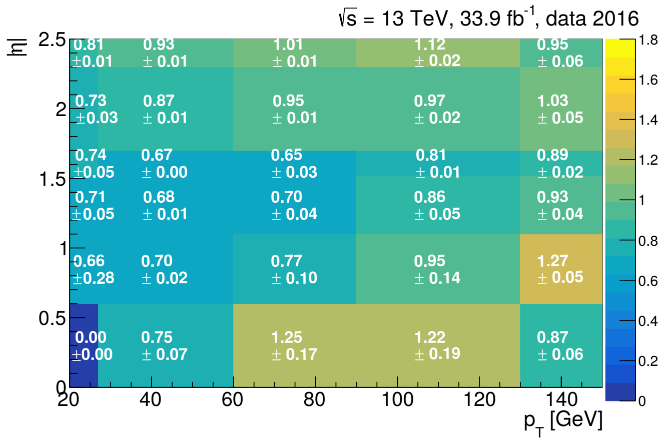


Figure 6.15: Charge misidentification scale factors with absolute systematic uncertainties for wrongly reconstructed electrons. Electrons are selected with the Tight identification and Gradient isolation operating point.

from correctly-reconstructed electrons are reported in Fig. C.6 of Appendix C.

7 Electroweak Signal Topology and Optimisation

The observation and measurement of the cross section of the electroweak-mediated $W^\pm W^\pm jj$ process is performed using data collected by the ATLAS detector at a centre-of-mass energy of 13 TeV during the LHC running period 2015-2016, which corresponds to an integrated luminosity of 36.1 fb^{-1}. This chapter and the following Chapter 8 and Chapter 9 are dedicated to this analysis. The signal definition and simulation is discussed in Section 7.1 and the other processes that enter the signal region but constitute the background to the signal are the object of Section 7.2, together with their event simulation. The selection criteria used to identify candidate events are detailed in Section 7.3. The object definition and event selection are the result of a dedicated optimisation procedure that aimed at increasing the significance of the process. This is presented in details in Section 7.4. When needed, references can be made to Chapter 8 for the estimation techniques of the background processes, and to Chapter 9 for the final fit and measurement.

7.1 Signal

7.1.1 Definition and Topology

Due to the complexity of the definition of the pure VBS contribution to the $pp \to W^\pm W^\pm jj \to \ell^\pm \nu \ell^\pm \nu jj$ process, the topology and modelling of the signal deserve special attention. An overview of the features of VBS processes has been given in Section 2.4.1. As discussed there, it is not possible to separate the genuine VBS contribution to the process, therefore the *signal* is defined in the analysis as all processes at $\mathcal{O}(\alpha_{\mathrm{EW}}^6)$ that contribute to the final state. The interference with the QCD production, discussed in Section 2.4.1 as well, can not be separated experimentally, and will be included in the measured cross section.

The distinctive feature of the $W^\pm W^\pm jj$ production via VBS is the presence of two highly energetic forward jets, the *tagging* jets, one in each hemisphere of the detector. The two W bosons are emitted from an initial quark, and the resulting quarks tend to be at very high absolute values of rapidity and to have a large momentum. This leads to a typically very large invariant mass and large rapidity separation between the two jets. The two same-charge leptons originating from the decay of the W bosons lay generally between the tagging jets. In this analysis only W decays to electrons and/or muons are considered, and this notation will be used to indicate electrons or positrons, as well as muons and antimuons. The characteristic topology of such VBS $W^\pm W^\pm jj$ events, depicted in Fig. 7.1, is exploited in the kinematic selection to distinguish the signal[1] events from other processes with the same final state. In particular, m_{jj} and Δy_{jj} have a high discriminating power between the electroweak and QCD production, as can be seen in Fig. 7.2, where the two distributions are shown fro the EW and QCD samples. A display of a $pp \to W^+W^+jj \to e^+\nu\mu^+\nu jj$ candidate event recorded by the ATLAS detector is shown in Fig. 7.3.

[1] Although the specific VBS topology is exploited, the signal definition will always indicate all VBS and non-VBS $\mathcal{O}(\alpha_{\mathrm{EW}}^6)$ contributions throughout this thesis, and the terms *signal* or *electroweak signal* will be used interchangeably.

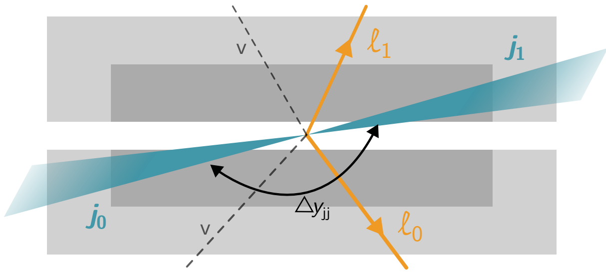


Figure 7.1: Topology of a typical $W^{\pm}W^{\pm}jj$ VBS event. The two tagging jets $j_{0,1}$ have a large rapidity separation Δy_{jj} and the products of the W boson decay, the same-charge leptons $\ell_{0/1}$ and neutrinos ν lie between the two jets.

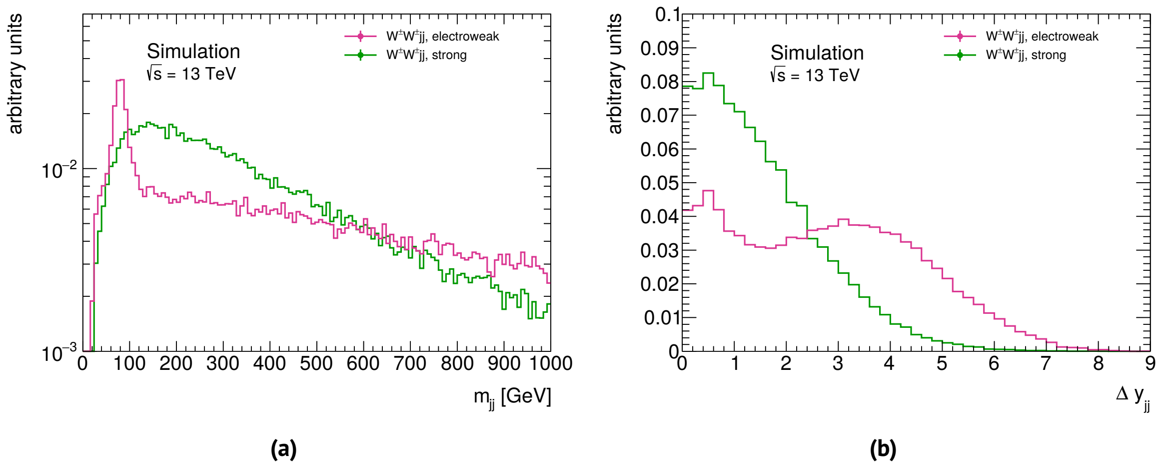


Figure 7.2: Distribution of the dijet invariant mass (a) and rapidity separation between the two jets of highest p_T (b) for the EW and QCD $W^{\pm}W^{\pm}jj$ processes, simulated separately with SHERPA 2.2.2 generator.

7.1.2 Simulation

The electroweak production of $W^{\pm}W^{\pm}jj$ is simulated using the SHERPA 2.2.2 generator [55,131–133], which uses an internal parton shower algorithm. It is generated by fixing $\mathcal{O}(\alpha_{EW}^{6})$ in the $pp \to W^{\pm}W^{\pm}jj \to \ell^{\pm}\nu\ell^{\pm}\nu jj$ process. The nominal PDF set is the NNPDF3.0 implemented at NNLO in QCD. The $W^{\pm}W^{\pm}jj$ process is generated with up to one additional parton at LO in QCD in the matrix element, and multi-leg processes are simulated combining the final-state topologies with the matrix-element parton shower merging algorithm [133, 134] and are matched to the SHERPA internal parton shower [132, 135].
Due to the discrepancies in the modelling of the $W^{\pm}W^{\pm}jj$ electroweak process among different event generators, deeper comparison studies are conducted, aiming at investigating the nature of the tensions.

Alternative signal samples

The modelling of the signal is studied for different event generators, which provide diverse descriptions of the $W^{\pm}W^{\pm}jj$ electroweak process. Two samples used for the comparison with the nominal SHERPA sample described above are generated with MADRAPH5 2.3.3 [136] and one with POWHEG-BOX

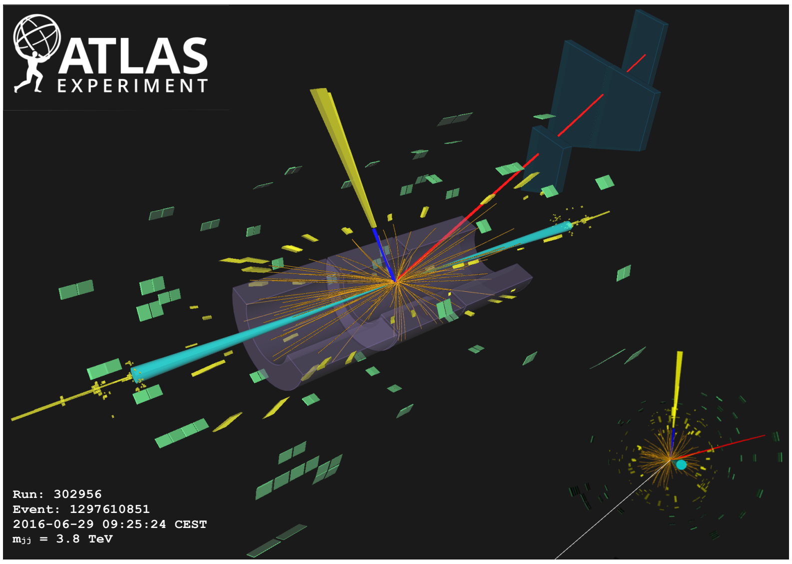


Figure 7.3: Display of a $pp \to W^+W^+jj \to e^+\nu\mu^+\nu jj$ candidate event at $\sqrt{s} = 13$ TeV recorder by the ATLAS detector. The red line represents the muon, while the blue line represents the electron traversing the inner detector. Its energy depositions inside the electromagnetic calorimeter are indicated by the yellow cone. The two jets are represented by cyan cones with yellow energy depositions in the forward calorimeters. The $E_{\mathrm{T}}^{\mathrm{miss}}$ is depicted in the azimuthal view by a dashed grey line. The electron and muon have transverse momenta of 48 GeV and 38 GeV, respectively. The jets have transverse momenta of 118 GeV and 104 GeV, are separated by Δy_{jj} = 7.1 and are reconstructed with an invariant mass of m_{jj} = 3.8 TeV. Adapted from Ref. [5].

2 [121–124] interfaced to the PYTHIA 8.230 algorithm [99] that uses the VBS approximation [31].

The MADRAPH5 samples are generated at LO QCD up to two jets, with the third jet modelled by the parton shower, while SHERPA is generated al LO QCD in the second and third jet. The POWHEG sample is the only one generated at NLO in QCD. However, the POWHEG-BOX generator is missing Feynman diagrams for the s-channel triboson production, and their interference with the other Feynman diagrams. Despite the contribution of these diagrams being below five percent at LO QCD in the region $m_{jj} > 200$ GeV, when moving to NLO it increases to 40% in the region $200 < m_{jj} < 400$ GeV [37], hence becoming non-negligible.

A major difference among generators lays in the choice of the factorisation and renormalisation scales μ_{F} and μ_{R}. One MADRAPH5 sample employs the sum of transverse masses of all final state particles of the matrix element, which is a dynamical event-by-event scale. The other one uses a scale fixed to the mass of the W boson, which is the same as used by POWHEG. The SHERPA generator employs a dynamical scale set at the true diboson invariant mass.

A quantitative comparison in the kinematic region of interest for the analysis is presented in Section 9.2.1.

The SHERPA signal sample is adopted by the analysis as the nominal one, while the one produced with POWHEG is used for systematic studies that will be detailed in Section 9.7.

7.2 Backgrounds

There are other SM processes that can present the same final state as the $pp \to W^{\pm}W^{\pm}jj \to \ell^{\pm}\nu\ell^{\pm}\nu jj$ process, or can mimic it, passing therefore the signal selection. Nevertheless they have a different topology, which can be exploited in order to separate them from the $W^{\pm}W^{\pm}jj$ electroweak process. The impact of these processes is reduced with dedicated requirements in the event selection, and ultimately their residual contribution in the signal region is estimated.
They can be divided into three categories:

- processes that produce genuine same-charge leptons, such as ZZ+jets, WZ+jets, $t\bar{t}$+W/Z and triboson;
- processes that produce two opposite-charge leptons, where the charge of one lepton is wrongly reconstructed. They constitute the charge misidentification background and are mainly caused by $t\bar{t}$, $W^{\pm}W^{\mp}$+jets and Z/γ^{*}+jets production;
- processes with *non-prompt* leptons in the final state, i.e. jets or photons misreconstructed as leptons, or leptons from decays of hadrons containing a b or c quark, so called heavy flavour hadrons decays. The main background processes sources of non-prompt leptons are W+jets, $t\bar{t}$, $V\gamma$+jets, single top or QCD multijet processes.

The estimation of background processes will be the subject of Chapter 8, where the techniques based on data and simulation will be discussed. In the following the event simulation for these processes is described.

7.2.1 Event Simulation

For the $W^{\pm}W^{\pm}jj$ QCD sample SHERPA 2.2.2 is used, fixing the α_{EW} to $\mathcal{O}(\alpha_{\text{EW}}^{4})$, while the order of α_{s} is not fixed but determined automatically by SHERPA from the matrix-element, based on the number of jets in the final state. The PDF set is the NNPDF3.0 [137] implemented at NNLO in QCD with $\alpha_{\text{s}} = 0.118$.
The semi-leptonic diboson processes are generated with SHERPA 2.2.1, while SHERPA 2.2.2 is used for the diboson processes with four charged leptons in the final state, three leptons and one neutrino or two charged leptons and two neutrinos, with the matrix element containing all diagrams with four electroweak vertices.
The triboson processes are generated with SHERPA 2.1.1 with the CT10NLO PDF set [126].
The Z/γ^{*}+jets samples are generated with MG5_aMC@NLO 2.2.2 [136] at LO, while SHERPA 2.2.1 is used for the W+jets process. Both are interfaced to PYTHIA 8.186 for the parton shower and are normalised to NNLO QCD cross section calculations.
The $t\bar{t}$+W/Z samples are generated at NLO accuracy with MG5_aMC@NLO with the parton shower performed by PYTHIA 8.210, and use the NNPDF3.0NLO PDF set [137].
The generator POWHEG-BOX V2 is used for the $t\bar{t}$ production, with the CT10 PDF set, while the electroweak t-channel, s-channel and Wt-channel of single-top processes are generated with

POWHEG-BOX V1 which uses the 4-flavour scheme, with CT10f4 PDF set. In all these cases the parton shower, hadronisation and underlying event are simulated with PYTHIA 6.428 [53].
The $V\gamma$+jets process is generated with SHERPA 2.1.1 with exactly three electroweak vertices at NLO, up to one parton, or LO, up to three partons, using the CT10NLO PDF set [126]. The electroweak process is generated with SHERPA 2.2.4 with NNPDF3.0NNLO with five orders of the electroweak coupling at LO QCD accuracy.
The interaction with the detector of the generated particles is performed by a full detector simulation [63] based on GEANT4 [64], and the reconstruction algorithms are the same used for data and explained in Chapter 5. The effects of multiple pp interactions are accounted for by overlaying the collisions with soft QCD processes generated with PYTHIA 8.186 [125] using the A2 tune and the MSTW2008LO PDF set [103].
A summary of the processes and their simulation details is given in Table 7.1.

Table 7.1: Simulated event samples used for signal and background processes. The category, the cross section times branching ratio $\sigma \cdot \mathcal{B}$, the information on generators, parton shower algorithm and PDF set used are listed.

Process	$\sigma \cdot \mathcal{B}$ [nb]	Generator+Parton shower	PDF set
Signal			
$WWjj \to \ell\nu\ell\nu jj$ EW	$2.5{\cdot}10^{-5}$	SHERPA 2.2.2	NNPDF3.0NNLO
QCD background			
$WWjj \to \ell\nu\ell\nu jj$ QCD	$4.08{\cdot}10^{-5}$	SHERPA 2.2.2	NNPDF3.0NNLO
Diboson semi-leptonic			
$WW \to \ell\nu jj$	$1.38{\cdot}10^{-1}$	SHERPA 2.2.1	NNPDF3.0NNLO
$WZ \to \ell\nu jj$	$1.14{\cdot}10^{-2}$	SHERPA 2.2.1	NNPDF3.0NNLO
$WZ \to jj\ell\ell$	$3.43{\cdot}10^{-3}$	SHERPA 2.2.1	NNPDF3.0NNLO
$WZ \to jj\nu\nu$	$6.79{\cdot}10^{-3}$	SHERPA 2.2.1	NNPDF3.0NNLO
$ZZ \to \ell\ell jj$	$1.56{\cdot}10^{-2}$	SHERPA 2.2.1	NNPDF3.0NNLO
$ZZ \to jj\nu\nu$	$1.56{\cdot}10^{-2}$	SHERPA 2.2.1	NNPDF3.0NNLO
Diboson leptonic			
$WW \to \ell\nu\ell\nu$	$1.25{\cdot}10^{-2}$	SHERPA 2.2.2	NNPDF3.0NNLO
$gg \to WW \to \ell\nu\ell\nu$	$8.54{\cdot}10^{-4}$	SHERPA	CT10
$WWjj \to \ell\ell\nu\nu jj$	$1.16{\cdot}10^{-4}$	SHERPA 2.2.2	NNPDF3.0NNLO
$WZjj \to \ell\nu\ell\ell jj$ EW	$4.67{\cdot}10^{-5}$	SHERPA 2.2.2	NNPDF3.0NNLO
$WZ \to \ell\nu\ell\ell$ QCD	$4.58{\cdot}10^{-3}$	SHERPA 2.2.2	NNPDF3.0NNLO
$ZZ \to 4\ell$	$1.25{\cdot}10^{-3}$	SHERPA 2.2.2	NNPDF3.0NNLO
$gg \to ZZ \to 4\ell$	$2.07{\cdot}10^{-5}$	SHERPA	CT10
$ZZjj \to 4\ell jj$	$1.05{\cdot}10^{-5}$	SHERPA 2.2.2	NNPDF3.0NNLO
Triboson			
$WWW \to 3\ell 3\nu$	$8.34{\cdot}10^{-6}$	SHERPA 2.1.1	CT10
$WWZ \to 4\ell 2\nu$	$1.73{\cdot}10^{-6}$	SHERPA 2.1.1	CT10
$WWZ \to 2\ell 4\nu$	$3.43{\cdot}10^{-6}$	SHERPA 2.1.1	CT10
$WZZ \to 5\ell\nu$	$2.18{\cdot}10^{-7}$	SHERPA 2.1.1	CT10
$WZZ \to 3\ell 3\nu$	$1.92{\cdot}10^{-6}$	SHERPA 2.1.1	CT10
$ZZZ \to 6\ell$	$1.71{\cdot}10^{-8}$	SHERPA 2.1.1	CT10
$ZZZ \to 4\ell 2\nu$	$4.41{\cdot}10^{-7}$	SHERPA 2.1.1	CT10
$ZZZ \to 2\ell 4\nu$	$4.45{\cdot}10^{-7}$	SHERPA 2.1.1	CT10
Single boson			
$Z \to \ell\ell$	$2.66{\cdot}10^{1}$	MADRAPH5 + PYTHIA 8	N30NLO and A14NNPDF23LO (low mass)
$W \to \ell\nu$	$1.85{\cdot}10^{2}$	SHERPA 2.2.1	NNPDF3.0NNLO
Top quark			
$t\bar{t}W$	$5.48{\cdot}10^{-4}$	MG5_aMC@NLO + PYTHIA 8	MEN30NLO
$t\bar{t}Z(\to \ell\ell)$	$1.10{\cdot}10^{-4}$	MG5_aMC@NLO + PYTHIA 8	MEN30NLO
$t\bar{t}$	$7.69{\cdot}10^{-2}$	POWHEG-BOX 2	CT10
Single top s-channel	$3.31{\cdot}10^{-3}$	POWHEG-BOX 1	CT10f4
Single top t-channel	$6.95{\cdot}10^{-2}$	POWHEG-BOX 1	CT10f4
Wt	$7.16{\cdot}10^{-3}$	POWHEG-BOX 1	CT10f4

7.3 Selection of Candidate Events

The nominal selection is the result of an optimisation process that affected kinematic quantities related to the definition of the objects, namely the transverse momenta of the leading jet, the sub-leading jet and the sub-leading lepton, as well as event-selection criteria, for the missing energy of the event and the rapidity-separation between the two jets. The method adopted for the optimisation is discussed in Section 7.4, nevertheless in the following a complete description of the object and event selection chain is offered. The description is divided in the pre-selection step, the criteria that define the optimisation region, and the final requirements for candidate events in the signal region.

7.3.1 Object Selection

The physics objects used in this analysis are electrons, muons, jets and missing transverse energy, reconstructed in ATLAS as described in Chapter 5, where details on the definition of identification and isolation operating point can be found. Leptons and jets undergo three steps of selection. Loose requirements for the pre-selection level, additional ones used as baseline for the optimisation studies and the nominal selection criteria used to select objects in the signal region.

At pre-selection level, muons requirements on the transverse and longitudinal impact parameters ensure that the track originates from the primary vertex. In particular, the longitudinal impact parameter multiplied by the sine of the polar angle of the candidate must satisfy $|z_0 \times \sin\theta| < 0.5$ mm, and the transverse impact parameter divided by its significance is required to be $|d_0/\sigma_{d_0}| < 10$. Both calorimeter and tracking isolation criteria are applied to reject muons from in-flight decays or from b- and c-jets. The pre-selection requires Loose quality muons and has loose kinematic requirements, with $p_\mathrm{T} > 6$ GeV and $|\eta| < 2.7$.

For electrons the Loose identification operating point with the requirement on one B-layer hit is used at the pre-selection level. The kinematic and acceptance requirements imposed are $p_\mathrm{T} > 6$ GeV and $|\eta| < 2.47$, while the longitudinal and transversal impact parameter variables are required to be $|z_0 \times \sin\theta| < 0.5$ mm and $|d_0/\sigma_{d_0}| < 5$.

Jets are considered within $|\eta| < 4.5$, and for the pre-selection are required to have $p_\mathrm{T} > 25$ for $|\eta_\mathrm{det}| < 2.4$ and $p_\mathrm{T} > 30$ GeV for $2.4 < |\eta_\mathrm{det}| < 4.5$. The $|\eta_\mathrm{det}|$ does not refer to the pseudorapidity of the reconstructed object, but refers instead to the detector geometry, so that the jet is entirely contained within the acceptance value, and not only its barycentre. In order to reject spurious jets from pile-up the *jet vertex tagger* (JVT)[2] [139] is used in the pre-selection, by requiring JVT > 0.59 for jet with $p_\mathrm{T} < 60$ GeV and $|\eta_\mathrm{det}| < 2.4$.

An overlap removal procedure is performed on objects passing the pre-selection to remove reconstructed duplications of the same physical objects that were wrongly reconstructed as different particles by two or more algorithms. This can be the case of electrons and jets, which both leave

[2] This is a discriminant constructed as a two-dimensional likelihood with track-based variables. Input tracks to this algorithm are reconstructed as described in Section 5.1 and associated to the jets using *ghost association* [138]. They are treated as four-vectors of infinitesimal magnitude during jet reconstruction and assigned to the jet with which they are clustered.

Table 7.2: Object selection criteria for the pre-selection level, the optimisation region and nominal criteria for the signal region.

		Pre-selection	Optimisation	Nominal
Muons	Quality	Loose	Medium	Medium
	Isolation	–	Gradient	Gradient
	$\|z_0 \times \sin\theta\|$ [mm]	< 0.5	< 0.5	< 0.5
	$\|d_0/\sigma_{d_0}\|$	< 10	< 3	< 3
	$\|\eta\|$	< 2.7	< 2.5	< 2.5
	p_T [GeV]	> 6	> 27 (μ_0), 20 (μ_1)	> 27 ($\mu_{0,1}$)
Electrons	Identification	Loose+IBL hit	Tight	Tight
	Isolation	–	Gradient	Gradient
	$\|z_0 \times \sin\theta\|$ [mm]	< 0.5	< 0.5	< 0.5
	$\|d_0/\sigma_{d_0}\|$	< 5	< 5	< 5
	$\|\eta\|$	< 2.47	< 2.47 excl. $1.37 < \|\eta\| < 1.52$	< 2.47 excl. $1.37 < \|\eta\| < 1.52$ $\|\eta^e\| < 1.37$ in ee channel
	p_T [GeV]	> 6	> 27 (e_0), 20 (e_1)	> 27 ($e_{0,1}$)
	e/γ ambiguity	–	Unambiguous e	Unambiguous e
Jets	$\|\eta\|$	< 4.5	< 4.5	< 4.5
	p_T [GeV]	> 25 ($\|\eta_{det}\| < 2.4$) > 30 ($2.4 < \|\eta_{det}\| < 4.5$)	> 25 ($\|\eta_{det}\| < 2.4$) > 30 ($2.4 < \|\eta_{det}\| < 4.5$)	> 35 (j_0), > 65 (j_1)

traces in the calorimeter. Jets are removed if the distance $\Delta R(e, j)$ is smaller than 0.2, while the electron is discarded if $0.2 < \Delta R(e, j) < 0.4$. To remove overlapping leptons, the electron is removed in case it shares an inner detector track with a muon, and kept only if the muon is calorimeter-tagged (see Section 5.3), in which case the muon is removed. To avoid the less-likely case of overlapping between a muon and a jet, if they lie within a cone of $\Delta R = 0.2$, the jet is discarded if it has less than three tracks with $p_T > 0.5$ GeV.

After the pre-selection step, tighter criteria are imposed on the identification, isolation, impact parameter and geometrical acceptance of the leptons used for the kinematic optimisation.

Muons are selected with Medium quality and Gradient isolation. They are considered within the range $|\eta| < 2.5$ and are required to have $|z_0 \times \sin\theta| < 0.5$ mm and $|d_0/\sigma_{d_0}| < 3$.

Electrons are required to pass the Tight identification criteria and the Gradient isolation is required. They are selected outside of the transition region of the calorimeter, $1.37 \leq |\eta| \leq 1.52$, within $|\eta| < 2.47$. A specific classification is applied to electrons, to identify whether to consider them only as electron candidates, or potential photons candidate as well. This is based on the candidate's E/p, the p_T^{track}, the presence of a pixel hit and secondary vertex information. A requirement on unambiguous electrons is imposed at the optimisation-level selection.

For both electrons and muons the transverse momentum of the leading lepton is required to be $p_T > 27$ GeV, while the selection on the sub-leading lepton is relaxed to $p_T > 20$ GeV and will be object of the kinematic optimisation.

In this region the jet requirements are the same as for the pre-selection, since their transverse momenta are affected by the optimisation studies. Finally, the kinematic requirements for objects

Table 7.3: Trigger chains required for electrons and muons for the 2015 and 2016 data-taking years. The High Level Trigger and Level 1 Trigger selections oh the p_{T} threshold are indicated, together with further requirements on identification (ID) and isolation (Iso) when present [141].

Object	Year	High Level Trigger		Level 1 Trigger	
		p_{T} threshold [GeV]	Further selection	p_{T} threshold [GeV]	Further selection
Muons	2015	20	Loose Iso ($p_{\mathrm{T}}^{\mathrm{cone}}20/p_{\mathrm{T}} < 0.12$)	15	–
		50	–	20	–
	2016	26	Medium Iso ($p_{\mathrm{T}}^{\mathrm{varcone30}}/p_{\mathrm{T}} < 0.07$)	20	–
		50	–	20	–
Electrons	2015	24	Medium ID	20	variable p_{T} thr., hadronic core and EM Iso
		60	Medium ID	22	
		120	Loose ID	22	
	2016	26	Tight ID (no d_0 info), Loose Iso ($p_{\mathrm{T}}^{\mathrm{varcone20}}/E_{\mathrm{T}} < 0.1$)	22	
		60	Medium ID (no d_0 info)	22	
		140	Loose ID (no d_0 info)	22	

in the signal region that differ from the optimisation region are the selection on the transverse momentum of the sub-leading lepton at $p_{\mathrm{T}} > 27$ GeV and of the sub-leading and leading jet at $p_{\mathrm{T}} > 35$ GeV and 65 GeV respectively. In addition, in the ee channel only, a requirement of $|\eta^e| < 1.37$ is applied to reduce the charge misidentification background.

The selection criteria for muons, electrons and jets are summarised in Table 7.2 for the pre-, optimisation and nominal selection.

7.3.2 Event Selection

The selection chain for the event candidates can be divided as well into a pre-selection level, which aims at selecting high-quality events, a further set of requirements that define the region used for the kinematic optimisation, and the final nominal criteria that define the signal region. The pre-selection criteria select only validated events, removing the ones collected with detectors in non-optimal conditions or with noisy calorimeter cells, together with events where jets fail quality criteria defined by dedicated algorithms [140].

At least one reconstructed vertex is required, where the primary one is defined as in Section 5.1. Only events with at least three tracks with $p_{\mathrm{T}} > 0.5$ GeV associated to the primary vertex are considered for the analysis.

Events have to pass at least one of several chains of trigger selections. They are based on requirements posed on one lepton, the single-lepton triggers introduced in Section 6.4.1, which requires at least one lepton to pass the trigger thresholds, and ensure that at least one of the signal leptons is the one matched to the trigger. The trigger chains used are different for electrons and muons, and for the data-taking years 2015 and 2016 and their corresponding simulated samples. They are summarised in Table 7.3.

The region used as baseline for the optimisation studies requires leptons and jets to pass the criteria on identification, isolation, impact parameter, geometrical acceptance and transverse momenta listed in Table 7.2. A lower selection on the invariant mass of the two highest-p_{T} leptons is set to

20 GeV, to reduce the uncertainties on the modelling of low-mass Drell-Yan events. At this stage of the selection, a veto on events with a third lepton that pass the above requirements is imposed.
In simulated events, only leptons classified as *prompt* are considered. The definition is based on the information from simulation on the origin and type of the leptons. Electrons and muons are labelled as prompt in the analysis if they are genuine leptons from direct decays, i.e. a category that includes leptons from bosons, mesons or top quark. For electrons, also the products of conversion of Final State Radiation (FSR) photons are taken into account, as well as particles labelled as photon. This definition of prompt electrons differs from the tighter one discussed in Section 6.2, since the scope in this case is broader than for the charge misidentification studies. Events in simulation which have both prompt leptons pass the so-called *truth matching* selection.
Exactly two leptons of same-charge and at least two jets passing the pre-selection criteria are required.
The two highest-p_T jets are the tagging ones, used to calculate dijet quantities. A requirement on the invariant dijet mass is imposed at $m_{jj} > 200$ GeV, while the two electrons in the ee channel need to satisfy a veto on the Z mass by having $|m_{ee} - m_Z| > 15$ GeV, to reduce the Drell-Yan background. This is not applied in the $\mu\mu$ channel because the charge misidentification is negligible for muons. In order to identify the b-jets the MV2c10 tagger [119] is used at the 85% efficiency operating point, and events are rejected if at least one b-tagged jet is found.
Further event selection criteria are imposed in the signal region, together with the different requirements imposed on leptons and jets discussed in Section 7.3.1 and summarised in Table 7.2 for the nominal selection. Two of them result from the optimisation process, and affect the missing transverse energy, required to be larger than 30 GeV, and the separation between the two tagging jets $\Delta y_{jj} > 2$. The latter, together with a tighter requirement on the invariant mass of the two jets of $m_{jj} > 500$ GeV, defines the electroweak signal region.
The selection chains are summarised in Table 7.4 for the pre-selection, optimisation and signal region. The optimisation for the selection criteria on the variables of interest, that led to the selection presented in this chapter, is detailed in the following.

Table 7.4: Events pre-selection, optimisation region and signal region selection. The dedicated criteria for leptons and jets refer to the different requirements applied for the optimisation studies or for the nominal selection, which are summarised in Table 7.2, while the the trigger selection is summarised in Table 7.3.

<table>
<tr><th></th><th>Optimisation region</th><th>Signal region</th></tr>
<tr><td rowspan="3">Pre-selection</td><td colspan="2">Event quality</td></tr>
<tr><td colspan="2">$N_{\text{primary vertex}} = 1$</td></tr>
<tr><td colspan="2">Trigger selection</td></tr>
<tr><td></td><td colspan="2">Leptons and jets dedicated selection</td></tr>
<tr><td></td><td colspan="2">$m_{\ell\ell} > 20$ GeV</td></tr>
<tr><td></td><td colspan="2">3$^{\text{rd}}$ lepton veto</td></tr>
<tr><td></td><td colspan="2">Truth matching (prompt lepton selection of genuine e/γ,μ)</td></tr>
<tr><td></td><td colspan="2">$N_{\ell\,\text{same-charge}} = 2$</td></tr>
<tr><td></td><td colspan="2">$N_j \geq 2$</td></tr>
<tr><td></td><td colspan="2">$m_{jj} > 200$ GeV</td></tr>
<tr><td></td><td colspan="2">$|m_{ee} - m_Z| > 15$ GeV</td></tr>
<tr><td></td><td colspan="2">$N_{b\text{-jet}} = 0$ with MV2c10 at 85% efficiency</td></tr>
<tr><td></td><td>–</td><td>$E_{\text{T}}^{\text{miss}} > 30$ GeV</td></tr>
<tr><td></td><td>–</td><td>$\Delta y_{jj} > 2$</td></tr>
<tr><td></td><td>$m_{jj} > 200$ GeV</td><td>$m_{jj} > 500$ GeV</td></tr>
</table>

7.4 Kinematic Optimisation

The kinematic optimisation is performed with the aim of finding the selection criteria that maximise the expected significance for the electroweak $W^{\pm}W^{\pm}jj$ EW process.

7.4.1 Optimisation Region

The region where the optimisation is performed is the one described in Section 7.3 and summarised in Table 7.4. The variables of interest for the optimisation are the transverse momenta of the leading and sub-leading jets p_{T}^{j0} and p_{T}^{j1}, the sub-leading lepton $p_{\text{T}}^{\ell 1}$, the missing energy of the event $E_{\text{T}}^{\text{miss}}$ and the rapidity-separation between the two tagging jets Δy_{jj}.
The samples used are the nominal ones described in Section 7.1.2 and 7.2.1, the non-prompt lepton and charge misidentification backgrounds are estimated with data and will be the objects of Section 8.1 and Section 8.4 respectively.

7.4.2 Multi-Dimensional Grid Scan Procedure

The technique used for the optimisation is a *multi-dimensional grid scan*, where a specific set of variables is considered, within defined ranges and with defined variation steps, called *bins* in the following. These define a multi-dimensional space of all possible selection configurations, made by varying the selections on all quantities, in all possible combinations. Each point of this space is therefore an n-dimensional point, n being the number of variables considered for the optimisation.

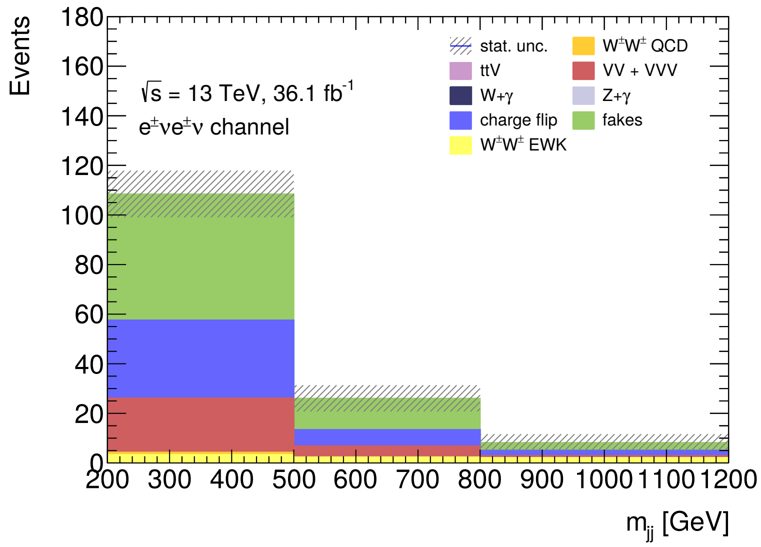


Figure 7.4: Dijet invariant mass distribution in bins used for the kinematic optimisation.

The choice of the best set of cuts relies on a ranking of the points, based on a figure of merit that quantifies the power of signal enhancement over background rejection. The chosen figure of merit is the expected Poissonian significance that includes the information on background uncertainties. It is demonstrated in Ref. [142] that this can be expressed as

$$Z = \sqrt{2\left((S+B)\ln\left[\frac{(S+B)(B+\sigma_B^2)}{B^2+(S+B)\sigma_B^2}\right]-\frac{B^2}{\sigma_B^2}\ln\left[1+\frac{\sigma_B^2 S}{B(B+\sigma_B^2)}\right]\right)}.$$

The terms S and B refer to the signal and background yields respectively, while σ_B is the total uncertainty on the background. At the time of the optimisation, a full evaluation of the systematic uncertainties related to each background was not available, hence systematic uncertainties have been assigned in the scope of this study. The backgrounds considered are grouped into three classes, namely non-prompt background, labelled as *fakes*, WZ, and the last one including all remaining background processes, labelled as *others*. For each of them, the σ_B^2 is hence defined as

$$\sigma_B = \sqrt{\sigma_{B,\text{stat.}}^2+\sigma_{B,\text{syst.}}^2}, \tag{7.1}$$

where an overall 20% is considered for the non-prompt lepton background yield, 20% for WZ and 10% for the other backgrounds, based on preliminary estimation available. Nevertheless, the impact of different uncertainty values is checked. For each point the significance is evaluated scanning a range of possible systematic values on each background at the time, while keeping the others fixed. The final fitting procedure is performed on the m_{jj} distribution. Therefore, in order for the optimisation to be as close as possible to the fit, the same variable is used to assess the figure of merit. This is evaluated for six channels, i.e. the lepton flavour-charge combinations (e^+e^+), (e^-e^-), ($e^+\mu^+$, μ^+e^+), ($e^-\mu^-$, μ^-e^-), ($\mu^+\mu^+$), ($\mu^-\mu^-$) in three m_{jj} bins:

- b_1: $200 < m_{jj} < 500$ GeV

Table 7.5: Set of variables used for the multi-dimensional optimisation. The range used for the grid scan which determines the boundary of the configuration space is detailed, together with the points scanned for each variable, the bin widths and the pre-optimisation cut values.

Variable	Range	Bin width	Pre-optimisation selection
p_T^{j0}	25-85 GeV	5 GeV	> 25 GeV
p_T^{j1}	25-60 GeV	5 GeV	> 25 GeV
$p_T^{\ell 1}$	20-27 GeV	3.5 GeV	> 27.0 GeV
E_T^{miss}	0-50 GeV	5 GeV	> 30 GeV
Δy_{jj}	1.0-3.0	0.5	> 2.4

- b_2: $500 < m_{jj} < 800$ GeV
- b_3: $m_{jj} > 800$ GeV.

The m_{jj} distribution in the optimisation region is shown in Fig. 7.4 with the described binning. The significance Z_b from each bin b is combined in quadrature to obtain the significance Z_c in each flavour-charge channel c

$$Z_c = \sqrt{\sum_{b=1}^{3} Z_b^2},$$

which is further combined to obtain the final overall significance estimation

$$Z = \sqrt{\sum_{c=1}^{6} Z_c^2}.$$

7.4.3 Scan and Results

The variation ranges for the five variables and the bin sizes are listed in Table 7.5, where the selection used before the optimisation is reported as well, mainly based on the Run 1 analysis [143]. The ranges determine the boundary of the configuration space, and the bins are the sizes of the steps used for scanning. In the table the number of one-dimensional points analysed for each variable are listed as well, taking into account that the starting and final points are analysed as well. The scan is run over a total of 17160 five-dimensional points.

Unphysical background fluctuations due to low statistical power can push a configuration to be higher in the ranking, due to unrealistic low background yields. In particular, this is the case of the non-prompt lepton background in the analysis. This background is highly fluctuating in the optimisation region from one bin to another. This is visible in Fig. 7.5 where the distributions for the optimised variables are shown in the optimisation region in the ee channel. Distributions for the $\mu\mu$ and $e\mu + \mu e$ channels are shown in Appendix B.

The non-prompt background presents a spiky trend, which can result in an artificial increase of the significance for certain bins, and hence for certain selections. This means that a mathematical ranking based univocally on a figure of merit would be blind to physics-related considerations. As a

Table 7.6: The ten best points as result from the multi-dimensional grid scan optimisation, ranked based on the binned Poissonian significance of each configuration.

	$p_{\mathrm{T}}^{j_0} >$ [GeV]	$p_{\mathrm{T}}^{j_1} >$ [GeV]	$p_{\mathrm{T}}^{\ell_1} >$ [GeV]	$E_{\mathrm{T}}^{\mathrm{miss}} >$ [GeV]	$\Delta y_{jj} >$	Z
Point 1	75	35	20	40	1.5	5.08
Point 2	75	25	20	40	1.5	5.07
Point 3	75	40	20	40	1.5	5.07
Point 4	75	30	20	40	1.5	5.07
Point 5	75	35	20	40	2.0	5.07
Point 6	75	40	20	40	2.0	5.07
Point 7	75	35	20	40	1.0	5.06
Point 8	75	25	20	40	1.0	5.06
Point 9	75	25	20	40	2.0	5.06
Point 10	75	30	20	40	1.0	5.06

consequence, the configuration of the selection criteria that leads to the highest significance might be a sub-optimal choice.

The 10 best points resulting from the optimisation are listed in Table 7.6. Based on those a tuning of the selected configuration is performed. A decision is made, to be conservative on the $E_{\mathrm{T}}^{\mathrm{miss}}$ requirement, given the highly fluctuating non-prompt background in its distribution (see Fig. 7.5e), and to maintain the $E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV selection of the Run 1 analysis despite the scan indication of rising the limit to 40 GeV.

A cautious choice is made on the Δy_{jj} selection as well. The highest points in the ranking present a looser selection on this variable with respect to the previous $\Delta y_{jj} > 2.4$, which aimed at enhancing the electroweak production modes with respect to the QCD ones. In this case a choice is made to lower the selection only to $\Delta y_{jj} > 2$, to maintain a good discriminating power between the $W^{\pm}W^{\pm}jj$ QCD background and the $W^{\pm}W^{\pm}jj$ electroweak signal process.

A similar situation holds for the $p_{\mathrm{T}}^{\ell 1}$, where a conservative choice to keep the 27 GeV limit is made, in order not to extend to a kinematic region below the trigger threshold and where the non-prompt lepton background is most likely larger.

For what concerns the jet p_{T} selection, the scan indicates an increase in significance when rising the lower threshold. This suggestion is followed, as this allows also for a better rejection of low-p_{T} jets from pile-up.

As a consequence of these considerations, none of the ten best-ranked points is considered reliable for the experimental scenario of a highly fluctuating background rich in non-prompt leptons. The configuration tuned on these arguments is

$$
\begin{aligned}
p_{\mathrm{T}}^{j_0} &> 65 \text{ GeV} \\
p_{\mathrm{T}}^{j_1} &> 35 \text{ GeV} \\
p_{\mathrm{T}}^{\ell_1} &> 27 \text{ GeV} \\
E_{\mathrm{T}}^{\mathrm{miss}} &> 30 \text{ GeV} \\
\Delta y_{jj} &> 2.
\end{aligned}
$$

Table 7.7: Expected signal and background yields and their statistical and assumed systematic uncertainties summed in quadrature for the optimised configuration split by flavour and charge. The systematic uncertainties assigned are 20% for the non-prompt background, 20% for WZ and 10% for other backgrounds. The Poissonian significance is reported for each m_{jj} bin.

Channel	m_{jj} [GeV]	EW $W^{\pm}W^{\pm}jj$	Fakes	WZ	Others	Z
e^+e^+	$[200, 500]$	0.65±0.03	2.62±1.76	2.05±0.23	2.51±0.11	0.20
	$[500, 800]$	0.87±0.04	4.44±2.32	0.86±0.12	1.22±0.18	0.24
	$[800, +\infty]$	2.99±0.07	1.84±1.62	1.05±0.09	0.98±0.08	1.01
e^-e^-	$[200, 500]$	0.48±0.05	6.79±3.20	1.59±0.14	2.33±0.18	0.10
	$[500, 800]$	0.48±0.03	0.99±1.70	0.72±0.08	1.22±0.23	0.19
	$[800, +\infty]$	1.03±0.04	2.84±2.21	0.59±0.07	0.66±0.06	0.32
$\mu^+\mu^+$	$[200, 500]$	1.86±0.09	2.88±0.93	5.79±0.27	1.81±0.07	0.53
	$[500, 800]$	2.16±0.06	0.49±0.27	2.65±0.18	1.06±0.05	0.97
	$[800, +\infty]$	7.12±0.11	0.87±0.41	2.88±0.17	1.13±0.04	2.64
$\mu^-\mu^-$	$[200, 500]$	1.13±0.04	3.08±1.15	3.98±0.21	1.02±0.05	0.36
	$[500, 800]$	1.18±0.04	2.19±0.89	1.80±0.13	0.52±0.03	0.49
	$[800, +\infty]$	2.37±0.06	0.71±0.34	1.85±0.18	0.44±0.03	1.20
$e^+\mu^+ + \mu^+e^+$	$[200, 500]$	3.01±0.07	15.72±4.65	13.97±0.57	7.78±0.35	0.38
	$[500, 800]$	3.93±0.08	3.93±2.69	7.32±0.38	4.72±0.71	0.77
	$[800, +\infty]$	12.82±0.15	3.73±1.95	6.68±0.27	4.58±0.44	2.56
$e^-\mu^- + \mu^-e^-$	$[200, 500]$	2.02±0.06	10.14±3.43	10.47±0.55	7.47±0.62	0.31
	$[500, 800]$	2.07±0.06	6.50±2.65	5.09±0.35	3.04±0.30	0.43
	$[800, +\infty]$	4.48±0.09	0.97±1.04	3.92±0.22	2.47±0.24	1.39

It occupies a lower position in the ranking, however is considered a more robust choice.

This leads to a significance of 4.31.

In order to check how each variable impact the significance, an $n-1$ scan is performed around the chosen configuration, with one variable at the time left free to change along the variation range. This is presented in Fig. 7.6 for all variables. The artificial increase and decrease of the significance emerges in particular in Fig. 7.6d. The impact of the variation of the systematic uncertainties assigned to each background is analysed, and the scan is shown in Fig. 7.6f for the chosen selection. The largest impact is observed by the systematic uncertainties related to the WZ background, which results in a decrease of significance for values higher than 30-40%.

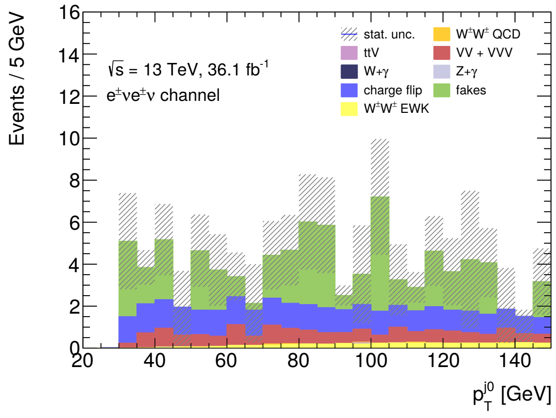
The signal and background yields, as well as the significance for each m_{jj} bin at the chosen selection are reported in Table. 7.7 for channels split in flavour and charge.

The modelling of the selection is checked in a region enriched in non-prompt lepton contribution, namely the *top fakes same-charge validation region*. The detailed definition will be given in Table 8.5, where all validation regions are described. Nevertheless, the modelling of the transverse momentum distribution of the two leptons is shown here in Fig 7.7. A disagreement between data and predictions is visible in certain bins, considered to originate from the estimation of the non-prompt lepton background, covered with the assigned systematic uncertainty of 20%.

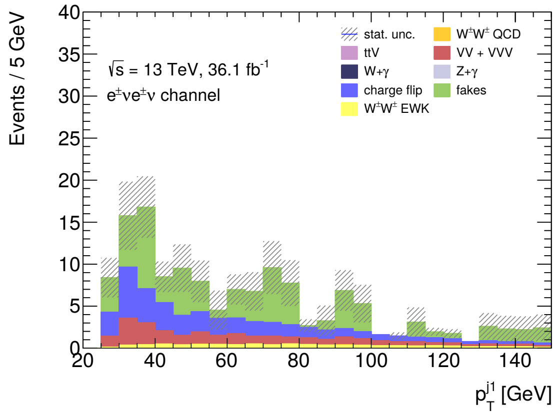
In order to quantify the improvement of the optimised selection with respect to that of the Run 1 analysis [143], the Poissonian significance is evaluated in the two cases. The improvement in the significance evaluated with the Poisson formulation is about 8%, going from $Z_{pre} = 4.00$ to $Z_{optim} = 4.31$. The selections and results are compared in Table 7.8.

Table 7.8: Selection criteria resulting from the kinematic optimisation compared to the previous selection of the Run 1 analysis [143]. The Poissonian significance is evaluated for the two scenarios.

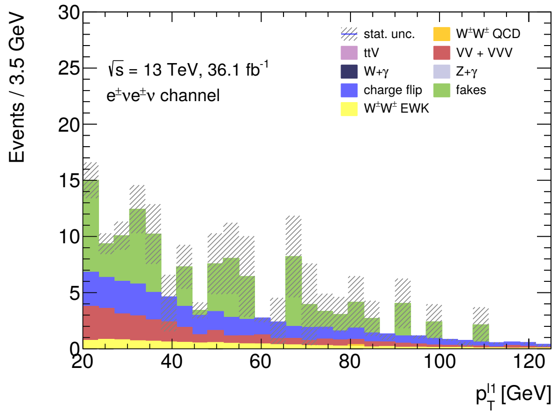
Variable	Optimised selection	Previous selection
p_{T}^{j0}	> 65 GeV	> 25 GeV
p_{T}^{j1}	> 35 GeV	> 25 GeV
$p_{\mathrm{T}}^{\ell 1}$	> 27 GeV	> 27 GeV
$E_{\mathrm{T}}^{\mathrm{miss}}$	> 30 GeV	> 30 GeV
Δy_{jj}	> 2.0	> 2.4
Z	4.31	4.00



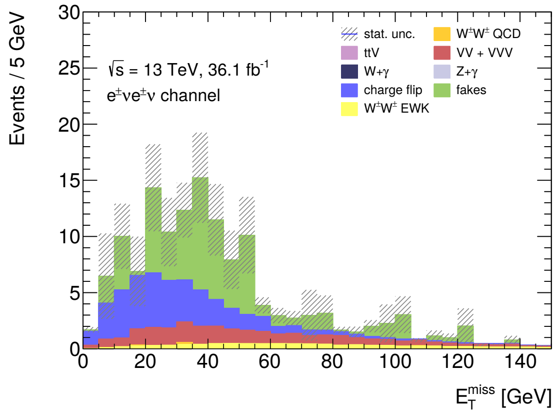
(a)



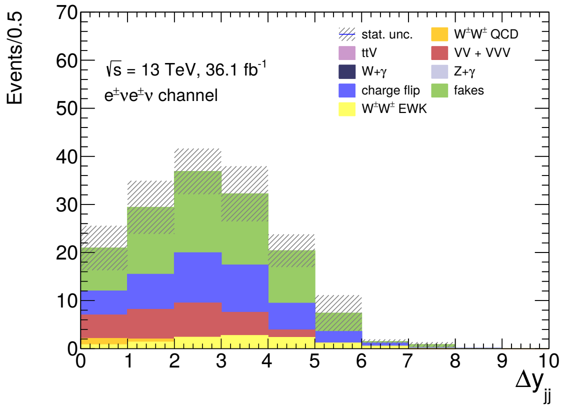
(b)



(c)



(d)



(e)

Figure 7.5: Distributions of transverse momentum of the leading (a) and sub-leading (b) jet, sub-leading lepton (c), missing transverse energy (d) and rapidity separation between the two tagging jets (e) in the optimisation region for the ee channel. Only statistical uncertainties are shown.

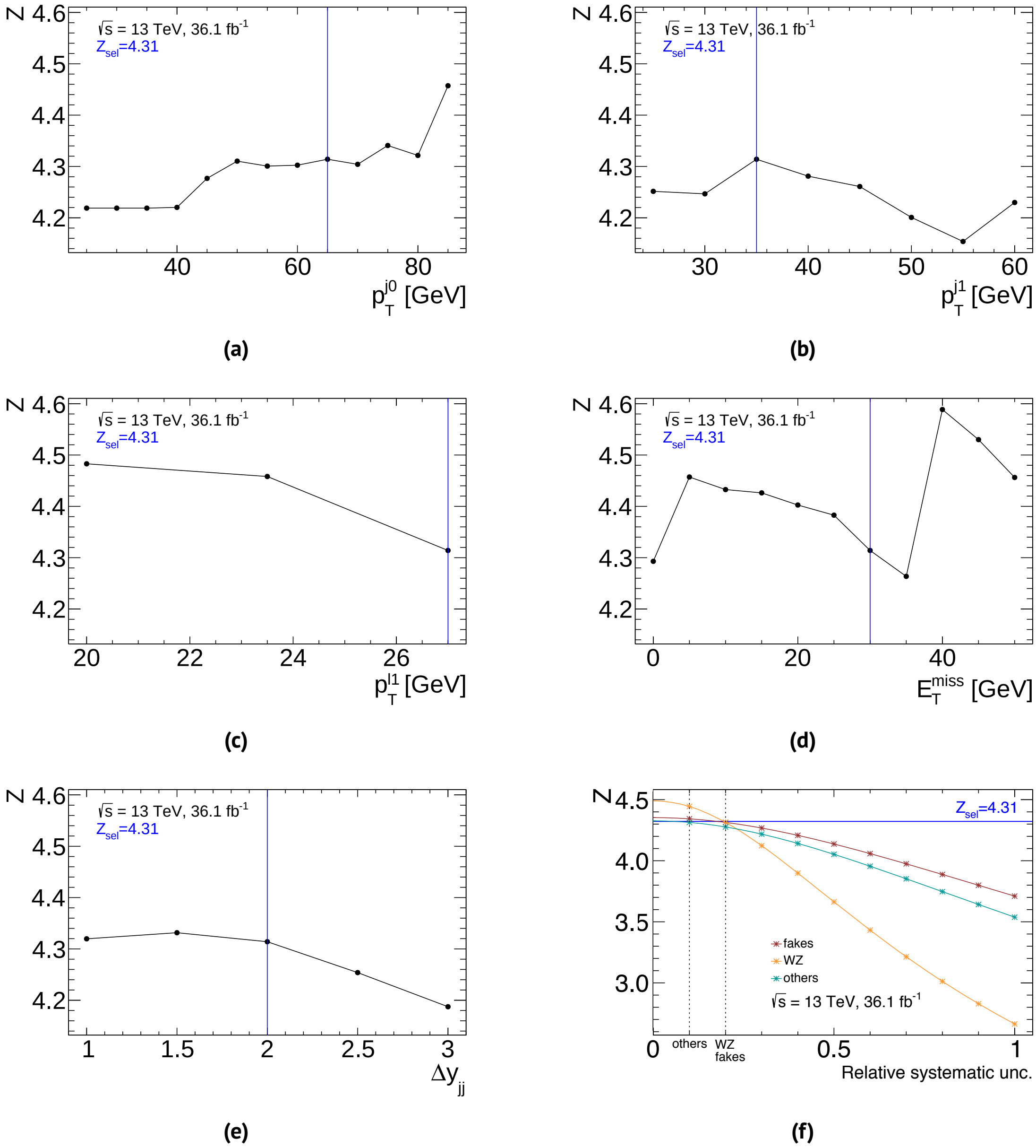
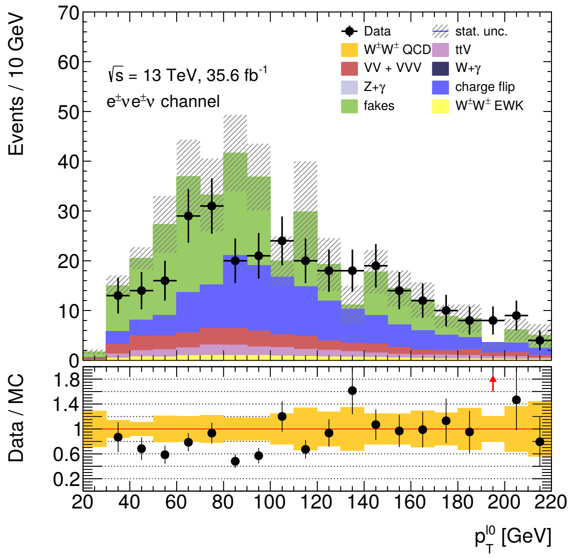
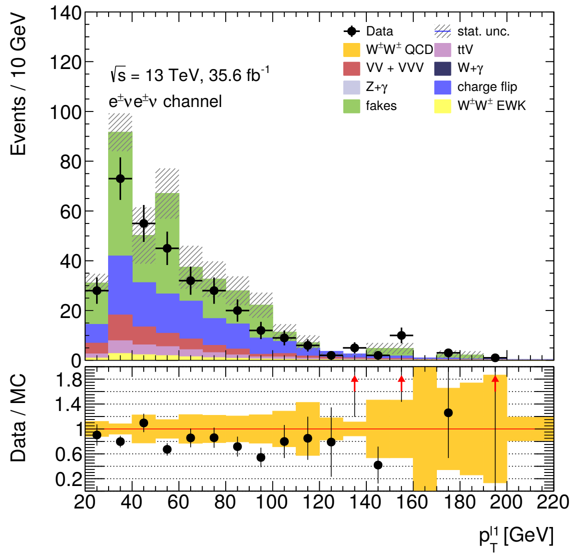


Figure 7.6: Scan of the impact on the significance Z of the selection on the transverse momentum of the leading (a) and sub-leading (b) jet, sub-leading lepton (c), the missing transverse energy (d) and the rapidity separation between the two tagging jets (e). The scans are shown for a point in the configuration space where all variables except the one plotted are fixed to the optimised selection. The variable on the x-axis is varied within the ranges and with the steps used in the optimisation process. The blue line indicates the selection corresponding to the chosen configuration, which has a significance of Z = 4.31. (f) Scan of the impact of the assumed systematic uncertainties on the evaluation of the overall significance Z. The dashed lines indicate the assumed values for each background that give the resulting significance of Z = 4.31.



(a)



(b)

Figure 7.7: Distribution of the transverse momentum of the leading (a) and sub-leading (b) lepton for the ee channel in the top fakes same-charge control region. Only statistical uncertainties are shown.

8 Background Estimation

Events from background processes which are left in the signal region after the selection are estimated with different techniques. The predictions for the non-prompt lepton background is performed using data with a method explained in Section 8.1. The background from WZ+jets is estimated deriving the normalisation in a dedicated control region, as discussed in Section 8.2. The estimation of the background from photon conversion is based on simulation of the $V\gamma$+jets process and presented in Section 8.3. The charge misidentification background is treated extensively in Section 8.4, where the technique for its estimation is described, both in the signal region and in the region where a looser selection is applied to one of the electrons. Finally the validation regions defined to check the correct modelling of the different predictions are discussed in Section 8.5.

8.1 Non-Prompt Lepton Background

In processes where a prompt lepton is produced together with a jet, the event can pass the signal region selection if the jet is wrongly reconstructed as a lepton. This can be the case for electrons, as both electrons and jets leave a trace in the calorimeter, and muons when they originate from decays of heavy-flavour hadrons, as well as in-flight decays of $\pi^{\pm}$ or $K^{\pm}$ mesons. In this case they can be classified as signal muons, and thus can wrongly enter the signal region. Processes that can lead to this class of events form the non-prompt lepton background, which due to its nature may not be properly modelled in simulations and need to be estimated with data. The processes sources of non-prompt leptons are W+jets, $t\bar{t}$ where one W boson decays leptonically and the other hadronically, and single top processes with a W boson decaying leptonically. Diboson processes such as WW, WZ, ZZ where one boson decays hadronically and the other leptonically have smaller contributions due to the smaller production cross sections.
The estimation of the non-prompt lepton background consists in weighting a sample designed to be enriched with jets, but kinematically similar to the signal region. To select this sample, a nominal lepton passing all signal requirements and a jet-like lepton passing looser conditions are required, in addition to all other nominal signal region selections. The factor that performs the weighting of this nominal+loose sample, namely *ID+antiID*, is the so called *fake-factor*. It is the ratio of the probability for a jet to satisfy the nominal lepton criteria to the probability to satisfy the loose ones. It is measured in a dijet sample, a dedicated jet-enriched sample where physics processes that produce prompt leptons are suppressed.
The estimation of the non-prompt background in the signal region is obtained by using

$$N_{\text{fake}} = \text{f}_{\text{lepton}} \times \left(N_{\text{data}} - N_{\text{prompt}}\right)_{\text{ID+antiID}} \tag{8.1}$$

in the ee and $\mu\mu$ channel, and

$$N_{\text{fake}} = \text{f}_{e} \times \left(N_{\text{data}} - N_{\text{prompt}}\right)_{\text{ID}\mu\text{+antiID}e} + \text{f}_{\mu} \times \left(N_{\text{data}} - N_{\text{prompt}}\right)_{\text{ID}e\text{+antiID}\mu} \tag{8.2}$$

in the $e\mu$ channel where non-prompt lepton background receives contributions from events with a

Table 8.1: Selection criteria for muons and electrons for the ID and antiID selection. The former is the nominal selection applied in the signal region, the latter is used to construct the dijet sample for the fake-factor measurement and the ID+antiID region for the estimation of the background from non-prompt leptons.

		ID	antiID
Muons	Quality	Medium	Medium
	Isolation	Gradient	–
	$\lvert z_0 \times \sin\theta \rvert$	< 0.5 mm	< 0.5 mm
	$\lvert d_0/\sigma_{d_0} \rvert$	< 3	< 10
	$\lvert \eta \rvert$	< 2.5	< 2.5
	p_T	> 27 GeV	> 15 GeV
	Additional	–	Fail other ID selection
Electrons	Identification	Tight	Medium
	Isolation	Gradient	–
	$\lvert z_0 \times \sin\theta \rvert$	< 0.5 mm	< 0.5 mm
	$\lvert d_0/\sigma_{d_0} \rvert$	< 5	< 5
	$\lvert \eta \rvert$	< 2.47 excl. $1.37 < \lvert \eta \rvert < 1.52$	< 2.47 excl. $1.37 < \lvert \eta \rvert < 1.52$
	p_T	> 27 GeV	> 15 GeV
	e/γ ambiguity	Unambiguous e	Unambiguous e
	Additional	–	Fail other ID selection

prompt muon and a fake electron or with a prompt electron and a fake muon.
The N_prompt refers to the contribution from events where two prompt leptons are present, one satisfying the ID selection and the other passing the antiID one. This needs to be subtracted, and it is done using simulation. The contribution of events from charge misidentification is subtracted as well, using instead a dedicated estimation on data explained in Section 8.4.6.

AntiID lepton definition The definition of the loose lepton criteria plays a central role in determining the purity and number of events of the two samples. It is of great importance to enhance the fake lepton contribution in the ID+antiID sample as well as to enrich the jet population and suppress the prompt leptons in the dijet sample used for the fake-factor measurement. In both samples muons are required to pass the same quality, acceptance and $|z_0 \times \sin\theta|$ criteria as the signal ones. The requirements on d_0/σ_{d_0} are relaxed, the p_T threshold is lowered to 15 GeV and muons need to fail all other signal requirements. For electrons the identification working point is moved to Medium and the nominal Tight identification and Gradient isolation selection are required to fail. Acceptance and selection on the impact parameter remain as in the nominal case, while the p_T is lowered to 15 GeV as for muons. The requirements for antiID leptons are summarised and compared to the ID ones in Table 8.1.

Fake-factor determination and uncertainties The fake-factor is generally expressed as the ratio between the number of candidates, likely jets, satisfying the nominal identification criteria and the ones satisfying the antiID lepton criteria, and is usually binned in p_T.

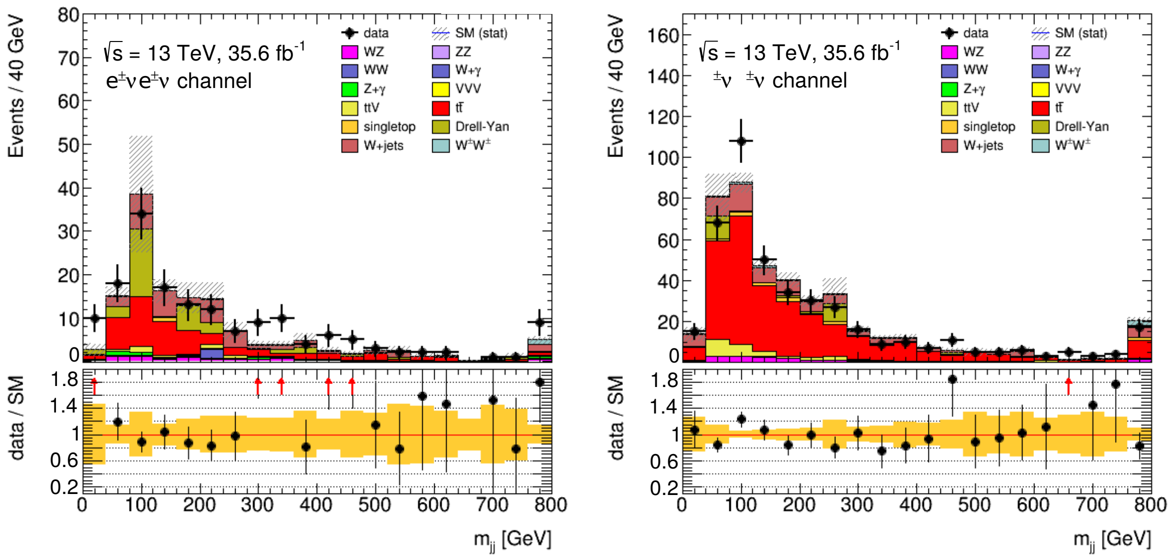
Table 8.2: Event selection for dijet sample used to determined the fake-factor. The trigger selection is dedicated to the dijet region and is explained in the text.

Dijet sample selection
Trigger selection
N_ℓ = 1 (ID or antiID) and $p_{\mathrm{T}}^{\ell} > 15$ GeV
$N_j > 0$
$p_{\mathrm{T}}^{\text{tag-jet}} > 25(30)$ GeV for $\lvert\eta\rvert < 2.5$ ($2.5 < \lvert\eta\rvert < 4.5$)
Tag-jet is b-tagged with MV2c10 at 85% efficiency
$\lvert\Delta\phi(\ell, \text{tag jet})\rvert > 2.8$
$E_{\mathrm{T,track}}^{\mathrm{miss}} + E_{\mathrm{T}}^{\mathrm{miss}} < 50$ GeV

The dijet data sample used to measure it is created with candidate events selected by the combination of different trigger chains, which select events with an electron or a muon plus jets. They differ from the nominal ones described in Table 7.3. A pre-scaled (see Section 4.2.6) trigger selects events with at least one electron with $p_{\mathrm{T}} > 12$ GeV and very loose identification operating point, without the d_0 information in it, seeded from the electromagnetic Level 1 trigger at $p_{\mathrm{T}} > 10$ GeV, with a variable p_{T} threshold and hadronic core isolation applied. The trigger chain used for muons is a combination of two, one selecting events with one muon with $p_{\mathrm{T}} > 14$ GeV and the other with $p_{\mathrm{T}} > 50$ GeV. In the sample selection exactly one ID or one antiID lepton are required and at least one jet with $p_{\mathrm{T}} > 25(30)$ GeV for $\lvert\eta\rvert < 2.5(2.5 < \lvert\eta\rvert < 4.5)$. This jet is required to be a b-jet in order to enrich the sample with heavy-flavour jets. Indeed, these are proven to be the primary source of fake leptons in the signal region, by checking their origin in simulation. The lepton is required to be back-to-back with the jet in the azimuthal plane with $\lvert\Delta\phi(\ell, \text{tag jet})\rvert > 2.8$, while a requirement on $E_{\mathrm{T,track}}^{\mathrm{miss}} + E_{\mathrm{T}}^{\mathrm{miss}} < 50$ GeV suppresses the contamination from W+jets with the W boson decaying leptonically. The full selection for the dijet sample is summarised in Table 8.2. In this sample contributions from prompt leptons from W+jets, Z/γ^*+jets, $t\bar{t}$ and single top are subtracted with simulation. Their impact is small, mainly deriving from W+jets and affecting only region where lepton p_{T} is above $\sim$80 GeV.

The fake-factor is applied to the ID+antiID region, whose composition is studied with simulation and shown in Fig. 8.1. The dominant process is the top quark pair production, followed by W+jets or Drell-Yan Z+jets production. A small fraction of non-prompt leptons originates also from $V\gamma$ samples. The Z/γ^*+jets sample in the ee channel is a source of events from charge misidentification, while the diboson processes generate prompt same-charge leptons in fully leptonic decays, and non-prompt leptons in semi-leptonic decays.

The fake-factor is hence calculated in dijet data where non-prompt lepton candidates are mainly from b and c decays, and applied in a region dominated by $t\bar{t}$ events, where almost all non-prompt leptons are from decays of b-hadrons. Given that depending on the process a different fraction of momentum of the underlying jet is carried by the non-prompt lepton, the variable used for the binning of the fake-factor should account for this difference. This is why the total transverse momentum inside a variable-size cone is chosen for the binning of the fake-factor denominator,



(a) **(b)**

Figure 8.1: Dijet invariant mass distribution in the ee (a) and $\mu\mu$ (b) channels in the ID+antiID sample. Only statistical uncertainties are shown [144].

instead of the usual transverse momentum. Indeed, the sum of the transverse momentum of the non-prompt lepton candidate p_{T} and the $p_{\mathrm{T}}^{\mathrm{varcone}}$ variable (see Section 5.2) can mitigate the dependency by recovering the jet momentum not carried by the non-prompt lepton. The fake-factor for a given bin b is defined as

$$\mathrm{f}_{\mathrm{lepton},b} = \frac{N_{\mathrm{ID},\,b(p_{\mathrm{T}})}}{N_{\mathrm{antiID},\,b(p_{\mathrm{T}}+p_{\mathrm{T}}^{\mathrm{varcone}})}}. \tag{8.3}$$

The fraction of the transverse momentum of the underlying jet carried by the lepton, that is different for ID and antiID leptons, is therefore restored by including the transverse momentum in the cone. The nominal fake-factor is thus determined in bins of the p_{T} + $p_{\mathrm{T}}^{\mathrm{varcone}}$ variable, and each bin now refers to loose non-prompt objects in the denominator which have the same p_{T} as the ID objects in the numerator.

A one-dimensional fake-factor is computed for muons while for electrons it is additionally measured in two η bins, splitting the range at $|\eta|$= 1.37 in order to separate electrons inside or outside of the barrel of the electromagnetic calorimeter. The fake-factor lays below 0.2 for muons with p_{T} + $p_{\mathrm{T}}^{\mathrm{varcone}}$ $<$ 75 GeV, and it reaches 0.5 in the highest bin 75 GeV $<$ p_{T} + $p_{\mathrm{T}}^{\mathrm{varcone}}$ $<$ 200 GeV. For electrons it rises up to 0.6 reaching 0.9 in the last p_{T} + $p_{\mathrm{T}}^{\mathrm{varcone}}$ bin for $|\eta| < 1.37$, while for $1.37 < |\eta| < 2.5$ it lays below 0.8 up to p_{T} + $p_{\mathrm{T}}^{\mathrm{varcone}}$ = 75 GeV, and is slightly larger than 1 in the last bin.

One systematic uncertainty on the fake-factor measurement is related to the variation of the selection in the dijet sample. In addition, the subtraction of the prompt lepton via W+jets simulation is varied by 10%, as this is the discrepancy found in data-simulation comparison of the W+jets sample in a W+jets enriched data control region. Also, due to the difference in jet flavour composition between the dijet sample and the ID+antiID one, the fake-factor is computed by switching off the requirement on the tagging jet to be a b-jet, and the difference is accounted for as systematic uncertainty. Finally, a systematic uncertainty addresses the residual dependence of the fake-factor on the underlying jet p_T. Weighting factors are derived using the true underlying jet p_T spectrum on events selected by requiring one lepton and one true b-jet. The factors aim at the matching between p_T distributions of $t\bar{t}$ and W+jets in dijet data, selected by requiring the tagging jet to be a b-jet. These factors are used to reweight the jet p_T spectra for the true b-jet that is used to determine an approximation for the true underlying jet p_T in the ID+antiID region. The p_T and $p_T + p_T^{\text{varcone}}$ distributions are reweighted with these factors and the fake-factor is calculated with and without the reweighting. This systematic affects the measurement mainly at high p_T, while at low p_T it is consistent with the other systematic uncertainties. The systematic variations lay in general below 10% for the muon case, and they reach up to 20-30% for electrons.

Non-prompt background estimation in the signal region Due to the use of the p_T^{varcone} variable, special care is needed when applying the fake-factor to the ID+antiID region with Eq. (8.1) and (8.2). The p_T of the loose object in this region is always substituted by the $p_T + p_T^{\text{varcone}}$, both when applying the momentum requirement and when applying the weight, that must be picked from the bin corresponding to the $p_T + p_T^{\text{varcone}}$ in the fake-factor map. Finally, the quantity has to be assigned as the p_T of the predicted non-prompt ID lepton when applying the fake-factor to get the final estimation.

The modelling of the estimation is checked in various validation regions, enriched in non-prompt leptons from W+jets or $t\bar{t}$ events. They are described in detail in Section 8.5, where an overview of all validation regions used in the analysis is given. Distributions will be presented as well in Fig. 8.14 and 8.18, showing the good agreement within the statistical uncertainties between data and non-prompt lepton background estimation.

The final estimation in the signal region is performed by applying the fake-factor with Eq. (8.1) and (8.2). The prompt-lepton contribution is subtracted by requiring both reconstructed lepton candidates to be matched to generated prompt leptons, as defined in the truth matching described in Section 7.3.2. Events from charge misidentification are removed via a dedicated estimation performed on data detailed in Section 8.4.

The yields estimated for each channel are reported in Table 8.3. The uncertainty is approximately 50% in the $\mu\mu$ channel and ranges between 40-90% in the ee and $e\mu$ ones. The systematic uncertainties included are the ones related to the fake-factor measurement and the statistical one of the control regions.

Table 8.3: Estimated yields for the non-prompt leptons background in the signal region for the six flavour and charge channels. The central value is shown together with its statistical uncertainty. The systematic yield variations are listed, deriving from the fake-factor variation within its statistical and systematic uncertainties. The labels f_e and f_μ indicate the fake-factors for electrons and muons, respectively.

	e^+e^+	e^-e^-	$\mu^+\mu^+$	$\mu^-\mu^-$	$e^+\mu^+$ + μ^+e^+	$e^-\mu^-$ + μ^-e^-
Central yield ($\pm$ stat.)	4.73 ± 1.90	6.68 ± 2.48	2.97 ± 0.73	1.84 ± 0.26	22.12 ± 3.94	14.96 ± 3.33
f_e stat. up	$+0.74$	$+0.95$	–	–	$+2.62$	$+2.29$
f_e stat. down	-0.74	-0.95	–	–	-2.62	-2.29
f_e syst. up	$+0.57$	$+1.10$	–	–	$+2.58$	$+3.01$
f_e syst. down	-1.79	-3.77	–	–	-8.50	-5.85
f_μ stat. up	–	–	$+0.41$	$+0.24$	$+0.93$	$+0.46$
f_μ stat. down	–	–	-0.41	-0.24	-0.93	-0.46
f_μ syst. up	–	–	$+3.22$	$+0.42$	$+11.68$	$+4.42$
f_μ syst. down	–	–	-0.43	-0.18	-1.39	-0.59

8.2 *WZ*+jets Background

Background processes from ZZ+jets, $t\bar{t}$+W/Z and triboson are estimated with simulation, whilst the WZ+jets modelled in simulation but its normalisation is derived from data. This is one of the main background processes, as events from $W^\pm Zjj \to \ell^\pm\nu\ell^\pm\ell^\mp jj$ can enter the signal region if one lepton from the decay of the Z boson is outside of the acceptance or not successfully reconstructed or identified. A control region is defined, and the normalisation is derived with a unique fit on the sum of contributions in all channels $e^+\mu^+$, μ^+e^+, $\mu^+\mu^+$, e^+e^+ and their charge conjugates, performed simultaneously in the signal region and in an additional low-m_{jj} control region. Details on the fitting methodology will be given in Chapter 9. The control region is defined to be kinematically similar to the signal region, but enriched in WZ+jets events by explicitly requiring a third lepton passing the signal selection. A lower p_T threshold at 15 GeV is used for this third lepton. Besides the same selection criteria as in the signal region for $E_\mathrm{T}^\mathrm{miss}$, number of jets, b-jet veto, m_{jj} and Δy_{jj}, additional requirements are applied to increase the purity of the region. A Z boson candidate is required by selecting an opposite-charge same-flavour dilepton pair and a lower limit on $m_{\ell\ell\ell}$ at 106 GeV is applied to reduce contributions from Z/γ^*+jets and $Z\gamma^*$ events. The distributions of the invariant mass of the three leptons and the pseudorapidity separation between the two jets are shown in Fig. 8.2. The prediction for the relative contribution between electroweak and strong production is performed with simulation using SM cross sections.

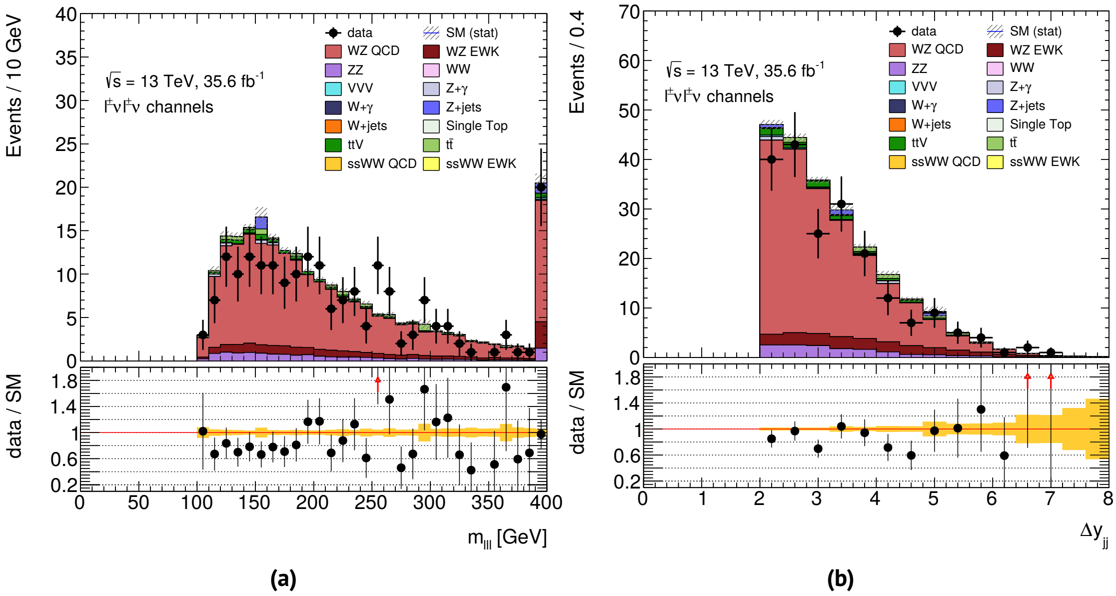


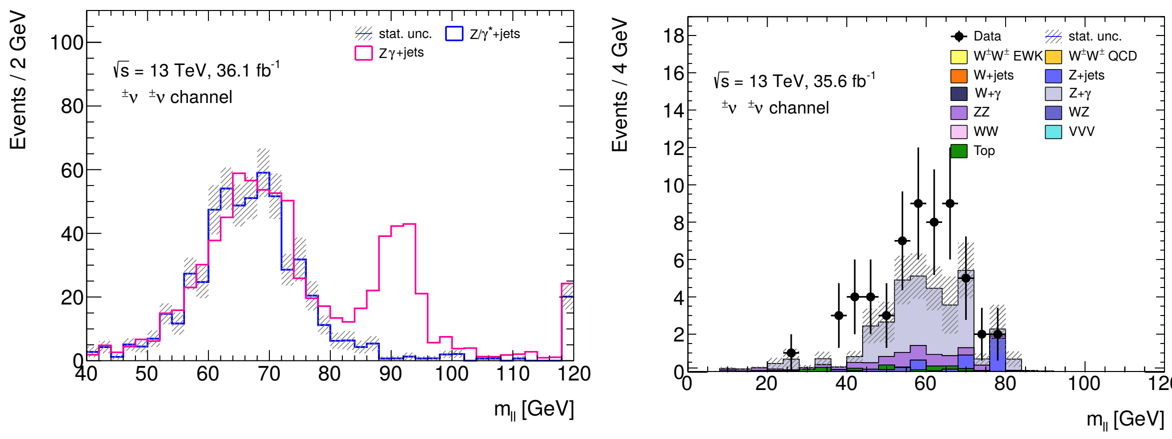
Figure 8.2: Distribution of the invariant mass of the three leptons (a) and rapidity separation between the two jets (b) in the WZ control region. Only statistical uncertainties are shown [145].

8.3 $V\gamma$+jets Background

If an electron from a photon conversion passes the nominal selection, events from $V\gamma$+jets could end up in the signal region. After the event selection no simulated $Z\gamma$+jets events are left in the signal region, but the modelling of photon conversions is studied in a data region enriched in $Z\gamma$+jets events and extended to $W\gamma$+jets, assuming a similar behaviour of the two processes.
This background is estimated by measuring a normalisation factor in a $Z\gamma \to \mu^{\pm}\mu^{\mp}\gamma$ control region and reweighting the simulated $W\gamma$+jets contribution in the signal region.
In order to construct the control region the nominal criteria detailed in Section 7.3 for the pre-selection and trigger matching are imposed together with the requirement of an electron in addition to the two muons from the Z. The selection for the region requires:

- $p_{\mathrm{T}}^{\mu} > 20(27)$ GeV for the leading (sub-leading) muon;
- $p_{\mathrm{T}}^{e} > 27$ GeV;
- $75 < m_{\mu\mu\gamma} < 100$ GeV;
- $E_{\mathrm{T}}^{\mathrm{miss}} < 30$ GeV.

Since a further sample is needed to cover the full p_{T}^{γ} and $\Delta R(\gamma, \ell)$ range not covered by the $Z\gamma$+jets sample, additional Drell-Yan Z/γ^{*}+jets samples are used, generated with POWHEG. An overlap between the two is present, in cases where an FSR isolated photon is produced in the Z/γ^{*}+jets sample, as shown in Fig. 8.3a in the distribution of the invariant mass of the two muons. The removal of the overlap is performed by identifying events which generated FSR photons in the Z/γ^{*}+jets sample which satisfy the same p_{T}^{γ} and $\Delta R(\gamma, \ell)$ requirements as photon generated in



(a) **(b)**

Figure 8.3: (a) Dimuon invariant mass for the $Z\gamma$+jets sample and the Z/γ^*+jets one from POWHEG. The overlap between the two samples is visible from the distribution in the region around 70 GeV. (b) Distribution of the dimuon invariant mass in the $Z\gamma$ control region. Only statistical uncertainties are shown [146].

the $Z\gamma$+jets sample, and removing such events from the $Z\gamma^*$+jets sample while keeping them in the $Z\gamma$+jets sample.

The final event yields in the $Z\gamma$ region are reported in Table. 8.4 for the $Z\gamma$+jets sample, the Z/γ^*+jets one and the other background processes contributing, i.e. other diboson processes, triboson and top production.

The normalisation factor obtained from the difference between the predicted yield and the data in the is $Z\gamma$ region is

$$\kappa_{W\gamma\text{+jets}} = \frac{N_{\text{data}} - N_{\text{other backgrounds}}}{N_{Z\gamma\text{+jets}} + N_{Z/\gamma^*\text{+jets}}} = 1.77 \pm 0.36,$$

which is applied as correction factor to the $W\gamma +$jets sample in the signal region. A systematic of 44% is attributed to it, which corresponds to a 0.77 absolute uncertainty, i.e. the full effect of the normalisation factor. This accounts for uncertainties in the region definition, which are visible in the not perfect modelling from the dimuon invariant mass shown in Fig. 8.3b. It takes also into account the overlap removal procedure and covers the differences between the $Z\gamma \to \mu^{\pm}\mu^{\mp}\gamma$ process where the scale factor is derived, and the $W\gamma$ region which is the target of the estimation[1].

[1]The rather large uncertainty assigned addresses the open points in the modelling of the region and in the estimation procedure which are currently being investigated in separated studies and not treated in the course of this thesis

Table 8.4: Event yields in the $Z\gamma$ region predicted by simulation and from data.

Process	Event yields
$Z\gamma$+jets	24.6±3.3
Z/γ^*+jets	3.0±1.5
$VV+VVV$	6.7±0.3
top	1.5±0.5
Total prediction	35.8±3.7
Data	57

8.4 Background from Charge Misidentification

Some events from opposite-charge processes can pass the selection if the charge of one lepton is wrongly reconstructed. The charge misidentification in general has a low probability, as explained in Chapter 6, but the large cross section of the background processes with respect to the signal amplifies the effect, making this background one of the major ones of the analysis. For muons, due to their higher mass compared to the electrons, it is less likely to undergo bremsstrahlung emission of photons. Furthermore tracks from muons can be reconstructed both in the ID and in the MS, improving the capability of track reconstruction compared to electrons that leave tracks in the ID only. In Fig. 8.4 the distribution for the leading lepton p_{T} is shown for opposite-charge and same-charge $Z \to \mu\mu$ simulated events and data dimuon events, in a region where a requirement of $|m_{\mu\mu} - m_Z| < 10$ GeV is imposed. As can be deduced from the plots, the relation between same-charge and opposite-charge is of the order of $\sim 10^{-5}$, therefore charge misidentification probability is considered negligible for muons.

The estimation of this background is hence performed in the ee, $e\mu$ and μe channels only.

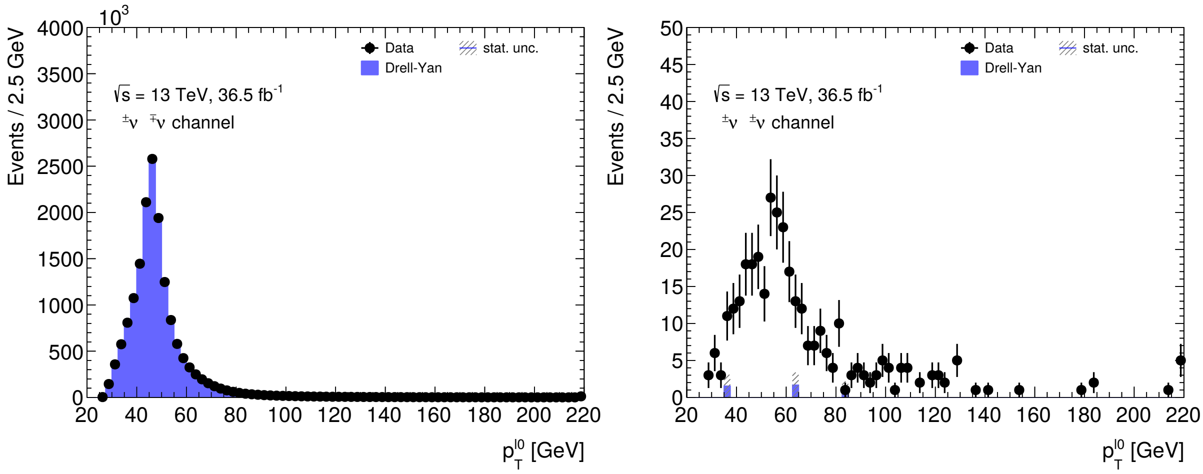
The background from charge misidentification is estimated in data, using the knowledge of the charge misidentification probabilities. These are extracted from simulation as discussed in Section 8.4.1. The actual background estimation is explained in Secs. 8.4.2 and 8.4.3, while the related systematic uncertainties are discussed in Section 8.4.5. Finally, Section 8.4.6 is dedicated to the estimation of the charge misidentification background in the ID+antiID region used for the non-prompt background estimation.

8.4.1 Probability Extraction

The measurement of the charge misidentification probabilities is the first step in the estimation, and is done exploiting the information from simulation. In order to do so, this needs to be corrected with simulation-to-data scale factors, so that the probability of misidentification well matches the one measured in data. Scale factors are measured as explained in Chapter 6 and specific results for the working point used in the analysis are presented in Section 6.4.4.

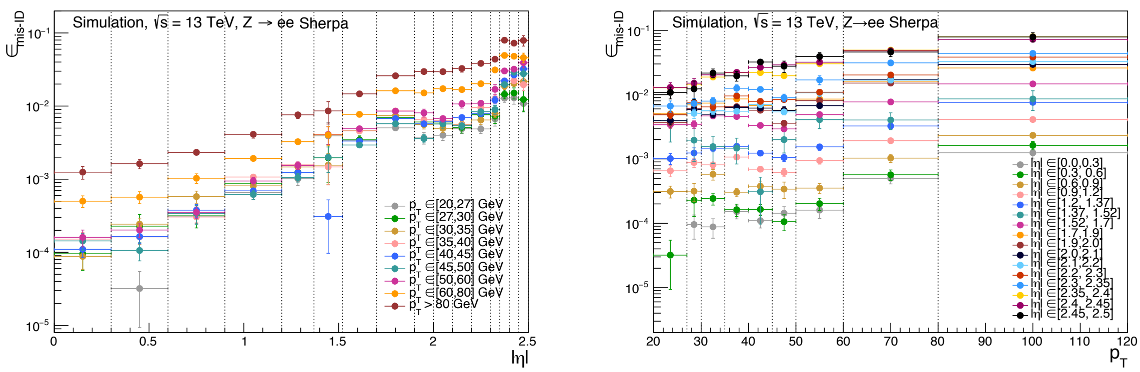
The extraction of the probabilities is performed on $Z \to ee$ events. This is the default process for such measurements due to its well-known modelling and the presence of only two electrons in the final state that makes it appropriate for charge misidentification studies.

After applying the scale factors to correct the $Z \to ee$ simulation for the mismodelling of the charge



(a) (b)

Figure 8.4: Distribution of the transverse momentum of the leading muon in a dimuon data sample and $Z \to \mu\mu$ simulated Drell-Yan sample for opposite-charge (a) and same-charge (b) events.



(a) (b)

Figure 8.5: Charge misidentification probabilities as a function of $|\eta|$ (a) and p_{T} (b) measured in simulated $Z \to ee$ events after applying the charge misidentification scale factors. Only statistical uncertainties are shown.

misidentification probability, the probabilities can be extracted by exploiting the information from simulation. The charge misidentification probability for a certain $(p_{\mathrm{T}}, |\eta|)$ bin i is hence

$$\varepsilon_i = \frac{(\text{wrong charge electrons})_i}{(\text{all prompt electrons})_i}. \tag{8.4}$$

The *wrong* charge and *prompt* definitions are those given in Chapter 6. The probabilities are shown in Fig. 8.5 as a function of $|\eta|$ and p_{T} separately.

8.4.2 Opposite-Charge Events Weighting

The estimation of the charge misidentification event yield in the signal region is obtained by weighting the opposite-charge data events, under the assumption that the opposite-charge events are dominated by prompt electrons. In this case the prompt requirements aim at selecting an electron not originating from meson or hadron decays, or from Dalitz decays. The weighting factor w used to weight the opposite-charge events is built using the probabilities obtained previously. Being $\varepsilon(p_{\mathrm{T}}, |\eta|) \equiv \varepsilon$ the probability of an electron to have its charge wrongly reconstructed and therefore $(1-\varepsilon)$ the probability of the correct charge reconstruction, for an event with genuine opposite-charge electrons:

- $(1-\varepsilon)^2$ is the probability of correct reconstruction, i.e. the probability that none of the two electrons undergoes charge misidentification;
- ε^2 is the probability of the electron-pair to be reconstructed as opposite charge, when both electrons undergo charge misidentification;
- $2\varepsilon(1-\varepsilon)$ is the probability to reconstruct the opposite-charge pair as same-charge pair, i.e. to wrongly reconstruct the charge of one electron.

The weight w is hence defined as the ratio between the probability of having a same-charge event and the one of having an opposite-charge event. The numerator is given by the combined probability of misidentifying the charge of one of the electrons, while the numerator accounts for the probability of misidentifying either both or none of the two, resulting in

$$w = \frac{\varepsilon_1(1-\varepsilon_2)+(1-\varepsilon_1)\varepsilon_2}{1-(\varepsilon_1+\varepsilon_2)+2\varepsilon_1\varepsilon_2}, \tag{8.5}$$

where ε_1 and ε_2 are the probabilities for the leading and sub-leading electron. In case of single-electron events only the probability for electrons ε_e is accounted for, and Eq. (8.5) simplifies to

$$w = \frac{\varepsilon_e}{1-\varepsilon_e},$$

which is used in the $e\mu$ or μe channel.

8.4.3 Energy Loss Correction

As extensively discussed in Chapter 6, the main cause of misidentification of electrons' charge is related to bremsstrahlung photon emission and subsequent production of an electron–positron pair. Electrons lose part of their energy in the material interaction and this leads to a leakage of the energy deposit outside the fixed electromagnetic calorimeter cluster and to a worsening of the energy resolution. For this reason, electrons reconstructed with the wrong charge have on average a lower energy and a wider energy resolution distribution compared to the correctly-reconstructed ones. The phenomenon is visible in Fig. 8.6, where the distribution for same-charge events both in data and simulation is shifted towards lower energies.

The weighting of the opposite-charge data with the charge misidentification probability is therefore not sufficient to obtain a correct kinematic distribution in the same-charge estimation, as it does not correctly describe the four-momentum of the wrongly-reconstructed electrons. To account for this, momenta are corrected with two terms derived with simulation and the correction is applied

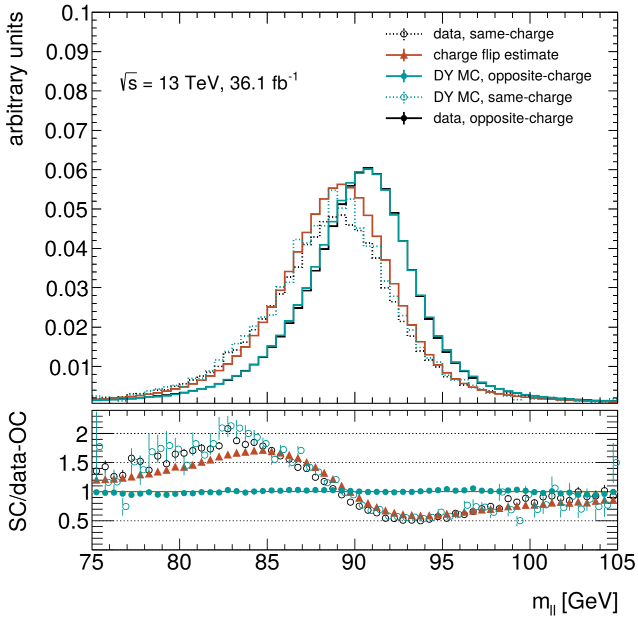


Figure 8.6: Distribution of the invariant mass of the two electrons for opposite-charge (black solid line) and same-charge (black dashed line) data, and opposite-charge (cyan solid line) and same-charge (cyan dashed line) simulation. The red line shows the estimation of the same-charge events based on the described estimation. Distributions are normalised to one. The shift towards lower p_{T} values and the broader shape for same-charge electrons are visible, for data, simulation prediction and the estimation. The lower pad shows the ratio of the same-charge distribution to the opposite charge data distribution. Only statistical uncertainties are shown.

to the electron's p_{T} as

$$p_{\mathrm{T}}^{\text{corrected}} = \frac{p_{\mathrm{T}}^{\text{original}}}{1+\alpha} + \Delta E, \tag{8.6}$$

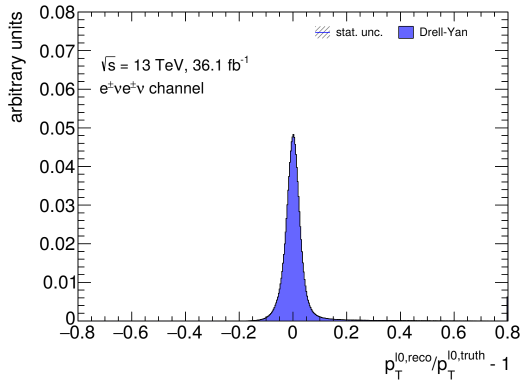
where $p_{\mathrm{T}}^{\text{original}}$ is the p_{T} of the electron used for the estimation. The term $(1+\alpha)$ is the *residual energy scale* while ΔE is the noise term used to widen the distribution.

Both terms are derived based on the *energy response*, defined as $\left(p_{\mathrm{T}}^{\text{reco}}/p_{\mathrm{T}}^{\text{truth}} - 1\right)$, which quantifies the difference between the p_{T} of the original electron and the reconstructed quantity.

This is depicted in Fig. 8.7 for correctly and wrongly reconstructed electrons, as a one-dimensional distribution profiled over the pseudorapidity of the leading electron. The information on the correct or wrong charge reconstruction is obtained from the comparison with the original charge of the electron as detailed in Chapter 6. The energy response distribution for wrong charge electrons is shifted towards values smaller than zero, which is mainly visible in Fig. 8.7b. From there, the narrower distribution that characterises correct-charge electrons compared to the wrong-charge ones emerges as well.

The term α is derived in bins of $|\eta|$ as

$$\alpha = \frac{\left(\frac{p_{\mathrm{T}}^{\text{reco}}}{p_{\mathrm{T}}^{\text{truth}}} - 1\right)^{\text{correct}}}{\left(\frac{p_{\mathrm{T}}^{\text{reco}}}{p_{\mathrm{T}}^{\text{truth}}} - 1\right)^{\text{wrong}}}, \tag{8.7}$$



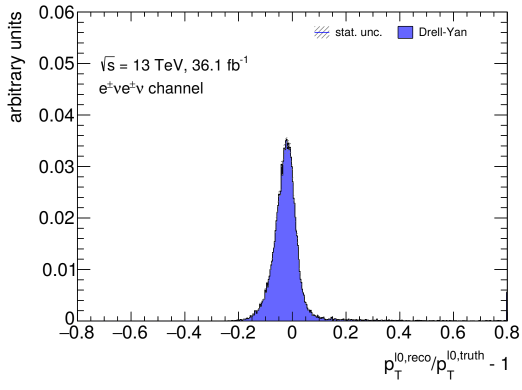


Figure 8.7: Energy response $\left(p_{\mathrm{T}}^{\mathrm{reco}}/p_{\mathrm{T}}^{\mathrm{truth}}-1\right)$ distribution for correct- (a) and wrong-charge (b) electrons.

which quantifies the difference in the energy response for wrongly and correctly reconstructed electrons.

The term ΔE is distributed according to a Gaussian distribution centred around 0, with its width corresponding to a *residual constant term* c. Its definition is also based on the energy response derived differentially in $|\eta|$ bins as

$$c=\left\langle\frac{p_{\mathrm{T}}^{\mathrm{reco}}}{p_{\mathrm{T}}^{\mathrm{truth}}}\right\rangle^{\mathrm{wrong}}-\left\langle\frac{p_{\mathrm{T}}^{\mathrm{reco}}}{p_{\mathrm{T}}^{\mathrm{truth}}}\right\rangle^{\mathrm{correct}}. \tag{8.8}$$

The term $1+\alpha$ and the residual constant term c are shown in Fig. 8.8 as a function of $|\eta|$.

Since in data it is not possible to identify the misreconstructed electron, the corrections of Eq. (8.6) are applied randomly on the leading or sub-leading electron. One of the two is randomly picked, based on the probability of both electrons to undergo charge misidentification, and the electron for which no energy correction is applied is considered the one correctly reconstructed. This procedure determines therefore the charge of the full same-charge event.

In Fig. 8.6 the m_{ee} distribution obtained after the scaling and the smearing of opposite-charge data events is shown compared to data and simulation distributions. The figure also shows that the statistical uncertainties of the estimation are reduced, compared to the charge misidentification estimated with simulated events. Some residual mismodelling is although visible in the estimation, which will need to be accounted for in the evaluation of the systematic uncertainties.

8.4.4 Validation

In order to validate the modelling of the estimation prior to performing it in the signal region, a validation region enriched in charge misidentified electrons is defined. This is the same-charge inclusive validation region. It requires two same-charge leptons that pass the nominal lepton selection criteria, having an invariant mass within 15 GeV of the nominal Z boson mass. This region has a high purity in background from charge misidentification. The nominal requirement in the ee channel of $|\eta^e|<1.37$ is not applied in this region, as the procedure needs to be validated also

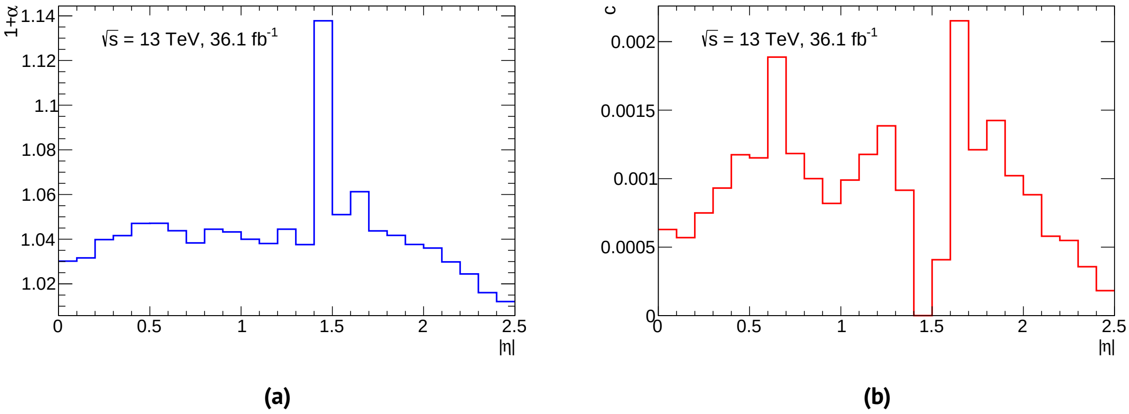


Figure 8.8: Residual energy scale correction $(1+\alpha)$ (a) and constant term (b) used for the energy correction procedure. They account for the shift and smearing of the electron's energy due to the differences in the energy reconstruction for electrons with correct and wrong charge, and are derived as a function of $|\eta|$.

for electrons in the $e\mu$ or μe channel which can have $|\eta^e| > 1.37$. The estimation is performed in the same way as for the signal region, i.e. by weighting and applying the energy correction to opposite-charge data. The validation is performed in the ee channel and some of the distributions are shown in Fig. 8.9. The estimation is shown with statistical uncertainties only, and data agree within 10-20% except for the high transverse momentum tails of the p_{T} distributions where the agreement is worsened by the presence of non-prompt leptons for which the estimation is not shown. The $m_{\ell\ell}$ distribution shows that the energy loss is overcorrected, leading to an underestimation of the charge misidentification yield, as more electrons fall outside of the kinematic acceptance.

8.4.5 Systematic Uncertainties

Systematic uncertainties in the estimation of the charge misidentification background are estimated from the systematic uncertainties in the scale factor measurement and from the energy correction procedure.

The sources of uncertainties related to the scale factor are explained in details in Section 6.4.3. The probabilities are determined with the scale factor varied up and down within their combined statistical and systematic uncertainty, and the estimate is repeated. The difference in the estimated background yields is assigned as a systematic uncertainty. To estimate the uncertainty related to the overcorrection of the energy loss, which caused the not perfect modelling of the estimation, the full correction effect is assigned as a systematic uncertainty. This is done in order to safely cover the uncertainties related to the energy correction technique, and it is considered adequate since the systematic uncertainties related to the estimation are not limiting the analysis precision.

The final background yields estimated in the signal region in each channel are reported in Table 8.5, together with their variation due to the considered systematic sources.

Table 8.5: Estimated yields for the background due to electron charge misidentification in the signal region. Central value and systematic uncertainties are listed in the table for the four channels, separately for positive and negative final states. The $\mu\mu$ channel is omitted as the muon charge misidentification probability is negligible.

	e^+e^+	e^-e^-	$e^+\mu^+$	$e^-\mu^-$	μ^+e^+	μ^-e^-
Central yield ($\pm$ stat.)	1.33 ± 0.08	1.17 ± 0.08	2.36 ± 0.28	1.85 ± 0.21	1.11 ± 0.14	1.10 ± 0.15
SF up	$+0.13$	$+0.10$	$+0.08$	$+0.06$	$+0.03$	$+0.03$
SF down	-0.10	-0.10	$+0.03$	-0.04	-0.02	-0.02
Energy correction	$\pm$ 0.05	$\pm$ 0.03	$\pm$ 0.20	$\pm$ 0.18	$\pm$ 0.01	$\pm$ 0.01

8.4.6 Charge Misidentification in ID-antiID Region

The estimation of the background from non-prompt leptons is presented in Section 8.1, which is the reference for details on the estimation procedure and the electron antiID selection used in the following. When estimating the background from non-prompt leptons, the contribution from prompt and charge misidentified leptons inside the ID+antiID sample has to be subtracted. The charge misidentification contribution is estimated from data, with the same technique as the one used for the estimation of the background from charge misidentification in the signal region explained in Section 8.4. However, the leptons in this sample are selected with different identification criteria. One of the electrons is an ID electron, the other an antiID one. To estimate the charge misidentification contribution weights have to be constructed as in Eq. (8.5), and hence additional charge misidentification probabilities are required for the electron that follows the antiID definition. These are measured on data with a tag-and-probe technique, but nevertheless preliminary test are performed on $Z \to ee$ simulated events to verify the validity and applicability of the method. The data and simulated event samples used are described in Section 6.1. In particular 33.9 fb^{-1} of $Z \to ee$ data events collected in 2016 at $\sqrt{s} = 13$ TeV has been used, and $Z \to ee$ POWHEG simulated samples are used for the validation tests.

Probabilities extraction with tag-and-probe method The likelihood method as used in Chapter 6 is not suitable for the probability measurement in the ID+antiID region. This is because the likelihood function assumes the same identification criteria for the two electrons, without differentiating between the two objects by construction. Therefore in this case, where two different selections are applied to the electrons, a tag-and-probe method is used, whose basic concepts are introduced in Section 5.2. The tag object is the electron that passes the ID selection, and the probe the antiID one. In the following the terms tag and ID electron will be used interchangeably, as well as probe and antiID.

One of the assumptions in order to use the tag-and-probe method to measure charge misidentification probabilities, is that the tag electron did not undergo charge misidentification. The tighter identification, isolation and kinematic requirements on the tag electron make its probability to undergo charge misidentification extremely low compared to the probe's one, nevertheless this assumption is to be proven.

This is done in simulation, where the probabilities are extracted for the probe by comparing the

Table 8.6: Selection for the ID+antiID validation region. The *nominal* criteria refer to the signal region objects and events selection as in Table 7.2 and 7.4.

ID+antiID validation region
Nominal pre-selection, p_{T}^{ℓ} and $m_{\ell\ell}$
$N_{\text{ID-}e}$ = 1 and $N_{\text{antiID-}e}$ = 1
Truth matching
$N_{\ell\,\text{same-charge}}$ = 2
$N_j < 2$ (nominal jet selection)
$80 < m_{ee} < 100$ GeV
$E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV
$N_{b\text{-jet}} = 0$

reconstructed charge to the one from simulation, following what is discussed in Chapter 6. They are shown in Fig. 8.10a and 8.10c, and compared to nominal probabilities for ID electrons of Fig. 8.5, where the standard methodology was followed to measured the probability for nominal, i.e. ID electrons. The probabilities for the probe electrons range from 10^{-2} to 10^{-1}, being one or two order of magnitude larger that the tag one, depending on the (p_{T}, $|\eta|$) bin.

An additional validation is performed as a *truth closure* test, by comparing the probabilities for the probe just obtained using the information from simulation with the tag-and-probe method applied to simulated events. The two are compared, the former shown in Fig. 8.10a and Fig. 8.10c, and the latter in Fig. 8.10b and 8.10d. The probabilities are considered to be compatible.

Finally, the probabilities for antiID electrons can be measured on $Z \to ee$ data events with the tag-and-probe. The measurement is performed by selecting same-charge events and the charge misidentification probabilities for the probe, i.e. the antiID electron in this case, can be measured as the number of same-charge events over all events

$$\varepsilon = \frac{N_{\text{wrong charge antiID electrons}}}{N_{\text{all antiID electrons}}} \equiv \frac{N_{\text{same-charge}}}{N_{\text{all}}}. \tag{8.9}$$

The measured probabilities for antiID electrons are shown in Fig. 8.11.

Weights construction and estimation in the ID+antiID region The weights for the ID+antiID region are constructed based on Eq. (8.5), but in this case either the ε_1 or ε_2 term for each event refers to the charge misidentification probability for antiID electron just obtained on data with the tag-and-probe method. The scaling of the opposite-charge events of the ID+antiID sample is ultimately performed, assuming a good purity in prompt electrons. Prompt electrons are defined as for the nominal selection discussed in Section 8.4.2 and are electrons not originating from meson, hadron or Dalitz decays. The purity of the region is checked evaluating the impact of the truth matching requirement defined in Section 7.3.2, by comparing distributions with and without this selection. This is shown in Fig. 8.12 where the distributions of the dilepton transverse momenta are displayed, with and without the prompt requirement. The opposite-charge region used for the estimation is considered pure enough, as the introduction of the truth matching has no impact on the yields.

The estimation is finally performed in the ID+antiID region, where opposite-charge events are weighted to obtain same-charge events to be used in the subtraction during the fake estimation procedure. The four momentum of the charge-misidentified electron in this region is affected by the energy correction due to the charge misidentification estimation in addition to the p_{T} + $p_{\mathrm{T}}^{\mathrm{varcone}}$ of the non-prompt estimation.
The estimation is checked in a dedicated validation ID+antiID region, constructed to be as pure as possible in events from charge misidentification. The region is described in Table 8.6. Fig. 8.13 shows the modelling in the validation region for the η and p_{T} of the leading lepton, which is good within the statistical uncertainties with a disagreement of 10% on average.
The systematic uncertainties on the background evaluation are estimated in the ee channel by considering the difference between data yields in the validation region and the predictions, coming from the estimated charge misidentification events, and further smaller contributions evaluated with simulation. In total 820.58 charge misidentification events are estimated in this region, with 5.28 events from other processes. Data amounts to 760 events, hence the discrepancy between data and predictions is of 10%. This is considered as systematic uncertainty on the method and propagated to the non-prompt background estimation. The propagation does not account for the transfer from the validation region to the signal region, since the impact of this systematic is considered small and not a limiting factor to the final precision of the measurement.

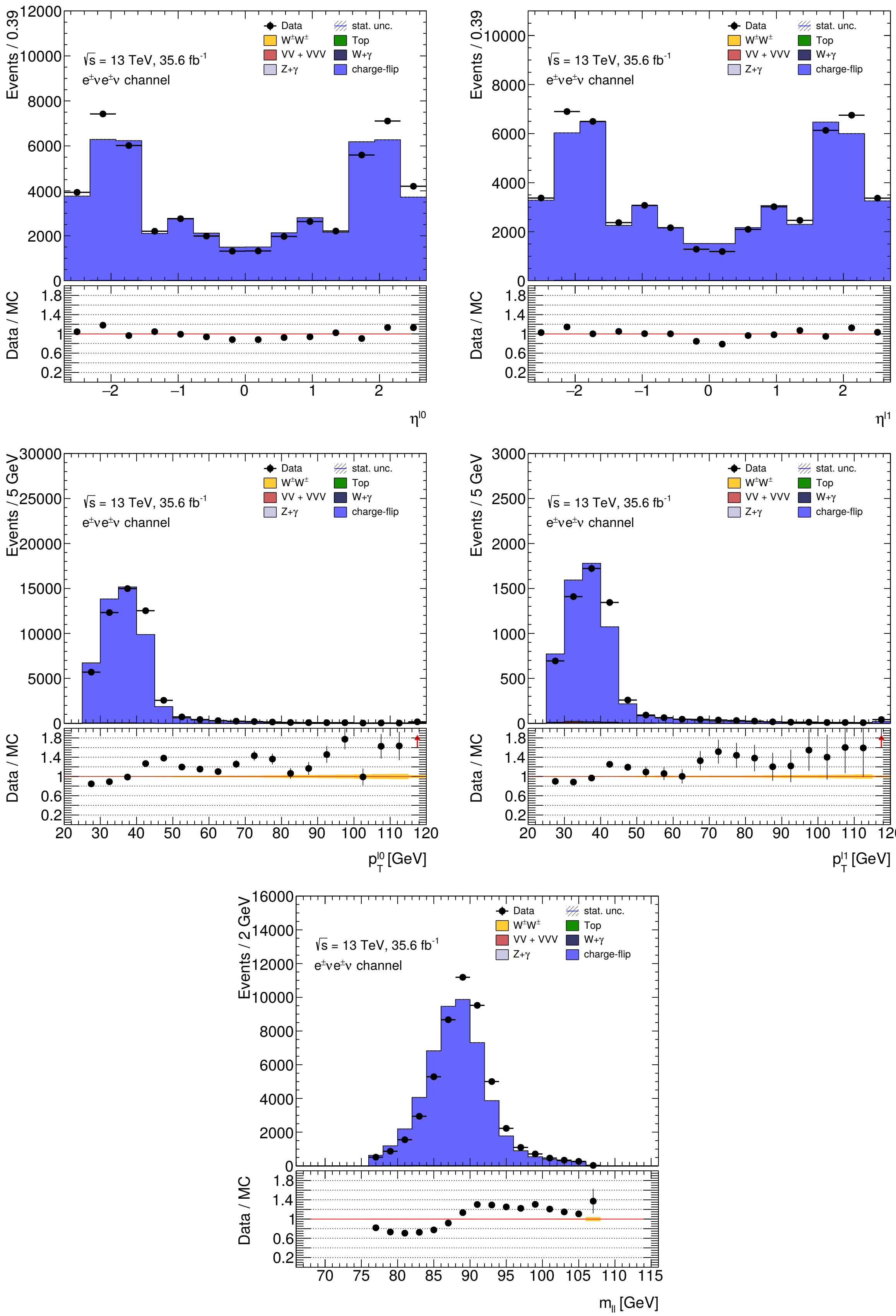


Figure 8.9: Distributions of η and p_T for the leading and sub-leading electron, and dilepton invariant mass for the ee channel in the same-charge inclusive validation region (see Table 8.7 for detailed description).

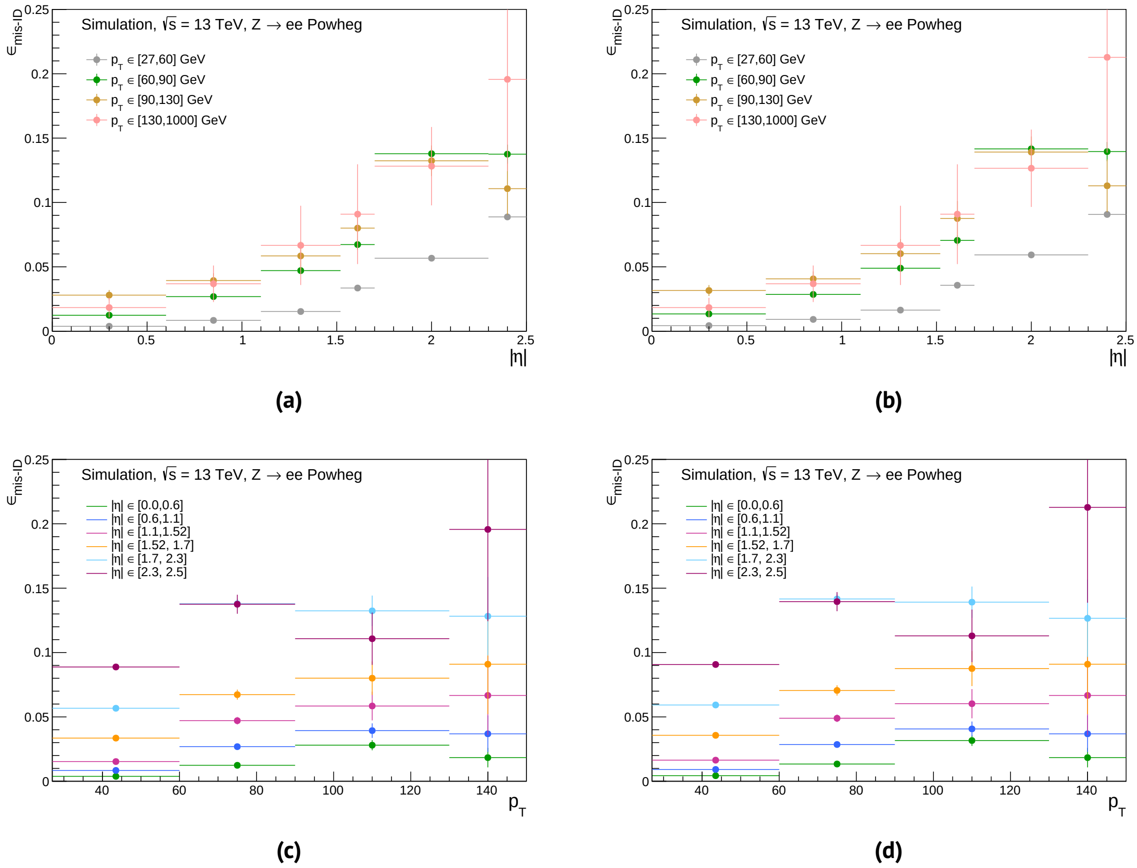


Figure 8.10: Charge misidentification probabilities measured for antiID electrons by exploiting the information from simulation (a,c) and with the tag-and-probe method (b,d) on simulated $Z \to ee$ events as a function of $|\eta|$ (top) and p_{T} (bottom). Only statistical uncertainties are shown.

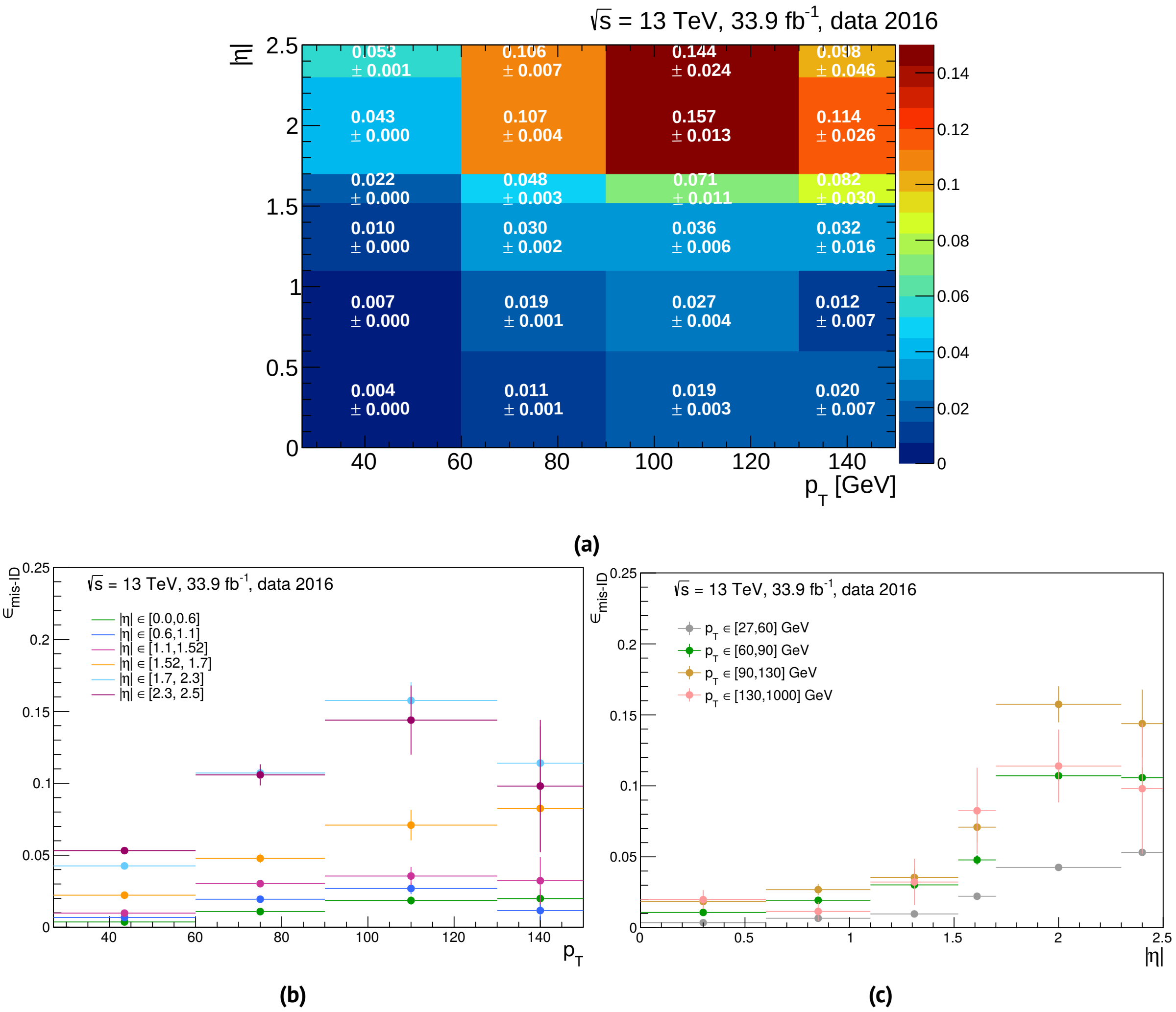
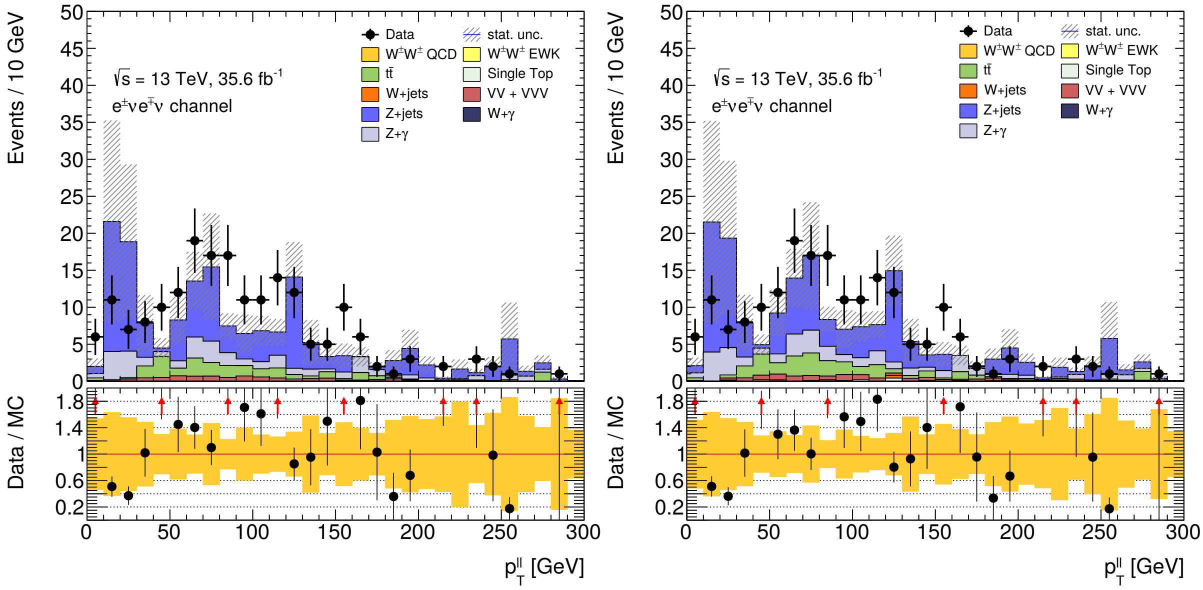


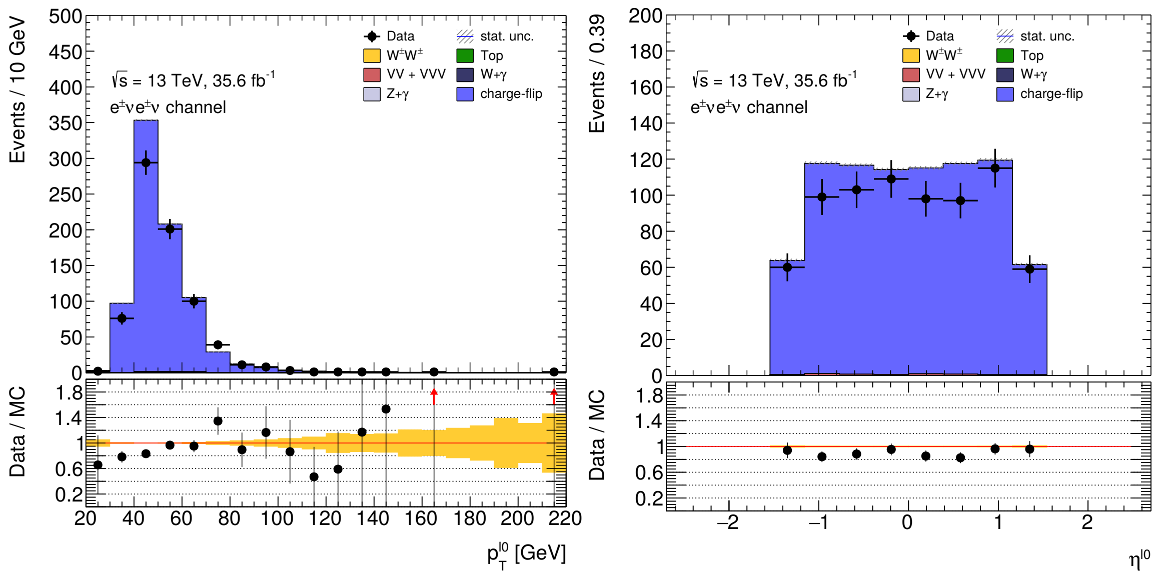
Figure 8.11: Charge misidentification probabilities for antiID electrons as a two-dimensional map in (p_T, $|\eta|$) (a) and as a function of the electron p_T (b) and $|\eta|$ (c). Only statistical uncertainties are shown.



(a)

(b)

Figure 8.12: Distributions in the opposite-charge ID+antiID region of the p_T of the dilepton pair in the ee channel without (a) and with (b) prompt lepton requirement (see Section 7.3.2).



(a)

(b)

Figure 8.13: Transverse momentum (a) and pseudorapidity distributions (b) for the leading lepton in the ID+antiID validation region in the ee channel. Only statistical uncertainties are shown.

8.5 Background Modelling Validation

The modelling of the estimated background processes is tested in validation regions, which are designed to have a negligible contribution from signal, and to be enriched with events of the processes that is to be checked. They are not used to derive normalisation factors, hence they do not play a role in the final fit, as control regions do.
Seven validation regions are defined, which are discussed in the following and summarised in Table 8.7. The modelling in each region is shown in some illustrative distributions, that compare data and estimated backgrounds before the fit.

- *Same-charge inclusive.* It is used to test the background from charge misidentification and non-prompt leptons. The selection criteria are the same as the signal region, and requiring no b-jets limits the presence of $t\bar{t}$ events which contribute to both backgrounds. The charge misidentification background is enhanced by requiring the invariant mass of the dilepton system to be within 15 GeV of the nominal Z boson mass in the ee channel only. The estimation in this region describes data well, and distributions can be seen in Fig. 8.14.

- *Same-charge low-Δy_{jj}.* It provides a test of the QCD-induced backgrounds, WZ, non-prompt lepton and charge misidentification backgrounds. The only difference with respect to the nominal signal region selection is the reversed requirement on Δy_{jj}, being $|\Delta y_{jj}| < 2$. A good agreement within the large statistical uncertainties is found in this region, and is shown in Fig. 8.15.

- *Same-charge low-m_{jj}.* It is used to test the non-prompt lepton background. The selection is the same as the signal region with the exception of the requirement on the dijet invariant mass which is reversed to $m_{jj} < 500$ GeV and the one on Δy_{jj} which is removed. The missing Δy_{jj} requirement is also the difference between this validation region and the low-m_{jj} control region used in the fit. Even without the lower bound on m_{jj}, which is removed to allow more events to pass the region selection, the statistical uncertainty in this region is quite large. Nevertheless it is possible to ensure that no significant deviation between data and predictions is present. Distributions are shown in Fig. 8.16.

- *Same-charge low-n_j.* It is used to test the modelling of the WZ, the non-prompt lepton and charge misidentification backgrounds, together with other prompt backgrounds. The selection on the jet multiplicity is inverted with respect to the signal region and only events with at most one jet are accepted, where the counting is done with jets selected with pre-selection requirements detailed in Section 7.3.1. Data are described well by the predictions, even within the small statistical uncertainties, as visible in Fig. 8.17.

- *Same-charge top fakes.* It is dedicated to the non-prompt lepton background validation. Events are accepted only if they have exactly one b-jet, hence they are mostly $t\bar{t}$ events, where the b-jet is identified as a lepton. The estimation well describes the data in this region almost everywhere, with a non-significant disagreement observed especially in the high-lepton-p_{T} region. Distributions are shown in Fig. 8.18.

- *Same-charge top-EW*. It is used to test the modelling of the charge misidentification and $t\bar{t}$+W/Z backgrounds. Events with exactly two b-jets are selected, and they are mainly events where one of the leptons undergoes charge misidentification or is a non-prompt lepton. The jet multiplicity is checked in this region. It is shown in Fig. 8.19a and is found to provide a good description of the data.

- *Trilepton inclusive*. It provides a test for the WZ modelling. Exactly three charged leptons are required, the third one passing the lepton-veto selection and having $p_{\mathrm{T}} > 15$ GeV. The selection criteria on jets are dropped and no b-jets requirement is applied to reject non-prompt leptons. For this region all flavour-charge channels are combined and a good agreement is found between data and predictions except for a mismodelling due to the small number of available events in the lower $m_{\ell\ell\ell}$ bins, as visible in Fig. 8.19b.

In general the validation regions prove that the estimation is reliable for all background processes, both modelled with simulation and estimated with data. The yields can hence be obtained for the signal region selection, which will be used as input in the fit for the cross section extraction. This will be the topic of the next chapter.

Table 8.7: Summary of validation regions. The differences with respect to the signal region are detailed, together with the background processes that the region is targeting and the channels where the validation is checked. The signal-region selection on $m_{jj} > 500$ GeV is not applied in any validation region. A requirement on $m_{jj} > 200$ GeV is applied everywhere unless specified otherwise.

Validation region	Differences wrt signal region	Backgrounds tested
Same-charge inclusive	No $E_{\mathrm{T}}^{\mathrm{miss}}$ requirement No Δy_{jj} requirement No m_{jj} requirement	Charge misidentification , non-prompt leptons
Same-charge low-Δy_{jj}	Reversed Δy_{jj} selection	QCD-induced (similar composition as signal region)
Same-charge low-m_{jj}	$m_{jj} < 500$ GeV	Non-prompt leptons, (similar composition as signal region)
Same-charge low-n_j	Jets at pre-selection $N_j < 2$ No jet-based quantities requirements	WZ, non-prompt leptons Charge misidentification, prompt leptons
Same-charge top-fakes	$N_{b\text{-jet}}$ = 1 No Δy_{jj} requirement No m_{jj} requirement	$t\bar{t}$ (mainly non-prompt leptons with charge misidentification contributions)
Same-charge top-EW	$N_{b\text{-jet}}$ = 2 No Δy_{jj} requirement No m_{jj} requirement	$t\bar{t}$+W/Z (mainly non-prompt leptons with charge misidentification contributions)
Trilepton inclusive	$p_{\mathrm{T}}^{\ell 2} > 15$ GeV 3-leptons requirement	WZ

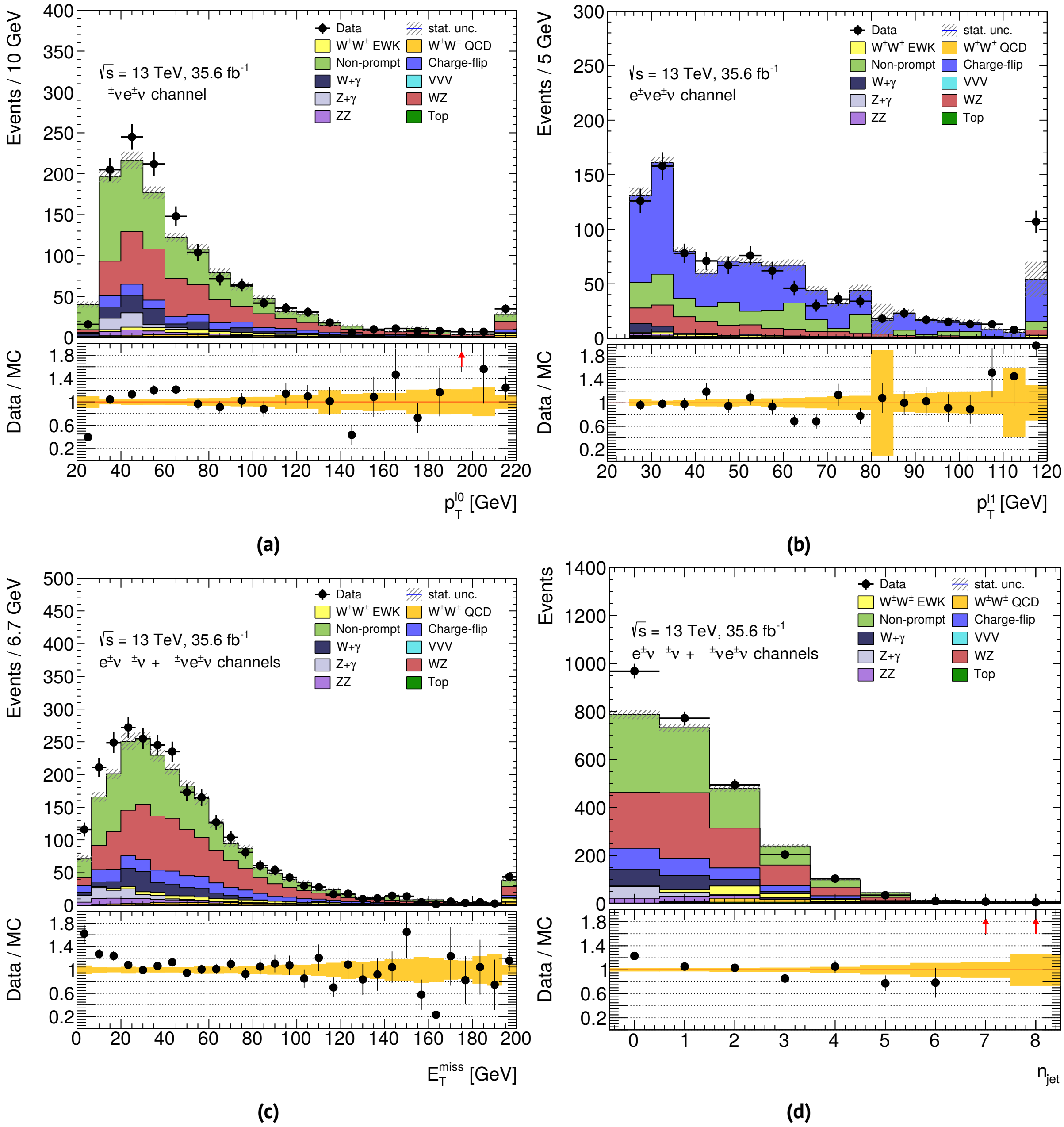


Figure 8.14: Distributions of the leading lepton p_{T} in the $\mu^{\pm}e^{\pm}$ channel (a), sub-leading lepton p_{T} in the $e^{\pm}e^{\pm}$ channel, $E_{\mathrm{T}}^{\mathrm{miss}}$ (c) and jet multiplicity (d) in the combined $e^{\pm}\mu^{\pm}$ + $\mu^{\pm}e^{\pm}$ channels in the same-charge inclusive validation region. Only statistical uncertainties are shown.

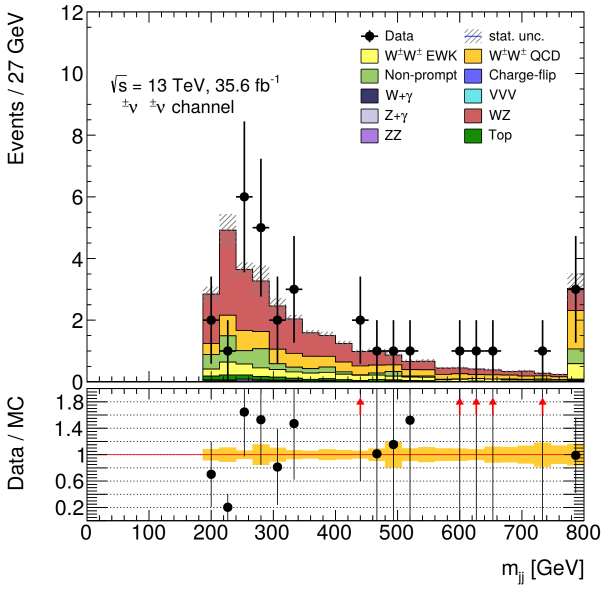
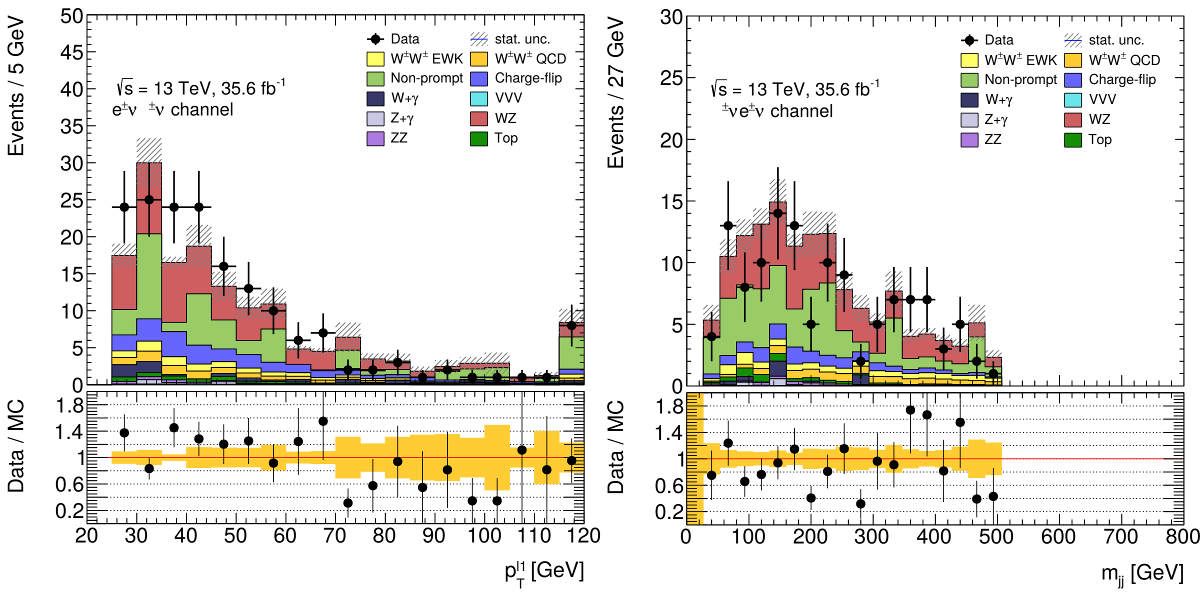
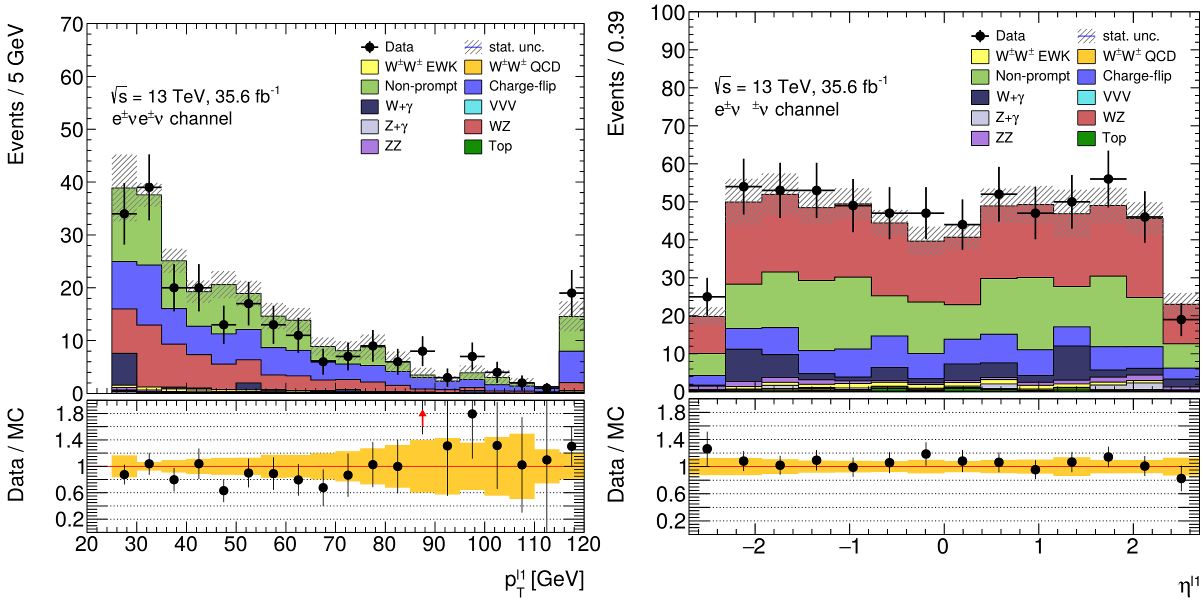


Figure 8.15: Distribution of the dijet invariant mass in the $\mu^{\pm}\mu^{\pm}$ channel in the same-charge low-Δy_{jj} validation region. Only statistical uncertainties are shown.



(a) **(b)**

Figure 8.16: Distributions of the sub-leading lepton p_{T} in the $e^{\pm}\mu^{\pm}$ channel (a) and dijet invariant mass in the $\mu^{\pm}e^{\pm}$ channel (b) in the same-charge low-m_{jj} validation region. Only statistical uncertainties are shown.



(a) (b)

Figure 8.17: Distributions of the sub-leading lepton p_T in the $e^\pm e^\pm$ channel (a) and the sub-leading lepton pseudorapidity in the $e^\pm\mu^\pm$ channel (b) in the low n_j validation region. Only statistical uncertainties are shown.

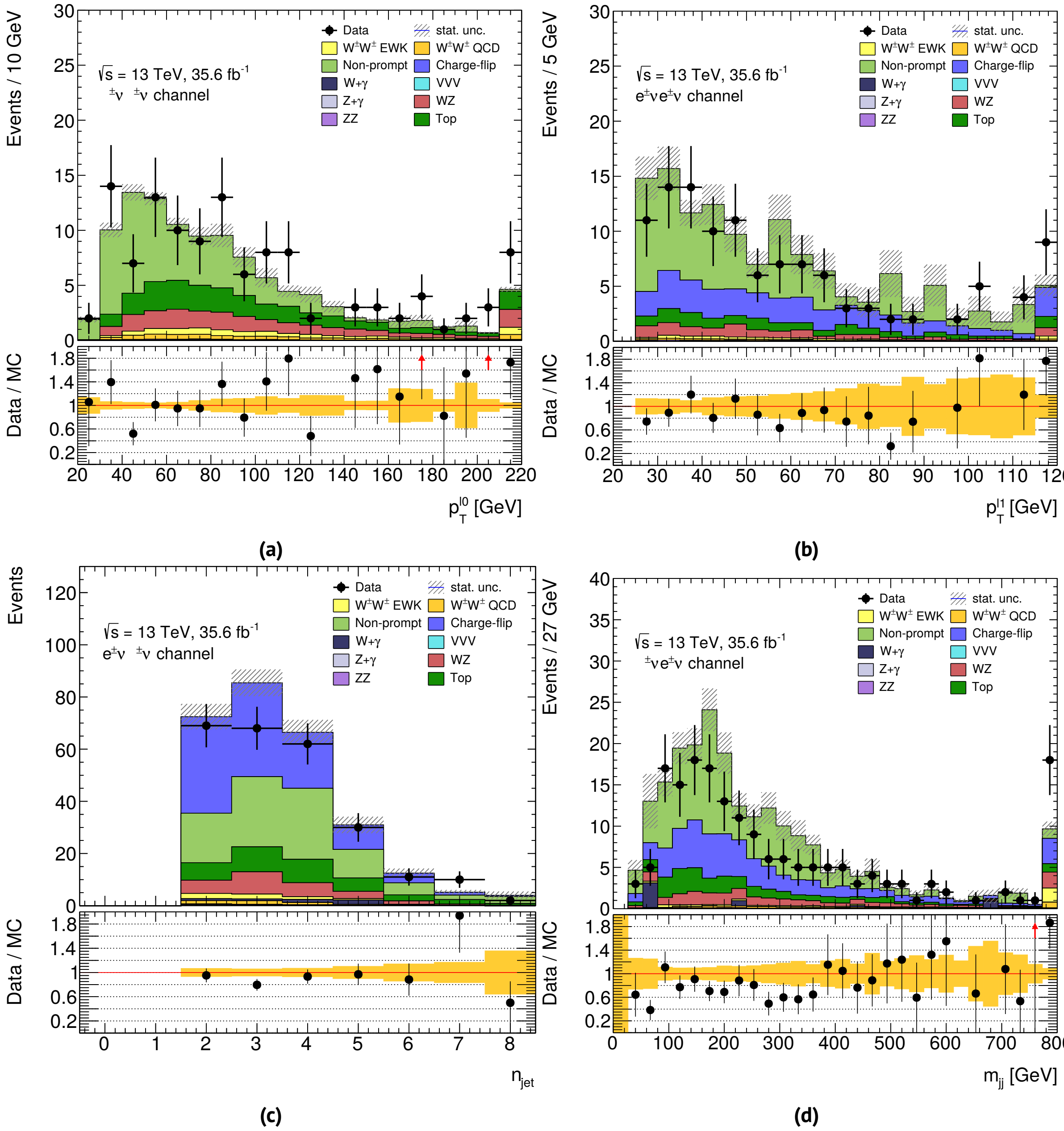
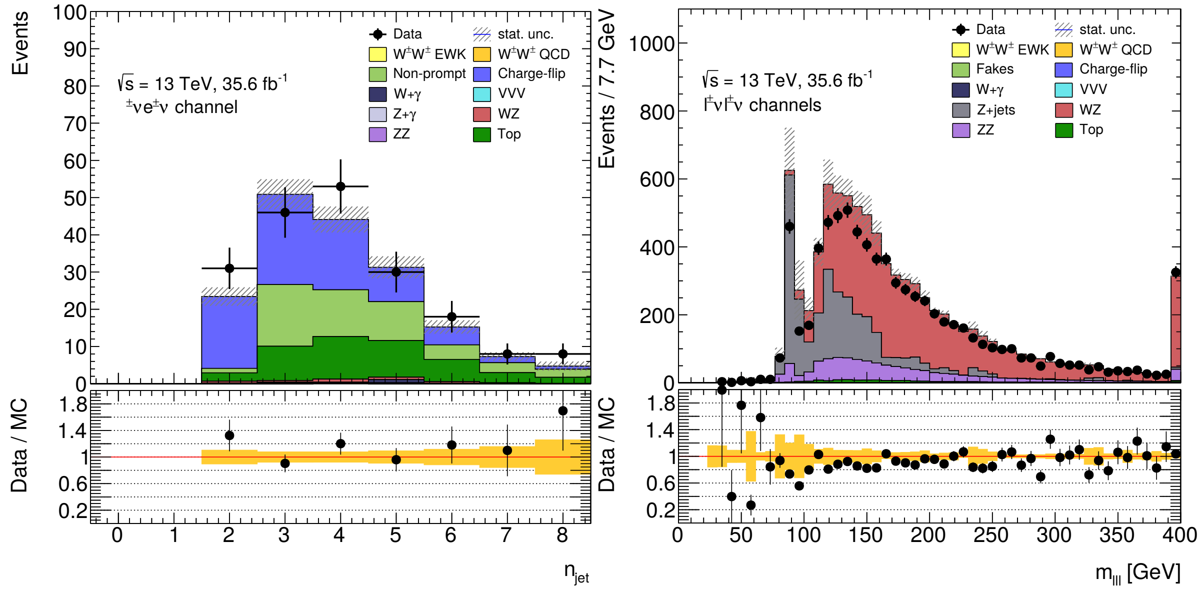


Figure 8.18: Distributions of the leading lepton p_T in the $\mu^\pm\mu^\pm$ channel (a), sub-leading lepton p_T in the $e^\pm e^\pm$ channel (b), jet multiplicity in the $e^\pm\mu^\pm$ channel (c) and dijet invariant mass in the $\mu^\pm e^\pm$ channel (d) in the same-charge top-fakes validation region. Only statistical uncertainties are shown.



(a) **(b)**

Figure 8.19: Distributions of the jet multiplicity for the $\mu^{\pm}e^{\pm}$ channel in the same-charge top-EW validation region (a) and the trilepton invariant mass for all channels combined in the trilepton inclusive validation region (b). Only statistical uncertainties are shown.

9 Cross Section Measurement

This chapter is dedicated to the procedure used to translate event yields into the measurement of the process cross section in a fiducial region. The bases of the methodology are given in Section 9.1 and Section 9.2 describes the predictions in the fiducial volume. The systematic uncertainties and the description of the fitting procedure are presented in Section 9.3 and Section 9.4 respectively. The expected results are discussed in Section 9.5, while Section 9.6 is dedicated to the results on observed data. In Section 9.7 studies are presented, which prove the independence of the measured cross section with respect to the signal predictions. Finally, a brief overview and outlook an the interpretation of results is given in Section 9.8.

9.1 Methodology

9.1.1 Maximum Likelihood Method

In this analysis a *profile likelihood approach* is used to extract the amount of signal and background in a signal region, and to accurately account for the related uncertainties. This is based on a maximum likelihood fit [147–149], where the effect of systematic uncertainties is taken into account by a set of parameters. These are the so-called *nuisance parameters* and need to be estimated from data or from some auxiliary measurement, but are not of primary interest in the fit result.
The *signal strength* μ is generally (one of) the parameter(s) of interest of the fit. This is defined as

$$\mu = \frac{\sigma_{\text{meas}}}{\sigma_{\text{theo}}}, \tag{9.1}$$

which is the ratio between the measured cross section of the process under investigation and its theoretical prediction, the SM one in the case of this analysis.
The signal strength is a continuous parameter that corresponds to zero in case of the background-only hypothesis, and 1 in case of the hypothesis of signal and background predicted by the theory.
Given S signal events and B background events, the likelihood function is built starting from the general definition

$$\mathcal{L}(S) = \frac{(S+B)^n}{n!} e^{-(S+B)}, \tag{9.2}$$

modified so that the signal is multiplied by the parameter μ. Furthermore, a binned kinematic space is usually considered, so that the likelihood function becomes a product of several Poisson terms, as events in each bin are expected to follow a Poissonian distribution.
Considering the regions r and the bins b, and neglecting for the moment the systematic uncertainties, the likelihood function is

$$\mathcal{L}_{\text{main}}(n_{rb}, \mu) = \prod_r \prod_b \frac{(\mu S_{rb} + B_{rb})^{n_{rb}}}{n_{rb}!} e^{-(\mu S_{rb} + B_{rb})}, \tag{9.3}$$

where n_{rb} is the number of observed events in each region and bin.
By introducing the systematic uncertainties, the signal and each background acquire a dependency

on the respective nuisance parameters $\vec{\theta}$, and the shape and normalisation of their templates is modified when varying the values for these parameters. Therefore, maximising the likelihood function yields to a simultaneous measurement of the nuisance parameters and the parameter(s) of interest μ, with best-fit values $\hat{\vec{\theta}}$ and $\hat{\mu}$ respectively.
Usually $\vec{\theta}$ are provided with a central value, determined from what is generally addressed as *auxiliary measurement*, and a $\pm 1\sigma$ variation, where σ is their related uncertainty derived from that measurement. This plus–minus variation is transferred into continuous distributions to interpolate intermediate variation values, and conventionally the nominal nuisance parameter is $\theta = 0$ and the $\pm\sigma$ variation corresponds to ± 1.
The likelihood function of Eq. (9.3) is modified to include the information of the nuisance parameters as

$$\mathcal{L}(n_{rb}, \mu, \vec{\theta}) = \mathcal{L}_{\text{main}}(n_{rb}, \mu, \vec{\theta}) \cdot \prod_p \mathcal{F}(\theta_p)\,. \tag{9.4}$$

The last term represents the probability distribution of the nuisance parameter that describes the previous knowledge on the parameter, i.e. the *prior*. Usually this is parametrised by a Gaussian distribution $Gaus(\theta_p|0,1)$, but other functional forms can be used as well, like a log-norm function. Nuisance parameters can be classified into three classes

- floating normalisation factors without *prior*, which include the parameter(s) of interest, such as μ;
- uncertainties with prior, that can have an overall normalisation effect on the distributions and can be correlated through the bins;
- scale factors for the statistical uncertainties in simulation, which are usually given bin-by-bin and are shared for all samples following an approximation of the Barlow-Beeston treatment [149].

The validity of Eq. (9.4) is ensured if pre-fit uncertainties are considered uncorrelated.
The probability of observing a certain number of events is maximised by minimising the negative logarithm of the likelihood function with respect to all parameters, keeping as free parameter the signal strength μ, which is then determined by the fit, and other free normalisation parameters if present. In this case regions dominated by a background of interest can be included in the formulation of the likelihood to obtain information on constraints or correlations among parameters in that region.
The observed data can be substituted by the *Asimov dataset* [150], which is an artificial dataset where all the observed quantities are fixed to their expectation values. This means that the parameter μ is fixed to the value relative to a certain signal hypothesis, and other free normalisation parameters, if present, are fixed to their expected values. Their expected value can be retrieved for example with a fit in a separated control region.
In general a fit is defined *conditional* when parameters are fixed to some values, which can be their expected value or can serve to test different hypotheses like the background-only one, where the signal strength is set to zero. In the *unconditional fit* free floating parameters are determined from the maximisation.

The significance is determined by defining a test statistic

$$q_s = -2\ln\frac{\mathcal{L}(n_{rb},\mu,\hat{\hat{\vec{\theta}}})}{\mathcal{L}(n_{rb},\hat{\mu},\hat{\vec{\theta}})}, \tag{9.5}$$

where $\hat{\hat{\vec{\theta}}}$ are the values that maximise the likelihood function for a given μ. In the denominator the values of both $\hat{\mu}$ and $\hat{\vec{\theta}}$ are the values that give the maximum of the likelihood. By setting the μ to zero the test statistic evaluates the compatibility of the observed data with the background-only hypothesis.

9.1.2 Cross Section Extraction

From the definition of the signal strength in Eq. (9.1), the measured cross section is extracted as $\sigma_{\text{meas}} = \hat{\mu}\sigma_{\text{theo}}$, where $\hat{\mu}$ is obtained from the profile fit. The kinematic regions involved in the measurement are defined at two levels, namely generation-level, where simulated quantities are considered before being reconstructed, and reconstruction-level. The so-called *fiducial region* is defined at generation-level with selection criteria that follow closely the reconstruction-level ones of the signal region. The cross section measurement is obtained in the fiducial region, and derived as

$$\begin{aligned}\sigma^{\text{fid}}_{\text{meas}} &= \hat{\mu}\cdot\sigma^{\text{fid}}_{\text{theo}}\\ &= \hat{\mu}\cdot\sigma^{\text{tot}}_{\text{theo}}\cdot\mathcal{A}.\end{aligned} \tag{9.6}$$

The factor $\mathcal{A}$ is the *acceptance*, and accounts for the restriction from the total phase to the fiducial region, due to the detector geometry and kinematic selection criteria. It is defined as

$$\mathcal{A} = \frac{N^{\text{fid}}_{\text{sig, MC, gen}}}{N^{\text{tot}}_{\text{sig, MC, gen}}}, \tag{9.7}$$

and its action is sketched in Fig. 9.1. The term $N^{\text{fid}}_{\text{sig, MC, gen}}$ represents the number of signal events from simulation at generation-level, counted in the fiducial region, while the term $N^{\text{tot}}_{\text{sig, MC, gen}}$ indicates the signal events from simulation at generation-level in the total phase space. Therefore, the total theoretical cross section multiplied by the acceptance in Eq. (9.6) gives the theoretical *fiducial cross section*, i.e. the cross section predicted in the phase space defined by the geometrical acceptance and the specific kinematic selection. Theoretical uncertainties enter the calculation of the total theoretical cross section and the acceptance, since the latter is derived with Eq. (9.7) using generation-level information. The ones affecting only the kinematic acceptance enter the acceptance factor, and are the only theoretical uncertainties entering the measurement together with the shape uncertainties, since normalisation uncertainties cancel in the calculation from Eq. (9.6). All experimental and remaining theory uncertainties are accounted for in the fit, and hence affect the measurement of μ.

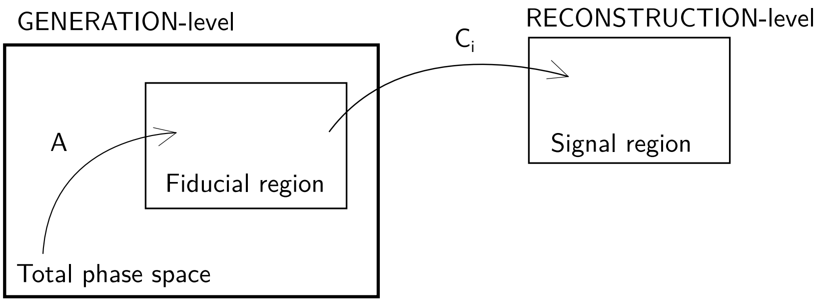


Figure 9.1: Kinematic regions involved in the cross section calculation. The acceptance is indicated with A, while C_i represents the vector of efficiency correction factors.

9.1.3 Generation- to Reconstruction-Level Propagation

Based on the definitions of signal strength of Eq. (9.1) and total cross section of Eq. (9.6), the number of expected signal events in a signal region defined at reconstruction-level can be expressed as

$$N^{\text{SR}}_{\text{sig, exp}} = \mu \cdot L \cdot \sigma^{\text{fid}}_{\text{theo}} \cdot \mathcal{A} \cdot \mathcal{C}. \tag{9.8}$$

The term L indicates the total integrated luminosity of the dataset under study, whilst the factor $\mathcal{C}$ is the so-called *efficiency correction factor*. It accounts for detector and reconstruction inefficiencies, resolution effects and migrations in and out of the fiducial region, i.e. for events which are generated outside of the fiducial region, and reconstructed in the signal region. Considering a binned kinematic phase space, it is defined in a bin i as

$$\mathcal{C}_i = \frac{N^{\text{SR}}_{\text{sig, MC, reco}\,i}}{N^{\text{fid}}_{\text{sig, MC, gen}\,i}}, \tag{9.9}$$

where the numerator indicates the number of simulated signal events counted at reconstruction-level in a bin i of the signal region, whilst the number of simulated signal events counted at generation-level in the same bin i of the fiducial region enters the denominator. The action of the efficiency correction factor is depicted schematically in Fig. 9.1. When considering differential distributions also the migration of events which are generated in a certain bin j at reconstruction-level and reconstructed in bin i can become sizeable, and needs to be accounted for. The vector of efficiency correction factors is substituted with a *response matrix* R_{ij} that describes both migration effects in and out of the fiducial region and through the bins, and accounts for efficiency effects as well. Each element of the matrix is defined as the probability p(reco|true) that an event is reconstructed in a bin i given that it was generated in bin j. It is hence calculated as a ratio between the number of events $N_{\text{reco}}(\text{bin}^{\text{reco}}_i, \text{bin}^{\text{gen}}_j)$ reconstructed in bin i that were generated in a generation-level bin j, over the total number of events $N_{\text{gen}}(\text{bin}^{\text{gen}}_j)$ generated in a generation-level bin j

$$R_{ij} = \frac{N_{\text{reco}}(\text{bin}^{\text{reco}}_i, \text{bin}^{\text{gen}}_j)}{N_{\text{gen}}(\text{bin}^{\text{gen}}_j)}.$$

In contrast with the C factors, also generation-level information are considered in the numerator of the response matrix.

9.2 Predictions in the Fiducial Region

9.2.1 Fiducial Region Definition

Based on the kinematic selection that defines the signal region, detailed in Section 7.3.2, the fiducial region is defined at generation-level with criteria that to follow closely the signal region ones. It is defined by requiring

- two prompt leptons (e, μ), with $p_{\mathrm{T}} > 27$ GeV and $|\eta| < 2.5$. Bremsstrahlung photons are added to the leptons within the cone $\Delta R(\ell, \gamma) < 0.1$;
- transverse momentum of the two-neutrino system of $p_{\mathrm{T}}^{\nu\nu} > 30$ GeV;
- at least two jets[1] with $p_{\mathrm{T}}^{j0} > 65$ GeV and $p_{\mathrm{T}}^{j1} > 35$ GeV, $|\eta| < 4.5$, reconstructed with the anti-k_t algorithm with a radius parameter R = 0.4;
- the minimum ΔR between leptons and jets must be $\Delta R(\ell, j) > 0.3$;
- $m_{jj} > 500$ GeV and $\Delta y_{jj} > 2.0$.

Since only electrons and muons are considered in the fiducial region, events which contain one or both leptons originated from τ decays are excluded. Nevertheless they can be reconstructed in the signal region, and these migrations need to be accounted for. This is done in the analysis by subtracting the predicted fraction of events involving tau decays when performing the propagation from generation- to reconstruction-level described in Section 9.1.3. This is done instead of subtracting the number of events predicted by simulation and it relies on the correct modelling of the $W \to \tau\nu$ branching fraction.

9.2.2 Theoretical Fiducial Cross Section

The theoretical uncertainties accounted for in the calculation of the fiducial cross section prediction are related to the choice of the simulation parameters, in particular the PDF set and α_{s}, and the renormalisation and factorisation scales μ_{R} and μ_{F}. Uncertainties from parton shower and its matching to the matrix element are accounted for as well.

The nominal expected cross section is evaluated with the nominal signal sample SHERPA 2.2.2 in the fiducial region detailed in Section 9.2.1 at LO to be

$$\sigma^{\mathrm{fid}}_{\mathrm{LO\ EW,\ SHERPA}} = 2.01 \pm 0.02(\mathrm{stat})\,^{+0.29}_{-0.23}(\mathrm{scale})\,^{+0.16}_{-0.02}(\mathrm{parton\ shower})\,^{+0.05}_{-0.03}(\mathrm{PDF}{+}\alpha_{\mathrm{s}})\ \mathrm{fb}. \tag{9.10}$$

[1] Jets are reconstructed clustering all final state particles excluding neutrinos, muons from b or c decays, prompt leptons from W or Z boson decays and prompt photons.

9.2.3 Treatment of Higher Order Corrections and Interference

The LO fiducial cross section prediction is the baseline for the analysis. This because the interference with the QCD background, as detailed in Section 2.4.1, is defined only at leading order [28] and the event generator existing at the time of the analysis did not include NLO EW corrections[2].
Nevertheless, corrections at NLO EW to the $pp \to \ell^{\pm}\nu\ell^{\pm}\nu jj$ EW process are calculated in Ref. [28] and are provided there at parton level in a phase space almost identical to the analysis' fiducial region, except for the m_{jj} requirement. The corrections provided correspond to the $\mathcal{O}(\alpha_{\text{EW}}^{7})$, since they are the dominant ones, as discussed in Section 2.4.1. These can be used to obtain an estimation of the cross section value at higher EW orders. A k-factor is calculated in the fiducial region as

$$\text{k} = \frac{\sigma_{\text{NLO}}^{\text{fid}}}{\sigma_{\text{LO}}^{\text{fid}}} = 0.842, \tag{9.11}$$

with a negligible statistical uncertainty. The NLO EW corrections include real emissions of photons in the final state, which are also accounted for in the parton shower in the SHERPA nominal sample. To estimate the double counting a sample is produced with SHERPA without the final state QED radiation. The difference between the two lays within the statistical uncertainty, the double counting is hence neglected and the nominal parton shower settings are used.
In addition to the NLO EW corrections, also the interference contribution is not accounted for in the calculation of the predicted cross section of Eq. (9.10). At reconstruction level, although, interference necessarily enters the signal region. As an important consequence the signal definition includes the interference, estimated to be around 6% of the electroweak cross section in the fiducial region of the analysis [37].

9.2.4 Signal Samples Comparison

The alternative two MADRAPH5 samples and the POWHEG sample introduced in Section 7.1.2 are compared to the nominal SHERPA sample in the fiducial region. Table 9.1 offers an overview of the generators characteristics, cross sections and acceptances.
The cross section predicted by SHERPA shows tensions with respect to the other generators. Detailed studies have been performed on the topic, which are documented in Ref. [152]. In particular this sheds light on the SHERPA mismodelling of the color flow in the parton shower. This increases the central jet emission, but since the matrix element includes up to one additional parton the effect is not very visible in differential distributions, but causes the lower cross section value. Nevertheless, the results are derived with the signal sample generated with SHERPA 2.2.2 and the independence of the physics measurement from this choice will be tested afterwards and presented in Section 9.7.

[2]A generator is now available for the $pp \to \ell^{\pm}\nu\ell^{\pm}\nu jj$ process, with $\ell = e, \mu$, that includes NLO EW corrections. It is based on POWHEG in combination with the matrix-element generator RECOLA [151].

Table 9.1: Comparison of signal samples. The different characteristics are summarised together with the predictions, calculated on the generated samples. The uncertainties are the statistical uncertainties on the event generation. The fiducial phase space is defined in Section 9.2.1, while the fiducial cross section with NLO electroweak correction is obtained by applying the k-factor from Eq. (9.11).

Generator	SHERPA 2.2.2	MADRAPH5 2.3.3 A	MADRAPH5 2.3.3 B	PowhegBox_V2
Order (QCD)	2,3j@LO	LO	LO	NLO
Order (EW)	LO	LO	LO	LO
Processes included	VBS, triboson, non-resonant	VBS, triboson, non-resonant	VBS, triboson, non-resonant	VBS, non-resonant
Scales μ_F, μ_R	dynamical: m_{WW}^{true}	dynamical: sum of transverse masses	fixed: m_W	fixed: m_W
Parton shower	SHERPA	PYTHIA8 8.212	PYTHIA8 8.212	PYTHIA8 8.230
Generator cross section (fb)	40.81 ± 0.05	40.16 ± 0.07	43.53 ± 0.09	28.70 ± 0.05
Acceptance in fid PS	0.0493 ± 0.0002	0.0680 ± 0.0005	0.0801 ± 0.0006	0.1074 ± 0.0006
Fid. cross section (fb)	2.01 ± 0.02	2.77 ± 0.02	3.49 ± 0.03	3.08 ± 0.01
Fid. cross section with NLO EWK corr. (fb)	1.69 ± 0.02	2.33 ± 0.02	2.94 ± 0.03	2.62 ± 0.01

9.3 Systematic Uncertainties

9.3.1 Theoretical uncertainties

The theoretical systematic uncertainties on the signal sample are the ones mentioned in Section 9.2.2 for the calculation of the theoretical fiducial cross section. Besides the standard recommended variations related to the simulation parameters, additional uncertainties have been considered. One accounts for the missing NLO EW corrections, discussed in Section 9.2.3, and one is related to the interference between the EW- and QCD-induced final states.

The recommended theoretical uncertainties are evaluated not only for the $W^{\pm}W^{\pm}jj$ EW sample, but also for the $W^{\pm}W^{\pm}jj$ QCD and WZ processes, for which EW and QCD modes are combined in the signal region and low-m_{jj} control region. They are evaluated in the WZ control region as well, for the $W^{\pm}W^{\pm}jj$ samples and for the WZ process. The parton shower uncertainty and the one due to the missing NLO EW corrections and to the interference for the $W^{\pm}W^{\pm}jj$ process are not considered in this region. The latter ones are not provided in the WZ control region as no reconstructed sample is available for parton shower evaluation. The results for the WZ control region are summarised in Table 9.2.

The theoretical uncertainties are calculated for all channels combined to maximise the number of events under study. The m_{jj} binning is the same as used in the final fit. For fiducial quantities all these components are added in quadrature to form what is called the *total theoretical uncertainty*. The parton shower uncertainty and the NLO and interference corrections are provided at generation-level and are propagated to reconstruction-level with a response matrix, as described in Section 9.1.3. This is derived as described there on the m_{jj} distribution using SHERPA 2.2.2, and is shown in Fig. 9.2. The evaluation of the theoretical uncertainties is detailed in the following.

Recommended simulation uncertainties

PDF including α_s As described in Chapter 3 the uncertainties in the determination of PDFs come from the experimental uncertainties from the dataset and the choice of the functional form of the PDF fit [51]. The nominal PDF set for the three samples is the NNPDF3.0 [137], and the uncertainty is assessed both by varying internal parameters and by considering the differences between different

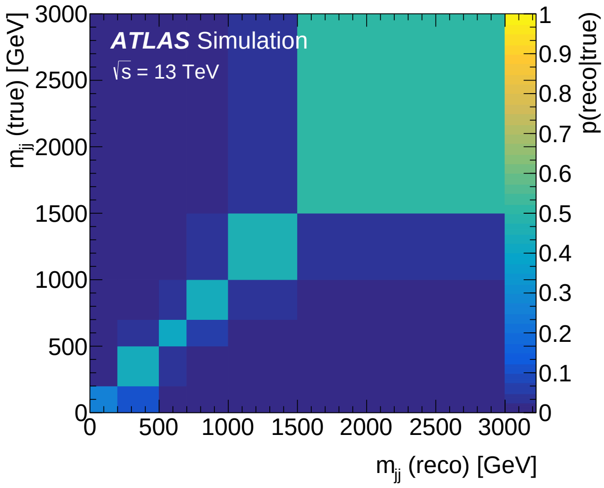


Figure 9.2: Response matrix for the m_{jj} distribution. The matrix elements represent the conditional probability p(reco|true) that a $W^{\pm}W^{\pm}jj$ electroweak event is reconstructed in a certain m_{jj} bin at reconstruction-level, given that it was generated with a certain value in a bin of m_{jj} at generation-level [5].

Table 9.2: Summary of relative theoretical uncertainties for the WZ and $W^{\pm}W^{\pm}jj$ process in the WZ control region combined in all channels. The EW and QCD contributions are listed separately for the WZ, but they are combined in the fit.

Process	WZ QCD	WZ EW	$W^{\pm}W^{\pm}jj$
MC statistics [%]	± 0.9	± 0.8	± 50
7-point scale [%]	$^{+28.8}_{-19.7}$	$^{+6.8}_{-5.9}$	± 33
PDF envelope [%]	$^{+2.02}_{-1.31}$		$^{+3.52}_{-3.52}$
α_s uncertainties [%]	$^{+0.71}_{-0.77}$		$^{+0.76}_{-1.26}$
Parton shower [%]	± 24.9		–

PDF sets, where CT14 and the MMHT2014 PDF sets are used for this purpose. The final PDF uncertainty is computed by taking the envelope of the two uncertainties in the m_{jj} distribution. It ranges from $\sim$0.5% to $\sim$1.4% for $W^{\pm}W^{\pm}jj$ EW in the signal region, while it affects more the $W^{\pm}W^{\pm}jj$ QCD and WZ in the same region going up to about 5.8%.
The uncertainties on α_s are related to the fact that it is determined using truncated fixed order calculations. The nominal value at the m_Z scale is 0.118, and the uncertainties are computed by evaluating the PDF set with varied α_s values. The size of the uncertainty related to the α_s variation is around 2% for $W^{\pm}W^{\pm}jj$ EW, $W^{\pm}W^{\pm}jj$ QCD and WZ combined in all fitting regions, which is of comparable size of the statistical uncertainty of the nominal prediction. The PDF and α_s uncertainties are combined in quadrature for the fit.

Scale variation The missing higher order contributions to the partonic cross section are estimated through an uncertainty, as stated in Chapter 3, that is assessed by varying the factorisation and renormalisation scales μ_F and μ_R by factors of 0.5 and 2 with the constraint $0.5 < \mu_F/\mu_R < 2$, following the 7-point scale variation method. A symmetrised envelope of all variations is then considered as systematic uncertainty. For all the regions this uncertainty is dominated by the μ_R variation. For the WZ sample this is the dominant uncertainty amounting to 20-30%, as reported in Table 9.2.

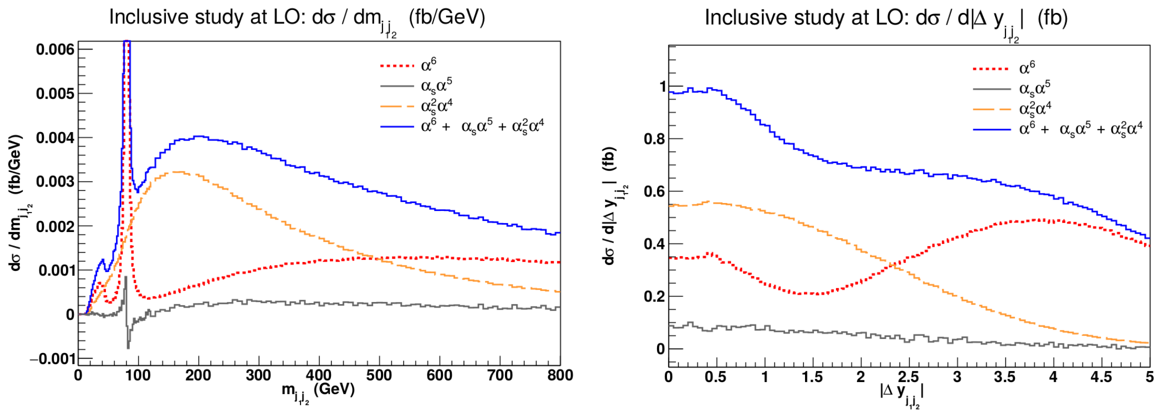
Parton shower To evaluate the parton shower uncertainties on the $W^{\pm}W^{\pm}jj$ EW and QCD samples, the envelope of five variations is considered, each of them deriving from a variation of internal shower parameters of the SHERPA generator, such as the shower merging scale, the shower end scale and the variation of the kinematic recoil scheme. They are available at particle level, and therefore need to be propagated to reconstruction level with the response matrix shown in Fig. 9.2, retrieved following what has been explained in Section 9.1.3. The variation for the $W^{\pm}W^{\pm}jj$ EW sample reaches up to 12% in the low-m_{jj} region and lays around 5-7% in the signal region.
For the WZ sample the uncertainty is derived by taking the relative difference between the sample generated with POWHEG and interfaced either to PYTHIA 8 or Herwig++, and then applying this difference to the nominal SHERPA sample. This reaches values around 25%, as reported in Table 9.2.
In general the largest uncertainties are related to the scale variation in all regions, followed by the parton shower and PDF+α_s summed in quadrature with the smallest contribution.

Next-to-leading order EW corrections

The NLO EW corrections for the $W^{\pm}W^{\pm}jj$ EW process provided in Ref. [28] are given as differential distributions of the cross section in m_{jj} bins of 40 GeV, from 0 GeV to 13000 GeV. They are provided at generation-level and applied in the analysis to the LO m_{jj} distribution to obtain the NLO one. The two distributions obtained in this way are propagated to reconstruction-level with the response matrix shown in Fig. 9.2. After normalising the NLO distribution to the LO one the relative difference between the two is symmetrised and considered as a variation of the nominal LO electroweak signal. They are computed for all channels combined since no different behaviour is expected for the NLO EW corrections for different lepton charge and flavour.

Interference between EW and QCD production

The interference between electroweak and QCD production is quantified in the analysis with MADRAPH5, which allows to produce samples based on the coupling order information at the Matrix-Element-Square level and can therefore produce the interference term alone. The interference term is generated for the process $pp \to \mu^+\nu\mu^+\nu jj$ in a phase space that requires $|\Delta\eta_{jj}| < 10$, $p_T^j > 20$ GeV and no selection on m_{jj}. The cross sections are listed in Table 9.3, and are derived from an inclusive sample generated in addition, that includes the EW, QCD and interference term. The electroweak production contributes around 57% to the total cross section, the QCD production 37%, and the interference around 6%.
In order to account for the interference as systematic uncertainty on the signal, the ratio between the interference part and the EW-only part is computed in bins of m_{jj}. This is considered as a systematic variation from the nominal distribution and is then symmetrised. This is evaluated in all



(a) (b)

Figure 9.3: Dijet invariant mass (a) and rapidity separation between the two jets (b) for the three LO contributions to the process $pp \to \mu^+\nu_\mu e^+\nu_e jj$. The EW component is depicted in red, the QCD in orange and the interference in grey. The blue line indicates the sum of all contributions. The kinematic region used to generated the distributions is similar to the fiducial region used in the analysis, but no selection criteria on m_{jj} and Δy_{jj} are applied [37].

Table 9.3: Cross section values for the process $pp \to \mu^+\nu\mu^+\nu jj$ for the inclusive EW, QCD and interference production σ_{inc}, and for the EW- (σ_{EW}), QCD- (σ_{QCD}) and interference-only (σ_{int}) production. The computation is performed in the phase space $|\Delta\eta_{jj}| < 10$, $p_{\text{T}}^{j} > 20$ GeV.

σ_{inc} [fb]	σ_{EW} [fb]	σ_{QCD} [fb]	σ_{int} [fb]
3.646±0.001	2.132±0.001	1.371±0.001	0.2270±0.0002

channels combined, as no dependency on lepton charge and flavour is expected. Fig. 9.3 shows the m_{jj} and Δy_{jj} distributions for the different contributions, as calculated in Ref. [37]. The interference contribution is concentrated in the same part of the phase space as the QCD background, which is clearly visible in the Δy_{jj} distribution.

9.3.2 Experimental uncertainties

Uncertainties in the procedure of object reconstruction, signal prediction and background estimation, luminosity measurement and pile-up are propagated in the estimation of the physics processes and affect the event yields. The systematic uncertainties considered in the analysis related to electrons concern energy resolution, energy scale, identification, isolation, reconstruction and trigger efficiencies. For muons, inner detector and track resolution, spectrometer resolution, momentum scale, trigger and isolation efficiencies are considered. The systematic uncertainties taken into account for jets are the energy resolution (JER), energy scale (JES), JVT efficiency and b-tag efficiency, whilst the resolution and scale uncertainties are considered for the missing transverse energy. Two additional uncertainties are related to the luminosity measurement and the pile-up reweighting. The uncertainties that refer to the estimation of backgrounds based on data are treated in Chapter 8 for each contribution. In particular four uncertainties are related to the estimation of the non-

Table 9.4: Experimental systematics uncertainties for the $W^{\pm}W^{\pm}jj$ EW processes in all channels before the fit. The effects of all nuisance parameters are added in quadrature.

$W^{\pm}W^{\pm}jj$ EW	ee % Yield	$e\mu$ % Yield	$\mu\mu$ % Yield
Jet-related Uncertainties	2.28	2.22	2.28
b-tagging efficiency	1.81	1.76	1.74
Pile-up	0.48	0.97	2.42
Trigger efficiency	0.02	0.08	0.47
Lepton reconstruction and identification	1.45	1.14	1.83
MET reconstruction	0.26	0.17	0.21

Table 9.5: Experimental systematics uncertainties for the $W^{\pm}W^{\pm}jj$ QCD processes in all channels before the fit. The effects of all nuisance parameters are added in quadrature.

$W^{\pm}W^{\pm}jj$ QCD	ee % Yield	$e\mu$ % Yield	$\mu\mu$ % Yield
Jet-related Uncertainties	3.41	3.04	2.85
b-tagging efficiency	2.56	2.48	2.48
Pile-up	4.99	0.45	0.33
Trigger efficiency	0.02	0.08	0.41
Lepton reconstruction and identification	1.62	1.19	1.89
MET reconstruction	0.41	0.22	0.34

prompt lepton background, that refer to the fake-factor measurement, and two to the charge misidentification background, related to the scale factors measurement and the energy correction. Each source is treated independently in the fit implementation.

Their effect on the events yield is summarised in Table 9.4 for the signal and in Tables 9.5 to 9.6 for the $W^{\pm}W^{\pm}jj$ QCD and WZ background processes. The main impact on the yield is given by the jet uncertainty, which is around 2% for the $W^{\pm}W^{\pm}jj$ EW process, 3% for the $W^{\pm}W^{\pm}jj$ QCD one and 9% for the dominant WZ background. It causes large uncertainties in the other backgrounds as well, but the impact on the analysis is small due to the lower contribution of processes like $t\bar{t}$+W/Z, $V\gamma$ and ZZ.

The largest impact in the yield variation due to uncertainties on the charge misidentification background estimation is given by the statistical uncertainty on the scale factors.

The uncertainty on the integrated luminosity, obtained following a calibration of the luminosity scale [153] is 2.1%.

9.3.3 Treatment in the Fit

The theoretical and experimental systematic uncertainties are accounted for with nuisance parameters in the fit, included as overall *normalisation factors* that change the yield of a certain physics process and as *shape variations* that affect the distribution of events with respect to a certain variable. In addition, a cross section uncertainty is assigned as normalisation uncertainty on the predictions for ZZ, triboson, $t\bar{t}$+W/Z, $W\gamma$ and $Z\gamma$ processes based on Refs. [154–156]. They vary from 20% to

Table 9.6: Experimental systematics uncertainties for the WZ process in all channels before the fit. The effects of all nuisance parameters are added in quadrature.

WZ	ee % Yield	$e\mu$ % Yield	$\mu\mu$ % Yield
Jet-related Uncertainties	9.58	5.03	8.45
b-tagging efficiency	2.49	2.23	2.40
Pile-up	2.99	3.49	3.33
Trigger efficiency	0.03	0.09	0.43
Lepton reconstruction and identification	1.52	1.24	3.07
MET reconstruction	0.93	0.79	1.63

Table 9.7: Cross section normalization uncertainties assigned to simulated background processes [154–156].

Process	Uncertainty
ZZ	20 %
Triboson	30 %
$t\bar{t}V$	30 %
$W\gamma$	20 %
$Z\gamma$	20 %

30% and are reported in Table 9.7. For the $W\gamma$ and the $Z\gamma$ backgrounds the additional uncertainty of 44% discussed in Section 8.3 is considered. Despite these high values, these uncertainties do not have a big impact on the measured signal cross section due to the extremely low event yields of these processes. An overview of all nuisance parameters is given in Table 9.8 for theoretical uncertainties, while the experimental ones are summarised in Table 9.9. The names are listed together with a short explanation of their meanings.

For the signal, the $W^{\pm}W^{\pm}jj$ QCD and the WZ samples, the recommended uncertainties related to PDF+α_s, renormalisation and factorisation scale and parton shower are computed as detailed in Section 9.3.1 and are included as shape variation. The normalisation effect for the signal sample is included in the uncertainties on the predicted fiducial cross section. For this reason the uncertainties on the signal sample are normalised to the nominal prediction to avoid double counting. The uncertainty related to the NLO EW correction is considered as a nuisance parameter applied as a shape uncertainty on the signal, as well as the interference between electroweak and QCD production.

The experimental uncertainties listed in Section 9.3.2 are included as shape and normalisation uncertainties. Also the two nuisance parameters related to the estimation of the charge misidentification background and the four nuisance parameters of the non-prompt background are accounted for as shape and normalisation uncertainties.

In order to account for the systematic uncertainties related to the limited number of simulated events, a single nuisance parameter per m_{jj} bin, region and channel is considered, instead of multiple ones for each simulated sample [149]. In each bin the total uncertainty is built from the ones related to each sample, and the parameter is defined by considering the relative uncertainty

Table 9.8: Name and description for nuisance parameters related to the theoretical systematic uncertainties.

	Systematic uncertainty	Short description
$W^{\pm}W^{\pm}jj$ EW	signal_EW6_TheoPDF	PDF shape uncertainty
	signal_EW6_TheoScale	Scale shape uncertainty
	signal_EW6_TheoShower	Parton Shower shape uncertainty
	signal_EW6_TheoAlphas	QCD coupling scale shape uncertainty
	TheoInterference	Interference between $W^{\pm}W^{\pm}jj$ QCD and EW shape uncertainty
	TheoEWCorr	missing NLO EW corrections shape uncertainty
$W^{\pm}W^{\pm}jj$ QCD	signal_EW4_TheoPDF	PDF shape and normalisation uncertainty
	signal_EW4_TheoScale	Scale shape and normalisation uncertainty
	signal_EW4_TheoShower	Parton Shower shape and normalisation uncertainty
	signal_EW4_TheoAlphas	QCD coupling scale shape and normalisation uncertainty
WZ	WZ_TheoPDF	PDF shape uncertainty
	WZ_TheoScale	Scale shape uncertainty
	WZ_TheoShower	Parton Shower shape uncertainty
	WZ_TheoAlphas	QCD coupling scale shape uncertainty

on the total event yield. The signal sample corresponds to a large integrated luminosity, so that including its uncertainty can artificially drive the expected significance. Therefore, it is usually not accounted with a nuisance parameter related to the amount of simulated events.

Table 9.9: Name and description for nuisance parameters related to the experimental systematic uncertainties.

Systematic uncertainty	Short description
	Dataset
Luminosity	uncertainty on total integrated luminosity
PRW_DATASF	uncertainty on pile-up reweighting
	Electrons
EL_EFF_Trigger_Total_1NPCOR_PLUS_UNCOR	trigger efficiency uncertainty
EL_EFF_Reco_Total_1NPCOR_PLUS_UNCOR	reconstruction efficiency uncertainty
EL_EFF_Iso_Total_1NPCOR_PLUS_UNCOR	isolation efficiency uncertainty
EL_EFF_ID_CorrUncertaintyNP	identification efficiency uncertainty
EL_EFF_ID_SIMPLIFIED_UncorrUncertaintyNP	identification efficiency uncertainty
EG_RESOLUTION_ALL	energy resolution uncertainty
EG_SCALE_ALLCORR	energy scale uncertainty
EG_SCALE_E4SCINTILLATOR	energy scale uncertainty
EG_SCALE_LARCALIB_EXTRA2015PRE	energy scale uncertainty
EG_SCALE_LARTEMPERATURE_EXTRA2015PRE	energy scale uncertainty
EG_SCALE_LARTEMPERATURE_EXTRA2016PRE	energy scale uncertainty
	Muons
MUON_EFF_TrigStatUncertainty	trigger efficiency uncertainty (stat)
MUON_EFF_TrigSystUncertainty	trigger efficiency uncertainty (syst)
MUON_EFF_STAT	reconstruction and ID efficiency uncertainty for muons with $p_T > 15$ GeV (stat)
MUON_EFF_SYS	reconstruction and ID efficiency uncertainty for muons with $p_T > 15$ GeV (syst)
MUON_EFF_STAT_LOWPT	reconstruction and ID efficiency uncertainty for muons with $p_T < 15$ GeV (stat)
MUON_EFF_SYST_LOWPT	reconstruction and ID efficiency uncertainty for muons with $p_T < 15$ GeV (syst)
MUON_ISO_STAT	isolation efficiency uncertainty (stat)
MUON_ISO_SYS	isolation efficiency uncertainty (syst)
MUON_TTVA_STAT	track-to-vertex association efficiency uncertainty (stat)
MUON_TTVA_SYS	track-to-vertex association efficiency uncertainty (syst)
MUON_ID	momentum resolution uncertainty from inner detector
MUON_MS	momentum resolution uncertainty from muon system
MUON_SCALE	momentum scale uncertainty
	Jets
JET_EffectiveNP	energy scale uncertainty from the in situ analysis
JET_EtaIntercalibration_Modeling	energy scale uncertainty on eta-intercalibrations
JET_EtaIntercalibration_TotalStat	energy scale uncertainty on eta-intercalibrations
JET_EtaIntercalibration_NonClosure	energy scale uncertainty on eta-intercalibrations
JET_Pileup_OffsetMu	energy scale uncertainty on pile-up
JET_Pileup_OffsetNPV	energy scale uncertainty on pile-up
JET_Pileup_PtTerm	energy scale uncertainty on pile-up
JET_Pileup_RhoTopology	energy scale uncertainty on pile-up
JET_Flavor_Composition	energy scale uncertainty on flavour composition
JET_Flavor_Response	energy scale uncertainty on samples' flavour response
JET_BJES_Response	energy scale uncertainty on b jets
JET_PunchThrough_MC15	energy scale uncertainty for punch-through jets
JET_SingleParticle_HighPt	energy scale uncertainty from the behaviour of high-p_T jets
JET_JER_SINGLE_NP	energy resolution uncertainty
JET_JvtEfficiency	JVT efficiency uncertainty
FT_EFF_Eigen_B	b-tagging efficiency uncertainties ("BTAG_MEDIUM"): 3 components
FT_EFF_Eigen_C	for b jets (`FT_EFF_Eigen_B`), 3 for c jets (`FT_EFF_Eigen_C`) and
FT_EFF_Eigen_L	5 for light jets (`FT_EFF_Eigen_L`)
FT_EFF_extrapolation	b-tagging efficiency uncertainty on the extrapolation to high-p_T jets
FT_EFF_extrapolation_from_charm	b-tagging efficiency uncertainty on tau jets
	MET
MET_SoftTrk_ResoPara	track-based soft term related longitudinal resolution uncertainty
MET_SoftTrk_ResoPerp	track-based soft term related transverse resolution uncertainty
MET_SoftTrk_Scale	track-based soft term related longitudinal scale uncertainty
	Fakes
FakeElSta	statistical uncertainty due to electron fakes
FakeElSys	systematic uncertainty due to electron fakes
FakeMuSta	statistical uncertainty due to muon fakes
FakeMuSys	systematic uncertainty due to muon fakes
	Charge Flip
CF_NoCorr	effect of not using energy corrections (one-sided)
CF_SFunc	effect of varying the scale factors

9.4 Fitting Methodology

In order to extract the signal strength and cross section for the $W^{\pm}W^{\pm}jj$ EW process, the maximum likelihood method explained in Section 9.1.1 is performed on the invariant mass distribution of the two tagging jets. In combination to the signal region which has lower edge at $m_{jj} > 500$ GeV, two additional regions are fit simultaneously, the low-m_{jj} and the WZ control regions. The low-m_{jj} control region is added to the signal region due to a non-negligible presence of signal events, but it receives large contributions from WZ and non-prompt lepton events as well, which are hence constrained in the fit. The low-m_{jj} control region differs from the signal region only in the m_{jj} range, being $200 < m_{jj} < 500$ GeV, where the upper bound is set to remove signal contributions.
The WZ control region is defined by requiring a third lepton with $p_{\mathrm{T}} > 15$ GeV, and a lower limit on the invariant mass of the three leptons is imposed at 106 GeV to reduce contributions from Z/γ^{*}+jets and $Z\gamma$+jets processes. An opposite-charge same-flavour dilepton pair is required to select Z bosons, while all other criteria reflect the signal region selection. However in the WZ control region the tagging jets could derive from different processes with respect to the signal region. The three leptons requirement in the WZ region makes it likely for the jets to be originated from initial partons, while the jets in the signal region could derive from hadronic W decays, τ leptons or leptons outside the acceptance. Therefore, the m_{jj} distribution is slightly harder in the WZ control region, as can be seen in Fig. 9.4. The normalisation μ_{WZ} of the WZ background is considered as a floating parameter in the final fit.
Four m_{jj} bins are considered in the signal region, optimised to increase the expected signal sensitivity, with lower boundaries at 500 GeV, 700 GeV, 1000 GeV and 1500 GeV. In the low-m_{jj} control region only one m_{jj} bin is considered, as well as for the WZ control region. The signal region is split into six channels based on lepton charge and flavour categorisation, namely $e^{\pm}e^{\pm}$, $\mu^{\pm}\mu^{\pm}$ and $e^{\pm}\mu^{\pm}$, to better exploit the different background compositions among the channels. The same splitting applies to the low-m_{jj} control region, as this is included to test the constraining power on the non-prompt background, and the ratio of WZ to non-prompt events varies between lepton channels. The lepton charge and flavour channels are combined in the WZ control region.

Significance extraction Before performing the final combined fit, the expected significance is extracted. This is done deriving the test statistic with Eq. (9.5), where the conditional fit in the numerator is evaluated for μ = 0, i.e. for the background-only hypothesis. To evaluate the expected significance a fit is performed first in the two control regions with observed data, to measure the normalisation of the WZ process. An Asimov dataset, defined in Section 9.1.1, is then regenerated taking into account this value. Finally the combined fit in all three regions is performed leaving μ_{WZ} as free floating. Usually only Asimov data are used to derive the expected significance, but in this case observed data are included in the WZ and low-m_{jj} control regions, since these are used to determine the normalisation of the WZ contribution.
The inclusion of the low-m_{jj} control region does not affect the expected significance, as the constrains on the nuisance parameters are not propagated to the combined fit. Nevertheless it is included in the fit of the control regions as it constrains the non-prompt background, reducing the

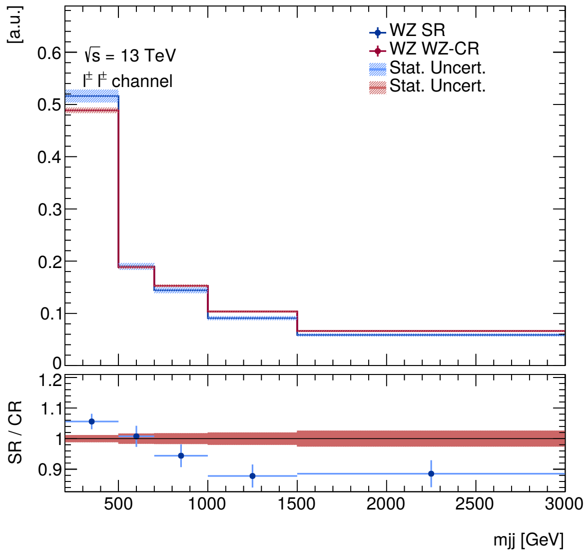


Figure 9.4: Dijet invariant mass for the WZ process compared in the signal region (blue) and in the WZ control region (red) for all channels combined. Both histograms are normalised to unity. Bands represent statistical uncertainties [157].

statistical and systematic uncertainties on μ_{WZ}.
In order to obtain the observed significance finally, a fit with observed data in all the regions is performed. No separate fit in the control regions is needed, since this is used only for the purpose of generating the Asimov data with the correct WZ normalisation.

9.5 Expected Performance

9.5.1 Signal Strength Extraction

Before the combined fit is performed in the three regions, the normalisation of the WZ process to be used for generating the Asimov dataset is retrieved with a fit in the WZ and low-m_{jj} control region, as explained in Section 9.4. The theoretical uncertainties on the WZ yields are computed as discussed in Section 9.3.1. As summarised in Table 9.2 the dominant source is represented by the perturbative scale variation amounting to $^{+\,29\%}_{-\,20\%}$, followed by the parton shower variation uncertainty with 25%. The variations related to the PDF set and α_s have smaller impact on the event yields, bringing 2% and 1% variation respectively. Only few nuisance parameters are slightly pulled by the fit, in particular the ones related to the non-prompt background, and less strongly the ones related to the WZ scale and parton shower theory uncertainties.
The fit obtains for the normalisation of the WZ process a value of

$$\mu_{WZ} = 0.91 \pm 0.07(\text{stat}) \pm 0.32(\text{syst}). \tag{9.12}$$

The pre- and post-fit distributions for the two control regions are shown in Figs. C.7 and C.8 of Appendix C.

The full combined fit is performed in the signal region and the two control regions on the Asimov dataset first. The main constraints on nuisance parameters are on the ones related to the systematic uncertainties on the electron and muon fake-factor measurement, but they are not significant. The low-m_{jj} control region region plays an important role in constraining these parameters, as shown by studies performed excluding the region, where no nuisance parameter is constrained by the fit in this case. The plots showing the pulls of nuisance parameters with and without the inclusion of the low-m_{jj} control region are shown in Figs. C.9 and Fig. C.10 of Appendix C.
The nuisance parameters are ranked based on their impact on $\hat{\mu}$, by performing four fits for each of them in addition to the nominal fit yielding to $\hat{\mu}$. The fit is run twice by shifting each nuisance parameter to its pre-fit uncertainties and again with the post-fit uncertainties, obtaining the respective values for μ. The differences between $\hat{\mu}$ and the one obtained with the post-fit variation of the parameter are ranked by decreasing variation, i.e. based on

$$\max(\mu_{\text{postF}}^{\text{up}} - \hat{\mu}, \mu_{\text{postF}}^{\text{down}} - \hat{\mu}).$$

The ranking is presented in Fig. 9.5, where the 20 parameters with the highest impact are shown and their respective impact on the signal strength is depicted. The nuisance parameter names and explanations are reported in Tables 9.8 and 9.9. The uncertainties related to the non-prompt background play a major role, followed by the parton shower uncertainties of the $W^{\pm}W^{\pm}jj$ EW and the scale variation of the $W^{\pm}W^{\pm}jj$ QCD sample. In general mainly jet-related, non-prompt background and some theoretical uncertainties populate the first 20 parameters in the ranking. The non-prompt lepton nuisance parameters show an asymmetry which is due to the systematic uncertainty related to the jet p_{T} reweighting, explained in Section 8.1. This uncertainty estimates a one-direction effect, and hence is not symmetrised in the fit. The parton shower nuisance parameter for the signal is one-sided, since despite the up and down variations of the single parameters that are varied, the effect of the envelope goes in one direction only. This is not symmetrised in the fit. The expected significance is calculated as discussed in Section 9.4 to be 4.4σ, 5.4σ when considering only the statistical uncertainties.
The expected signal strength and its uncertainties are

$$\mu_{\text{exp}} = 1.00^{+0.25}_{-0.23}(\text{stat}) \pm 0.09(\text{theo}) \,^{+0.12}_{-0.10}(\text{syst}). \tag{9.13}$$

The normalisation of the WZ process is not fixed from the control-region-only fit, and therefore is obtained from the full combined fit and determined to be

$$\mu_{WZ} = 0.91 \pm 0.07(\text{stat}) \,^{+0.39}_{-0.24}(\text{syst}). \tag{9.14}$$

9.5.2 Cross Section Measurement

The fiducial cross section is measured on Asimov data with the procedure detailed in Section 9.1.2, following Eq. (9.6), i.e. multiplying the parameter of interest of the fit $\hat{\mu}$ by the expected theoretical cross section in the fiducial region. The SHERPA predictions are considered for the moment, with the theoretical value $\sigma^{\text{fid}}_{\text{LO EW, SHERPA}}$ given in Eq. (9.10), while μ_{exp} from Eq. (9.13) is used to obtain the

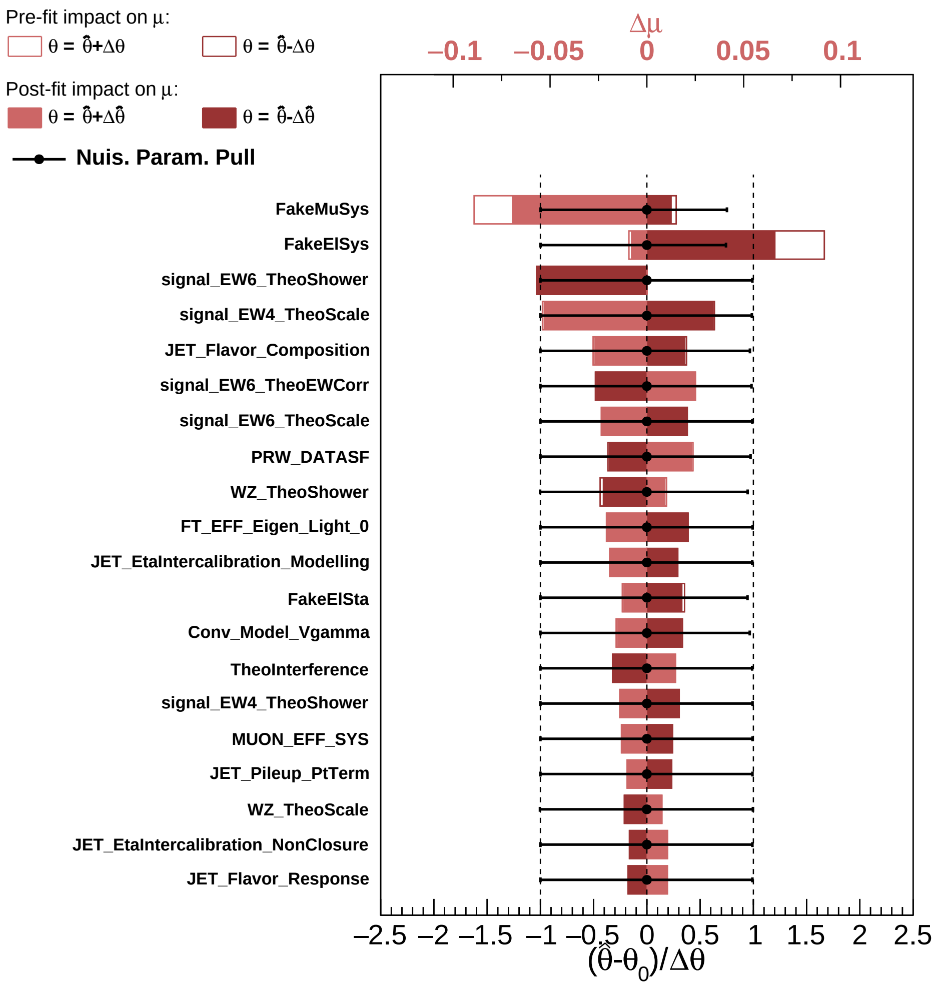


Figure 9.5: Ranking of nuisance parameters included in the fit based on their impact on the signal strength measured on Asimov data. Only the top 20 parameters are shown and the statistical uncertainties from simulation are not taken into account in the ranking. The empty (filled) brown rectangles correspond to the pre-(post-)fit impact on μ, referred to by the upper scale. The impact of each nuisance parameter $\Delta\mu$ is given comparing the nominal $\hat{\mu}$ with the best fit value of the parameter $\hat{\theta}$ to the fit result where the nuisance parameter is shifted to its pre-(post-)fit uncertainties $\pm\Delta\theta$ ($\Delta\hat{\theta}$). The black points show the pulls of the nuisance parameters with respect to the nominal value θ_0 with the relative post-fit error $\Delta\hat{\theta}/\Delta\theta$, referred to by the lower scale [158].

measured expected fiducial cross section. The expected fiducial cross section is

$$\sigma^{\text{fid}}_{\text{meas}} = 2.01\,^{+0.49}_{-0.24}(\text{stat}) \pm 0.18(\text{theo})\,^{+0.24}_{-0.20}(\text{syst})\text{ fb}. \tag{9.15}$$

As discussed in Section 9.3.3 the theoretical uncertainties include the normalisation theoretical uncertainties for the signal predictions deriving from the theoretical fiducial cross section from Eq. (9.10), the shape theoretical uncertainties for the $W^{\pm}W^{\pm}jj$ EW sample which are accounted for in the fit to extract μ_{exp}, together with the shape uncertainties for the WZ background and the normalisation and shape uncertainties for the $W^{\pm}W^{\pm}jj$ QCD one. The systematic uncertainties refer to the experimental uncertainties accounted for in the fit to extract the signal strength.

9.6 Results with Observed Data

9.6.1 Signal Strength Extraction

Once the reliability of the fit is proven on the Asimov dataset, observed data can be investigated. The number of events in the signal region before performing the fit is reported in Table 9.10 for the background processes, the predicted signal and the observed data for the six channels separately. The yields for the signal and background processes as obtained after the fit are compared to data in Table 9.11. After the fit 122 data events are counted in the signal region, compared to the 69$\pm$7 background events.

The pulls on the nuisance parameters after the fit do not exceed the 1-σ variation. The 15 parameters with the largest pulls are shown in Fig. 9.6, where the names and explanation of the nuisance parameters refer to those given in Tables 9.8 and 9.9. The plot with all nuisance parameters entering the fit is shown in Fig. C.11 of Appendix C. The nuisance parameter related to the systematic uncertainty on the electron fake-factor, together with the ones relative to the jet resolution and jet flavour composition uncertainties show the largest pulls, followed by the parameter related to the missing NLO corrections on the signal sample with a less significant pull. The action of the fit is visible in Fig. 9.7, where the m_{jj} distribution is shown before and after the fit in the low-m_{jj} control region and in the signal region for the charge- and flavour-combined channel. The action of the fit affects mainly the non-prompt lepton background, in particular in the channels with electrons. As can be see in Tables 9.10 and 9.11, the non-prompt lepton background in the signal region is reduced in the ee channel of about 50%, and in the $e\mu$ channel of about 20-30%, while it remains stable in the $\mu\mu$ channel. The decrease of the non-prompt lepton background affects the low-m_{jj} region as well, as visible from Fig. 9.7. The nuisance parameter related to the modelling of the $V\gamma$ background is slightly pulled up, and the small increase of the relative background is mainly visible in the 1$^{\text{st}}$ and 3$^{\text{rd}}$ m_{jj} bin in the signal region. Small pulls are also affecting the nuisance parameters related to the signal. In particular the down-pulls of the nuisance parameters related to the NLO EW corrections and interference and the up-pull of the scale variation parameter are reflected in the decrease in the signal yield in the low-m_{jj} region and the increase in the high-m_{jj} region.

The fit mainly constrains the nuisance parameters related to the systematic uncertainties in the determination of the electron and muon fake-factor, which were overestimated in the measurement.

Table 9.10: Summary of the data event yields, and the expected signal and background event yields in the signal region before the fit. The numbers are shown for the six individual channels and for all channels combined. The backgrounds from $V\gamma$ production and electron charge misidentification are combined in the "e/γ conversions" category. The "Other prompt" category combines ZZ, triboson and $t\bar{t}$+W/Z background contributions.

	e^+e^+	e^-e^-	$e^+\mu^+$	$e^-\mu^-$	$\mu^+\mu^+$	$\mu^-\mu^-$	Combined
WZ	1.9±0.6	1.3±0.4	14±4	8.9±2.5	5.5±1.6	3.6±1.1	35±10
Non-prompt	4.0±2.3	2.3±1.7	9±5	6±4	0.55±0.15	0.67±0.25	23±10
e/γ conversions	1.74±0.29	1.8±0.4	6.1±1.6	3.7±0.8	–	–	13.4±2.5
Other prompt	0.17±0.05	0.14±0.04	0.90±0.19	0.60±0.14	0.36±0.10	0.19±0.05	2.4±0.5
$W^\pm W^\pm jj$ QCD	0.38±0.13	0.16±0.05	3.0±1.0	1.2±0.4	1.8±0.6	0.76±0.25	7.3±2.4
Expected background	8.2±2.4	5.7±1.8	33±7	21±5	8.2±1.7	5.3±1.1	81±14
$W^\pm W^\pm jj$ EW	3.8±0.6	1.49±0.22	16.5±2.4	6.5±1.0	9.1±1.4	3.5±0.5	41±6
Data	10	4	44	28	25	11	122

Table 9.11: Summary of the data event yields, and the signal and background event yields in the signal region as obtained after the fit. The numbers are shown for the six individual channels and for all channels combined. The backgrounds from $V\gamma$ production and electron charge misidentification are combined in the "e/γ conversions" category. The "Other prompt" category combines ZZ, triboson and $t\bar{t}$+W/Z background contributions.

	e^+e^+	e^-e^-	$e^+\mu^+$	$e^-\mu^-$	$\mu^+\mu^+$	$\mu^-\mu^-$	Combined
WZ	1.48±0.32	1.09±0.27	11.6±1.9	7.9±1.4	5.0±0.7	3.4±0.6	30±4
Non-prompt	2.2±1.1	1.2±0.6	5.9±2.5	4.7±1.6	0.56±0.05	0.68±0.13	15±5
e/γ conversions	1.6±0.4	1.6±0.4	6.3±1.6	4.3±1.1	–	–	13.9±2.9
Other prompt	0.16±0.04	0.14±0.04	0.90±0.20	0.63±0.14	0.39±0.09	0.22±0.05	2.4±0.5
$W^\pm W^\pm jj$ QCD	0.35±0.13	0.15±0.05	2.9±1.0	1.2±0.4	1.8±0.6	0.76±0.25	7.2±2.3
Expected background	5.8±1.4	4.1±1.1	28±4	18.8±2.6	7.7±0.9	5.1±0.6	69±7
$W^\pm W^\pm jj$ EW	5.6±1.0	2.2±0.4	24±5	9.4±1.8	13.4±2.5	5.1±1.0	60±11
Data	10	4	44	28	25	11	122

The post-fit yields are shown in Fig. 9.8 for the low-m_{jj} region separately for the six channels, together with the combined channel for the WZ control region.

The ranking of the top 20 nuisance parameters is performed as detailed in Section 9.5, results are shown in Fig. 9.9. The main differences between the fit on observed and Asimov dataset, shown in Fig. 9.5, relate to the nuisance parameters for the theoretical uncertainties on the signal and the systematic uncertainties on the fake-factor measurement. The former acquire importance compared to the fit on the Asimov dataset, since the signal strength is now $\mu \gg 1$, while the uncertainties related to the the fake-factor have a lower impact due to the decrease of the non-prompt lepton background after the fit.

The correlations among nuisance parameters are shown in Table 9.12 for absolute values up to 15%. Both the parton shower uncertainty and the scale uncertainty for the WZ are anti-correlated with the normalisation parameter for the WZ background, as their up(down)-pull induces a large down(up)-shift on the μ_{WZ} floating normalisation.

The WZ normalisation is scaled with $\mu_{WZ} = 0.86 \pm 0.07(\text{stat})\,^{+0.31}_{-0.23}(\text{theo})\,^{+0.18}_{-0.08}(\text{syst})$, which derives mainly from the constraints on this background from the dedicated control region.

The signal strength obtained in the m_{jj} fit and the nuisance parameters are propagated in order to

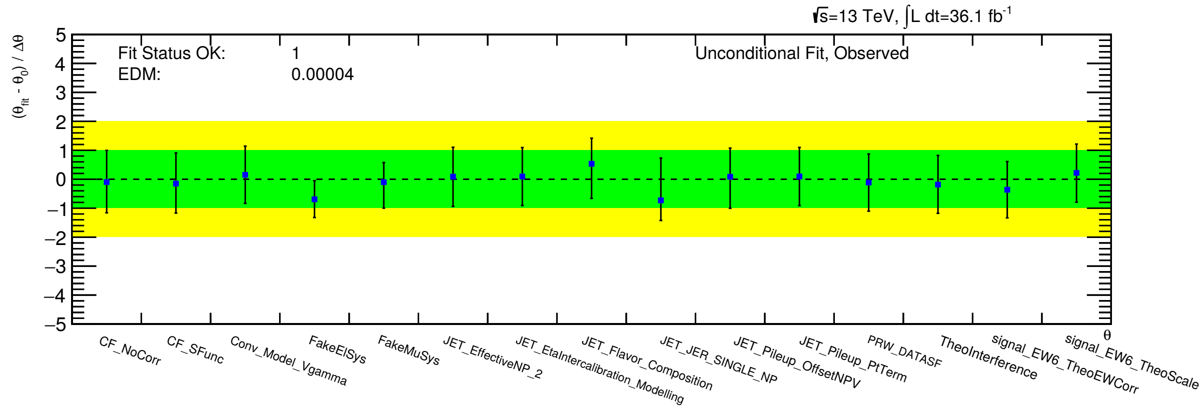


Figure 9.6: The fifteen nuisance parameters with largest pulls for the combined unconditional fit to observed data in the signal region and WZ and low-m_{jj} control regions [157].

Table 9.12: Correlation between nuisance parameters calculated for the unconditional fit to observed data. Only correlations with an absolute value equal or larger than 15% are listed.

Parameter 1	Parameter 2	Correlation in %
WZ_TheoShower	μ_WZ	-67
WZ_TheoScale	μ_WZ	-67
FakeElSta	FakeElSys	-17
JET_Flavor_Composition	μ_WZ	-16
FakeElSys	μ	-16
WZ_TheoShower	μ	$+16$
signal_EW6_TheoShower	μ	$+15$
FakeElSys	JET_Flavor_Composition	-15
Conv_Model_Vgamma	FakeElSys	-15
Conv_Model_Vgamma	JET_JER_SINGLE_NP	-15

obtain the distributions for variables different from the fitted one. The only exception concerns the parameters for the interference and electroweak corrections for which a constant uncertainty is assigned. Post-fit distributions are shown in Fig 9.10 for $m_{\ell\ell}$ and Δy_{jj}.
The signal strength resulting from the fit using the SHERPA signal sample is

$$\mu_{\text{obs}} = 1.44^{+0.26}_{-0.24}(\text{stat})\ ^{+0.07}_{-0.08}(\text{theo})\ ^{+0.13}_{-0.11}(\text{syst}), \tag{9.16}$$

while the observed significance is 6.5σ for an expected 4.4σ. The scan of the observed likelihood is shown in Fig. 9.11.

9.6.2 Cross Section Measurement

The fiducial cross section of the $W^{\pm}W^{\pm}jj$ EW process can be extracted first with the signal strength resulting from the fit performed on the SHERPA sample. At a later time, the same result can be extracted with an alternative sample to confirm that the mismodelling of the signal sample does not affect the measurement of the physical quantity of interest.

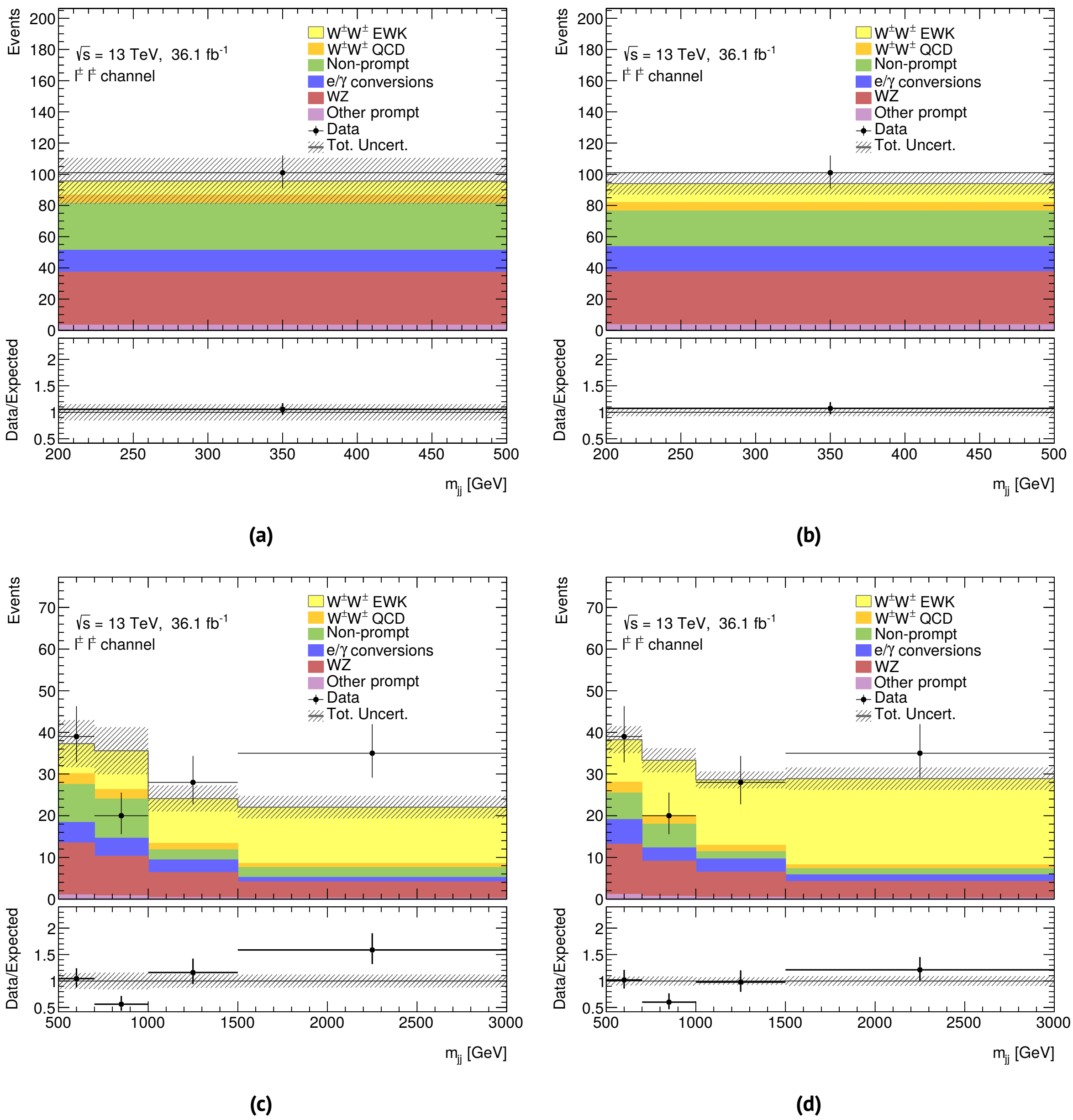


Figure 9.7: Pre-(a) and post-fit (b) m_{jj} distribution in the low-m_{jj} region, and pre-(c) and post-fit (d) m_{jj} distribution in the high-m_{jj} region in the flavour- and charge-combined channel. The hatched band represents the total uncertainty, including the statistical, cross section and systematic uncertainties [157].

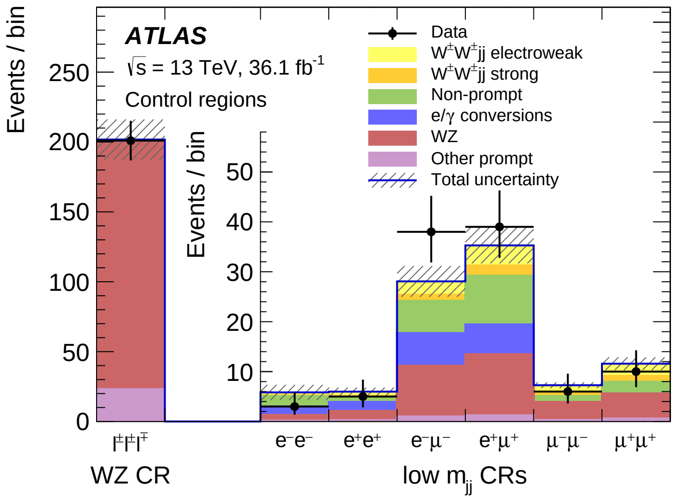


Figure 9.8: Event yields in the WZ and low-m_{jj} control regions as predicted after the combined fit. The bands represent the sum in quadrature of the statistical and systematic uncertainties of the background. The "e/γ conversion" category includes the background from $V\gamma$ processes and electron charge misidentification. The category "Other prompt" combines ZZ, VVV and $t\bar{t}$+W/Z contributions [5].

The cross section is measured with the procedure detailed in Section 9.1.2, following Eq. (9.6), as done with the expected cross section but using the μ_{obs} obtained fitting the observed data. The resulting cross section in the fiducial region is

$$\sigma^{\text{fid}}_{\text{meas,obs}} = 2.89\ ^{+0.51}_{-0.48}(\text{stat})\ ^{+0.14}_{-0.16}(\text{theo})\ ^{+0.24}_{-0.22}(\text{exp syst})\ ^{+0.08}_{-0.06}(\text{lumi})\ \text{fb}, \tag{9.17}$$

and includes the interference between electroweak and strong production. As discussed for the expected cross section measurement in Section 9.5, the theoretical uncertainties include the normalisation theoretical uncertainties for the signal predictions deriving from the theoretical fiducial cross section from Eq. (9.10), the shape theoretical uncertainties for the $W^{\pm}W^{\pm}jj$ EW sample which are accounted for in the fit to extract μ_{exp}, together with the shape uncertainties for the WZ background and the normalisation and shape uncertainties for the $W^{\pm}W^{\pm}jj$ QCD one. The systematic uncertainties refer to the experimental uncertainties accounted for in the fit to extract the signal strength. The impact of each systematic uncertainty on the measured cross section is presented in Table 9.13. Without accounting for correlations, the impact of each source of systematic uncertainty is computed by first performing the fit with the corresponding nuisance parameter fixed to one standard deviation up or down from the value obtained in the nominal fit, and then symmetrising these up and down variations. For each component, the impacts of several sources of systematic uncertainty are added in quadrature. The greatest impact on the experimental uncertainties side derives from the uncertainties related to the estimation of the non-prompt lepton background, jet and $E_{\text{T}}^{\text{miss}}$ reconstruction and from the statistical uncertainty on the background processes. The theoretical uncertainties on the WZ modelling have a similar impact, followed by the theoretical uncertainties on the QCD background and the recommended theoretical uncertainties on the signal process.

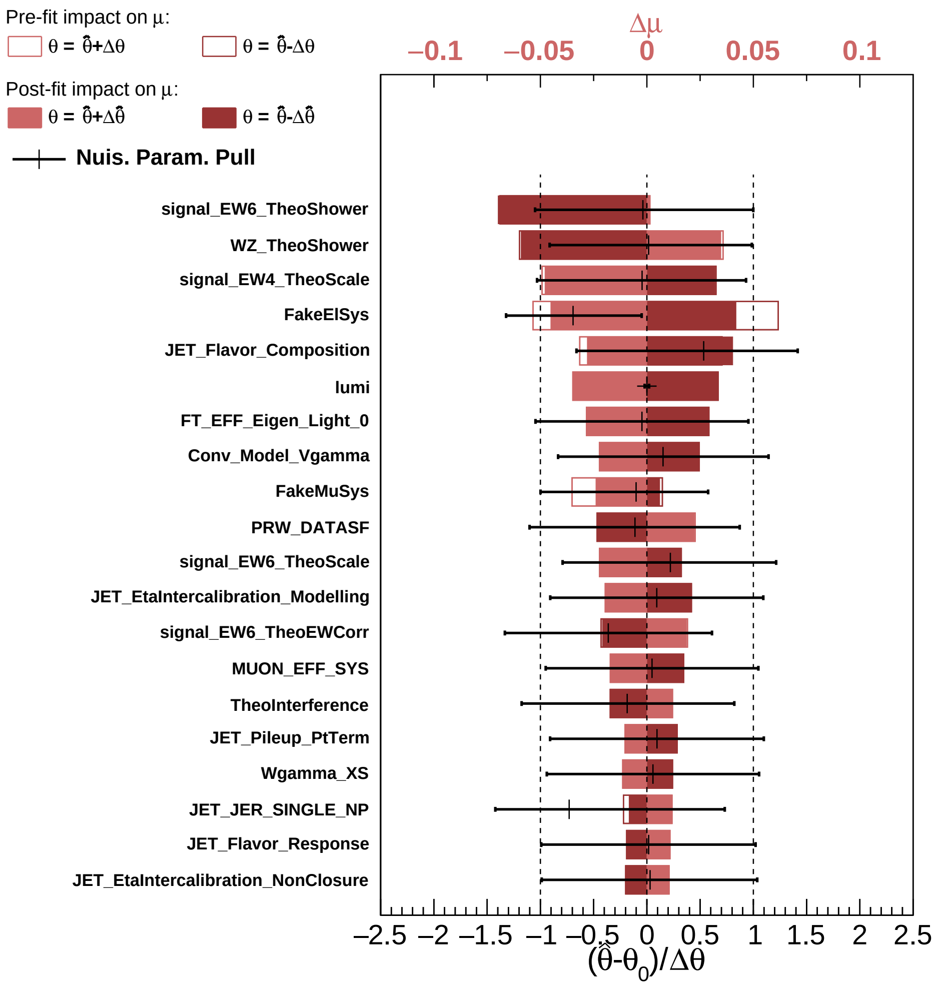


Figure 9.9: Ranking of nuisance parameters included in the fit based on their impact on the signal strength measured on observed data. Only the top 20 parameters are shown and the statistical uncertainties from simulation are not taken into account in the ranking. The empty (filled) brown rectangles correspond to the pre-(post-)fit impact on μ, referred to by the upper scale. The impact of each nuisance parameter $\Delta\mu$ is given comparing the nominal $\hat{\mu}$ with the best fit value of the parameter $\hat{\theta}$ to the fit result where the nuisance parameter is shifted to its pre-(post-)fit uncertainties $\pm\Delta\theta$ ($\Delta\hat{\theta}$). The black points show the pulls of the nuisance parameters with respect to the nominal value θ_0 with the relative post-fit error $\Delta\hat{\theta}/\Delta\theta$, referred to by the lower scale [158].

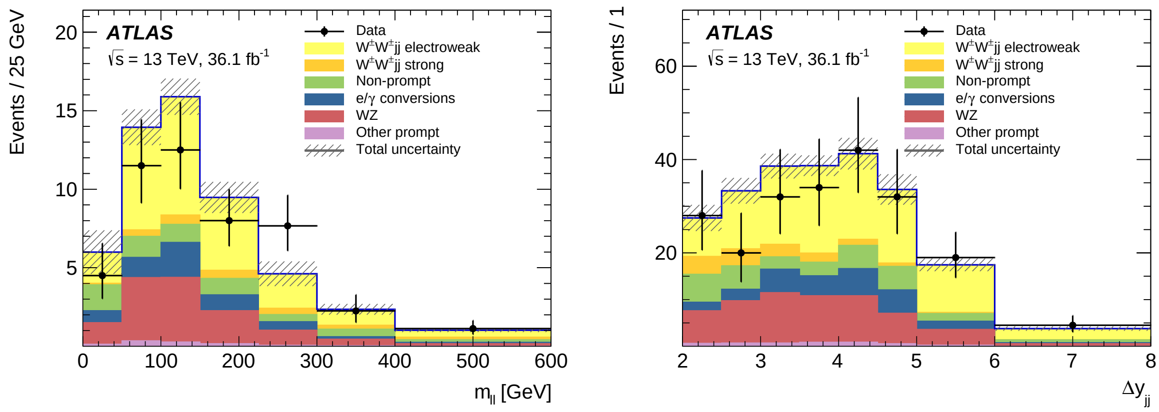


Figure 9.10: Distribution of $m_{\ell\ell}$ (a) and Δy_{jj} (b) in the signal region as obtained after the combined fit for all channel combined. The bands represent the sum in quadrature of the statistical and systematic uncertainties of the backgrounds. The "e/γ conversion" category includes the background from $V\gamma$ process and electron charge misidentification. The category "Other prompt" combines ZZ, VVV and $t\bar{t}$+W/Z contributions. The last bin in the distributions contains the overflow [5].

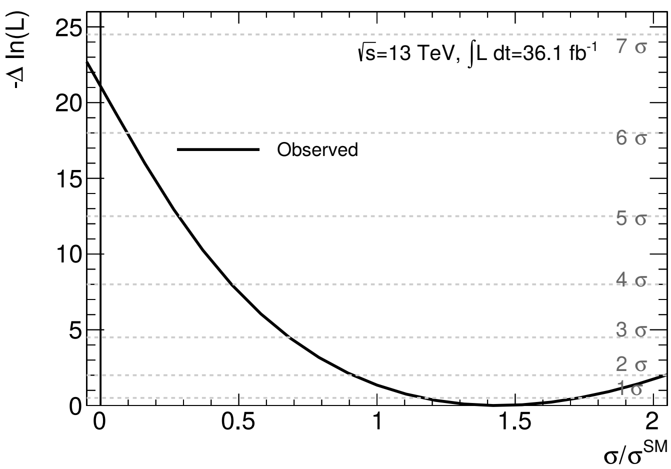


Figure 9.11: Likelihood scan for observed data for the six lepton channels combined using all nuisance parameters [157].

Table 9.13: Impact of different components of systematic uncertainty on the measured fiducial cross section, without accounting for correlations. For each component, the impacts of several sources of systematic uncertainty are added in quadrature. The sources of systematic uncertainties are classified in experimental and theory modelling following the categorisation used for the measured fiducial cross section.

Source	Impact [%]
Experimental	
Electrons	± 0.6
Muons	± 1.3
Jets and $E_{\mathrm{T}}^{\mathrm{miss}}$	± 3.2
b-tagging	± 2.1
Pile-up	± 1.6
Background, statistical	± 3.2
Background, misid. leptons	± 3.3
Background, charge misrec.	± 0.3
Background, other	± 1.8
Theory modelling	
$W^{\pm}W^{\pm}jj$ electroweak-strong interference	± 1.0
$W^{\pm}W^{\pm}jj$ EW, EW corrections	± 1.4
$W^{\pm}W^{\pm}jj$ EW, shower, scale, PDF & α_s	± 2.8
$W^{\pm}W^{\pm}jj$ QCD	± 2.9
WZ	± 3.3
Luminosity	± 2.4

9.7 Studies with Alternative Signal Sample

In order to verify the independence of the measured cross section from the signal template, studies are finally performed using the alternative sample produced with POWHEG+PYTHIA8. The differences between the two predictions have been discussed in Section 9.2.4. Two studies are conducted. One consists in performing the fitting procedure with an additional systematic uncertainty that account for the signal mismodelling, whilst the other uses the alternative POWHEG sample for the signal, without additional systematic uncertainties. These tests have been performed with a version of the fitting framework slightly different from the one used to obtain the nominal results shown in Eqs.(9.16) and (9.17). For this reason, the values for the signal strength and cross section obtained with the previous version are reported, in order to allow for a consistent comparison with the results obtained from the studies.
The signal strength obtained with the previous version of the fitting framework is

$$\mu_{\mathrm{obs,Sherpa}} = 1.45^{+0.26}_{-0.24}(\mathrm{stat})\ ^{+0.13}_{-0.14}(\mathrm{syst}), \tag{9.18}$$

while the cross section is measured to be

$$\sigma^{\mathrm{fid}}_{\mathrm{meas,Sherpa}} = 2.91^{+0.51}_{-0.47}(\mathrm{stat})\ \pm 0.27(\mathrm{syst})\ \mathrm{fb}, \tag{9.19}$$

for an observed significance of 6.9σ, when 4.6σ was expected.

Nominal fit with additional systematic uncertainty The shape difference between the SHERPA and POWHEG sample is assigned as an additional systematic uncertainty on the SHERPA signal. To do so the m_{jj} POWHEG distribution is scaled by

$$\frac{\sigma^{\text{fid}}_{\text{theo,Sherpa}}}{\sigma^{\text{fid}}_{\text{theo,Powheg}}} \approx 0.65$$

and the one-sided shape variation is symmetrised by the fit and accounted for with a nuisance parameter of the $W^{\pm}W^{\pm}jj$ EW SHERPA signal in the signal region. The variation is shown in Fig. 9.12 in the signal region for the six channels combined, since no substantial difference is expected among different channels.

A significance of 7.0σ is obtained, where 4.6σ was expected. The signal strength is measured to be

$$\mu_{\text{obs,shape syst}} = 1.44^{+0.26}_{-0.24}(\text{stat}) \pm 0.14(\text{syst}).$$

The value of the signal strength in this case agrees with the one of Eq. (9.18), with a slight increase in the systematic uncertainties, going from $^{+0.13}_{-0.14}$ to ±0.14, since here an additional contribution is involved.

The fiducial cross section measured is

$$\sigma^{\text{fid}}_{\text{meas,shape syst}} = 2.89^{+0.51}_{-0.48}(\text{stat})^{+0.29}_{-0.27}(\text{syst})\ \text{fb},$$

where differences with respect to the result of Eq. (9.19) lay within 1% of the cross section value.

The pulls on the nuisance parameters does not present major differences with respect to the nominal ones, both shown in Appendix C in Figs. C.12 and C.11 respectively. The additional nuisance parameter relative to the POWHEG-SHERPA difference is pulled down of about 1-σ variation, which corresponds to the variation towards the POWHEG shape. The parameter is slightly constrained by the fit. The fact that the other nuisance parameters are not affected by the change in the signal shape, and that the new one absorbs the effect is an important additional confirmation of the stability of the result.

Full fit with alternative signal sample A further test of the fit independence from the signal sample is given by performing the full fit with the alternative POWHEG $W^{\pm}W^{\pm}jj$ EW signal sample, and its related theoretical uncertainties. For the $W^{\pm}W^{\pm}jj$ QCD process as well as for the other backgrounds the nominal predictions hold.

The full fit is performed with the alternative signal, and the observed significance results in 7.0σ compared to a 6.9σ expected. The higher expected significance with respect to the nominal SHERPA sample, and consequent better agreement between the expected and observed values, is related to the higher expected cross section of the POWHEG sample. The signal strength obtained from the fit

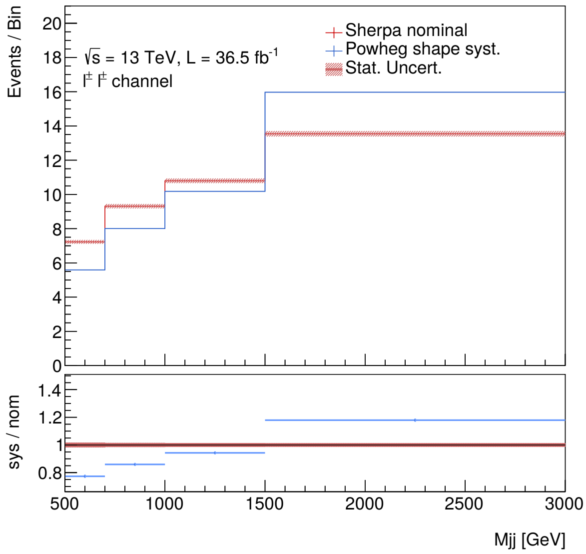


Figure 9.12: One-sided variation on the SHERPA signal due to the additional systematic given by the shape difference between SHERPA and POWHEG sample in the six lepton channels combined.

is

$$\mu_{\text{obs,Powheg}} = 0.94\ ^{+0.17}_{-0.16}(\text{stat})\ ^{+0.08}_{-0.07}(\text{syst}),$$

which in this case as expected is in better agreement with μ = 1, with respect to the nominal one of Eq (9.18).

The theoretical expected fiducial cross section for the POWHEG signal sample reported in Table 9.1 is

$$\sigma^{\text{fid}}_{\text{theo,Powheg}} = 3.08^{+0.45}_{-0.46}\ \text{fb},$$

and is used to extract the cross section

$$\sigma^{\text{fid}}_{\text{meas,Powheg}} = 2.89\ ^{+0.52}_{-0.49}(\text{stat})\ ^{+0.25}_{-0.22}(\text{syst})\ \text{fb}.$$

Also in this case, differences between this result and the nominal one of Eq. (9.19) lay within 1%.

The results for the $W^{\pm}W^{\pm}jj$ EW cross sections measured with the SHERPA 2.2.2 signal sample, with an additional systematic uncertainty accounting for the shape differences between the SHERPA and POWHEG sample and with the POWHEG signal sample are

$$\begin{aligned}
\sigma^{\text{fid}}_{\text{meas,Sherpa}} &= 2.91^{+0.51}_{-0.47}(\text{stat})\ \pm 0.27(\text{syst})\ \text{fb} \\
\sigma^{\text{fid}}_{\text{meas,shape syst}} &= 2.89\ ^{+0.51}_{-0.48}(\text{stat})\ ^{+0.29}_{-0.27}(\text{syst})\ \text{fb} \\
\sigma^{\text{fid}}_{\text{meas,Powheg}} &= 2.89\ ^{+0.52}_{-0.49}(\text{stat})\ ^{+0.25}_{-0.22}(\text{syst})\ \text{fb}.
\end{aligned}$$

The three values are found to be in agreement, demonstrating the negligible impact of the signal template on the physical quantity measured.

The POWHEG prediction is overlaid to the post-fit m_{jj} distribution, after being normalised to the

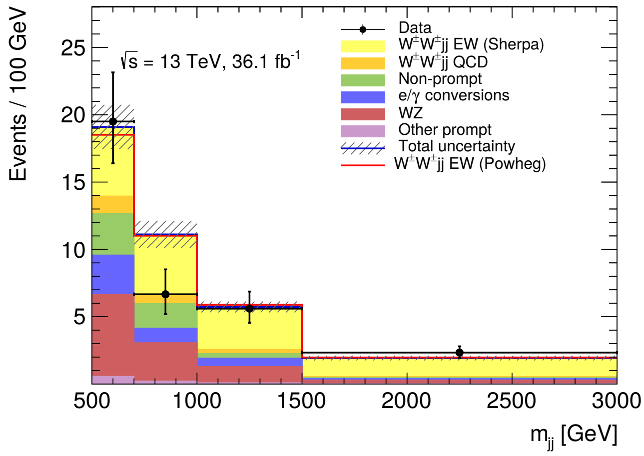


Figure 9.13: Post-fit m_{jj} distribution where the POWHEG signal prediction (red line) is shown overlaid, normalised to the SHERPA post-fit distribution (yellow area) in each of the six channels [157].

post-fit SHERPA signal, in Fig. 9.13. It is visible how the POWHEG shape is very close to the one obtained for SHERPA from the fit.

Finally a comparison between the measured cross section and the two different theoretical predictions is shown in Fig. 9.14.

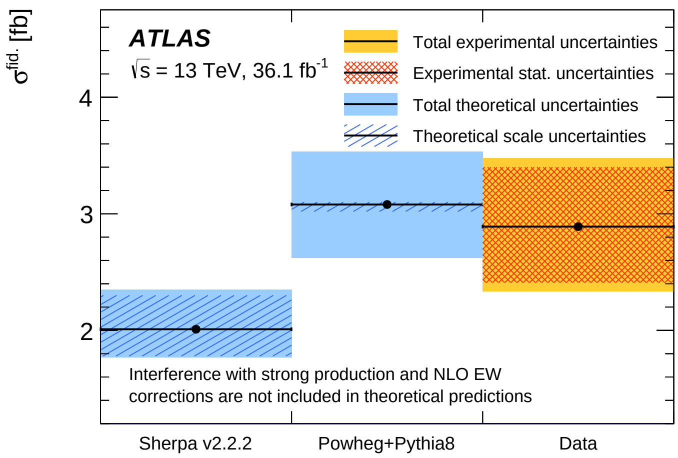


Figure 9.14: Measured fiducial cross section compared to the theoretical values calculated on the nominal SHERPA 2.2.2 sample and POWHEG+PYTHIA8. The light orange band represents the statistical and experimental uncertainties combined, while the checked orange band indicates the statistical-only uncertainty. The total theoretical uncertainties are depicted by a light blue band, and include the PDF and parton shower uncertainty, in addition to the scale dependence uncertainties which are shown also separately as a dashed blue band. Neither the interference of $W^{\pm}W^{\pm}jj$ EW and strong production, nor the NLO electroweak corrections are included in the theoretical predictions [5].

9.8 Outlook: Interpretation of Results

The result obtained is the first ATLAS observation and measurement of the electroweak $W^{\pm}W^{\pm}jj$ process. This is also the first time that a process which involves a diagram with quartic gauge coupling is accessed with the ATLAS detector. As stated in Section 2.4.1, measurements of VBS and of W boson scattering in particular, are a favourable way for the investigation of the electroweak sector of the SM and of the role of the Higgs boson.

While the existence of trilinear gauge couplings has been proven at the LEP experiment [159], no test of the existence of the quartic coupling has been performed up to now. A possible approach is carried on in this analysis, by investigating scenarios different from the SM predictions. Already in Ref. [160] the cross section of the $W^{+}W^{-}$ scattering is calculated for a set of Feynman diagrams removing the quartic-coupling contribution. A similar strategy is utilised, by generating different sets of events with calculations based on the complete set of diagrams predicted by the SM, or removing contributions from the quartic $WWWW$ coupling or the ones involving the Higgs boson couplings. However, it is not possible to arbitrary subtract the contributions from those diagrams without violating the gauge invariance. Two gauge choices are tested for both scenarios where Feynman diagrams are removed, namely the unitary gauge and the Feynman gauge. The cross sections in the two cases agree within the statistical uncertainties, being

$$\sigma^{\text{unitary}}_{\text{noHiggs}} = 0.0358 \pm 0.0001\,\text{pb} \qquad \sigma^{\text{Feynman}}_{\text{noHiggs}} = 0.0359 \pm 0.0001\,\text{pb}$$

$$\sigma^{\text{unitary}}_{\text{noQGC}} = 242.0 \pm 1.5\,\text{pb} \qquad \sigma^{\text{Feynman}}_{\text{noQGC}} = 239.2 \pm 1.3\,\text{pb}.$$

This is certainly not to be intended as a prove of gauge invariance, since no conclusion can be inferred about the general behaviour with other gauge choices. The two scenarios are chosen as a qualitative test, in order to produce an educational plot showing the expectations in these two specific cases. This is presented in Fig 9.15a, where the distributions of the dilepton invariant mass is shown for the alternative predictions of the $W^{\pm}W^{\pm}jj$ electroweak process. These have been generated with MADRAPH+PYTHIA8 and propagated to reconstruction-level with a response matrix obtained with POWHEG-BOX+PYTHIA 8. The SM prediction and the no-Higgs boson scenario ones are compatible within the statistical uncertainties, while the two scenarios without the quartic gauge coupling show a clear difference in the trend, which are however in agreement with each other.

The SM predictions are shown as well in Fig. 9.15b. Data are also shown, after the subtraction of the background contribution obtained after the fit. The distribution obtained with MADRAPH+PYTHIA8 is scaled to the $W^{\pm}W^{\pm}jj$ EW signal predictions obtained after the fit.

This qualitative study shows how the SM scenario is compatible with data in each bin within the respective uncertainties.

A robust, gauge-invariant way to interpret the results is the effective field theories approach, whose formalism is introduced in Section 2.4.2. In order to obtain the information on the Wilson coefficients of Eq. (2.27), larger amount of data is needed. This would allow for producing differential distributions, and fits can be performed to extract the parameters related to the operators. The dataset used to perform the analysis is a subset of the one recorded during the full Run 2. An addi-

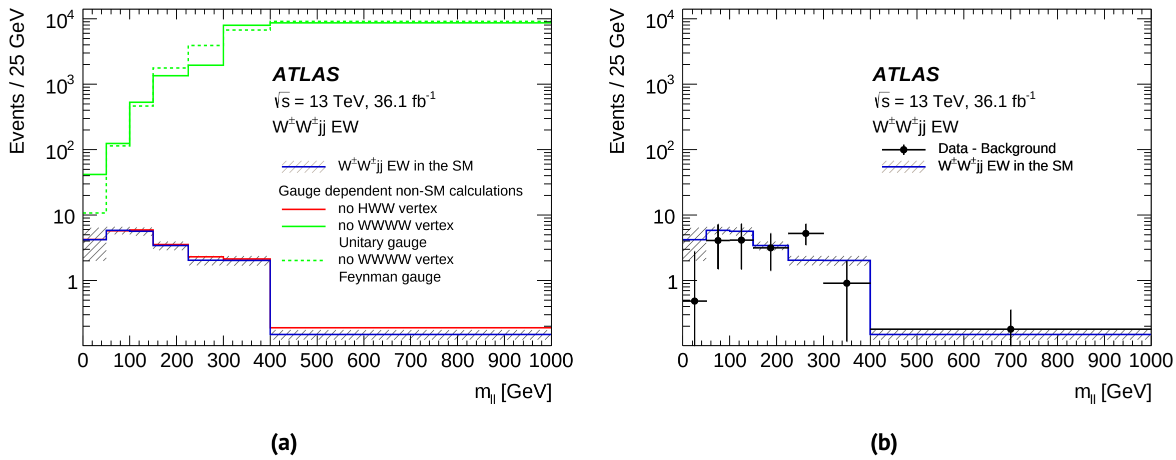


Figure 9.15: Distributions of the dilepton invariant mass for alternative predictions of the $W^{\pm}W^{\pm}jj$ electroweak process (a) and for the background subtracted data after the fit (b). The prediction in the SM (blue) are compared with calculations in the absence of the quartic $WWWW$ vertex (green) and in the absence of the Higgs boson (red). The latter two calculations are gauge dependent and shown in unitary gauge at leading-order QCD. The SM and non-SM predictions are simulated with MADRAPH+PYTHIA8. The uncertainties in the background subtracted data are the quadratic sum of the statistical uncertainties in the data, and the statistical and systematic uncertainties of the backgrounds after the fit. The uncertainties in the SM prediction include experimental and theoretical systematic uncertainties before the fit.

tional $\sim$ 100 fb^{-1} is to be included considering the full 2015-2018 run. Furthermore, approximately 300 fb^{-1} are expected for the Run 3 data taking at $\sqrt{s}$ = 14 TeV, which would allow to increase the statistical precision and do a big step towards model-independent gauge-invariant interpretations of the measurements on gauge boson couplings.

10 Summary

A measurement of the electroweak production of two same-charge W bosons decaying leptonically in association with two jets in a fiducial phase space has been performed in the course of this thesis. The analysis is conducted using a dataset corresponding to an integrated luminosity of 36.1 fb^{-1} recorded by the ATLAS detector at the Large Hadron Collider (LHC) in proton-proton collisions at $\sqrt{s}$ = 13 TeV, and has been published in Ref. [5]. It represents the first observation at ATLAS of a process involving significant contributions from Vector Boson Scattering (VBS) diagrams, and in particular from diagrams involving quartic couplings of gauge bosons, predicted by the Standard Model (SM).

The signal is defined as the $\mathcal{O}(\alpha_{\mathrm{EW}}^6)$ electroweak production of $W^\pm W^\pm jj$. The contribution with $\mathcal{O}(\alpha_s^2, \alpha_{\mathrm{EW}}^4)$, namely strong production, is considered as background, while the interference between the two modes is included in the signal measurement, but not in the signal simulation, and is addressed as theoretical uncertainty on the signal sample. The main sources of background events are primarily WZ production, processes involving non-prompt leptons and events with oppositely charged leptons that are reconstructed with the same charge due to misidentification of the charge of one electron. The latter is one of the dominant backgrounds, and is estimated in the course of this thesis with a methodology based on reweighting opposite-charge data. The reweighting procedure is based on the measurement of charge misidentification probabilities, which is performed using $Z \to ee$ events, with 33.9 fb^{-1} of data from $\sqrt{s}$ = 13 TeV proton-proton collisions. The probabilities range from less than 1% to a few percent, reaching $\sim$10% for increasing pseudorapidity values up to $|\eta|$ = 2.5, and transverse momentum larger than 120 GeV. Scale factors are obtained that correct the modelling of this effect in simulation. The corrections are sizeable for wrongly reconstructed electrons, and deviate from unity by up to 30%, with relative systematic uncertainties that range from 2% to 20%. The correction factors have been provided to the ATLAS Collaboration to be used in physics analyses sensitive to charge misidentification, and the work has been published in Ref. [6]. Electrons reconstructed with wrong charge have a lower reconstructed momentum in general, since the charge misidentification is mainly due to the interaction with the detector material that causes the electron to lose energy. This effect is accounted for in the estimation of the background from charge misidentification, and electrons from reweighted opposite-charge events are corrected with η-dependent correction factors to compensate the energy loss. The background from charge misidentification is estimated with a relative systematic uncertainty of 15%.

The kinematic criteria that define the signal region are optimised in order to enhance the expected significance of the measurement. The final signal extraction is performed with the profile likelihood method using the dijet invariant mass distributions. The cross section of the electroweak $W^\pm W^\pm jj$ process is measured in a fiducial volume, defined at generation-level by kinematic selections on the final state objects which are similar to the ones imposed at reconstruction-level in the signal region. The $W^\pm W^\pm jj$ electroweak process is observed with a significance of 6.5σ, and its fiducial cross section is measured to be

$$\sigma^{\mathrm{fid}}_{\mathrm{meas,obs}} = 2.89^{+0.51}_{-0.48}(\mathrm{stat})\,^{+0.29}_{-0.28}(\mathrm{syst})\ \mathrm{fb},$$

where the total uncertainty is dominated by statistical uncertainties. The largest systematic uncertainties are theory uncertainties on the WZ background as well as experimental uncertainties on the non-prompt lepton background estimation and on the jet and missing transverse momentum reconstruction.
Two different generators are considered for the signal, SHERPA 2.2.2 and POWHEG+PYTHIA 8.230. They present different predictions partially caused by an incorrect modelling of the parton shower in SHERPA 2.2.2, which has been identified during the course of the analysis. The measurement of the cross section is nevertheless largely independent from the simulated signal sample, and studies are presented in this thesis that prove that the mismodelling of the signal does not impact the result. In fact, the signal theoretical predictions affect only the expected significance and signal strength.
Analyses of VBS processes lay the foundations for an era of measurements that become feasible thanks to the higher centre-of-mass energies and integrated luminosities provided by the LHC, and can probe the inner structure of the SM electroweak sector and its spontaneous symmetry breaking. An interpretation of the results is given, that shows how data favours the SM in particular with respect to non-gauge-invariant scenarios without quartic couplings of the W boson. The impact of the Higgs contribution to the VBS amplitude can not be quantified with the current data and sensitivity.
In the near future even more favourable experimental conditions will allow to measure these rare processes with higher precision. Larger datasets are foreseen with the inclusion of 100 fb^{-1} from the full LHC Run 2, and the expected $\sim$ 300 fb^{-1} of data from $\sqrt{s}$ = 14 TeV collisions in Run 3. One of the most interesting extensions of this measurement is related to the polarisation of the scattering W bosons, since the restoration of the delayed unitarity by the Higgs boson is linked to their longitudinal component. Furthermore, the increased statistical power will permit to quantitatively assess the information on the gauge bosons self-couplings and parametrise it in terms of Effective Field Theories. This will allow to interpret possible deviations in a model-independent way, potentially providing hints on a theoretical extension of the Standard Model.

Contributions

Most of the material presented in this thesis is based on work performed within the ATLAS Collaboration. Thus, it is typically created from the contributions of many people and cannot be attributed to a single person. The figures and plots containing an ATLAS or ATLAS Preliminary label were created in the context of ATLAS papers or conference notes. The contributions from the author of this thesis are indicated in this section. All plots without dedicated ATLAS or ATLAS Preliminary labels or not referenced are private work done by the author of this thesis.

- Electron charge misidentification:
  - classification in simulation, Section 6.2;
  - measurement of misidentification probabilities and scale factors, Section 6.4.
- Analysis of $W^{\pm}W^{\pm}jj$ process:
  - optimisation of kinematic selection, Section 7.4;
  - estimation of the charge misidentification background for nominal selection and for subtraction in non-prompt lepton background estimation, Section 8.4;
  - studies for fit and measurement with alternative signal sample, Section 9.7.

The work on electron charge misidentification has been published, Ref. [6]: *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at* $\sqrt{s}$ *= 13 TeV*.
The analysis of the $W^{\pm}W^{\pm}jj$ process has been submitted for publication[1], Ref. [5]: *Observation of electroweak production of a same-sign* W *boson pair in association with two jets in* pp *collisions at* $\sqrt{s}$ *= 13 TeV with the ATLAS detector.*
Work has been done in the contest of the publication *The HiggsTools handbook: A beginners guide to decoding the Higgs sector* [38], related to the overview of concepts, use and limitations of the κ-framework. This is not treated in the scope of the thesis, but referenced in Section 2.4.2.

[1]Meanwhile published on October 25, 2019.

Bibliography

[1] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[2] ATLAS Collaboration, *Evidence for Electroweak Production of $W^{\pm}W^{\pm}jj$ in pp Collisions at $\sqrt{s} = 8$ TeV with the ATLAS Detector*, Phys. Rev. Lett. **113** (2014) 141803, arXiv:1405.6241 [hep-ex].

[3] CMS Collaboration, *Study of vector boson scattering and search for new physics in events with two same-sign leptons and two jets*, Phys. Rev. Lett. **114** (2015) 051801, arXiv:1410.6315 [hep-ex].

[4] CMS Collaboration, *Observation of Electroweak Production of Same-Sign W Boson Pairs in the Two Jet and Two Same-Sign Lepton Final State in Proton–Proton Collisions at* 13 *TeV*, Phys. Rev. Lett. **120** (2018) 081801, arXiv:1709.05822 [hep-ex].

[5] ATLAS Collaboration, *Observation of Electroweak Production of a Same-Sign W Boson Pair in Association with Two Jets in pp Collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector*, Phys. Rev. Lett. **123** (2019) 161801, https://link.aps.org/doi/10.1103/PhysRevLett.123.161801.

[6] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at $\sqrt{s}$ = 13 TeV*, Eur. Phys. J. (2019), arXiv:1902.04655 [hep-ex].

[7] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995. http://www.slac.stanford.edu/~mpeskin/QFT.html.

[8] F. Halzen and A. D. Martin, *Quarks and Leptons: An Introductory Course In Modern Particle Physics*. 1984.

[9] G. Sterman, *An Introduction to Quantum Field Theory*. Cambridge University Press, 1993.

[10] D. J. Griffiths, *Introduction to elementary particles; 2nd rev. version*. Physics textbook. Wiley, New York, NY, 2008. https://cds.cern.ch/record/111880.

[11] A. Djouadi, *The Anatomy of electro-weak symmetry breaking. I: The Higgs boson in the standard model*, Phys. Rept. **457** (2008) 1–216, arXiv:hep-ph/0503172 [hep-ph].

[12] Particle Data Group Collaboration, M. Tanabashi et al., *Review of Particle Physics*, Phys. Rev. D **98** (2018) 030001, https://link.aps.org/doi/10.1103/PhysRevD.98.030001.

[13] K. Olive, *Review of Particle Physics*, Chinese Physics C **40** (2016) 100001, https://doi.org/10.1088%2F1674-1137%2F40%2F10%2F100001.

[14] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, *Experimental Test of Parity Conservation in Beta Decay*, Phys. Rev. **105** (1957) 1413–1415, https://link.aps.org/doi/10.1103/PhysRev.105.1413.

[15] S. L. Glashow, *Partial-symmetries of weak interactions*, Nuclear Physics **22** (1961) 579 – 588, http://www.sciencedirect.com/science/article/pii/0029558261904692.

[16] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264–1266, https://link.aps.org/doi/10.1103/PhysRevLett.19.1264.

[17] A. Salam and J. Ward, *Electromagnetic and weak interactions*, Physics Letters **13** (1964) 168 – 171, http://www.sciencedirect.com/science/article/pii/0031916364907115.

[18] H. Fritzsch, M. Gell-Mann, and H. Leutwyler, *Advantages of the color octet gluon picture*, Physics Letters B **47** (1973) 365 – 368, http://www.sciencedirect.com/science/article/pii/0370269373906254.

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

[20] H. D. Politzer, *Reliable Perturbative Results for Strong Interactions?*, Phys. Rev. Lett. **30** (1973) 1346–1349, https://link.aps.org/doi/10.1103/PhysRevLett.30.1346.

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

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

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

[24] P. W. Anderson, *Plasmons, Gauge Invariance, and Mass*, Phys. Rev. **130** (1963) 439–442, https://link.aps.org/doi/10.1103/PhysRev.130.439.

[25] M. Szleper, *The Higgs boson and the physics of WW scattering before and after Higgs discovery*, arXiv:1412.8367 [hep-ph].

[26] Horejsi, J., *Introduction to Electroweak Unification — Standard Model From Tree Unitarity*. https://www.worldscientific.com/doi/abs/10.1142/9789814354011_fmatter.

[27] A. Ballestrero, E. Maina, and G. Pelliccioli, *W boson polarization in vector boson scattering at the LHC*, JHEP **03** (2018) 170, arXiv:1710.09339 [hep-ph].

[28] B. Biedermann, A. Denner, and M. Pellen, *Complete NLO corrections to W^+W^+ scattering and its irreducible background at the LHC*, JHEP **10** (2017) 124, arXiv:1708.00268 [hep-ph].

[29] C. F. Anders et al., *Vector boson scattering: Recent experimental and theory developments*, Rev. Phys. **3** (2018) 44–63, arXiv:1801.04203 [hep-ph].

[30] B. Biedermann, A. Denner, and M. Pellen, *Large Electroweak Corrections to Vector-Boson Scattering at the Large Hadron Collider*, Phys. Rev. Lett. **118** (2017) 261801, https://link.aps.org/doi/10.1103/PhysRevLett.118.261801.

[31] B. Jäger, C. Oleari, and D. Zeppenfeld, *Next-to-leading order QCD corrections to W^+W^+jj and W^-W^-jj production via weak-boson fusion*, Phys. Rev. D **80** (2009) 034022, https://link.aps.org/doi/10.1103/PhysRevD.80.034022.

[32] A. Denner, L. Hošeková, and S. Kallweit, *Next-to-leading order QCD corrections to W^+W^+jj production in vector-boson fusion at the LHC*, Phys. Rev. D **86** (2012) 114014, https://link.aps.org/doi/10.1103/PhysRevD.86.114014.

[33] F. Campanario, M. Kerner, L. D. Ninh, and D. Zeppenfeld, *Next-to-leading order QCD corrections to W^+W^+ and W^-W^- production in association with two jets*, Phys. Rev. D **89** (2014) 054009, https://link.aps.org/doi/10.1103/PhysRevD.89.054009.

[34] T. Melia, K. Melnikov, R. Röntsch, and G. Zanderighi, *Next-to-leading order QCD predictions for W+W+jj production at the LHC*, Journal of High Energy Physics **2010** (2010) 53, https://doi.org/10.1007/JHEP12(2010)053.

[35] T. Figy, D. Zeppenfeld, and C. Oleari, *Next-to-leading order jet distributions for Higgs boson production via weak-boson fusion*, Phys. Rev. D **68** (2003) 073005, https://link.aps.org/doi/10.1103/PhysRevD.68.073005.

[36] C. Oleari and D. Zeppenfeld, *QCD corrections to electroweak $\ell\nu_\ell jj$ and $\ell^+\ell^-jj$ production*, Phys. Rev. D **69** (2004) 093004, https://link.aps.org/doi/10.1103/PhysRevD.69.093004.

[37] A. Ballestrero et al., *Precise predictions for same-sign W-boson scattering at the LHC*, Eur. Phys. J. **C78** (2018) 671, arXiv:1803.07943 [hep-ph].

[38] M. Boggia et al., *The HiggsTools handbook: A beginners guide to decoding the Higgs sector*, Journal of Physics G: Nuclear and Particle Physics **45** (2018) .

[39] R. Alonso, E. E. Jenkins, A. V. Manohar, and M. Trott, *Renormalization Group Evolution of the Standard Model Dimension Six Operators III: Gauge Coupling Dependence and Phenomenology*, JHEP **04** (2014) 159, arXiv:1312.2014 [hep-ph].

[40] B. Henning, X. Lu, T. Melia, and H. Murayama, *Higher dimension operators in the SM EFT*, JHEP **08** (2017) 016, arXiv:1512.03433 [hep-ph].

[41] J. M. Campbell, J. W. Huston, and W. J. Stirling, *Hard Interactions of Quarks and Gluons: A Primer for LHC Physics*, Rept. Prog. Phys. **70** (2007) 89, arXiv:hep-ph/0611148 [hep-ph].

[42] M. H. Seymour and M. Marx, *Monte Carlo Event Generators*, pp. , 287–319. 2013. arXiv:1304.6677 [hep-ph].

[43] S. Höche, *Introduction to parton-shower event generators*, pp. , 235–295. 2015. arXiv:1411.4085 [hep-ph].

[44] G. Salam, *Ingredients for accurate collider Physics I*,. PSI Summer School Exothiggs, August 2016, https://www.psi.ch/particle-zuoz-school/Zuoz2016EN/salam1.pdf.

[45] G. Salam, *Ingredients for accurate collider Physics II*,. PSI Summer School Exothiggs, August 2016, https://www.psi.ch/particle-zuoz-school/Zuoz2016EN/salam2.pdf.

[46] J. Benecke, T. T. Chou, C. N. Yang, and E. Yen, *Hypothesis of Limiting Fragmentation in High-Energy Collisions*, Phys. Rev. **188** (1969) 2159–2169, https://link.aps.org/doi/10.1103/PhysRev.188.2159.

[47] J. Gao, M. Guzzi, J. Huston, H.-L. Lai, Z. Li, P. Nadolsky, J. Pumplin, D. Stump, and C. P. Yuan, *CT10 next-to-next-to-leading order global analysis of QCD*, Phys. Rev. **D89** (2014) 033009, arXiv:1302.6246 [hep-ph].

[48] V. N. Gribov and L. N. Lipatov, *Deep inelastic e p scattering in perturbation theory*, Sov. J. Nucl. Phys. **15** (1972) 438–450, [Yad. Fiz.15,781(1972)].

[49] 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, [Zh. Eksp. Teor. Fiz.73,1216(1977)].

[50] G. Altarelli and G. Parisi, *Asymptotic freedom in parton language*, Nuclear Physics B **126** (1977) 298 – 318, http://www.sciencedirect.com/science/article/pii/0550321377903844.

[51] S. Forte and G. Watt, *Progress in the Determination of the Partonic Structure of the Proton*, Ann. Rev. Nucl. Part. Sci. **63** (2013) 291–328, arXiv:1301.6754 [hep-ph].

[52] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, J. Pumplin, and C. P. Yuan, *New parton distributions for collider physics*, Phys. Rev. **D82** (2010) 074024, arXiv:1007.2241 [hep-ph].

[53] T. Sjöstrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (2006) 026, arXiv:hep-ph/0603175.

[54] M. Bahr et al., *Herwig++ Physics and Manual*, Eur. Phys. J. C **58** (2008) 639, arXiv:0803.0883 [hep-ph].

[55] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann, et al., *Event generation with SHERPA 1.1*, JHEP **02** (2009) 007, arXiv:0811.4622 [hep-ph].

[56] T. Kinoshita, *Mass Singularities of Feynman Amplitudes*, Journal of Mathematical Physics **3** (1962) 650–677, https://doi.org/10.1063/1.1724268, https://doi.org/10.1063/1.1724268.

[57] T. D. Lee and M. Nauenberg, *Degenerate Systems and Mass Singularities*, Phys. Rev. **133** (1964) B1549–B1562, https://link.aps.org/doi/10.1103/PhysRev.133.B1549.

[58] J. C. Collins, *Sudakov form-factors*, Adv. Ser. Direct. High Energy Phys. **5** (1989) 573–614, arXiv:hep-ph/0312336 [hep-ph].

[59] *Parton fragmentation and string dynamics*, Physics Reports **97** (1983) 31 – 145, http://www.sciencedirect.com/science/article/pii/0370157383900807.

[60] D. Amati and G. Veneziano, *Preconfinement as a property of perturbative QCD*, Physics Letters B **83** (1979) 87 – 92, http://www.sciencedirect.com/science/article/pii/0370269379908967.

[61] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. A **462** (2001) 152.

[62] S. Jadach, J. H. Kuhn, and Z. Was, *TAUOLA - a library of Monte Carlo programs to simulate decays of polarized τ leptons*, Computer Physics Communications **64** (1991) 275 – 299, http://www.sciencedirect.com/science/article/pii/001046559190038M.

[63] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823, arXiv:1005.4568 [physics.ins-det].

[64] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4: A Simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250.

[65] ATLAS Collaboration, *Athena*, Apr., 2019. https://doi.org/10.5281/zenodo.2641997.

[66] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001.

[67] CMS Collaboration, *The CMS experiment at the CERN LHC*, JINST **3** (2008) S08004.

[68] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002.

[69] LHCb Collaboration, *The LHCb Detector at the LHC*, JINST **3** (2008) S08005.

[70] MoEDAL Collaboration, *Technical Design Report of the MoEDAL Experiment*, Tech. Rep. CERN-LHCC-2009-006. MoEDAL-TDR-001, Jun, 2009. http://cds.cern.ch/record/1181486.

[71] TOTEM Collaboration, *The TOTEM experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08007.

[72] LHCf Collaboration, *The LHCf detector at the CERN Large Hadron Collider*, JINST **3** (2008) S08006.

[73] *LEP design report*. CERN, Geneva, 1984. https://cds.cern.ch/record/102083. Copies shelved as reports in LEP, PS and SPS libraries.

[74] E. Mobs, *The CERN accelerator complex - August 2018. Complexe des accélérateurs du CERN - Août 2018*, https://cds.cern.ch/record/2636343, General Photo.

[75] O. S. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock, *LHC Design Report*. CERN Yellow Reports: Monographs. CERN, Geneva, 2004. https://cds.cern.ch/record/782076.

[76] https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2#Luminosity_summary_plots_for_AN2.

[77] S. van der Meer, *Calibration of the effective beam height in the ISR*, Tech. Rep. CERN-ISR-PO-68-31. ISR-PO-68-31, CERN, Geneva, 1968. https://cds.cern.ch/record/296752.

[78] ATLAS Collaboration, *The ATLAS Inner Detector commissioning and calibration*, Eur. Phys. J. C **70** (2010) 787, arXiv:1004.5293 [hep-ex].

[79] ATLAS Collaboration, *Study of the material of the ATLAS inner detector for Run 2 of the LHC*, JINST **12** (2017) P12009, arXiv:1707.02826 [hep-ex].

[80] ATLAS Collaboration, *ATLAS pixel detector electronics and sensors*, JINST **3** (2008) P07007–P07007.

[81] ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report*, Atlas-tdr-19, 2010, https://cds.cern.ch/record/1291633.

[82] ATLAS Collaboration, *Track Reconstruction Performance of the ATLAS Inner Detector at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2015-018, 2015, https://cds.cern.ch/record/2037683.

[83] ATLAS Collaboration, *Readiness of the ATLAS liquid argon calorimeter for LHC collisions*, Eur. Phys. J. C **70** (2010) 723, arXiv:0912.2642 [hep-ex].

[84] ATLAS Collaboration, *Readiness of the ATLAS Tile Calorimeter for LHC collisions*, Eur. Phys. J. C **70** (2010) 1193, arXiv:1007.5423 [hep-ex].

[85] ATLAS Collaboration, *ATLAS liquid-argon calorimeter: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1996. http://cds.cern.ch/record/331061.

[86] A. Artamonov et al., *The ATLAS Forward Calorimeter*, JINST **3** (2008) P02010, http://stacks.iop.org/1748-0221/3/i=02/a=P02010.

[87] ATLAS Collaboration, *Commissioning of the ATLAS Muon Spectrometer with cosmic rays*, Eur. Phys. J. C **70** (2010) 875, arXiv:1006.4384 [hep-ex].

[88] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, JINST **13** (2018) P07017–P07017, https://doi.org/10.1088%2F1748-0221%2F13%2F07%2Fp07017.

[89] Abdel Khalek, S. and others, *The ALFA Roman Pot detectors of ATLAS*, JINST **11** (2016) P11013–P11013, https://doi.org/10.1088%2F1748-0221%2F11%2F11%2Fp11013.

[90] ATLAS Collaboration, *A New ATLAS ZDC for the High Radiation Environment at the LHC*, https://cds.cern.ch/record/2319880.

[91] ATLAS Collaboration, *Proton tagging with the one arm AFP detector*, ATL-PHYS-PUB-2017-012, 2017, https://cds.cern.ch/record/2273274.

[92] A Ruiz Martínez and ATLAS Collaboration, *The Run-2 ATLAS Trigger System*, Journal of Physics: Conference Series **762** (2016) 012003, http://stacks.iop.org/1742-6596/762/i=1/a=012003.

[93] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, Eur. Phys. J. C **77** (2017) 317, arXiv:1611.09661 [hep-ex].

[94] ATLAS Collaboration, *Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2*, Eur. Phys. J. C **77** (2017) 673, arXiv:1704.07983 [hep-ex].

[95] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **262** (1987) 444 – 450, http://www.sciencedirect.com/science/article/pii/0168900287908874.

[96] ATLAS Collaboration, *A neural network clustering algorithm for the ATLAS silicon pixel detector*, JINST **9** (2014) P09009, arXiv:1406.7690 [hep-ex].

[97] T. Cornelissen, M. Elsing, I. Gavrilenko, W. Liebig, E. Moyse, and A. Salzburger, *The new ATLAS track reconstruction (NEWT)*, J. Phys. Conf. Ser. **119** (2008) 032014.

[98] ATLAS Collaboration, *Early Inner Detector Tracking Performance in the 2015 Data at* $\sqrt{s} = 13$ *TeV*, ATL-PHYS-PUB-2015-051, 2015, https://cds.cern.ch/record/2110140.

[99] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, *An Introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159, arXiv:1410.3012 [hep-ph].

[100] ATLAS Collaboration, *Charged-particle multiplicities in pp interactions measured with the ATLAS detector at the LHC*, New J. Phys. **13** (2011) 053033, arXiv:1012.5104 [hep-ex].

[101] ATLAS Collaboration, *Vertex Reconstruction Performance of the ATLAS Detector at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2015-026, 2015, https://cds.cern.ch/record/2037717.

[102] ATLAS Collaboration, *Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC*, Eur. Phys. J. C **77** (2017) 332, arXiv:1611.10235 [hep-ex].

[103] ATLAS Collaboration, *Summary of ATLAS Pythia 8 tunes*, ATL-PHYS-PUB-2012-003, 2012, https://cds.cern.ch/record/1474107.

[104] ATLAS Collaboration, *Performance of primary vertex reconstruction in proton–proton collisions at $\sqrt{s} = 7$ TeV in the ATLAS experiment*, ATLAS-CONF-2010-069, 2010, https://cds.cern.ch/record/1281344.

[105] W. Lampl et al., *Calorimeter Clustering Algorithms: Description and Performance*, ATL-LARG-PUB-2008-002, 2008, https://cds.cern.ch/record/1099735.

[106] T. G. Cornelissen, M. Elsing, I. Gavrilenko, J.-F. Laporte, W. Liebig, M. Limper, K. Nikolopoulos, A. Poppleton, and A. Salzburger, *The global χ^2 track fitter in ATLAS*, Journal of Physics: Conference Series **119** (2008) 032013, http://stacks.iop.org/1742-6596/119/i=3/a=032013.

[107] ATLAS Collaboration, *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung*, ATLAS-CONF-2012-047, 2012, https://cds.cern.ch/record/1449796.

[108] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, Eur. Phys. J. C **77** (2017) 490, arXiv:1603.02934 [hep-ex].

[109] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using data collected in 2015 at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2016-015, 2016, https://cds.cern.ch/record/2203514.

[110] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data*, Eur. Phys. J. C **74** (2014) 3071, arXiv:1407.5063 [hep-ex].

[111] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **76** (2016) 292, arXiv:1603.05598 [hep-ex].

[112] J. Illingworth and J. Kittler, *A survey of the hough transform*, Computer Vision, Graphics, and Image Processing **44** (1988) 87 – 116, http://www.sciencedirect.com/science/article/pii/S0734189X88800331.

[113] ATLAS Collaboration, *Monte Carlo Calibration and Combination of In-situ Measurements of Jet Energy Scale, Jet Energy Resolution and Jet Mass in ATLAS*, ATLAS-CONF-2015-037, 2015, https://cds.cern.ch/record/2044941.

[114] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Rev. D **96** (2017) 072002, arXiv:1703.09665 [hep-ex].

[115] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063, arXiv:0802.1189 [hep-ph].

[116] S. D. Ellis and D. E. Soper, *Successive combination jet algorithm for hadron collisions*, Phys. Rev. **D48** (1993) 3160–3166, arXiv:hep-ph/9305266 [hep-ph].

[117] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, *Better jet clustering algorithms*, JHEP **08** (1997) 001, arXiv:hep-ph/9707323 [hep-ph].

[118] ATLAS Collaboration, *Performance of b-jet identification in the ATLAS experiment*, JINST **11** (2016) P04008, arXiv:1512.01094 [hep-ex].

[119] ATLAS Collaboration, *Measurements of b-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV*, JHEP **08** (2018) 089, arXiv:1805.01845 [hep-ex].

[120] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **78** (2018) 903, arXiv:1802.08168 [hep-ex].

[121] B. Jäger and G. Zanderighi, *NLO corrections to electroweak and QCD production of W+W+ plus two jets in the POWHEG BOX*, JHEP **2011** (2011) 55, https://doi.org/10.1007/JHEP11(2011)055.

[122] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040, arXiv:hep-ph/0409146.

[123] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, JHEP **11** (2007) 070, arXiv:0709.2092 [hep-ph].

[124] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043, arXiv:1002.2581 [hep-ph].

[125] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852–867, arXiv:0710.3820 [hep-ph].

[126] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, J. Pumplin, and C.-P. Yuan, *New parton distributions for collider physics*, Phys. Rev. D **82** (2010) 074024, https://link.aps.org/doi/10.1103/PhysRevD.82.074024.

[127] ATLAS Collaboration, *Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, JHEP **09** (2014) 145, arXiv:1406.3660 [hep-ex].

[128] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, *Parton distributions for the LHC*, Eur. Phys. J. **C63** (2009) 189–285, arXiv:0901.0002 [hep-ph].

[129] M. Muskinja, B. Kersevan, and A. Gorisek, *Search for new physics processes with same charge leptons in the final state with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13$ TeV.*, Aug, 2018. https://cds.cern.ch/record/2643902. Presented 20 Sep 2018.

[130] ATLAS Collaboration. *ATLAS MCTruthClassifier*, https:612//svnweb.cern.ch/trac/atlasoff/browser/PhysicsAnalysis/MCTruthClassifier.

[131] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, JHEP **12** (2008) 039, arXiv:0808.3674 [hep-ph].

[132] S. Schumann and F. Krauss, *A Parton shower algorithm based on Catani-Seymour dipole factorisation*, JHEP **03** (2008) 038, arXiv:0709.1027 [hep-ph].

[133] S. Höche, F. Krauss, S. Schumann, and F. Siegert, *QCD matrix elements and truncated showers*, JHEP **05** (2009) 053, arXiv:0903.1219 [hep-ph].

[134] S. Höche, F. Krauss, M. Schönherr, and F. Siegert, *QCD matrix elements + parton showers: The NLO case*, JHEP **04** (2013) 027, arXiv:1207.5030 [hep-ph].

[135] M. Schönherr and F. Krauss, *Soft Photon Radiation in Particle Decays in SHERPA*, JHEP **12** (2008) 018, arXiv:0810.5071 [hep-ph].

[136] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **07** (2014) 079, arXiv:1405.0301 [hep-ph].

[137] NNPDF Collaboration, R. D. Ball et al., *Parton distributions for the LHC Run II*, JHEP **04** (2015) 040, arXiv:1410.8849 [hep-ph].

[138] M. Cacciari and G. P. Salam, *Pileup subtraction using jet areas*, Phys. Lett. **B659** (2008) 119–126, arXiv:0707.1378 [hep-ph].

[139] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*, Eur. Phys. J. C **76** (2016) 581, arXiv:1510.03823 [hep-ex].

[140] ATLAS Collaboration, *Selection of jets produced in 13 TeV proton–proton collisions with the ATLAS detector*, ATLAS-CONF-2015-029, 2015, https://cds.cern.ch/record/2037702.

[141] S. M. Weber, *Operation and Performance of the ATLAS Level-1 Calorimeter and Topological Triggers in Run 2*, Tech. Rep. ATL-DAQ-PROC-2017-026, CERN, Geneva, Oct, 2017. https://cds.cern.ch/record/2287401.

[142] Cowan, G., *Discovery Sensitivity for a Counting Experiment with Background Uncertainty*, Tech. Rep. Royal Holloway, London, May, 2012. http://www.pp.rhul.ac.uk/~cowan/stat/medsig/medsigNote.pdf.

[143] ATLAS Collaboration, *Measurement of $W^{\pm}W^{\pm}$ vector-boson scattering and limits on anomalous quartic gauge couplings with the ATLAS detector*, Phys. Rev. D **96** (2017) 012007, arXiv:1611.02428 [hep-ex].

[144] Liqing Zhang, *Private communication.*

[145] Wenhao Xu, *Private communication.*

[146] Chilufya Mwewa, *Private communication.*

[147] W. Verkerke and D. Kirkby, *The RooFit toolkit for data modeling*, 2003, arXiv:physics/0306116 [physics.data-an].

[148] L. Moneta, K. Cranmer, G. Schott, and W. Verkerke, *The RooStats project*, p. , 57. Jan, 2010. arXiv:1009.1003 [physics.data-an].

[149] ROOT Collaboration, K. Cranmer, G. Lewis, L. Moneta, A. Shibata, and W. Verkerke, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, Tech. Rep. CERN-OPEN-2012-016, New York U., New York, Jan, 2012. https://cds.cern.ch/record/1456844.

[150] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. C **71** (2011) 1554, arXiv:1007.1727 [physics.data-an].

[151] M. Chiesa, A. Denner, J.-N. Lang, and M. Pellen, *An event generator for same-sign W-boson scattering at the LHC including electroweak corrections*, arXiv:1906.01863 [hep-ph].

[152] ATLAS Collaboration, *Modelling of the vector boson scattering process $pp \to W^{\pm}W^{\pm}jj$ in Monte Carlo generators in ATLAS*, ATL-PHYS-PUB-2019-004, 2019, https://cds.cern.ch/record/2655303.

[153] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC*, Eur. Phys. J. C **76** (2016) 653, arXiv:1608.03953 [hep-ex].

[154] ATLAS Collaboration, *Multi-boson simulation for 13 TeV ATLAS analyses*, ATL-PHYS-PUB-2016-002, 2016, https://cds.cern.ch/record/2119986.

[155] ATLAS Collaboration, *Multi-Boson Simulation for 13 TeV ATLAS Analyses*, ATL-PHYS-PUB-2017-005, 2017, https://cds.cern.ch/record/2261933.

[156] LHC Higgs Cross Section Working Group Collaboration, D. de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, arXiv:1610.07922 [hep-ph].

[157] Analysis Fitting Team, *Private communication*.

[158] Karolos Potamianos, *Private communication*.

[159] *Electroweak measurements in electron–positron collisions at W-boson-pair energies at LEP*, Physics Reports **532** (2013) 119 – 244, http://www.sciencedirect.com/science/article/pii/S0370157313002706, Electroweak Measurements in Electron-Positron Collisions at W-Boson-Pair Energies at LEP.

[160] A. Denner and T. Hahn, *Radiative corrections to $W^+W^- \to W^+W^-$ in the electroweak standard model*, Nuclear Physics B **525** (1998) 27 – 50, http://www.sciencedirect.com/science/article/pii/S0550321398002879.

[161] https://www.maplesoft.com/.

[162] https://www.higgstools.org/.

[163] B. Siciliano and O. Khatib, eds., *Springer Handbook of Robotics*. Springer Handbooks. Springer, Berlin, 2 ed., 2016.

[164] D. Stewart, *A Platform with Six Degrees of Freedom*, Proceedings of the Institution of Mechanical Engineers **180** (1965) 371–386, https://doi.org/10.1243/PIME_PROC_1965_180_029_02.

[165] *Kuka*, https://www.kuka.com/en-de.

Appendix

A Charge Misidentification Truth Studies

The need of measuring the charge misidentification probabilities carries the broader issue of a correct *truth-level* definition of electrons to be used to assess the charge reconstruction. In particular, the problem rises in case of bremsstrahlung emission of a photon that converts into an electron–positron pair.

In such cases, the first aim is being able to reach the parent electron of the bremsstrahlung photon, instead of the product of the conversion, usually matched as *true* counterpart of the reconstructed electron. The second point concerns the cases of multiple bremsstrahlung and conversion processes, where a way to recover the *generated* electron is needed, by tracking back the chain and retrieving the first electron not coming from the detector simulation step, but from the generation.

The studies performed aimed at finding an unambiguous way of defining such electron, and provide a baseline definition to be used for charge misidentification measurements.

The $Z \to ee$ sample used was generated with Powheg [122], whith the parton shower process performed by PYTHIA8 [99] and the particle decays by EVTGEN [61].

The studies rely on the `MCTruthClassifier` tool [130], the ATLAS Athena [65] algorithm that classifies the particles in simulation, i.e. the *truth* particles, according to where they originated from. It provides attributes to each particle, in particular the particle's origin `TruthOrigin`, type `TruthType`and final state. The composition of the $Z \to ee$ sample in terms of `TruthOrigin` and `TruthType` is shown in Fig. A.1, and the values of interest for the two variables are summarised in Table. A.1.

Table A.1: `TruthType` and `TruthOrigin` classification for electrons in a $Z \to ee$ sample as from Fig. A.1 [130].

`TruthType`	Classification
0	unknown
1	unknown electron
2	isolated electron
3	non-isolated electron
4	background electron

`TruthOrigin`	Classification
0	undefined
5	photon conversion
6	Dalitz decay
7	electromagnetic process
8	muon
13	Z boson
23	light meson
25	charmed meson
26	bottom meson
27	$c\bar{c}$ meson
29	$b\bar{b}$ meson
33	bottom baryon

The main contribution comes from electrons of `TruthType` *isolated* [2] with `TruthOrigin` from decays of a Z boson. The second greatest contribution comes from electrons of `TruthType` background. This class refers to electrons from photon conversions, Dalitz or meson decays, and in

[2] The term *isolated* in this contest indicates “signal-like” electrons, and is not to be confused with the lepton isolation of Chapter 5

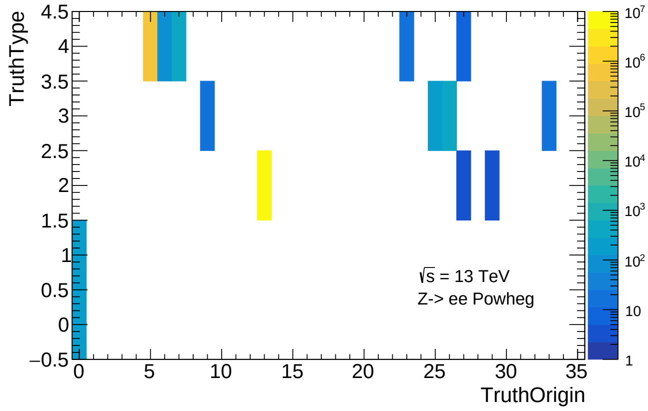


Figure A.1: TruthType and TruthOrigin of electrons in a $Z \to ee$ sample. The biggest contribution comes from electrons from Z boson as expected, i.e. TruthType = 2 and TruthOrigin = 13. The second largest contribution refers to background electrons from photon conversion, i.e. TruthType = 4 and TruthOrigin = 5. Details on the reference values for the two variables are given in Table A.1.

this particular case their TruthOrigin is photon conversion. From Fig. A.1 it is visible how the requirement on the TruthType = 2 is sufficient to define the electrons from the boson decay, and the TruthOrigin = 5 to define the electrons from conversion. Hence, the expression

$$\texttt{TruthType == 2 || TruthOrigin == 5} \tag{A.1}$$

covers the need of excluding all the background electrons which are not from conversion.
However, to address the issue of the multiple bremsstrahlung production and conversion chain a new variable is introduced, the FirstEgMother. This univocally identifies the electron at the origin of the conversion chain, being able to reach the first non-GEANT particle, as depicted in Fig. A.2.
The corresponding classification of the MCTruthClassifier is performed on this particle, and a selection based on its TruthType and TruthOrigin, as well as its PDG identification number (PDG ID) [12] ensures that the particle comes from a boson decay. The FirstEgMother is found by looping back through the decay chain of each truth particle, starting at its production vertex [3] and going back to parent vertices till the first non-electron or non-photon one, or till the first generator-vertex, i.e. not coming from the GEANT simulation. The daughter particle of this vertex, the FirstEgMother, is the electron or photon which holds the information of the original.
The definition of prompt electron used for charge misidentification studies of Eq.(A.1) becomes

$$\texttt{TruthType == 2 || (TruthOrigin == 5 \&\& |FirstEgMotherPdgID| == 11)}. \tag{A.2}$$

where the second part of the definition ensures the existence of an original electron from simulation at the origin of the conversion chain if the electron is classified as background. An electron that satisfies the criteria from Eq.(A.2) is what is defined in the main body as *prompt*, in the broader

[3] In this contest this is not intended as the reconstructed vertex as in Chapter 5, but as the splitting point where a particle is created in simulation.

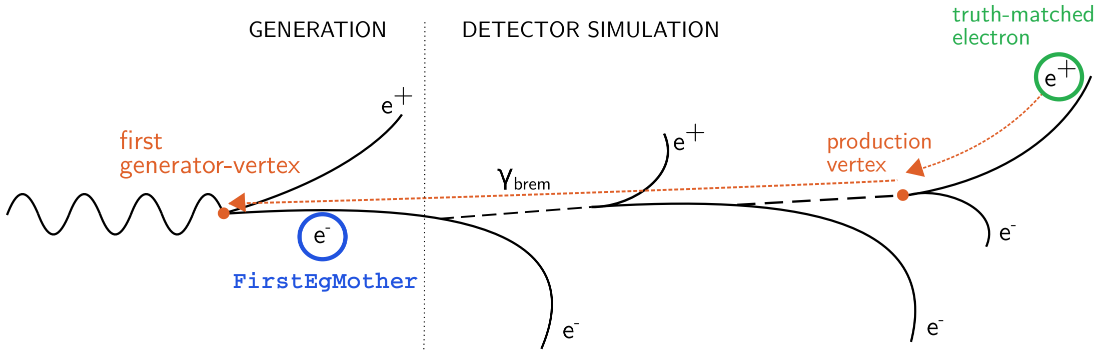


Figure A.2: Scheme of the chain to trace the original true electron that underwent charge misidentification. The decay chain is looped back from the production vertex of the truth-matched electron from detector simulation GEANT, to the first generator-level vertex. The daughter particle of this vertex is identified as the new `FirstEgMother` variable, and holds the information on the true original charge to be used for charge misidentification studies.

meaning of an electron originated in the generator for which it is meaningful to measure the charge misidentification probability.
A prescription for the definition of a wrong charge-electron can now be outlined on electrons that match the baseline prompt requirement of Eq.(A.2). That is

$$\texttt{recoCharge x firstEgMotherPdgID > 0}, \tag{A.3}$$

being the electron PDG ID $+11$ and the positron one -11.
These results have been introduced in the ATLAS framework and used for the studies and measurements of charge misidentification in simulation starting from the first ATLAS Run 2 analyses, including the one presented in this thesis.

B Multi-Dimensional Grid Scan

The distributions of the variables affected by the optimisation process in the multi-dimensional grid scan are shown in Fig. B.3 and B.4 for the $\mu\mu$ and the $e\mu + \mu e$ channels. The kinematic region used for the optimisation is described in the main body in Table 7.4.

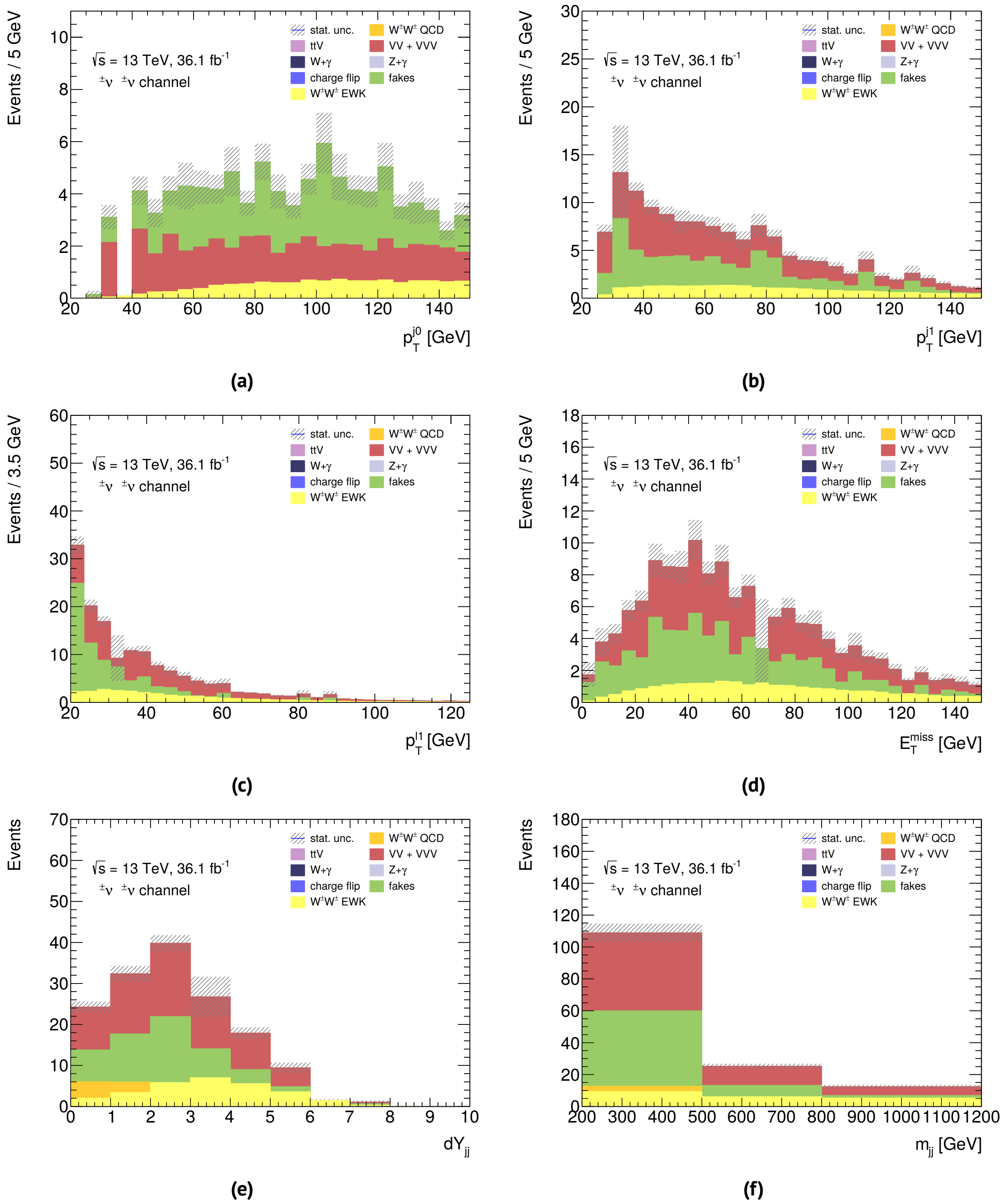


Figure B.3: Distributions of transverse momentum of the leading (a) and sub-leading (b) jet, sub-leading lepton (c), missing transverse energy (d), rapidity separation between the two tagging jets (e) and dijet invariant mass (f) in the optimisation region for the $\mu\mu$ channel. Only statistical uncertainties are shown.

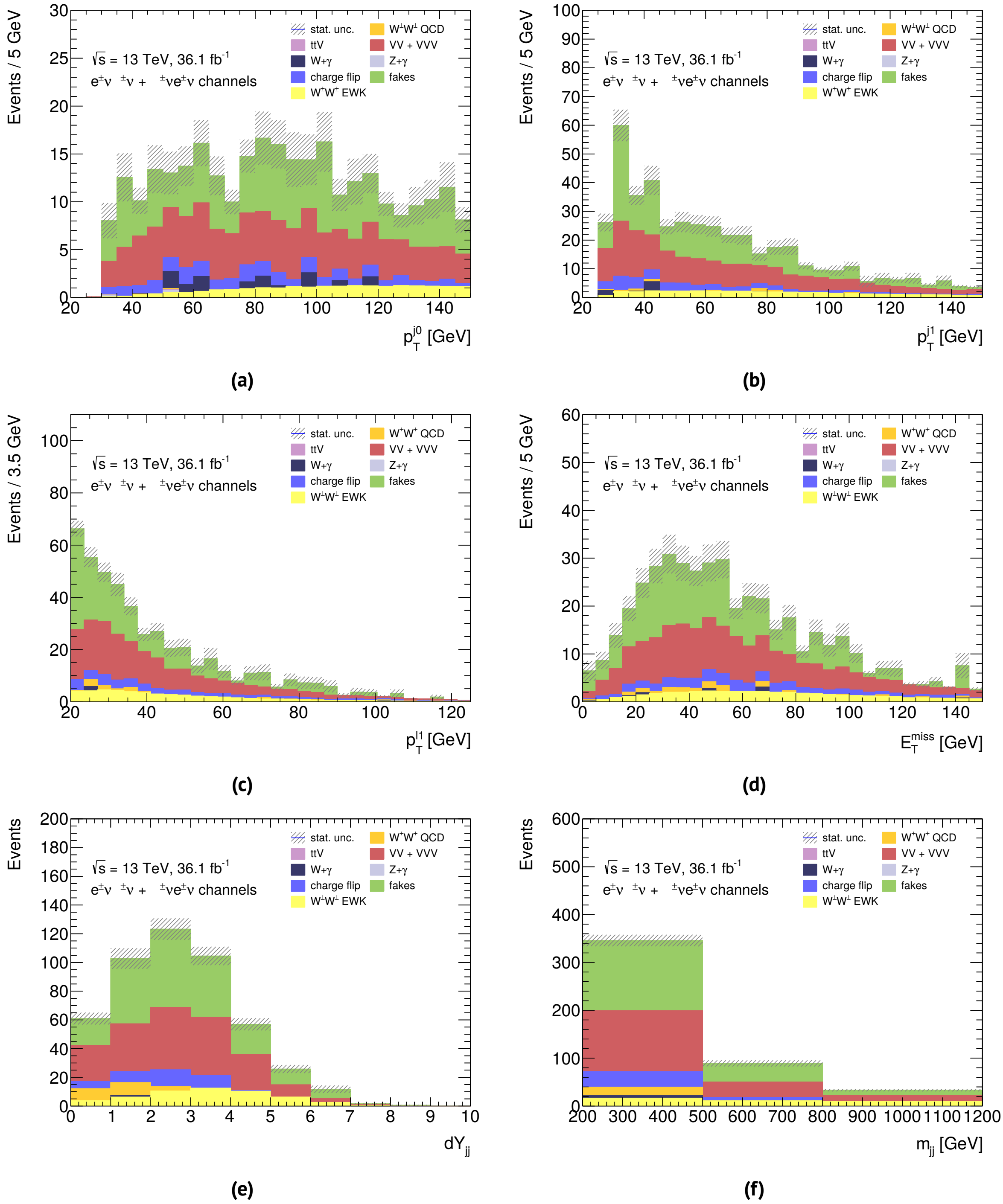


Figure B.4: Distributions of transverse momentum of the leading (a) and sub-leading (b) jet, sub-leading lepton (c), missing transverse energy (d), rapidity separation between the two tagging jets (e) and dijet invariant mass (f) in the optimisation region for the $e\mu + \mu e$ channel. Only statistical uncertainties are shown.

C Additional Plots

Charge Misidentification Scale Factors

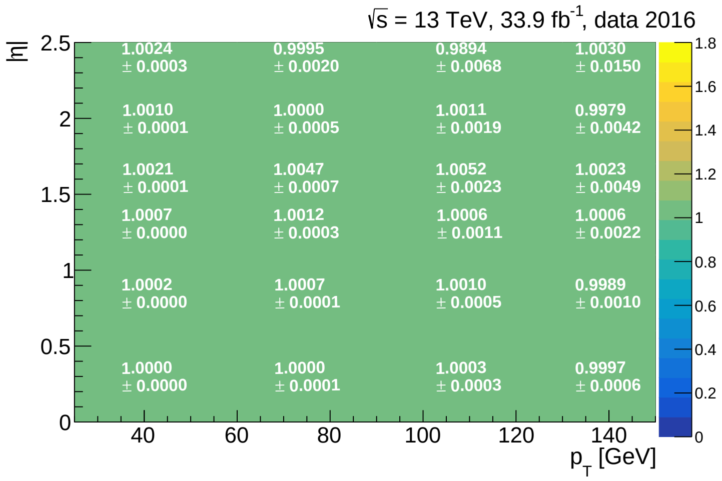


Figure C.5: Charge misidentification scale factors for correctly reconstructed electrons. Electrons selected with the Tight and FixedCutTight identification and isolation operating point. Only statistical uncertainties are shown.

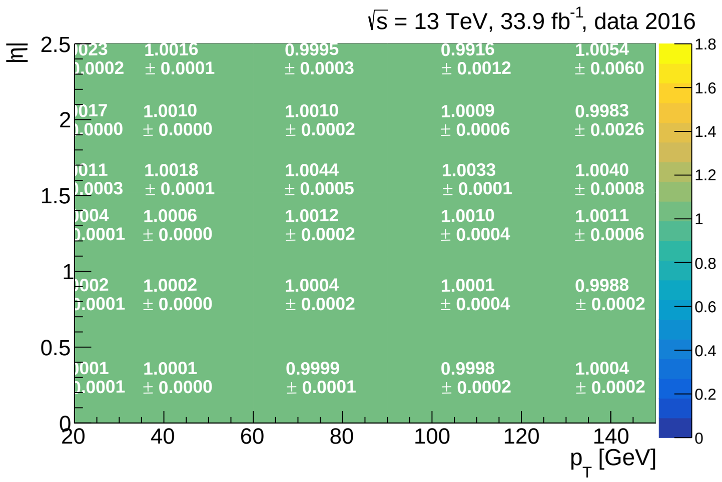


Figure C.6: Charge misidentification scale factors for correctly reconstructed electrons. Electrons are selected with the Tight and isoGradient identification and isolation operating point, and are dedicated to the kinematic selection used in the $W^{\pm}W^{\pm}jj$ analysis, detailed in Table. 7.2 of Ch. 7. Absolute systematic uncertainties are shown.

Control Region Fit

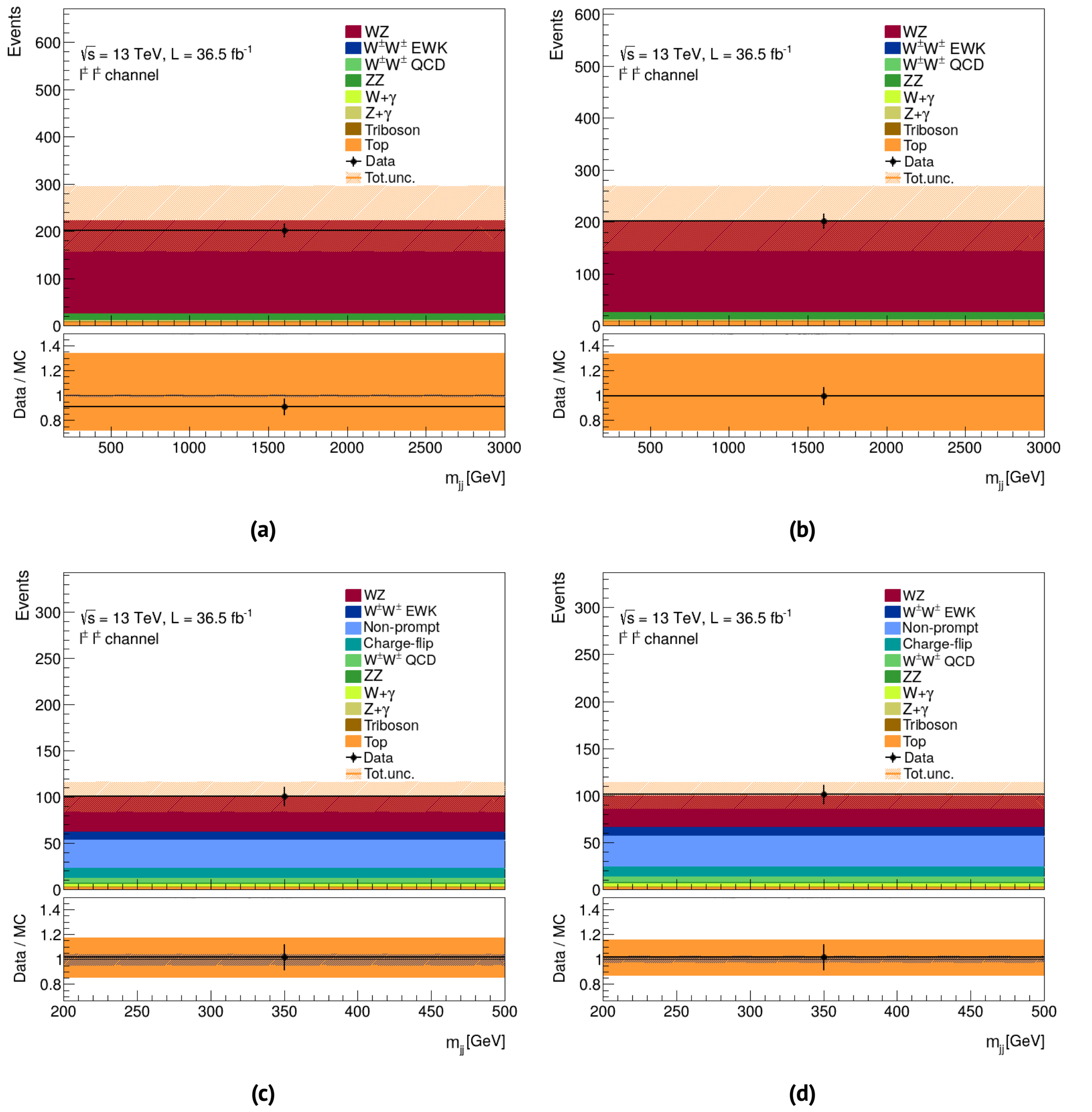


Figure C.7: Pre-(a,c) and post-(b,d) fit distributions for the WZ (top) and low-m_{jj} (bottom) for the combined $\ell^{\pm}\ell^{\pm}$ channel. In the ratio plots the orange band represents the systematic uncertainties while the statistical uncertainties are depicted in blue.

Control Region Fit

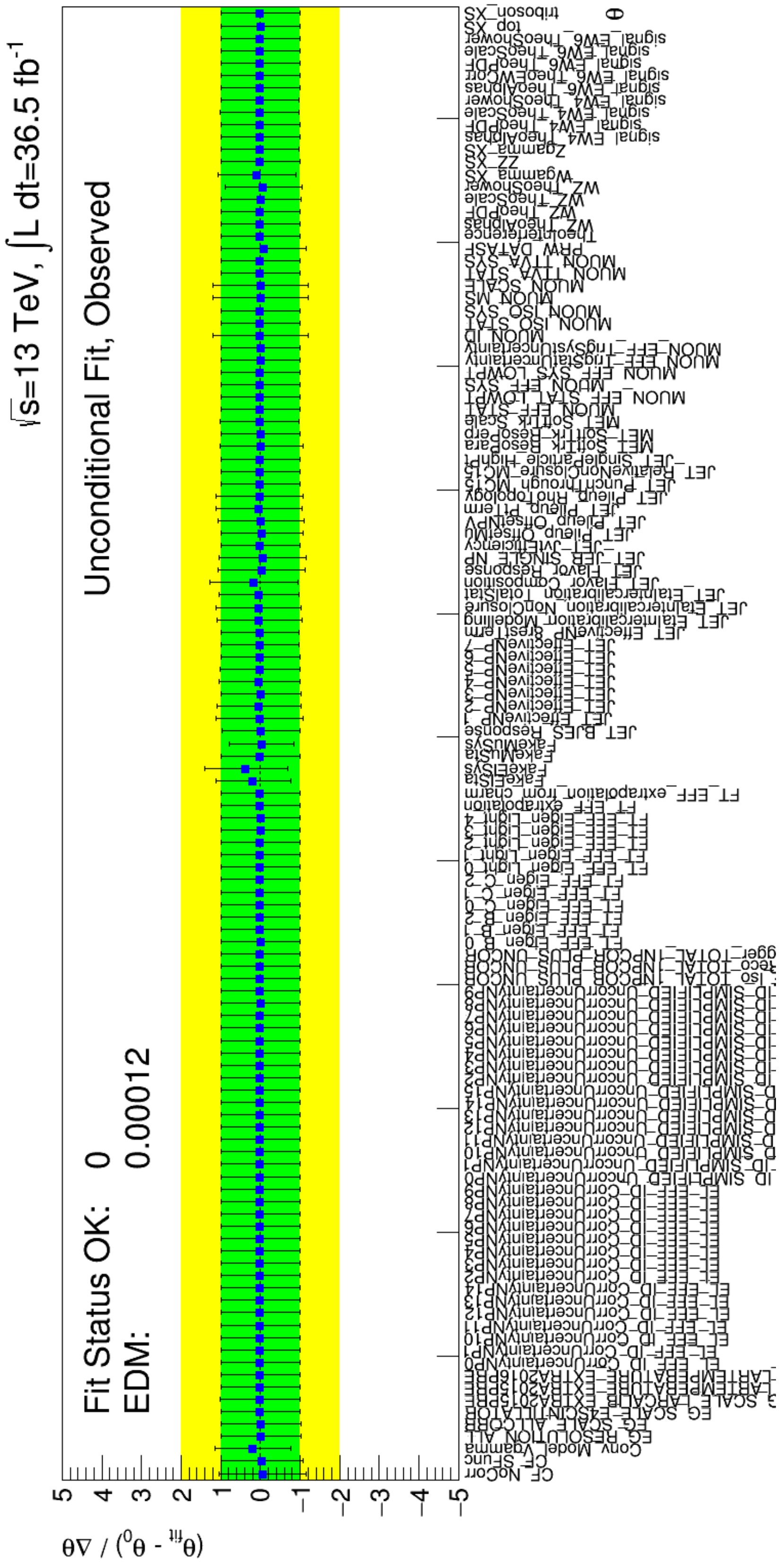


Figure C.8: Constrains on nuisance parameters extracted from the combined fit to observed data in the WZ and low-m_{jj} control regions. The error on each nuisance parameter obtained after the fit is reported in unit of the pre-fit uncertainty.

Expected Performance of the Full Fit

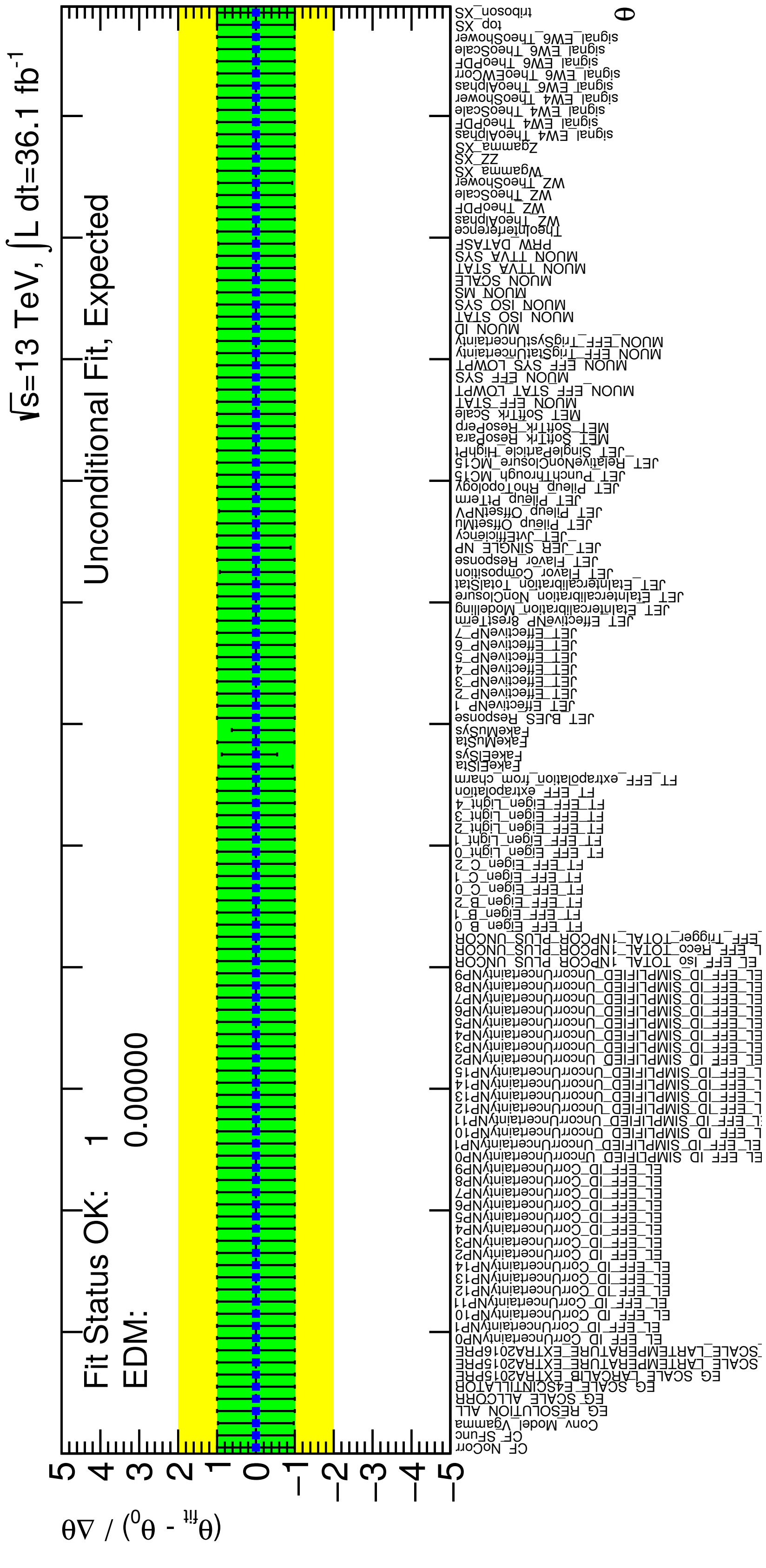


Figure C.9: Constrains on nuisance parameters extracted from the combined fit to Asimov data in the signal region, WZ and low-m_{jj} control regions. The error on each nuisance parameter obtained after the fit is reported in unit of the pre-fit uncertainty.

Expected Performance of the Full Fit

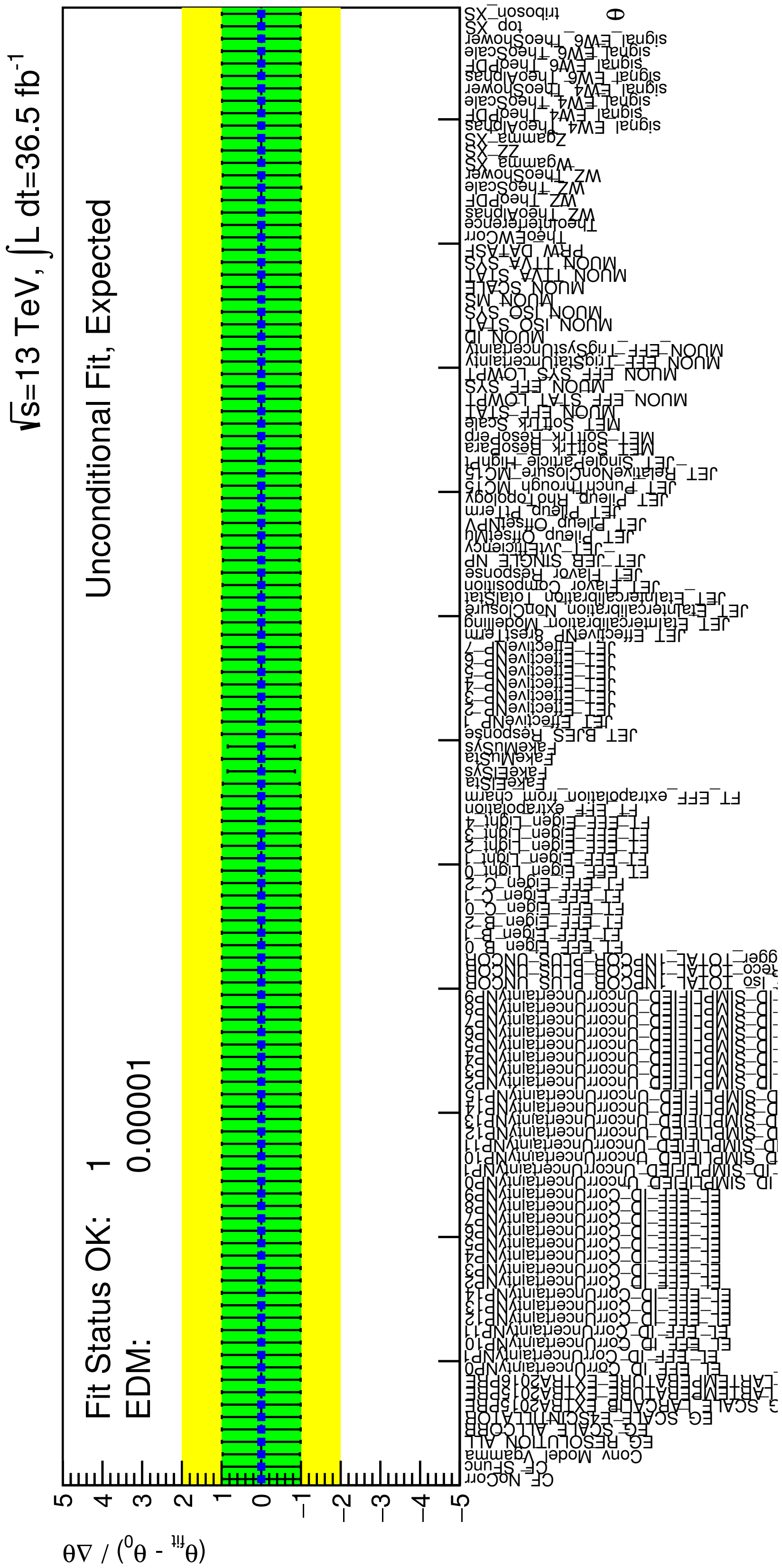


Figure C.10: Constrains on nuisance parameters extracted from the combined fit to Asimov data in the signal region and WZ control region excluding the low-m_{jj} control region. The error on each nuisance parameter obtained after the fit is reported in unit of the pre-fit uncertainty.

Fit with Observed Data

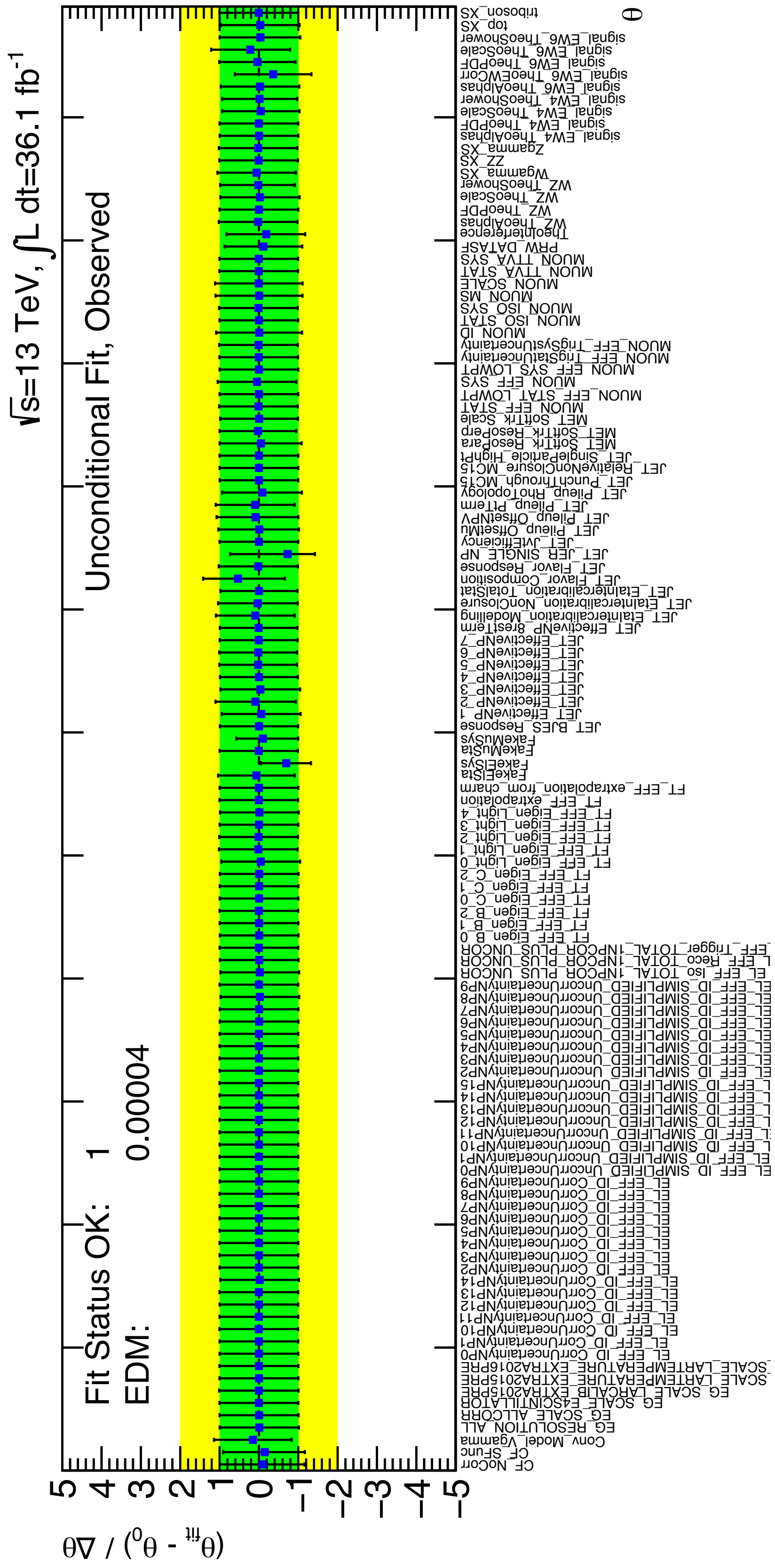


Figure C.11: Nuisance parameters pulls for the combined unconditional fit to observed data in the signal region and WZ and low-m_{jj} control regions. The error on each nuisance parameter obtained after the fit is reported in unit of the pre-fit uncertainty.

Studies with Alternative Signal Sample

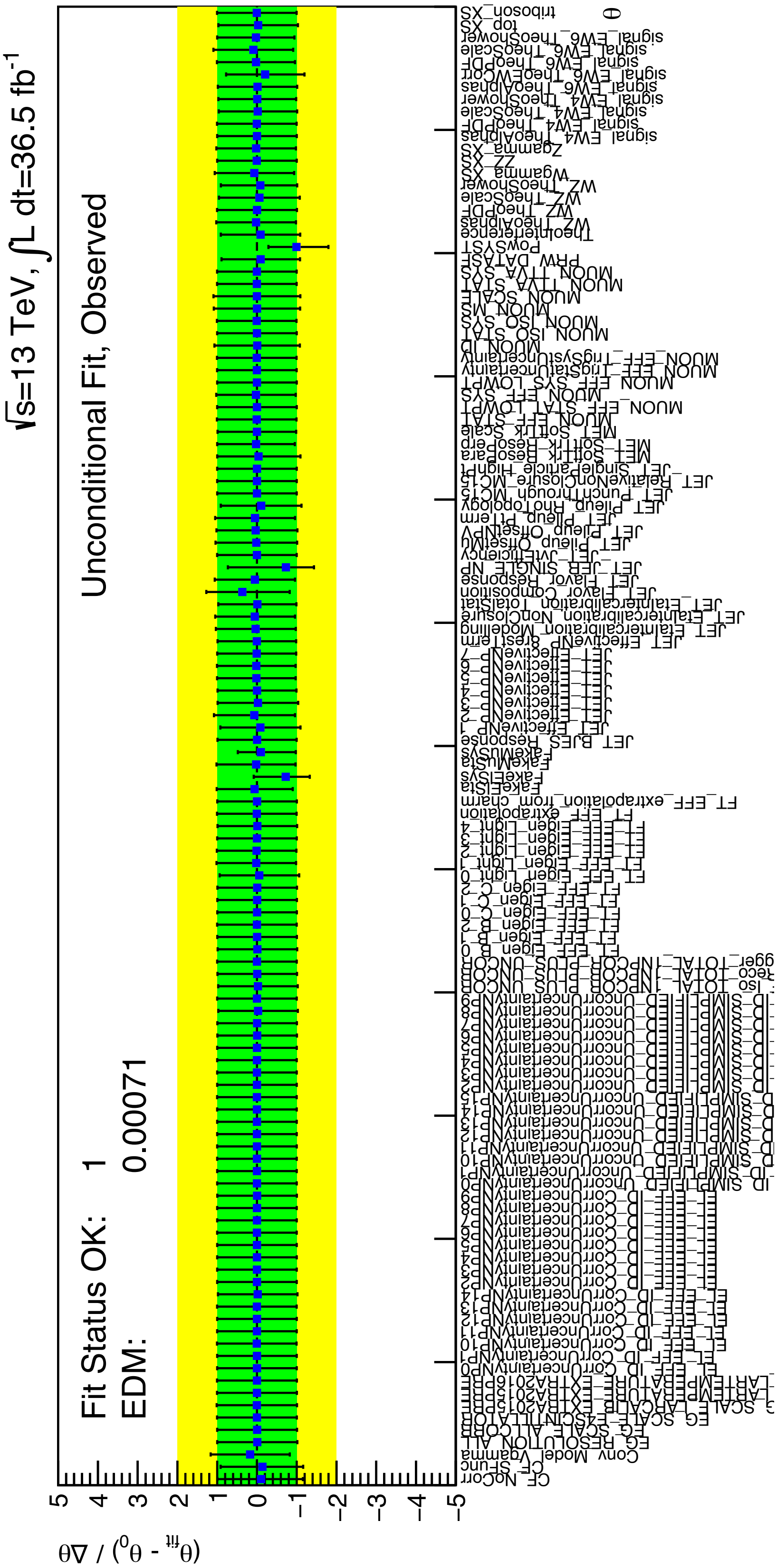


Figure C.12: Nuisance parameters pulls for the combined unconditional fit to observed data in the signal region and WZ and low-m_{jj} control regions with the addition of a nuisance parameter on the signal to account for the shape difference between SHERPA and POWHEG, named PowSYST. The error on each nuisance parameter obtained after the fit is reported in unit of the pre-fit uncertainty.

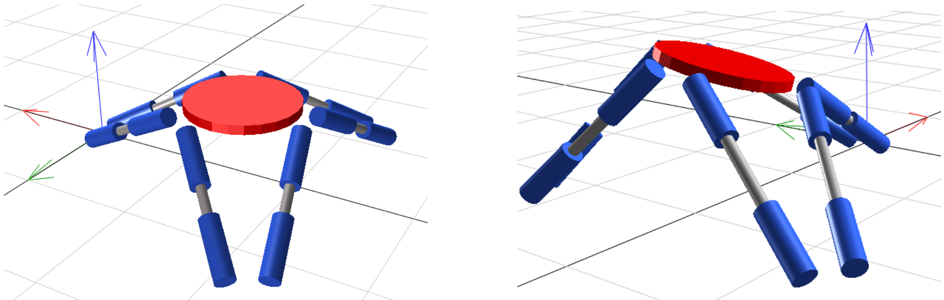
D Modelling of Robots Kinematic at Maplesoft

Maplesoft™ [161] is a software company born and based in Waterloo, Ontario, Canada. It provides high-performance software tools for engineering, science, and mathematics. It is best known as the manufacturer of the MAPLE computer algebra system and MAPLESIM physical modelling and simulation software. MAPLE has over 5000 functions covering virtually every area of mathematics, including calculus, algebra, differential equations, statistics, linear algebra, geometry with high-performance numeric computations. It includes code generation features and a full featured programming language to create scripts, programs, and full applications. MAPLESIM provides insight into systems behaviour and produces high-fidelity simulations. It includes over 700 built-in components from many different domains. The components are equation-based and visualisation capabilities include 2-D simulations and 3-D animations.
This appendix is dedicated to the description of the project conducted at Maplesoft during the Internship program under the *HiggsTools* grant[4]. The main aim of the project is to implement a generalised workflow to get the direct kinematics for a user-defined robotic model. Some basic definitions related to robotics nomenclature are outlined in Sec. D, the preliminary step of the project is presented in Sec. D while the main result is summarised in Sec. D. This appendix aims at giving an overview of the project, therefore a detailed treatment of the bases of robotics is not given. A comprehensive description and precise definitions are provided in Ref. [163].

Basic Definitions and Technicalities

A robot can be idealized as a rigid body where rigid parts, *links*, are connected by *joints* which allow for the motion and can induce different kind of movements (rotary, linear, etc.) depending on their type. The final component, *end effector*, is designed to make a specific task, and is the one to be controlled. A complete specification of the location of every point on the robot is its *configuration*, and since the links are rigid, the joints variables are enough to define its configuration. A robot is said to have *n* degrees of freedom (dof) when the configuration can be minimally specified by *n* parameters.
The *kinematics problem* for a robot consists in describing the motion of the end effect without considering forces and torques causing the motion, so it is reduced to a geometric problem. This description is usually divided into *forward* and *inverse kinematics*. The former deals with the issue of determining the position and orientation of the end effector given the values of the joint variables of the robot. The latter is computed to determine the values of the joint variables given the end effector's position and orientation.
The motion of a robot's end-effector is regulated by equations which are coordinate transformations between the ground reference frame and the frame of the end-effector. These can be rotations, translations or their combinations. In particular in Maplesoft's softwares, equations are obtained in MAPLE, and a worksheet is linked to the MAPLESIM program in order to simulate the model. The MAPLESIM program is fed with equations and the components for the model are generated out of it.

[4] Grant agreement PITN-GA-2012-316704 ("HiggsTools") from the Research Executive Agency (REA) of the European Union, in the contecxt of the Initial Training Network program supported by the 7th Framework Programme [162].



(a) (b)

Figure D.13: Stewart Platform for the flat 3-dof case (a) and non-level 6-dof case (b) shown in the MAPLESIM environment and generated with MAPLESIM via equations of motion derived with the MAPLE software.

This allows to have a programmatic access to the model, since these components are objects able to manipulate the model itself, as they generate the equations to obtain generalised coordinates and speeds, positions and constraints.

Preliminary Example: the Stewart Platform

In order to get familiar with the robotics environment and work methodology, the robotic model of the Stewart Platform [164] was chosen. This is a 6-dof model composed by a plate that can be moved by six joints acting with linear and rotational movements.

The Stewart Platform model was implemented originally for the 3-dof case. This means that the complete model allows for three linear movements, defined by (x, y, z), and three rotatory ones, defined by (η, ξ, ζ), which are instead set to 0 for the 3-dof case. This is shown in Fig. D.13a. The model is extended to the 6-dof case, i.e from a flat-platform model to a non-level one. The generalised coordinates defined for this model are

$$s, \alpha, \beta(\times 6) + (x, y, z, \eta, \xi, \zeta)$$

that can be reduced by 18 algebraic constraints.

The forward and inverse kinematic is implemented for the complete model as

$$(s, \alpha, \beta) \to (x, y, z, \eta, \xi, \zeta)$$

$$(x, y, z, \eta, \xi, \zeta) \to (s, \alpha, \beta)$$

The equations are solved and the components for forward and inverse kinematic generated in MAPLESIM, were they can be tested in the virtual model (Fig. D.13b).

The 3-dof case presents simplifications due to the symmetry of its geometry and motion. These simplifications are not possible in a non-level platform, which is hence not the best candidate to use as a starting point for the implementation of a general workflow. A simpler model is chosen for this purpose, the Kuka robot.

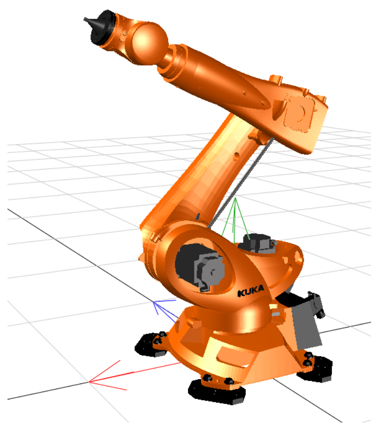


Figure D.14: Kuka Robot model as shown in the MAPLESIM environment.

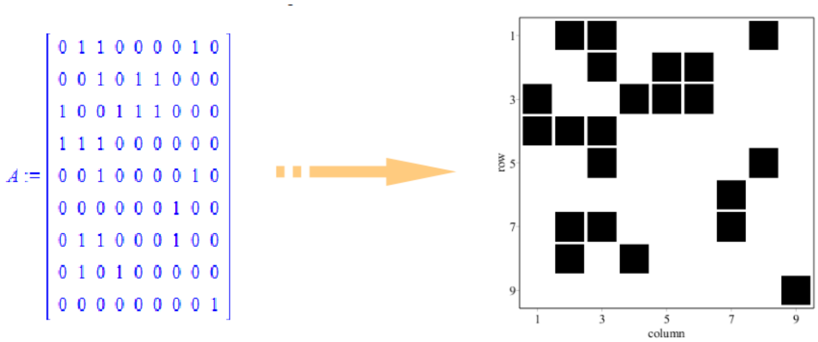


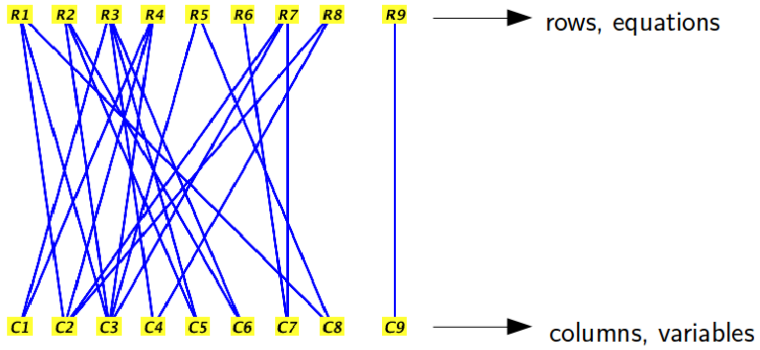
Figure D.15: A system with some sparsity and its incidence matrix.

Solving Big System of Equations

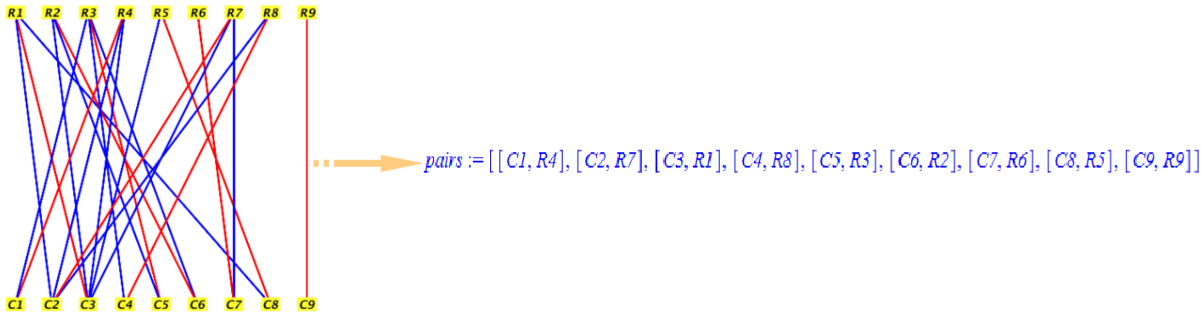
The model chosen to study the actual generalisation for the forward kinematics is the Kuka Robot [165], shown in Fig. D.14, which with its three links presents a simpler geometric structure that allows for an easier description.

The complexity that derives from the generalisation of even a simple model, requires the implementation of specific mathematical tools. In particular the most general system of equations that is needed to completely describe its motion can become very large and overdetermined when trying to explicitly express all the dependent variables in terms of the independent one. This moves the issue from the strictly robotic-related to a most general one, i.e. implementing in MAPLE a generalised workflow to solve big systems of equations.

The basic idea is to find the correct permutation of equations, so that the order used allows to always get to the solutions when existent, using the fewest number of steps. Given a system with some sparsity, a flow is implemented aiming at finding the correct order of equations and variables, which transforms the sparse system in a triangularised one. To achieve this Graph Theory has been used, so that given a system and its incidence matrix (Fig. D.15), finding the right permutations so that it has non-zero diagonal means finding the perfect matching of a bipartite graph (Fig. D.16). Applying the permutations given by the rows-columns matching allows to transform the matrix into a block-diagonal one and the blocks to be solved are obtained (Fig. D.17).



(a)



(b)

Figure D.16: Bipartite graph corresponding to the system in Fig. D.15 (a) and the matching obtained as a result to get the row-column pair combination to be used to solve the system (b).

The procedure has been incorporated in the workflow to get the forward kinematics component for a user defined model. This is shown in Figure D.18, where the flow given by the introduced functions is described, and the user's inputs highlighted, which indicates the level of generalisation and customisation achieved in the project.

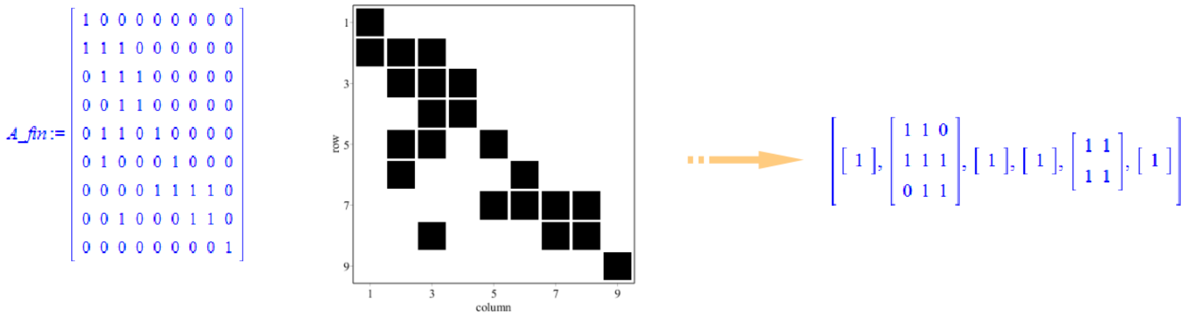


Figure D.17: The final system and matrix in the block-triangular form and the output blocks.

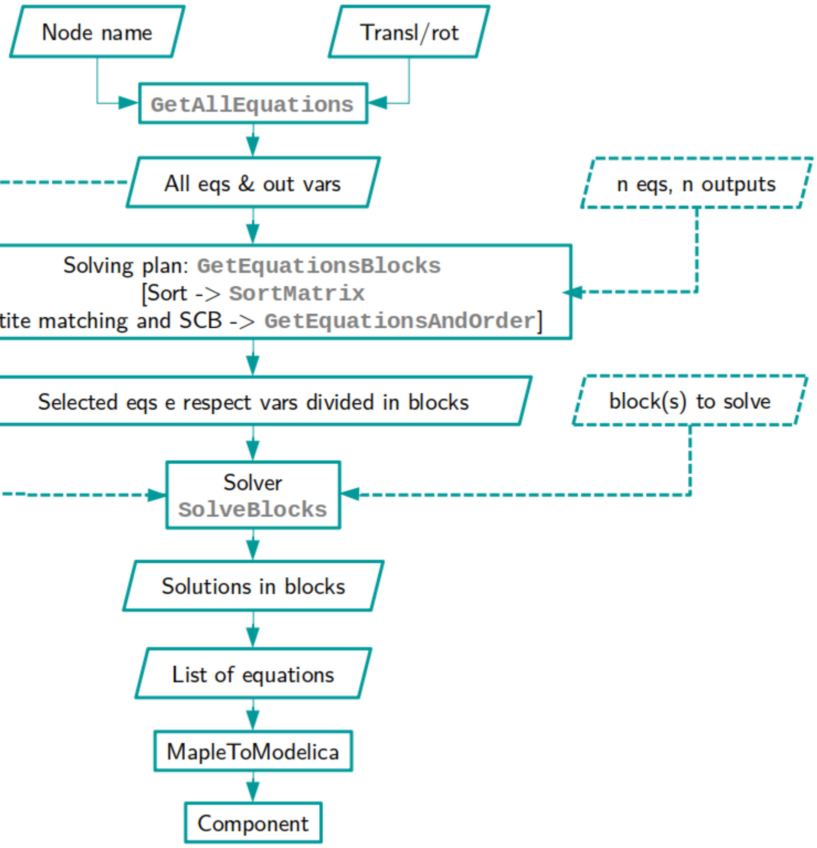


Figure D.18: Global flow created with and implemented in MAPLE for getting a component for forward kinematics and solving big systems of equations.

Acknowledgements

First of all, I would like to thank Prof. Karl Jakobs, whose decision to welcome me in his group marked the beginning of the most formative and intense experience I had so far. Thanks to Dr. Christian Weiser for patiently leading me during the last and most challenging years of my PhD and to Dr. Karsten Köneke to whom is due a big part of my growth.
Thanks to Prof. Nigel Glover, Prof. Frank Krauss and Prof. Giampiero Passarino, because I can not imagine a better path for my physics and personal life, than that of being part of the HiggsTools team.
A big thanks goes to Dr. Christopher Schwan, who carefully guided me through the world of Theory, for the long discussions, the work done together and the help with the theoretical part of this thesis.
Thanks to the people that were with me in the everyday battlefield. Benedict, who has always an answer and whose incredible support saved me in the most demanding period of my PhD. Philip, who showed me what doing Physics really means, and from whom I learnt how to improve from my mistakes. And Valerio, who doesn't know the difference between weekdays and weekends, when it comes to helping me. Thank you for feeding me not only with huge amounts of food during my visits to CERN, but with pills of your infinite knowledge and for being there, all the time, no matter if in Freiburg or elsewhere.
Thanks to my friends in the group. Thanks Hannah for being so supporting, I wish I was strong as you are, and thanks to Théo, Frank, Arturo, Leena, Ralf, Cedric, Sven, Philipp, Julian, Moritz, Krzysztof, Stephen, Liv, Frederik, Marc and Ines. Thank you Christina, for solving my problems before I see them.
Thanks to the people in the analysis team, Steffi, Franzi, Chilufya, Tim for the support during my writing, and Rustem and Karolos for the fruitful chats on Physics and not only.
Thank you to the people in Maplesoft that allowed me to live three incredible months in Canada, Jürgen, Kevin and Behzad, and thanks to Hadeer for all the food we shared, and for being a friend.
Thank you Savonas. If it's going to be so hard to leave Freiburg, it is your fault. I am talking to Vlad and your attention to my feelings, especially when they're hidden. To Giac(h)omo and Costanza and the happiness you give me, to Stefano and Valentina, the columns that make me feel like I've always lived here. I am talking to Edo and your being so similar to me, and to Lisa, because I feel like I've always known you. To Danny, both to your German and Italian soul, to Thor for your God's strength. To Veronika, who shared the boat with me in the stormy days, enjoying together the sunny ones, and to Michi and the beers drank together. To Sun Won, to your lovely attitude and your big, big heart, and to Maria, to your passion and your cheerfulness which I should learn. To Simona, because I shared with you the flat, the office, every pain, joy, worry, thought in such a deep way, that I can hardly tell who is who between us anymore. And I am talking especially to you, Riccardo. You are a special human being, and I learnt something from every word you said or action you made.
And thank you Savonas who already left Freiburg, watch out, I can reach you sooner or later. Thank you Luisa, for holding my hand when I arrived, and keep doing it whenever I need. Thanks to Sara and your incredible power and to Michele, that learnt immediately how to deal with my stress. To Manfre and all the jokes I can do *soloconmanfre*, I will keep doing them. Thanks Fabio for the coffee breaks together, to Simone and everything we have created in a little bit more than a year, that

will last. Peter, you are the *Friend* that everyone dreams about, who knows my fears and kills them before they hurt. Scody says thank you.
Thanks to Chris and Mel, dreaming about our next barbecue saved me in this period.
Thanks to all the incredible HiggsTools people, we were a great team from the very first day in Durham, and I knew that was just the the beginning.
And thanks Raquel and Pietro, my Teppa and ninja Turtles.
Sono passati più di dieci anni da quelle due settimane in Irlanda, e mi ritrovo a ringraziarvi Lu, Ali, Sele, Polly, Pollons e Disprez, perché nulla è cambiato, e siamo sparsi per il mondo, più vicini che mai. E grazie a quella fantastica schiera di persone dal cuore grande con cui ho condiviso i dolori della vita al CERN e le gioie delle serate ginevrine, Giuliano, Arianna, Daniele, Camilla, Giulia, Francesco, Elena, Lucrezia, Francesco, Betto, Fabio, Adele, Marco, ad esempio. Ma siete tantissimi, e sapete davvero cos'è l'amicizia. Grazie Mastro, per il pollo e il Vesuvio, e per tutto cosa verrà con te.
Grazie Bibi, per aspettarmi a Ginevra a braccia aperte e farmi passare la paura.
Grazie Will, Mett e Marco B., perché è bellissimo capirci quando pochi altri lo farebbero.
Carlotta, grazie. Per essere con me ogni giorno da sempre, e per quegli abbracci fortissimi quando ci vediamo.
Grazie a Frensis, Eli e GB, per i vostri innumerevoli viaggi a Friburgo. Quelli accorciano distanze che sono dolorose, ma che non cambieranno mai quello che siamo noi quattro. E grazie Biondo, che nn ti fermi davanti a nulla se è per il nostro bene.
Grazie alle Sgonelle, per essere così in simbiosi da avere un nome unico, e perché tutte, Valeria, Luisa, Silvia, Giulia, Francesca, Aleksandra, Marta, Ilaria, vi siete prese un pezzo della mia stanchezza, che divisa in nove è più leggera. Grazie a voi, ai vostri Sgonelli, e ai tantissimi viaggi che mi avete fatto fare in questi mesi di scrittura, per poterli chiamare mariti. Questi viaggi sono il ricordo più bello che mi resterà. Pazzi e Pazze, sogno una gita in tenda tutti insieme, per potervi ringraziare uno ad uno.
Grazie mamma e papà perché mi avete visto partire in pullman con un trolley, e deve fare effetto, ma eravate felici per me e non l'avete mai dato a vedere. Il grazie più grande va a voi, e alla *Buonanotte* che arriva puntuale ogni sera da quasi cinque anni come se fossi nella mia stanza. Grazie Paola, per il lindy hop a Friburgo, per le nostre diversità che si completano a vicenda, e per essere la certezza che anche se non le vedi sempre *sisters are always there*. Da zia Anna, zio Dodo, zia Mary, zio Max, zia Carla, zio Meo, zia Giuly, zio Maury ai cugini coraggiosi e favolosi che ho, tutti avete avuto una parte importante nel farmi andare avanti. E grazie a Emma, Franco e Valeria, i nonni che passano sopra alla tristezza del vedermi partire, perché sono orgogliosi di cosa provo a conquistare. Grazie anche a nonno Beppe, perché so che per lui sarebbe la stessa cosa. Questa tesi è dedicata a voi quattro.
Grazie Agata e Pino, per sopportare la distanza e non smettere di incoraggiarmi, anche con bonus di pizza e pasta al forno.
E invece grazie non basta per te, Stefano. Che hai scritto con me ogni parola di questa tesi. Che hai insistito per farmi allenare e arrivare a correre per un'ora di fila, quando io volevo mollare, per dimostrarmi che *non esistono condizioni sfavorevoli, ma solo atleti arrendevoli*.
Devi sapere, è solo grazie a te che ho imparato a non arrendermi.