Faculteit Wetenschappen
Departement Fysica

STUDY OF ELECTROWEAK $W^{\pm}W^{\pm}jj$ PRODUCTION WITH THE CMS DETECTOR

STUDIE VAN ELECTROZWAKKE $W^{\pm}W^{\pm}jj$ PRODUCTIE MET DE CMS DETECTOR

Proefschrift voorgelegd tot het behalen van de graad van doctor in de wetenschappen: fysica aan de Universiteit Antwerpen te verdedigen door

JASPER LAUWERS

PROMOTOR:

Prof. dr. Nick van Remortel

SUPERVISOR:

dr. Xavier Janssen

Antwerpen, 2018

DOCTORAATSJURY:
Prof. dr. Nick van Remortel (Promotor, Universiteit Antwerpen)
Prof. dr. Michiel Wouters (Voorzitter, Universiteit Antwerpen)
Prof. dr. Jo Verbeeck (Universiteit Antwerpen)
Prof. dr. Fabio Maltoni (Université catholique de Louvain)
dr. Pietro Govoni (Universita' degli Studi Milano Bicocca)

Nature uses only the longest
threads to weave her patterns,
so that each small piece of her fabric
reveals the organization of the entire tapestry.

— Richard P. Feynman

PREFACE

This thesis gives an overview of the research I performed in the Compact Muon Solenoid (CMS) collaboration between 2013 and 2017. It is centered around the experimental study of vector boson scattering (VBS), a process realized through electroweak triple and quartic boson interactions. The process is characterized by a specific event topology in proton-proton collisions at the Large Hadron Collider (LHC), enabling its measurement in the data collected with the CMS detector.
In the first part of this thesis, an overview is presented of the theory and experimental setup. The second part gives a detailed description of the performed analyses and the obtained results.

A chronological overview of the research I performed is given here to clarify my contribution to each of the studies described in this thesis. The main topics I worked on are:

- The measurement of the vector boson fusion (VBF) produced Higgs boson decaying to bottom quarks in Run I data.
- A feasibility study of vector boson scattering at the High Luminosity (HL) LHC.
- The measurement of the electroweak $W^\pm W^\pm jj$ production in Run II data.

During the first year of my research in experimental particle physics, I contributed to the study of VBF $H \to b\bar{b}$. The measurement of this process in Run I data lead to a publication in Physical Review D:

- Vardan Khachatryan et al. "Search for the standard model Higgs boson produced through vector boson fusion and decaying to bb." In: Phys. Rev. D92.3 [1]

The study is not described in this thesis, but a brief summary of the statistical procedure I developed for the analysis is given in Appendix A.

The second topic, a feasibility study of VBS at the HL-LHC, is described in the technical proposal of the CMS Phase II detector (2015):

- D. Contardo et al. "Technical Proposal for the Phase-II Upgrade of the CMS Detector." Tech. rep. CERN-LHCC-2015-010 [2]

And was later extended in a public analysis summary (2016):

- CMS Collaboration. "Prospects for the study of vector boson scattering in same sign WW and WZ interactions at the HL-LHC with the upgraded CMS detector." Tech. rep. CMS-PAS-SMP-14-008 [3]

The study is described in detail in Chapter 6 of this thesis. My most important contributions to this collaborative effort are the estimation of the largest background in the analysis containing jets misreconstructed as leptons, and the anomalous coupling study in an effective field theory framework.

The main result in this thesis is the measurement of the electroweak $W^{\pm}W^{\pm}jj$ production in 2016 (Run II) data, described in Chapter 5. Two versions of the analysis exist which provide an independent cross-check of the results, with independent meaning that they use a completely separate analysis framework but with an agreement on the general analysis strategy and based on the same simulation samples.
I performed one of these analyses, from the initial stages of producing the simulation samples and setting up a dedicated analysis framework to the final results described in this thesis. The measurement lead to the observation of the electroweak $W^{\pm}W^{\pm}jj$ process decaying to leptons and has been submitted for publication to Physical Review Letters:

- Albert M Sirunyan et al. "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 $\sqrt{s} = 13$ TeV." Submitted to Phys. Rev. Lett. [4]

A measurement of the opposite sign electroweak $W^{\pm}W^{\mp}jj$ production in Run I data was also performed, as well as preparations for its measurement in Run II data. The results of these studies are not described in this thesis.

ABSTRACT

The standard model of particle physics (SM) provides a theoretic framework for fundamental particles and their interactions. The recent discovery of the Higgs boson completes the model, in that all the elementary particles in the standard model have been observed and the model accurately describes their behavior as measured in experiments. Despite its successes, many observations can not be explained by the standard model in its current form. Searches for new particles to explain these phenomena, as performed at the Large Hadron Collider (LHC) at unexplored center-of-mass energies, are crucial to push the limits of our knowledge. This is complemented with precision measurements of the standard model and the Higgs boson in particular.

By studying the scattering of weak vector bosons (VBS), an important contribution is made to both fronts. The scattering of the weak vector bosons is strongly connected to electroweak symmetry breaking (EWSB) in the SM, regulated by the Higgs mechanism. It is through the EWSB that the weak vector bosons acquire mass, turning the massless Goldstone bosons in their longitudinal polarization. Measuring VBS provides information on the degrees of freedom of the Higgs field that are absorbed by the weak bosons. The VBS process also contains information on the quartic vector boson interactions, a part of the SM which has not been well tested yet and could be modified by new physics.

The vector boson scattering process can be studied at the LHC from the electroweak production of vector bosons in association with jets. It can be measured with the highest sensitivity in the case of two W bosons with the same charge decaying to leptons. The measurement of this process in data recorded with the Compact Muon Solenoid (CMS) detector during 2016 is described in this thesis, which lead to the first discovery of a process dominated by vector boson scattering. The electroweak production of two same-charge W bosons at the LHC operating at a center-of-mass energy of 13 TeV has a cross section of only 0.35 pb, which is two orders of magnitude smaller than the total production cross section of the Higgs boson discovered in 2012. The small cross section in combination with the backgrounds from strong diboson production and misreconstructions make this a challenging analysis.

To uncover the properties of VBS, the study of this process during the further operation of the LHC remains of great importance. A feasibility study for measuring VBS properties is described in the last chapter, with $3\,\mathrm{ab}^{-1}$ of data that will be collected with the CMS detector in the next decades at the upgraded High Luminosity LHC.

SAMENVATTING

Het standaard model van de deeltjesfysica (SM) biedt een theoretisch kader voor fundamentele deeltjes en hun interacties. De recente ontdekking van het Higgs boson vervolledigt het model, in de betekenis dat alle elementaire deeltjes in het standaard model zijn geobserveerd en het model hun gedrag accuraat beschrijft zoals gemeten in experimenten. Desondanks zijn successen, kunnen veel waarnemingen niet worden verklaard door het standaard model in zijn huidige vorm. Het zoeken naar nieuwe deeltjes om deze fenomenen te beschrijven, zoals uitgevoerd bij de Large Hadron Collider (LHC) met onverkende massamiddelpuntsenergieën, zijn van cruciaal belang om de grenzen van onze kennis uit te breiden. Dit wordt gecomplementeerd door precisiemetingen van het standaardmodel en het Higgsdeeltje in het bijzonder.

Door de verstrooiing van zwakke vectorbosonen (VBS) te bestuderen, wordt op beide fronten een belangrijke bijdrage geleverd. De verstrooiing van zwakke vectorbosonen is sterk verbonden met elektrozwakke symmetriebreking (EWSB) in het SM, uitgevoerd door het Higgs-mechanisme. Het is via de EWSB dat de zwakke vectorbosonen massa krijgen, door de massaloze Goldstone-bosonen in hun longitudinale polarisatie om te zetten. Het meten van VBS geeft informatie over de vrijheidsgraden van het Higgsveld die worden geabsorbeerd door de zwakke bosonen. Het VBS-proces bevat ook informatie over de vierpuntsinteracties tussen vectorbosonen, een onderdeel van het SM dat nog niet goed is getest en dat zou kunnen gemodificeerd worden door nieuwe fysica.

Het vectorbosonverstrooiingsproces kan worden bestudeerd bij de LHC door de elektrozwakke productie van vectorbosonen samen met jets. Het kan worden gemeten met de hoogste gevoeligheid in het geval van twee W-bosonen met dezelfde lading die vervallen naar leptonen. De meting van dit proces in data genomen met de Compact Muon Solenoid (CMS) detector in 2016 wordt beschreven in deze thesis, en leidde tot de eerste ontdekking van een proces gedomineerd door vectorbosonverstrooiing. De elektrozwakke productie van twee W-bosonen met dezelfde lading bij de LHC en een massamiddelpuntsenergie van 13 TeV heeft een werkzame doorsnede van slechts 0.35 pb, wat twee grootteordes kleiner is dan de totale werkzame doorsnede voor de productie van het in 2012 ontdekte Higgs-boson. De kleine werkzame doorsnede in combinatie met de achtergrond afkomstig van sterke dibosonproductie en misreconstructies maken dit een uitdagende analyse.

Om de eigenschappen van VBS te achterhalen, blijft de studie van dit proces tijdens de verdere operatie van de LHC van groot belang. Een haalbaarheidsstudie voor het meten van VBS-eigenschappen is beschreven in het laatste hoofdstuk, met $3\,\mathrm{ab}^{-1}$ aan data die met de CMS detector wordt verzameld in de komende decennia bij de geüpgrade LHC met hogere luminositeit.

CONTENTS

Part I

INTRODUCTION

Part I of this thesis meanders through the field of experimental particle physics, passing by the foundations of the field needed to understand and appreciate the main work described in part II.

A general overview of the standard model of particle physics is given, with a more specific section to identify the role of vector boson scattering in the theory. This is followed by an overview of the experimental setup and of the methods used for translating the experimental findings into theoretic insights.

1 THE STANDARD MODEL OF PARTICLE PHYSICS

Discoveries in the last century have led to the insight that all visible matter in the universe is made up of only a few constituents, called fundamental particles. Our understanding of these particles and three of the fundamental forces that mediate interactions among them, is described by the Standard Model of particle physics (SM). Since its formulation in the 1960s, the standard model has been able to provide an accurate description of fundamental particle interactions. With the recent discovery of the Higgs boson [5, 6], all particles that are predicted by the SM have been observed and no significant deviations have been found, including in the many high precision tests performed at the Large Hadron Collider (LHC). Also at low energies, the model performs well; the anomalous magnetic moment of the electron has been verified up to 12 significant figures. This makes the standard model one of the most well-tested theories in physics.

Despite its success in its field of applicability, the standard model is far from a complete theory. One of the missing pieces, that has been receiving continuous efforts since its formulation, is that the model does not include gravitation. Other open questions include the excess of matter over anti-matter that is observed in the universe, the origin of dark matter and the incredible fine-tuning on the order of 10^{-28} that is required to give the Higgs boson a mass so much lighter than the Planck mass.

1.1 PARTICLES AND FORCES

The fundamental particles in the standard model are naturally divided in two groups: fermions, following Fermi-Dirac statistics and generally described as "matter particles", and bosons, following Bose-Einstein statistics, that mediate interactions.

An overview of the fundamental particles is given in Figure 1. The fermions split up in quarks and leptons, both appearing in 3 generations. It is only the first generation that makes up everyday matter consisting of electrons, and the up and down quarks that are the components of protons and neutrons. Matter consisting of the heavier, higher generation fermions is unstable and will quickly decay to a lower generation. The strong, weak and electromagnetic interactions are mediated by fundamental bosons. They are accompanied in the standard model by the recently discovered scalar (spin zero) Higgs boson. The graviton is the corresponding force-carrying particle of gravity, although it has not been observed and is not described by the SM. For the interactions studied at the LHC, gravitation is extremely weak compared to the other forces and hence is completely negligible.

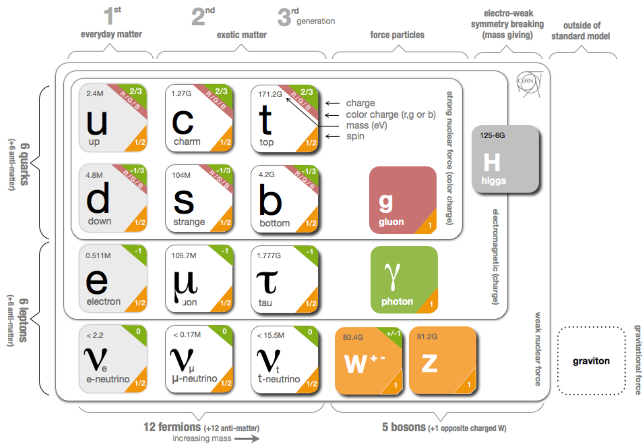


Figure 1: Fundamental particles of the standard model.

1.1.1 *Strong interaction*

The strong interaction is the force keeping quarks confined in nucleons (protons and neutrons). The particle that mediates the strong force is the gluon, adequately named as it "glues" together the quarks. Next to gluons, quarks are the only particles that experience the strong force. This is equivalently expressed by saying they have a non-zero *color charge*, the conserved quantity associated with this force. The color charge, unlike the electromagnetic charge, comes in 3 types (red, blue and green). The different structure of the strong interaction also expresses itself through other remarkable properties, one of which is color confinement. Due to the strength of the strong interaction – which does not decrease with distance – when quarks are separated, it will be energetically more favorable to form new quark pairs from the vacuum. This means no free quarks exist in nature, and only color-neutral states of multiple quarks (hadrons) are observed. Color-neutral states can be formed by combining three quarks with a different color charge (baryons) or color anti-color combinations (mesons).
Next to keeping nucleons together, the weaker residual strong force between color-neutral hadrons holds the nucleons together in atomic nuclei.

1.1.2 *Weak interaction*

The weak interaction operates at short distances due to its heavy carrier particles: the W and Z bosons with a mass around $80-90$ GeV. The interaction does not hold together bound states but it does interact with all particles in the standard model and plays an important role in radioactive decay and nuclear fusion. The most common example of the weak interaction is beta decay, a type of nuclear decay where an electron and neutrino are emitted from an atomic nucleus. At the subatomic level, a down quark converts to an up quark through the weak interaction by emitting a virtual W boson. The W boson then decays to an electron and a neutrino. This kind of interaction where a particle changes flavour (e. g. down to up quark) is only possible through the weak interaction.

1.1.3 *Electromagnetism*

Electromagnetism needs less introduction, as it is a force that is abundantly experienced in everyday life. The photon mediates the electromagnetic force and interacts with quarks as well as leptons that have an electric charge (electron, muon, tau). Although electromagnetism seems very different from the weak interaction at the low energies we are used to, they are unified in the electroweak theory [7–9].

1.2 STANDARD MODEL LAGRANGIAN

The Standard Model is formulated in Quantum Field Theory (QFT) [10], a theoretical framework that reconciles quantum mechanics and special relativity. The standard model particles appear in this framework as excited states of the underlying quantum fields defined at all points of space-time. As any theory satisfying special relativity, the SM is invariant under Poincaré symmetry, including translations, rotations and boosts. Besides the trivial Poincaré symmetry, the SM is a gauge theory, which means the Lagrangian is invariant under a continuous group of local transformations. The structure of the electroweak and strong interactions is determined by the non-abelian gauge theory with symmetry group:

$$SU(3)_C \times SU(2)_L \times U(1)_Y . \tag{1}$$

The subscripts indicate the conserved quantities that are associated with the symmetry groups: the color charge (C), weak isospin (L) and weak hypercharge (Y). The equivalence between symmetries and conservation laws is expressed by Noether's theorem [11]. $SU(3)_C$ is the symmetry group of the strong interaction and $SU(2)_L \times U(1)_Y$ determines the structure of the electroweak theory which is spontaneously broken by the Higgs mechanism [12–15].

The SM Lagrangian can be summarized as:

$$\mathcal{L}_{SM} = L_{EW} + L_{QCD} + L_{Higgs} + L_{Yukawa} , \tag{2}$$

with terms describing the electroweak (EW) theory and quantum chromodynamics (QCD) that follow from requiring invariance under their respective symmetry groups (Equation 1), as well as terms for the Higgs mechanism and Yukawa couplings.

In the following sections, it is shown how symmetry requirements lead to the standard model Lagrangian, why the Higgs mechanism is needed to add mass, and how the resulting Lagrangian gives rise to the frequently used Feynman diagrams. The focus lies on the ideas behind the theory, without deriving the full Lagrangian with the exact properties and particle content of the standard model [11, 16, 17].

1.3 QUANTUM ELECTRODYNAMICS

In this section the Lagrangian of quantum electrodynamics (QED) will be derived from symmetry principles.
Let us start from the Dirac Lagrangian, describing a spin-1/2 massive particle:

$$\mathcal{L} = i\overline{\psi}\not{\partial}\psi - m\overline{\psi}\psi\,, \tag{3}$$

with $\psi(x)$ the wave function of the electron with mass m and spacetime coordinates x, $\overline{\psi} \equiv \psi^\dagger\gamma^0$ with γ_μ the Dirac matrices and using the Feynman slash notation $\not{A} = \gamma_\mu A^\mu$ and natural units ($c = \hbar = 1$). By applying the Euler-Lagrange equations on this Lagrangian, one retrieves the Dirac equation. It is clear that the Lagrangian is invariant under global phase transformations $\psi \rightarrow e^{i\alpha}\psi$, which is equivalent to stating that the phase α is unmeasurable. Just as in quantum mechanics, only phase differences are physically observable. The phase transformation is global because α is a constant, it is specified for all space and time. We can generalize this symmetry by allowing α to differ from spacetime point to point: $\alpha(x)$ and require our theory to be invariant under the transformation:

$$\psi \rightarrow e^{i\alpha(x)}\psi\,. \tag{4}$$

Applying the local phase transformation to the Lagrangian shows that the mass term is invariant under local phase transformations, but the first term is not:

$$\partial_\mu\psi \rightarrow e^{i\alpha(x)}\partial_\mu\psi + ie^{i\alpha(x)}\psi\;\partial_\mu\alpha(x) \tag{5}$$

an additional term with $\partial_\mu\alpha(x)$ remains after the transformation. From a physical point of view, one can understand that a local phase transformation introduces phase differences which can be observed and will need to be compensated somehow. To see how to solve this, we can take a look at the definition of the derivative in the direction of a vector n^μ:

$$n^\mu\partial_\mu\psi = \lim_{\epsilon\rightarrow 0}\frac{1}{\epsilon}\left[\psi(x+\epsilon n) - \psi(x)\right]\,. \tag{6}$$

The two fields that are subtracted transform differently under the symmetry (Equation 4), which means that $\partial_\mu\psi$ loses its geometric meaning.

To define a derivative in a sensible way we need to introduce a field that compensates for the difference in phase transformation. This is called the covariant derivative:

$$D_\mu \equiv \partial_\mu - ieA_\mu \,, \tag{7}$$

which introduces a new field $A_\mu(x)$ that transforms as:

$$A_\mu \to A_\mu + \frac{1}{e}\partial_\mu \alpha \,. \tag{8}$$

It is easy to check that the covariant derivative transforms covariantly, like the field ψ itself:

$$D_\mu \psi \to e^{i\alpha(x)} D_\mu \psi \,. \tag{9}$$

For the Lagrangian we get:

$$\mathcal{L} = i\overline{\psi}\not{D}\psi - m\overline{\psi}\psi = i\overline{\psi}\gamma_\mu \partial^\mu \psi - m\overline{\psi}\psi + e\overline{\psi}\gamma^\mu \psi A_\mu \,. \tag{10}$$

The vector field $A_\mu(x)$, called the gauge field, couples to the Dirac particle in exactly the same way as the photon field. If $A_\mu(x)$ is the physical photon field, we must also add a kinetic energy term containing derivatives of $A_\mu(x)$. Because this term has to be invariant under $A_\mu \to A_\mu + \frac{1}{e}\partial_\mu \alpha$ it can only involve the gauge invariant field strength tensor:

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_n u A_\mu \,. \tag{11}$$

If we discard terms that are not invariant under time reversal and parity symmetries, there is only one kinetic term possible:

$$F_{\mu\nu}F^{\mu\nu} \,. \tag{12}$$

By requiring the Dirac Lagrangian to be locally gauge invariant, or invariant under the U(1) symmetry group, we are naturally lead to the Lagrangian of QED:

$$\mathcal{L}_{QED} = \overline{\psi}\left(i\not{D} - m\right)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \,. \tag{13}$$

1.4 ELECTROWEAK THEORY AND QCD

The same procedure can be followed using the other groups that make up the standard model. Requiring SU(2) invariance leads to the introduction of 3 additional fields. Linear combinations of these fields together with $A_\mu(x)$ gives the description of the photon, $W^\pm$ and Z bosons in the electroweak Lagrangian. Because the weak interaction is not symmetric under parity transformations ($\bar{x} \to -\bar{x}$) [18], ψ is split in left- and right-handed chiral components. The Lagrangian of the electroweak theory is given by:

$$\mathcal{L}_{EW} = i \sum_{f_L, f_R} \overline{\psi}_f \not{D} \psi_f - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}W^a_{\mu\nu}W^{a\ \mu\nu} \,, \tag{14}$$

where a sum is made over the left-handed doublets ψ_f^L and right-handed singlets ψ_f^R containing leptons with flavour f. The covariant derivative D_μ contains the bosonic fields B_μ and W_μ associated with the U(1) and SU(2) symmetry groups respectively:

$$D_\mu = \partial_\mu - \frac{ig}{2} W_\mu^a \sigma_a - ig' \frac{Y}{2} B_\mu, \tag{15}$$

with σ_a the generators of SU(2) and Y the weak hypercharge. The photon and weak bosons are a linear combination of the B and W fields. The kinetic terms of these fields contain both 3- and 4-point interactions of the electroweak bosons, due to the non-abelian nature of the SU(2) gauge theory. It are these terms that give rise to the vector boson scattering interactions described in the next chapter.
Requiring invariance under SU(3) leads to the introduction of 8 more physical fields $G_{\mu\nu}^a$ that describe the gluons carrying a color and anti-color charge. The Lagrangian of quantum chromodynamics is given by:

$$\mathcal{L}_{QCD} = i \sum_f \overline{\psi}_f \not{D} \psi_f - \frac{1}{4} G_{\mu\nu}^a G^{a\ \mu\nu}, \tag{16}$$

with $\overline{\psi}_f$ the quark fields in triplets of the color charge, λ_i the generators of SU(3) and the covariant derivative:

$$D_\mu = \partial_\mu - i g_s \frac{\lambda_a}{2} G_\mu^a. \tag{17}$$

1.5 HIGGS MECHANISM

An important aspect that is still missing in the electroweak model, is that the W and Z bosons are massive. It turns out giving mass to the weak bosons is not trivial. The most easy way, would be to add a mass term to the Lagrangian of the form:

$$\frac{m^2}{2} A_\mu A^\mu. \tag{18}$$

The problem with adding this term to the Lagrangian is that it leads to unrenormalizable divergences; the theory can not be formulated to all orders in perturbation theory by adjusting a finite number of couplings, turning the theory meaningless. This is connected to the term not being gauge invariant, it breaks the symmetry and also the renormalizability of the theory. Gauge theories, on the other hand, are renormalizable as shown by 't Hooft, Veltman, and others [19].
To give mass to the weak bosons, we would need to add a mass term, without breaking gauge invariance. This is where spontaneous symmetry breaking comes in. By introducing a potential that is invariant under the gauge symmetry of the Lagrangian, but the symmetry is not present in the ground state,

one can introduce mass for the weak bosons. This is known as the Higgs mechanism and the scalar field ϕ that is introduced is called the Higgs field. The extra term that needs to be added to the SM Lagrangian is of the form:

$$\mathcal{L} = (\partial_\mu \phi)^* (\partial^\mu \phi) + m^2 \phi^* \phi - \lambda (\phi^* \phi)^2 , \tag{19}$$

with $\phi \equiv (\phi_1 + i\phi_2)/\sqrt{2}$ a complex scalar field, $m^2 > 0$ and $\lambda > 0$. The shape of the Higgs potential is shown in Figure 2.

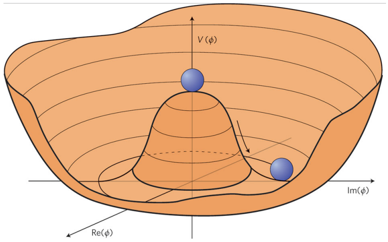


Figure 2: The Higgs potential as a function of $Re(\phi) = \phi_1$ and $Im(\phi) = \phi_2$. The ground state does not possess the global U(1) symmetry of the Lagrangian.

The ground states correspond to the minima of the potential $\phi_1^2 + \phi_2^2 = \frac{m^2}{\lambda}$ and the perturbation series should be expanded around one of these minima. This is done by choosing a ground state and substituting:

$$\phi(x) = \sqrt{\frac{1}{2}} \left[v + \eta(x) + i\xi(x) \right] , \tag{20}$$

with $v^2 = \frac{m^2}{\lambda}$ the vacuum expectation value. This yields:

$$\begin{aligned} \mathcal{L}' =& \frac{1}{2} (\partial_\mu \eta)^2 + \frac{1}{2} (\partial_\mu \xi)^2 - m^2 \eta^2 \\ & + \text{const. and higher order interaction terms in } \eta, \phi . \end{aligned} \tag{21}$$

The η field has a mass, while the ξ field only has a kinetic energy term but no mass. It can be shown that for each broken continuous symmetry, there appears a massless excitation called a Goldstone boson. This massless scalar particle emerges because it does not require any energy to change the ground state of the system to a different minima of the potential.
At first sight, this could be a problem because if there is a massless particle, we should detect it, massless particles are very easy to make. However, when applying this procedure to a *local* gauge theory, the Goldstone boson is turned into the longitudinal polarization of the massive gauge boson. An example of spontaneous symmetry breaking of a local gauge theory is given here for U(1) gauge symmetry. As seen before, this means we need to introduce the covariant derivate:

$$\mathcal{L} = (\partial_\mu + ieA_\mu) \phi^* (\partial^\mu - ieA^\mu) \phi + m^2 \phi^* \phi - \lambda (\phi^* \phi)^2 . \tag{22}$$

The Lagrangian can again be expanded around the vacuum. This time we can absorb one field in the gauge freedom:

$$\phi(x) = \sqrt{\frac{1}{2}}\,[v + \eta(x)]\ . \tag{23}$$

This yields:

$$\mathcal{L}' = \frac{1}{2}\,(\partial_\mu \eta)^2 - \lambda v^2 \eta^2 + \frac{1}{2} e^2 v^2 A_\mu^2 + \text{higher order interaction terms.} \tag{24}$$

The massless Goldstone boson has been used to generate the longitudinal polarization of the massive gauge boson. The Higgs mechanism makes it possible to give mass to the weak bosons in the SM without breaking the gauge invariance and in this way keeping the theory renormalizable.
By applying the same procedure on the electroweak lagrangian, one can give mass to the W and Z bosons. The Higgs Lagrangian is given by:

$$\mathcal{L}_{\text{Higgs}} = (D_\mu \phi)^\dagger\,(D^\mu \phi) + m^2 \phi^\dagger \phi - \lambda(\phi^\dagger \phi)^2\,, \tag{25}$$

with ϕ the Higgs doublet, that after fixing the gauge is given by:

$$\phi(x) = \sqrt{\frac{1}{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}. \tag{26}$$

Substituting the covariant derivative of QED and expanding around the vacuum lead to the mass terms for the W and Z boson, their interaction terms with the Higgs boson and the Higgs boson mass and self-interaction. Only two additional free parameters λ and m were introduced, relating the couplings of different terms.
Next to giving mass to the weak bosons, the Higgs boson also plays a role in providing mass to fermions through Yukawa couplings: terms containing the scalar Higgs and fermion fields of the form:

$$L_{\text{Yukawa}} = -\sum_{f,f'} \left(G_{f,f'}\ \overline{\psi}_f^L\ \phi\ \psi_{f'}^R + G'_{f,f'}\ \overline{\psi}_f^L\ \widetilde{\phi}\ \psi_{f'}^R + \text{hermitian conj.} \right), \tag{27}$$

with $G_{f,f'}^{(\prime)}$ free parameters that determine the strength of the Yukawa couplings, ψ_f^L and $\psi_{f'}^R$, the left-handed doublets and right-handed singlets of the fermion fields and $\widetilde{\phi} = i\sigma_2 \phi^*$.

Neutrino's are massless in the standard model as the exact properties of the neutrino and the mechanism through which it gains mass have not yet been experimentally verified. In reality, neutrinos have a very small mass observed indirectly by neutrino oscillations [20–22]. Although its mass is negligible for describing high energy interactions, the neutrino changes flavour as it propagates through space because the mass eigenstates are a mix of its flavour states.

1.6 FEYNMAN DIAGRAMS

To compare the predictions of the SM to nature, one needs to calculate observable quantities like interaction cross sections and decay rates. Because there is no exact solution for Lagrangians with interacting fields, these quantities are calculated perturbatively; the effect of the interaction is calculated as a correction to the non-interacting system. The free, non-interacting part of the Lagrangian can be solved analytically and gives the description of free particles. The interaction is then calculated up to a certain order in the perturbation series. This is possible when the disturbance is weak, or the coupling constant is $<< 1$. The resulting expression contains terms which can be visualized by Feynman diagrams. For example, the scattering of an electron and positron in QED can be represented by diagrams as in Figure 3.

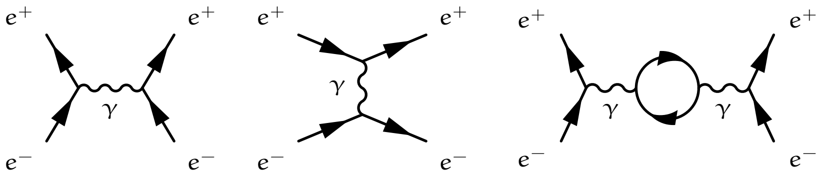


Figure 3: With time progressing from left to right, these diagrams represent e^+e^- scattering.

The left and middle diagrams in Figure 3 are the only possible diagrams for this process at leading-order (LO), their amplitudes are proportional to the coupling constant squared. The diagram on the right is one of the next-to-leading order (NLO) diagrams proportional to the coupling constant to the forth power. The lines and vertices that can be used in these diagrams are derived by applying *Feynman rules* to the Lagrangian. The amplitude of each diagram is then calculated by assigning a factor to each vertex and propagator. Four-momentum is conserved at each vertex and only the initial and final state particles need to be on-shell ($p^2 = m^2$). The cross section for electron positron scattering at leading-order (LO) is proportional to the squared sum of the LO amplitudes:

$$\sigma_{LO} \propto |\sum_i A_i|^2, \tag{28}$$

with A_i the amplitudes of the left and middle LO diagrams in Figure 3.
The probability amplitude for a certain process can be calculated from drawing all allowed Feynman diagram up to a certain order and summing their amplitudes. Measurable quantities are proportional to the squared amplitude which has contributions from all possible intermediate states and their

interference.
Physical quantities are more easily calculated using the Feynman rules, rather than the traditional approach of quantizing the Lagrangian and evaluating the perturbation series in the interaction Lagrangian. Next to being a useful mathematical tool, Feynman diagrams are viewed as the more concise description of fundamental particle scattering than the underlying canonical formalism.

2 VECTOR BOSON SCATTERING

The term *Vector Boson Scattering* is commonly used in literature, but requires some clarification. With *Vector Boson* one means electroweak vector bosons (W, Z and γ)[1]. Although here we will focus on the weak bosons (W and Z) specifically. The *Scattering*, as in classical physics, refers to a $2 \to 2$ process, but without requiring the particles to satisfy the Einstein energy-momentum relationship. In proton-proton collisions, the interacting vector bosons are virtual particles ($p^2 \neq m^2$).

The scattering of the weak vector bosons is strongly connected to electroweak symmetry breaking (EWSB) in the SM, operated by the Higgs mechanism. It is through the EWSB that the weak vector bosons acquire mass, turning the massless Goldstone bosons in their longitudinal polarization. In addition to studying the Higgs boson properties in the Higgs production and decay channels, measuring VBS is a complementary test of the nature of EWSB, providing information on the additional degrees of freedom of the Higgs field that are absorbed by the weak bosons. Next to probing the EWSB, the VBS process is also sensitive to quartic vector boson interactions, allowing to test an unexplored region of the standard model.

2.1 VBS IN pp COLLISIONS

In pp colliders – as the LHC – vector boson scattering occurs when two proton constituents, each from a different proton, radiate a vector boson and the two vector bosons interact. The W and Z bosons, having a half-life of about $3 \cdot 10^{-25}$ seconds, will subsequently decay to either a pair of quarks, or a pair of leptons. Depending on the decay channel of the two vector bosons, the VBS process is called *hadronic*, *leptonic*, or *semi-leptonic*. The two partons (gluons and quarks that make up hadrons) that initiated the interaction will be knocked out of the proton and can be measured as well. As already mentioned in Section 1.1, partons do not exist by themselves, but hadronize into color-neutral composite particles. The object we can observe is not the parton, but a *jet*: a spray of particles collimated along the direction of the initial quark or gluon.
In this chapter the scattering of W bosons with the same electric charge $W^{\pm}W^{\pm}$ – referred to as *same-sign* W bosons – will be used as an example

1 In principle also the gluon and non-fundamental mesons with non-zero spin are vector bosons.

to introduce the general properties of VBS. More specifically, the W^+W^+ scattering process in the leptonic decay channel:

$$q\,q \to W^+W^+\,q\,q \to \ell^+\nu\;\ell^+\nu\;q\,q\,. \tag{29}$$

The final state objects are the charged leptons, the neutrinos and the quarks. Neutrinos can not be measured and their presence is only indicated by an imbalance in the momentum of the detected particles. The identification of the process thus relies on the measurement of the charged leptons and the jets originating from the final state quarks.

The scattering of vector bosons happens at lowest order by a quartic interaction, or an exchange of an electoweak (W, Z, γ) or Higgs (H) boson. The allowed Feynman diagrams at leading order depend on the vector bosons under study. The Feynman diagrams for the leptonic W^+W^+ scattering are shown in Figure 4.

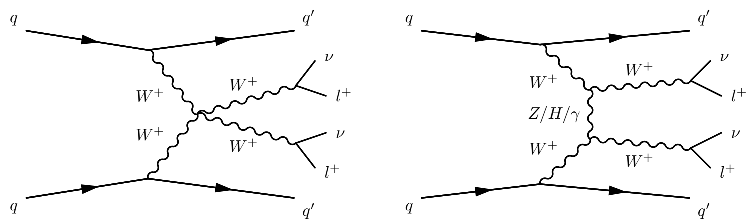


Figure 4: VBS Feynman diagrams at leading order representative for leptonic W^+W^+ scattering.

For other VBS processes also an s-channel[2] exchange is possible, when the vector bosons merge into an intermediate particle.

2.1.1 *Electroweak production*

The set of diagrams defined by Figure 4 is not gauge invariant. To define a physically meaningful set of diagrams, we need to add diagrams where one or both of the outgoing W^+ bosons decay directly without interacting.

2 The nomenclature follows from the Mandalstam variables for $2 \to 2$ scattering processes with p_1, p_2 the incoming and p_3, p_4 the outgoing particles:

$$s = (p_1 + p_2)^2 = (p_3 + p_4)^2$$
$$t = (p_1 - p_3)^2 = (p_4 - p_2)^2$$
$$u = (p_1 - p_4)^2 = (p_3 - p_2)^2$$

The exchange of an intermediate particle can happen in three ways, where the squared four-momentum of the particle equals s, t or u.

Examples of these diagrams are shown in Figure 5.

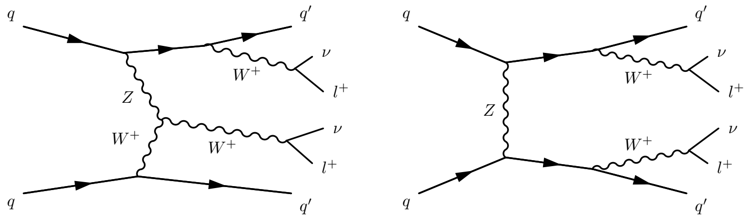
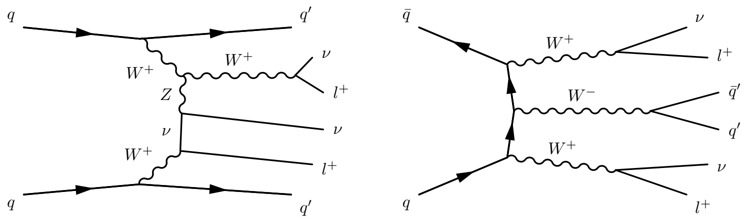


Figure 5: Non-VBS Feynman diagrams that contribute to the electroweak W^+W^+ production in the leptonic final state.

As already indicated in the previous Feynman diagrams, the weak bosons are unstable particles. The leptons they decay to have a longer lifetime, and are the on-shell particles that define the final state. There are other diagrams that lead to the exact same final state particles as defined by Equation 29. Two additional classes of electroweak diagrams contribute when considering the leptons and quarks as final states, shown in Figure 6.



(a) Non-resonant (b) Triboson

Figure 6: Additional diagrams that contribute to the electroweak $\ell^+\nu\ \ell^+\nu\ q\,q$ production.

One of the possible non-resonant diagrams is shown in Figure 6a and a diagram for triboson production in Figure 6b. It is the sum of the electroweak contributions in Figures 4, 5 and 6a that is gauge invariant. The triboson diagrams have the same initial and final state but can be seperated due to the different characteristics of the outgoing quark pair.
The non-resonant and triboson contributions are negligible in a typical VBS phase space because of the non-resonant suppression factor and the invariant di-quark mass around the W boson mass respectively [23]. The signal in VBS searches at pp colliders can therefore be defined as the combined set of contributions in Figures 4 and 5, called the electroweak production of same-sign W boson pairs in the two jet and two same-sign lepton final state. Note

that to define a physical, measurable process also diagrams where no vector boson scattering takes place have to be included.

2.1.2 *Strong Production*

There are also QCD mediated diagrams that produce the same final state, called the strong VVjj production. It is separated from EW production at LO by requiring two QCD vertices, i. e. the amplitude squared is of the order $\mathcal{O}(\alpha_{em}^4 \alpha_s^2)$ instead of $\mathcal{O}(\alpha_{em}^6)$ for the EW production. This means it are not vector bosons but the quarks that interact, via the strong interaction. For the same-sign W boson pair production in the leptonic decay channel (Equation 29) the additional diagram is shown in Figure 7.

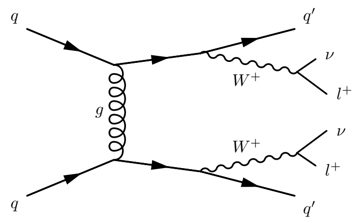


Figure 7: Feynman diagrams for strong W^+W^+ production in the leptonic final state.

When considering other combinations of two vector bosons, there are also diagrams with gluons in the initial state.
It is only the combination of two same-sign W bosons that can be produced exclusively by interactions with two quarks in both the initial and final state. The EW and strong contributions are compared for different vector boson scattering processes at $\sqrt{s} = 8$ and 13 TeV in Figure 8 after a basic VBS event selection. The figure shows that $W^\pm W^\pm$ scattering has the smallest relative strong contribution compared to other VBS processes. In the leptonic decay channel the presence of two same-sign charged leptons in the final state reduces the other SM backgrounds.
These features make the same-sign vector boson scattering a good candidate for measuring VBS. The second most sensitive measurable VBS process is the WZ boson scattering which has a larger strong production component, but has less background events from other processes because of the additional charged lepton in the final state.

2.1.3 *Interference*

The QCD induced production is a distinct gauge invariant process which can be separated from the electroweak production process. This is however only

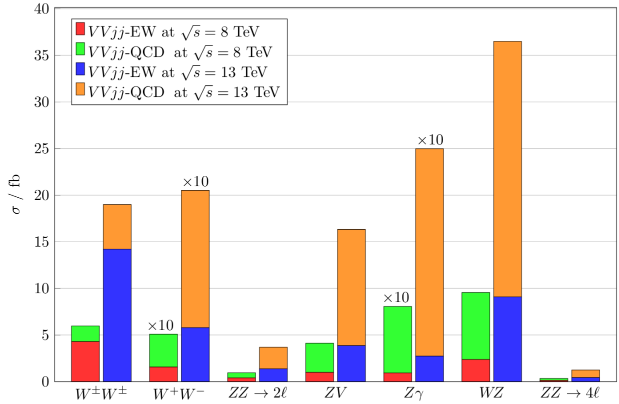


Figure 8: Production cross sections σ for the electroweak and the strong $VVjj$ production. The values for $W^{+}W^{-}$ and $Z\gamma$ final states are scaled down by a factor of ten to increase the visibility. The decay channels contain at least two charged leptons. Figure from [24].

allowed if the interference between the EW and strong production is small. As the transition probability between the pp initial state and the final state of VBS contains the squared sum over the amplitudes of all possible intermediate diagrams (Equation 28), interference between the diagrams has to be taken into account. In the phase space region of interest, the interference effects are usually small and the EW and strong $VVjj$ production can be defined as separate physical meaningful processes. The EW, strong and interference contributions to the same-sign W boson production cross section as a function of the invariant di-jet mass (m_{jj}) is shown in Figure 9.

2.1.4 *Next-to-Leading Order*

So far, only the LO calculation was considered. At next-to-leading order additional Feynman diagrams appear with extra strong or electroweak interactions. The QCD NLO corrections to both the EW and strong production process have been implemented in simulation software for a few years and causes an increase in the electroweak $W^{\pm}W^{\pm}jj$ production cross section of about 5% [26, 27]. Their contribution is not included in the analyses of Part II of this thesis, but the difference is covered by the theoretical scale uncertainties [25, 28].
In the last year, theoretical calculations of the NLO electroweak corrections have been performed. They turn out to be larger than the QCD NLO corrections. Normally it is the other way around and first studies indicate this is a specific property of VBS related to the large EW scale in the process. The electroweak NLO corrections reduce the cross-section by about 15% in a typ-

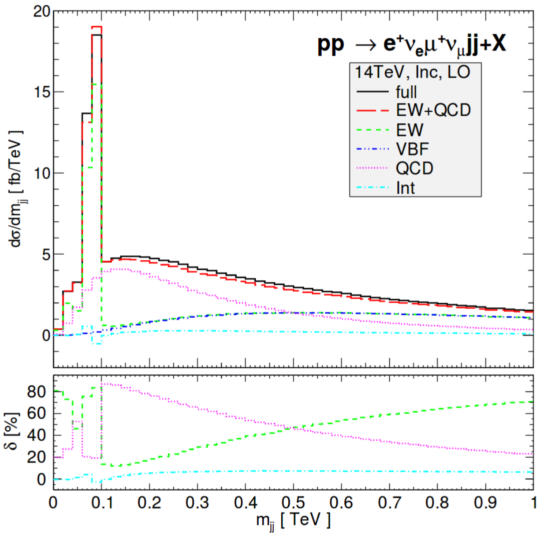


Figure 9: The differential cross section of the same-sign W boson production in the $e^+\mu^+\nu\nu jj$ final state as a function of the invariant mass of the two jets. The triboson contribution is shown as the difference between EW and VBF. The relative EW, QCD and interference contributions compared to the full LO results is plotted in the small panel. Figure taken from [25].

ical VBS phase space and its effect increases with the invariant mass of the di-quark and di-lepton system [29–31]. At the time the studies described in Part II were performed the electroweak NLO contribution was not yet available in event generators and it is only the LO calculations that are used in this thesis.

2.2 VBS TOPOLOGY

The scattering of vector bosons is accessible at the LHC when two incoming quarks interact through radiated vector bosons, leading to the observation of the decay products of the vector bosons, as well as the jets originating from the initial quarks that get deflected from the beam direction and enter the detector. This is schematically depicted in Figure 10.

The two highly energetic jets originating from the outgoing quarks are called the "tag jets" and are characterized by a large separation in pseudorapidity[3]

3 The pseudorapidity $\eta \equiv -\ln\left[\tan\left(\frac{\theta}{2}\right)\right]$, with θ the angle between the three-momentum of the particle and the positive direction of the beam axis. The relation between θ and η is visualized in Figure 11.

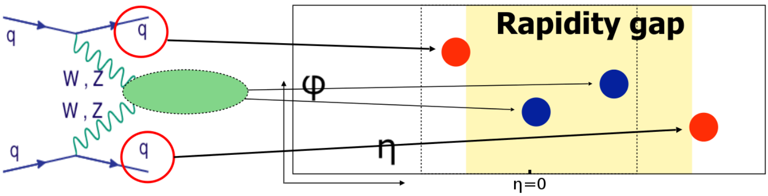


Figure 10: The event topology of VBS at pp colliders. The event signature is shown in a rolled out representation of the detector in the η-ϕ plane.

$\Delta\eta_{jj}$ and a high invariant mass m_{jj}. The W bosons and their decay products on the other hand are produced more centrally.

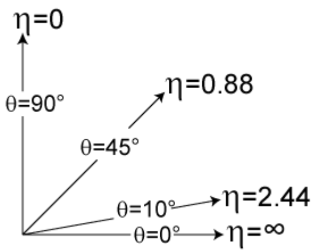


Figure 11: The relation between the polar angle θ and the pseudorapidity η, with $\eta = 0$ corresponding to the positive direction of the beam axis.

The vector boson that are radiated from the quarks are virtual particles. This is easy to understand in the rest frame of the incoming quark; the mass of the quark is not large enough to produce a heavy on-shell vector boson. The emitted virtual bosons tend to carry a small fraction of the incoming parton energy [32], while at the same time they need to carry enough energy to create the pair of outgoing on-shell bosons. This indicates that the outgoing jets are highly energetic. Their transverse momentum on the other hand is supressed, due to the factor that is introduced by the weak boson propagators:

$$\frac{1}{(p_q - p_{q'})^2 - M^2}, \tag{30}$$

with M the weak boson mass and $p_q - p_{q'}$ the four-momentum of the virtual boson, or equivalently the difference between the four-momenta of the incoming and outgoing quarks as defined in Figure 4. By approximating the relativistic quarks as massless, it follows that [23]:

$$\frac{1}{(p_q - p_{q'})^2 - M^2} \simeq -\frac{1}{p_{T,q'}^2 + M^2}, \tag{31}$$

which shows that outoing quarks with a large transverse momentum ($p_{T,q'} \gg M$) introduce an additional suppression factor.

2.2.1 *EW-Strong Separation*

The invariant mass spectrum of the two jets (m_{jj}) and their difference in pseudorapidity ($\Delta\eta_{jj}$) can be used to separate the electroweak from the strong VVjj production. Their distributions are shown in Figure 12 for the $W^{\pm}W^{\pm}jj$ production in the leptonic final state. The pseudorapidity distributions of the jet and lepton with the highest transverse momentum are shown in Figure 13. A variable that quantifies the relative pseudorapidity of the lepton with index i with respect to the di-jet system is the Zeppenfeld variable [33]:

$$z_i^{\text{lep}} = \frac{1}{\Delta\eta_{jj}} \left(\eta_i^{\text{lep}} - \frac{\eta_{j1} + \eta_{j2}}{2} \right), \tag{32}$$

which has an absolute value smaller than 0.5 if lepton i is in between the jets in pseudorapidity. The Zeppenfeld variable of the highest p_T lepton is shown in Figure 14.

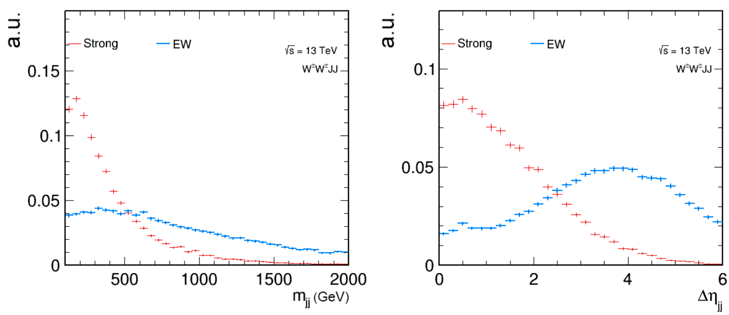


Figure 12: The m_{jj} and $\Delta\eta_{jj}$ distributions for the EW and QCD induced $W^{\pm}W^{\pm}jj$ production in the leptonic final state. The distributions are normalized to one and are generated with the MG5aMC [34] generator with $m_{jj} > 100\,\text{GeV}$, $p_T^{\text{lep},1} > 18\,\text{GeV}$ and $p_T^{\text{lep},2} > 8\,\text{GeV}$.

2.3 LONGITUDINAL POLARIZATION

There is more than one way to the conclude that a Higgs-like boson is needed in the standard model. The most common reasoning, as was followed in Section 1.5, is to add a particle that can provide mass to the vector bosons in a gauge invariant way. Another approach, is to restore the divergencies that

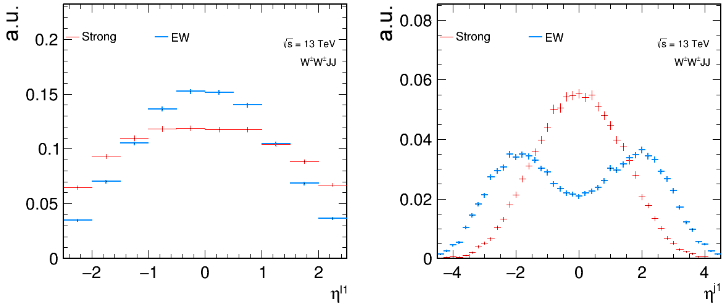


Figure 13: The pseudorapidity and transverse momentum distributions of highest p_T lepton and jet for the EW and QCD induced $W^{\pm}W^{\pm}jj$ production in the leptonic final state. The distributions are normalized to one and are generated with the MG5aMC generator with $m_{jj} > 100\,\text{GeV}$, $p_T^{\text{lep},1} > 18\,\text{GeV}$ and $p_T^{\text{lep},2} > 8\,\text{GeV}$.

appears in the electroweak sector of the SM. For a consistent quantum theory the sum of all probabilities, corresponding to the allowed evolution of the system, should be equal to one. This very basic requirement, called unitarity, is violated in the electroweak theory without the Higgs mechanism. When calculating the cross section for the scattering of longitudinal vector bosons, one finds that it increases with energy and as a consequence violates unitarity at high energy [35, 36]. As example, we will use the longitudinal same-sign W^+ boson scattering:

$$W_L^+ W_L^+ \to W_L^+ W_L^+ \,. \tag{33}$$

The LO amplitude without Higgs contribution in the high energy limit [37–39] is equal to:

$$\mathcal{M}_{\text{gauge}} = -g^2 \frac{s}{4m_W^2} + \mathcal{O}(s^0)\,, \tag{34}$$

with Mandelstam variable s equal to the square of the center-of-mass energy, m_W the W boson mass and g the SU(2) coupling constant (Equation 15). This amplitude corresponds to the sum of the quartic interaction and Z, γ exchange diagrams (as in Figure 4, but without the quark lines). The amplitude grows with energy, leading to unitarity violation at sufficiently high energy.

If the unitarity is not restored, the breakdown of perturbation theory at high energy would mean that the longitudinal vector bosons become strongly coupled in this region and multiple rescattering is likely to occur [40]. It also

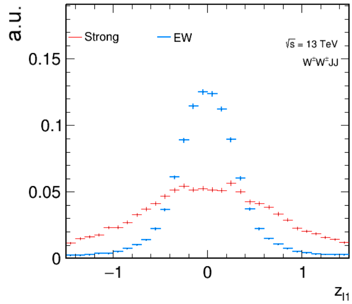


Figure 14: The Zeppenfeld variable (Equation 32) of the highest p_T lepton for the EW and QCD induced $W^{\pm}W^{\pm}jj$ production in the leptonic final state. The distributions are normalized to one and are generated with the MG5AMC generator with $m_{jj} > 100\,\text{GeV}$, $p_T^{\text{lep},1} > 18\,\text{GeV}$ and $p_T^{\text{lep},2} > 8\,\text{GeV}$.

means the predictability of the theory is lost. Preferably, we keep the coupling small and introduce additional particles and interactions to cancel the divergencies. The simplest solution consists of introducing a neutral, scalar particle H that can be exchanged between the W bosons (Figure 4). This adds a term to the LO amplitude that in the high energy limit becomes:

$$\mathcal{M}_H = g_{HWW}^2 \frac{s}{4m_W^4} + \mathcal{O}(s^0)\,. \tag{35}$$

The leading divergences cancel ($\mathcal{M}_{\text{gauge}} + \mathcal{M}_H = \mathcal{O}(s^0)$) when the following requirement is exactly fulfilled:

$$g_{HWW} = g \cdot m_W\,. \tag{36}$$

The HWW coupling predicted by the Higgs mechanism satisfies this requirement. To restore unitarity we need a scalar particle that couples in the exact same way as the Higgs boson that was derived in the electroweak symmetry breaking approach. The same calculation can be done for longitudinal VBS with other vector bosons and shows that the Higgs mechanism unitarizes the VBS process.

Requiring unitarity on the partial wave expansion of the scattering amplitude leads to a limit on the mass of the particle that unitarizes VBS. The particle needs to have a mass $\lesssim 1.2\,\text{TeV}$. Before the LHC was build, it was guaranteed that new physics would be discovered when reaching center of mass energies of $\sim 1.2\,\text{TeV}$. Either in the form of a Higgs-like boson or strongly interacting vector bosons.

In models that predict new physics in the EWSB sector, for example in the form of multiple Higgs bosons, the exact cancellation with the discovered Higgs boson does not apply [41]. New physics processes could disturb the delicate balance of the strong cancellations at high energy and lead to potentially large enhancements of the VBS rate. It is therefore crucial to test the exact cancellation as predicted by the SM and Higgs mechanism. The effect of a non-exact cancellation is shown in Figure 15.

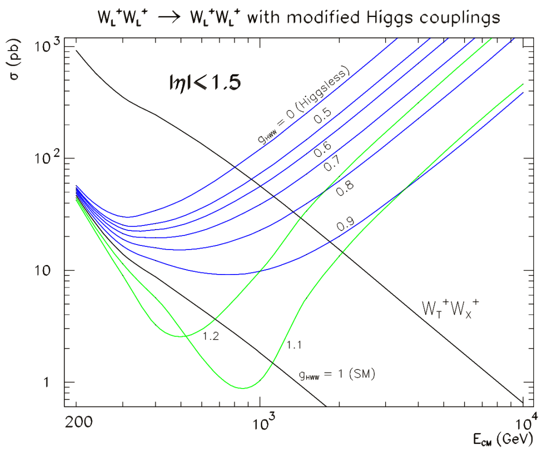


Figure 15: The $W_L^+W_L^+$ scattering cross sections as a function of the center of mass energy for different values of the HWW coupling. The multiplicative factor to the coupling is shown on the figure, with $g_{HWW} = 1$ corresponding to the Standard Model. Assumed are two colliding on-shell, unpolarized W^+ beams. Plot taken from [38].

In practice, measuring the longitudinal scattering component is far from trivial. It makes up no more than 5% of the total cross section containing all polarization states. Contamination from initial scattering states containing transverse polarized vector bosons ($V_T V_X \to V_L V_L$) can further complicate the matter. Generally, this background is small. For $W_L^\pm W_L^\pm$ scattering specifically it is negligible, but for this process information on the polarization is partially lost with the undetected neutrinos.

2.4 BEYOND THE STANDARD MODEL PHYSICS

The Higgs mechanism is not the only way to unitaritize the VBS process. Many extensions to the Higgs sector have been proposed and a variety of these models are still consistent with data. Measurements of the VBS process can be used to exclude – or place constraints on – extensions to the SM

describing physics beyond the standard model (BSM).
In one class of BSM models the Higgs boson is a composite particle, namely a bound state of new strong interactions. In many of these models the light composite Higgs emerges from the strongly-interacting sector as a pseudo-Goldstone boson [42, 43]. A different possibility is the existence of a strongly interacting fourth family [44].
In another class of theories the observed Higgs boson is part of an extended Higgs sector, as in the electroweak singlet [45] and two Higgs doublet model (2HDM) [46].

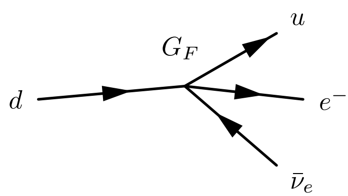
The VBS process is also sensitive to triple and quartic gauge couplings. Triple gauge couplings have been measured in non-VBS diboson production at the LEP collider [47, 48], TeVatron [49] and with highest precision recently at the LHC [50]. Their measurements show no significant deviation from the standard model expectation. The quartic couplings have not been measured with high precision yet and could be modified by BSM physics.

2.4.1 *Effective field theory*

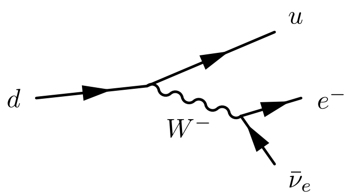
Searches for new physics can be performed by looking directly for the production of new particles, for example by searching for a resonant bump in an invariant mass spectrum, or by performing precision measurements of the interactions between SM particles. The second method allows to measure particles that are too heavy to be created at the available center-of-mass energy, but modify interactions by its virtual corrections. Since many BSM models exist, it would be useful to have a model independent framework in which sensitivity to new physics at a high energy scale can be quantified. An Effective Field Theory (EFT) approach is used to accomplish this, where the SM is assumed to be the low energy approximation of the full theory.

2.4.1.1 *Fermi Theory*

The most common example of an EFT in particle physics is Fermi theory. It was postulated in 1933 by Fermi [51] to describe nuclear beta decay and introduces quartic fermion interactions with coupling G_F:



The theory gives a good description at the low energies available at the time, but leads to non-renormalizable divergencies at high energy. This can be resolved by smearing the interaction point in space and time by the exchange of a virtual massive boson W:



The W boson introduces a propagator factor to the diagram: $\sim \frac{g^2}{k^2 - m_W^2}$ that at low energy ($k \ll m_W$) is indistinguishable from a point-interaction with coupling $\sim \frac{g^2}{m_W^2}$.

2.4.1.2 *Standard Model Effective Field Theory*

The SM is generally regarded as an incomplete theory, as already discussed in the previous chapter. It is reasonable to assume there exists new physics in the large unexplored energy range of $\sim 10^{12}$ TeV between the highest energy reached at collider experiments and the maximum allowed energy of a point particle indicated by the Planck mass. In this case, the SM is the low energy approximation of a larger, *full* theory. The new physics is then characterized by a higher energy scale Λ, usually describing the mass of new particles.

At low energy ($E \ll \Lambda$) the full theory can be approximated by adding higher order operators to the SM Lagrangian, describing new and modified interactions among the light SM particles. In Fermi theory, the higher order operators are the quartic fermion interactions that approximate the weak boson exchange.
One can write down all operators containing SM fields that obey the symmetries of the SM, like Lorentz invariance, gauge symmetry and conservation of baryon and lepton number [52]. From dimensional analysis, these operators ($\mathcal{O}^{(N)}$ with dimension N) have coefficients of inverse powers of mass and are suppressed by the high energy scale Λ, leading to a Lagrangian of the form:

$$\mathcal{L} = \mathcal{L}_{SM} + \sum_i \frac{c_i^{(6)}}{\Lambda^2} \mathcal{O}_i^{(6)} + \sum_i \frac{c_i^{(8)}}{\Lambda^4} \mathcal{O}_i^{(8)} + ... \tag{37}$$

with the SM Lagrangian containing dimension 4 terms. Higher order operators will be suppressed by higher powers of Λ.

Dimension 8 operators

We will consider only the operators that introduce gauge interactions. Since the dimension 6 operators introduce triple gauge interactions, one would expect that they can be better constrained by other measurements. It has been confirmed with simulation samples that the effect of dimensions 6 operators is very small on the same-sign W boson pair production, for coupling values that have been excluded in other measurements. Instead, we will focus on the dimension 8 operators, which are the lowest dimension operators that lead to quartic interactions without also producing two or three weak gauge boson vertices. Even though the dimension 8 operators are suppressed by powers of Λ with regards to the dimension 6 operators, they can still be dominant over the dimension 6 operators depending on the symmetry and particle content of the BSM model. For example, the exchange of a heavy boson can generate a direct, *tree level* contribution to quartic gauge-boson couplings (as in Fermi theory) while the triple gauge vertex would only be effected at *one-loop level* [53].

The dimension eight operators that give rise to quartic WWWW couplings [54], called anomalous quartic gauge couplings (aQGC), are listed here:

$$L_{S,0} = \left[(D_\mu\phi)^\dagger D_\nu\phi\right] \times \left[(D_\mu\phi)^\dagger D_\nu\phi\right]$$
$$L_{S,1} = \left[(D_\mu\phi)^\dagger D^\mu\phi\right] \times \left[(D_\nu\phi)^\dagger D_\nu\phi\right]$$

$$L_{M,0} = Tr[W_{\mu\nu}W^{\mu\nu}] \times \left[\left(D_\beta\phi\right)^\dagger D^\beta\phi\right]$$
$$L_{M,1} = Tr[W_{\mu\nu}W^{\nu\beta}] \times \left[\left(D_\beta\phi\right)^\dagger D^\mu\phi\right]$$
$$L_{M,6} = \left[(D_\mu\phi)^\dagger W_{\beta\nu}W^{\beta\nu}D^\mu\phi\right]$$
$$L_{M,7} = \left[(D_\mu\phi)^\dagger W_{\beta\nu}W^{\beta\mu}D^\nu\phi\right]$$

$$L_{T,0} = Tr\,[W_{\mu\nu}W^{\mu\nu}] \times Tr\left[W_{\alpha\beta}W^{\alpha\beta}\right]$$
$$L_{T,1} = Tr\left[W_{\alpha\nu}W^{\mu\beta}\right] \times Tr\,[W_{\mu\beta}W^{\alpha\nu}]$$
$$L_{T,2} = Tr\left[W_{\alpha\mu}W^{\mu\beta}\right] \times Tr\,[W_{\beta\nu}W^{\nu\alpha}]\ .$$

These operators are naturally divided into three categories. Two of them only contain Higgs doublets (S-operators) and parametrize physics that couple to the SM through the Higgs field only. Three others only contain SU(2) field strengths (T-operators) and the remaining four contain both SU(2) field strengths and Higgs doublets (M-operators). Each operator is scaled by a corresponding *Wilson* coefficient ($c_i^{(8)}/\Lambda^4$).

Unitarity

The SM EFT is not valid at high energy. For energies at the order of Λ there is no suppression of higher dimension operators and it would not make sense to only consider the lowest order operators. Nevertheless, multiple extensions to the SM EFT have been developed to unitarize the operators up to arbitrarily high energy [55, 56]. As the EFT theory is fit to data, it can not violate unitarity in the energy region that is measured. Requiring unitarity in energies beyond the measured region seems overly restrictive, as the EFT theory is invalid at high energy by construction [57]. In addition, different unitarization methods are relevant for different BSM models, while the use of a general method is crucial to compare the sensitivity between the data analysis of different processes. Recent discussions have lead to a proposal of applying cuts on the center-of-mass energy $\sqrt{s}$, but no straight-forward recipe has been proposed for W boson scattering where it is impossible to reconstruct this variable due to the escaping neutrinos. For these reasons, no unitarization method will be applied to the SM EFT in this thesis.

3 EXPERIMENTAL SETUP

Particle colliders have been the driving force behind discoveries and theoretical advancements in particle physics for the last 50 years. The quest for larger, higher energy colliders has culminated in the Large Hadron Collider (LHC). Protons are accelerated to record energies and brought to collision, allowing the study of rare interaction processes, occurring with a probability down to ~ 15 orders of magnitude smaller than the total interaction probability. Large detectors are build around the collision points and measure the stable final state particles that are produced in the interactions. The Compact Muon Solenoid (CMS) is one of the general purpose detectors, build to optimally measure the outgoing particle types, energies and directions. Planned upgrades to the LHC and CMS detector in the near future guarantee a significant increase in the physics reach for an additional decade of operation.

3.1 PARTICLE COLLIDERS

When ordinary particles collide at high energy, more exotic particles with a higher mass can be created. Increasing the center-of-mass energy of the collisions is therefore a straightforward way to improve a particle collider. But discoveries can also be made by improving the precision of the measurement, which can be done by increasing the number of collisions, by using a setup with a better determined initial state, or by decreasing other uncertainties.

The number of collisions is expressed by the interaction rate (W) at a collider, equal to the number of interactions per time. The interaction rate can be split in two factors: a process-dependent factor σ and a factor L that depends on the beam parameters of the collider:

$$W = \sigma \cdot L \,. \tag{38}$$

The total interaction cross section σ is the effective area for collision: the area transverse to the two particles relative motion within which they must meet in order to scatter from each other. The instantaneous luminosity L is by consequence expressed in units of $1/(\text{area} \cdot \text{time})$. The unit of area that is commonly used for these quantities is the barn (b) equal to $10^{-28}\,\text{m}^2$, or the femtobarn ($10^{-43}\,\text{m}^2$).
The main characteristics of a collider, next to the type of accelerator and type of particles that are used, are the center-of-mass energy ($\sqrt{s}$) and the instantaneous luminosity (L).

Multiple types of particles have been used and proposed for use in accelerators, the most obvious candidates being the electron and proton. The advan-

tage of using protons, is that they can be accelerated to very high energies in circular accelerators, as the particles can be accelerated with each revolution. Lighter particles - as the electron - lose more of their energy to synchrotron radiation when accelerated in a curved path, which is inversely proportional to the particles mass to the fourth power. The disadvantage of using protons, is that they are not fundamental particles. Only part of the proton takes part in the interaction, the remaining energy in the proton continues along the beampipe and is not measured. As a consequence, the center-of-mass energy of the colliding partons is a priori not precisely known.
Also heavy ions are collided in particle accelerators. Their collision can lead to the creation of a quark-gluon plasma: a state of matter consisting of asymptotically free strong-interacting quarks and gluons. The study of this state can bring new insights to the non-perturbative regime of QCD and the early evolution of our universe.

3.2 THE LHC

The first mention of a superconducting proton collider at CERN originates from 1977 [58]. After the successes of lepton colliders, colliding hadron beams at an increased center-of-mass energy would allow to study an unexplored energy region. The construction of the Large Hadron Collider was approved in 1994, after the cancellation of the similar Super-Conducting Super Collider project in the US in 1993, and having the advantage of reusing the existing infrastructure from the Large Electron-Positron (LEP) collider at CERN. Fifteen years later, during which many technological advancements were made for the construction of the LHC and its experiments, the LHC went into operation as the world's largest and highest-energy particle accelerator.

The LHC is a hadron accelerator and collider housed in a roughly circular 27 km tunnel at the border of Swiss and France [59]. A small portion of the LHCs running time is devoted to the collision of lead ions, while most of the running time two proton beams are collided after being accelerated in opposite direction inside vacuum beam pipes up to a maximum design energy of 7 TeV per proton. When the operational energy is reached, the beams are squeezed and brought to collide in 4 interaction points where the experiments are set up. Only a very small fraction of protons interact in each crossing and the proton beams are kept in circulation for multiple hours until the intensity of the beams is significantly reduced due to small losses accumulating over time. At this point, the beams are dumped and new proton bunches are injected.
Oscillating electric fields in radio frequency cavities accelerate the protons and keep them packed in separate bunches. About 2800 bunches of protons, that are separated by 25 ns in time, can be stored in each beam. Superconducting dipole electromagnets are used to keep the protons on their circular

path inside the 27 km ring, while higher order magnets are used to focus the beams. The magnets are cooled to below 2 K using superfluid helium, and operate at fields above 8 T.

Before the protons are injected in the LHC, they have already traveled through a chain of smaller accelerators that boost them to an energy of 450 GeV. A schematic overview of the accelerator complex and the experiments at CERN are shown in Figure 16.

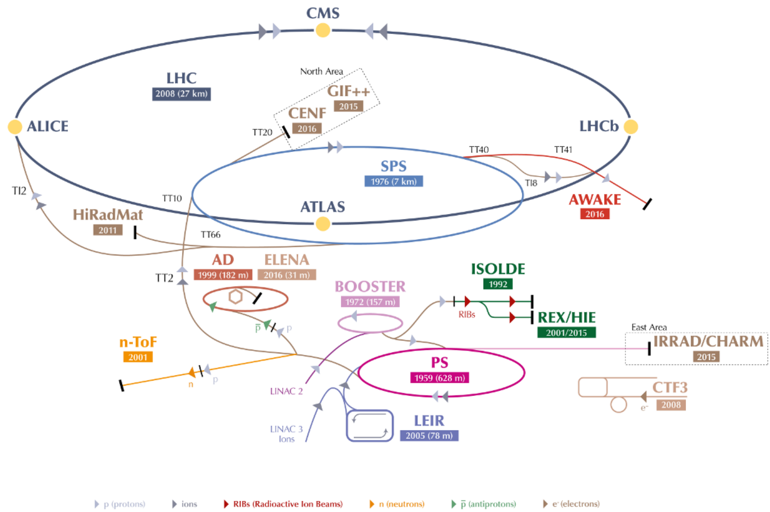


Figure 16: The accelerator complex and experiments at CERN [60].

3.2.1 *LHC Schedule*

The LHC operation alternates running periods of ~ 3 years with long shutdowns of ~ 2 years that are used for maintenance and upgrades to the collider and experiments. A summary of the schedule is shown in Figure 17, indicating the different experiment phases, center-of-mass energies, integrated luminosity and run periods as a function of time. The Run I data-taking period started in 2011 at a center-of-mass energy of 7 TeV, which was increased to 8 TeV in 2012, and during which almost 30 fb^{-1} of data was recorded. The data used in this thesis is recorded in 2016 during the second run of data-taking, with a center-of-mass energy of 13 TeV and a peak luminosity of 15 Hz/nb. The integrated luminosity of this dataset is 35.9 fb^{-1}.
After 300 fb^{-1} of data is collected by 2024 a major upgrade to the LHC is planned, called the High Luminosity (HL) upgrade, which is described in Section 3.5.

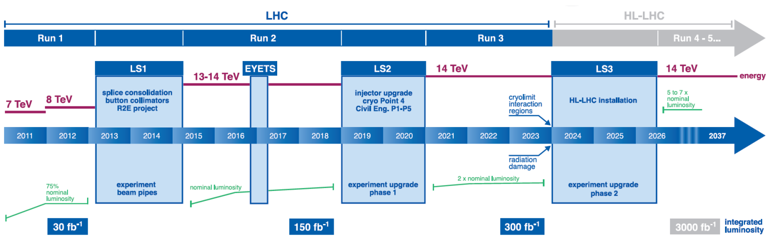


Figure 17: The timeline of the LHC programme between 2011 and 2037 [61].

3.2.2 *Experiments at the LHC*

Four large experiments, and three smaller experiments are build at the interaction points. Two large, general-purpose detectors (ATLAS [62] and CMS [63]) were designed to study the electroweak symmetry breaking mechanism, perform precision tests of the SM, and search for signs of new physics (e. g. supersymmetry, dark matter candidates). More specialized large experiments study the quark-gluon plasma produced in heavy-ion collisions (ALICE [64]) and CP violation and rare decays of b- and c-hadrons (LHCb [65]). The smallest experiments are used to study forward physics (TOTEM [66] and LHCf [67]) and to search for magnetic monopoles and other highly ionizing stable massive particles (MoEDAL [68]).

3.3 pp COLLISIONS AT THE LHC

Not all proton-proton interactions that take place at the LHC are head-on collisions. A significant fraction are glancing blows that leave the protons either intact (elastic collisions), or dissociated in only a few particles (diffractive collisions). In non-diffractive collisions an exchange between the protons of a particle with quantum numbers different from the vacuum takes place, and the outgoing particles produced in the collision gain a transverse momentum large enough to enter the detector acceptance. The non-diffractive cross section makes up about half of the total interaction cross section [69]. In these non-diffractive interactions, a very small fraction contains interesting events where high energetic or exotic particles are produced. An overview of the final states that are produced at the LHC and their cross sections is shown in Figure 18. The electroweak production of $W^{\pm}W^{\pm}$, the main subject of the second part of this thesis, has a cross section of about $0.35\,\text{pb}$ at $\sqrt{s} = 13$ TeV which is an order of magnitude smaller than the lowest line on Figure 18 for associated WH production.

Next to the proton-proton interactions we want to measure, interactions between other protons also leave signals in the detector at the same time. These additional interactions are called pileup, and originate from two sources. The

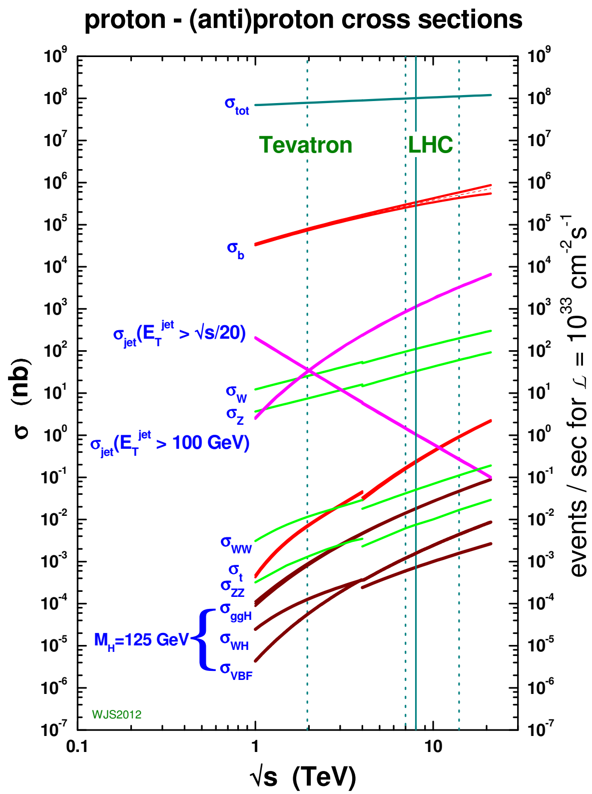


Figure 18: Standard Model production cross sections at hadron colliders as a function of the center-of-mass energy. The discontinuities in the different curves denote the change from $p\bar{p}$ collisions at the Tevatron to pp collisions at the LHC. The current center-of-mass energy for pp collisions at the LHC is indicated by the rightmost vertical dashed line ($\sqrt{s} = 13$ TeV) [70].

additional proton-proton interactions in the same bunch crossing, caused by the high luminosity at the LHC, are called in the in-time pileup. Particles produced in previous and next bunch crossings can also contribute to the energy measurement, and make up the out-of-time pileup.

3.3.1 *Final State Objects*

Heavy particles produced in the interaction will decay shortly after their production to lighter, stable particles with a longer lifetime. It are the stable particles (photons, electrons, muons, neutrinos, and charged and neutral hadrons) that travel through the detector. Except for the neutrinos (and possibly new particles as predicted in theories beyond the standard model), we can identify them from their interaction with the detector. The stable particles are reconstructed as physics objects, approximating the final state particles produced in the hard collision, divided in the following categories:

- Photon
- Electron
- Muon
- Missing transverse energy (E_T^{miss}):
  Neutrinos are stable particles that do not participate in the electromagnetic or strong interactions and as a consequence do not interact with the detector. In a hermetic detector, their presence can still be detected by an imbalance of energy in the transverse plane, called the missing transverse energy. It must be noted that the imperfect energy reconstruction of the other particles, particles not within the detector acceptance, or undiscovered weakly interacting particles can also contribute to the E_T^{miss}.
- Jet:
  Energetic quarks and gluons produced in the event lead to a spray of particles, consisting mostly of light hadrons, that are collimated along the direction of the initial parton. Most of the jet energy is carried by pions, the lightest mesons build from combinations of up and down quarks. The remainder of the jet energy is mostly carried by kaons (mesons containing a strange quark) and light baryons (protons and neutrons) [71].
  Jets can be divided further according to the particle they originate from, which can be a gluon, light quark, b-quark or a hadronic decayed tau.

3.4 CMS DETECTOR

The Compact Muon Solenoid (CMS) is a general-purpose detector, designed to study a wide range of particles and phenomena produced in the high energy collisions at the LHC. Although it is 15 meters high and 22 meters long, it is a compact detector given its total weight of 14,000 tonnes and all the detector material it contains. The second letter in CMS refers to the large muon detectors at the outside of CMS, that allow an excellent muon identification and reconstruction. Finally, the solenoid is the heart of the detector, around which all other components are build. With a length of 13 m and a diameter of 6 m it the largest of its kind, producing a magnetic field of 3.8 tesla. The strong magnetic field makes it possible to measure charged particle momenta with high resolution, despite the compactness of the detector. The solenoid is complemented by a iron return yoke to shape and confine the magnetic field. Different layers of dedicated detectors inside the solenoid are designed to optimally measure the different categories of stable particles; the most central layers are used to measure the charged particle trajectories, surrounded by calorimeters that perform destructive energy measurements. CMS has a hermetic design, consisting of a barrel and two closing end-caps. An overview of the CMS detector is shown in Figure 19.

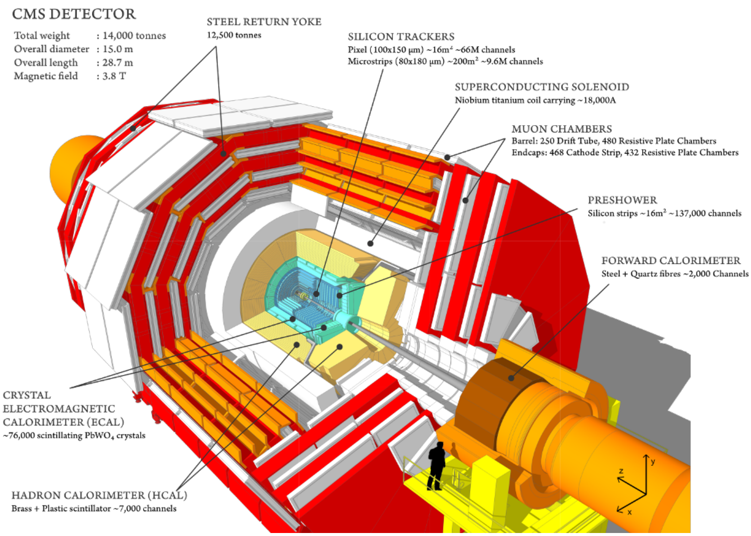


Figure 19: Schematic view of the CMS detector.

A righthanded coordinate system is used in CMS with the origin centered at the nominal collision point inside the experiment. The z-axis points along the counterclockwise direction of the proton beam (as seen from the top), the x-axis points to the center of the LHC ring and the y-axis points vertically upward. The pseudorapidity angle η is a monotonic function of the polar angle θ between the direction of the particle and the positive direction of the z-axis, as defined in the footnote of Section 2.2. The azimuthal angle ϕ is measured from the x-axis in the x-y plane. The coordinate axes are drawn on the views of the CMS detector (Figures 19 and 20).

3.4.1 *Stable particle signatures*

The stable particles interact with the CMS detector in a specific way, so that different particles can be distinguished from the set of subdetectors they interact with. The typical signatures of the different categories of stable particles in CMS are shown in Figure 20. The energy of photons and electrons is measured with the electromagnetic calorimeter. Electrons also leave hits in the tracker along their trajectory, while photons have zero charge and do not interact with the tracker. Hadrons deposit most of their energy in the hadronic calorimeter and charged hadrons also leave signals in the tracker and electromagnetic calorimeter. Muons are not stopped inside the detector, but leave behind a small fraction of their energy as hits in the tracker and muon detectors. The momentum of muons, as for all charged particles, can be measured from the curvature of their trajectory in the magnetic field.

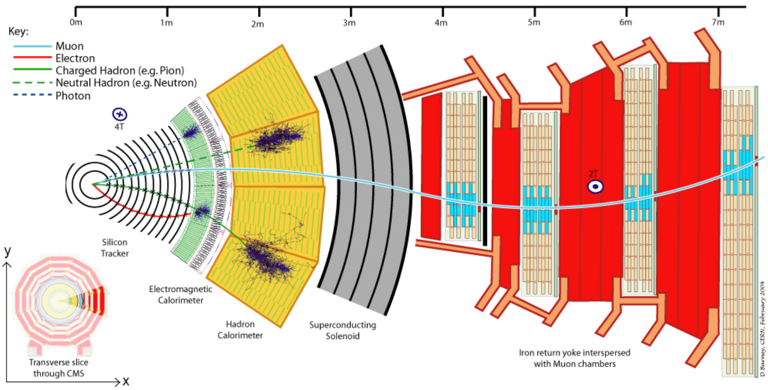


Figure 20: Overview of the detectors and interactions of the stable particles in a transverse slice of CMS.

High energetic charged particles that pass through the tracker will ionize atoms in the tracker material, losing only an insignificant fraction of their energy along the way. A light particle as the electron can lose a large fraction of its energy through bremsstrahlung, when the interaction with the material causes it to radiate a photon ($e \to e\ \gamma$) [72]. High energetic photons can decay in their turn to electrons when interacting with the material ($\gamma \to e^+e^-$), leading to the formation of particle showers: an avalanche of particles from radiated photons and pair productions, as sketched in Figure 20. This process is exploited in the electromagnetic calorimeter to measure the energy of electrons and photons as they gradually give off their energy. A similar effect takes place for hadrons in the hadronic calorimeter, where hadronic particle showers develop through multiparticle production and nuclear decay of excited nuclei. Hadronic showers take longer to develop, as they rely on nuclear interactions with a small probability compared to the interactions that cause electromagnetic showers.

3.4.2 *Tracker*

The first detector that the outgoing particles encounter is the tracker. It is used to reconstruct the charged particle trajectories and interaction vertices. To perform this task in the high pileup environment at the LHC with bunch crossings every 25 ns, a high granularity and fast response are required. In addition, the detector needs to be radiation hard and lightweight to avoid inducing particle showers.
The tracker, with a total length of 5.8 m and a diameter of 2.5 m, has a coverage in pseudorapidity up to $|\eta| < 2.5$. It is build from several subsections which are all silicon-based. Closest to the interaction point, a high granularity is assured by 3 layers of 150×100 μm silicon pixel detectors in the barrel and 2 layers in the endcap. This is complemented by silicon-strip detectors with a width of 80-180 μm, a length between 10 and 25 cm, and a geometry as shown in Figure 21.

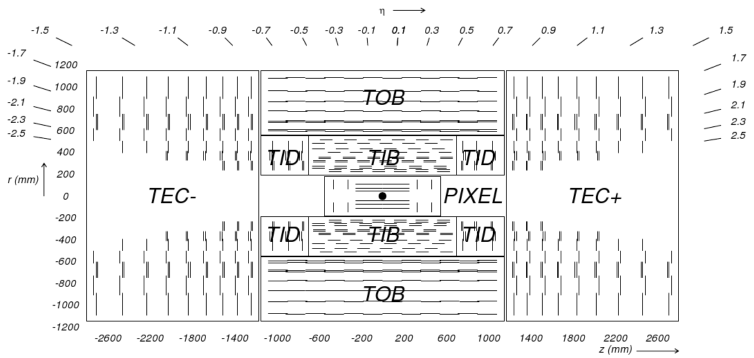


Figure 21: Overview of the modules that make up the tracker. Double lines indicate silicon strip modules that are mounted back-to-back with one of them rotated to obtain a 3D position measurement [73].

When charged particles pass through the silicon semiconductor [74], electron-hole pairs are created. An electric field is applied to separate and transport them, after which the collected charge is read out and amplified by the 75 million separate read-out channels.

The large amount of material inside the tracker volume (of detectors, support, cables and cooling) increases the probability for particle showers to be initiated inside the tracker, complicating the reconstruction (Section 4.3). This is quantified by the radiation length X_0, equal to both the mean distance an electron has to travel through a material to reduce its energy by a factor e, and $7/9$ of the photons mean free path before engaging in pair production. The number of radiation lengths a particle crosses as it passes through the tracker is shown in Figure 22 as a function of the pseudorapidity. The equivalent quantity for hadronic showers is the nuclear interaction length, equal to the mean distance before undergoing an inelastic nuclear interaction. The nuclear interaction length is about 4 times as large as the radiation length for the tracker material.

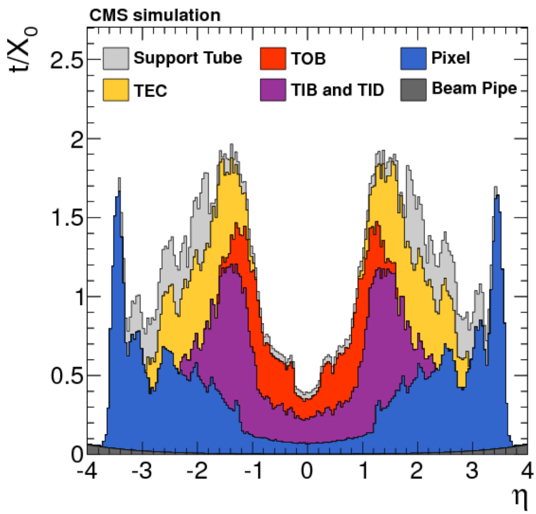


Figure 22: The number of radiation lengths X_0 a particle covers when transversing the tracker material as a function of the pseudorapidity [73].

3.4.3 *Electromagnetic Calorimeter*

The energy of electrons and photons is measured in the electromagnetic calorimeter (ECAL), up to a pseudorapidity of $|\eta| < 3.0$. The detector contains $\sim 76,000$ high density and highly transparent lead tungstate crystals ($PbWO_4$). Each crystal has a length of 22-23 cm, corresponding to about 25 radiation lengths. This means that when an energetic electron or photon with energy up to 1 TeV enters the crystal, more than 98% of the electromagnetic particle showers longitudinal energy is contained. The transverse size of the crystals matches the Moliére radius of 2.2 cm for $PbWO_4$ corresponding to the radius of a cylinder containing on average 90% of the showers transverse energy deposition.

When a charged particle passes through the crystals that make up the ECAL, it excites atoms in the material which relax almost instantly by emitting scintillation light. The total amount of scintillation light is proportional to the particles energy and is measured with photodetectors. Avalanche photodiodes made of semiconducting silicon are used in the barrel, while in the endcap vacuum phototriodes are used to withstand the high amount of radiation. The electric signal is then amplified, digitized and transported through optical fibres.
In order to improve the separation between single photons and double photons from neutral pion decays, a preshower detector is placed in front of the endcap calorimeter. Two planes of lead that induce electromagnetic showers are each followed by silicon sensors, enhancing the spacial resolution. An overview of the ECAL is shown in Figure 23.

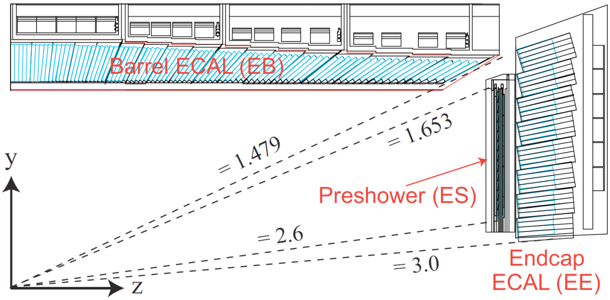


Figure 23: Geometric view of one quarter of the ECAL [75].

3.4.4 *Hadron Calorimeter*

The energy of hadrons is measured with a sampling calorimeter, called the hadron calorimeter (HCAL). It alternates layers of active detector material with layers of high density absorber material that maintain the hadronic particle showers. Sampling calorimeters can not measure the full energy of the shower but allow a reduction in size and cost of the detector.
The unusual choice of placing the calorimeters inside the solenoid magnet has as a consequence that the calorimeters can not be that large, mainly the hadronic calorimeter inside the solenoid is not large enough to contain the full energy of jets. The remaining energy is measured in additional barrel calorimeters placed outside of the solenoid. Coverage of the forward region ($3.0 < |\eta| < 5.2$) is provided by forward calorimeters, located at 11 m from the interaction point. Together with the usual barrel and endcap detectors, they make up the HCAL as shown in Figure 24. The full pseudorapidity acceptance is covered by about 10 nuclear interaction lengths. This means more than 99% of the energy is contained when a particle with an energy

up to 300 GeV enters the HCAL.

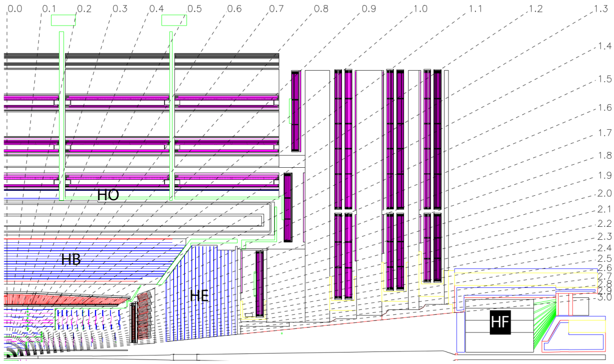


Figure 24: Longitudinal view of the CMS detector showing the locations of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters [63].

The barrel and endcap calorimeters alternate 4-9 mm plates of plastic scintillators as the active detector material, with plates of steel and brass with a thickness of 40-80 mm as absorber. In the outer calorimeter, the return yoke also serves as absorber material. The scintillation light is captured by wavelength-shifting fibres, after which is it optically added with successive tiles to form energy *towers* proportional to the shower energy. Finally, the light is converted with silicon-based hybrid photodiodes to electric signals.
The forward calorimeters use different materials to cope with the high amount of deposited energy in the forward region. Quartz (SiO_2) fibres, parallel to the beam line, are embedded in 5 mm thick steel plates. When particles travel through a fibre at a speed faster than the speed of light in the material, Cherenkov radiation is generated and transported through the fibre. Two lengths of fibres are used to discriminate between electromagnetic particle showers and the longer hadronic particle showers.

3.4.5 *Muon detectors*

The muon detectors measure the trajectories of charged particles in four stations placed inside the return yoke. As the return yoke filters out the remaining shower particles, it are almost exclusively muons that interact with these detectors, yielding an excellent muon identification. When a muon passes through, it ionizes the gas inside the muon chambers. A large electric field is applied that accelerates the charged particles and creates a local ionization avalanche inside the gas which is subsequently read out.
In the barrel, drift tubes (DT) measure the accumulated electron charge with

anode wires strung inside the tube. The drift time of the charges gives a position measurement with good spacial resolution. In the endcap, where the magnetic field is uneven and particle rates are high, cathode strip chambers (CSC) are used up to $|\eta| < 2.4$. Arrays of positively-charged anode wires and orthogonal cathode strips measure the charge of respectively electrons and positive ions. In both regions, the detectors are complemented by double-gap resistive plate chambers (RPC) with an excellent time resolution. Resistive parallel-plate electrodes create the electric field in two gaps and the charges are read out by a metallic strip in between. The time resolution of 1 ns is useful for associating the measurement to a specific bunch crossing and making fast decisions on if the event should be saved. An overview of the muon detector is shown in Figure 25.

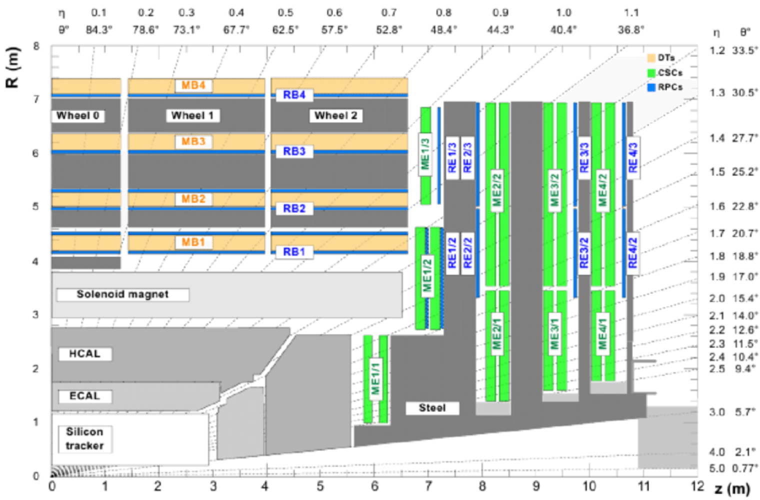


Figure 25: A quadrant of the muon detectors in CMS, consisting of Drift Tubes (DT), Cathode Strip Chambers (CSC) and Resistive Plate Chambers (RPC) [76].

3.4.6 *Trigger and Data Acquisition system*

With proton-proton collisions occurring every 25 ns, the CMS detector produces about 50 TB of data every second. As it is technically not possible to store and process all this data, a fast selection has to be performed to reduce it. The event rate is reduced by a factor 10^7 in two successive steps: a level 1 (L1) trigger and a High Level Trigger (HLT). The L1 trigger reduces the event rate of 40 MHz to about 100 kHz, and is performed close to the detectors with hardwired processors while the data is temporarily stored in pipelines. As the L1 trigger only has a few microseconds to make a decision, only the

less complicated information from the muon detectors and calorimeters can be used.
The *event builder* reads the data selected by the L1 trigger from the front-end drivers, sorts it in physical events, and passes it on to the HLT. These events are then processed with commercial hardware in a computer farm. Successive filters reconstruct part of the event and decide if the event should be discarded, using information from the full detector. One second later, a final decision is made whether the event is thrown away or stored on disk, reducing the event rate further to a frequency at the order of 100 Hz. An overview of the trigger and data acquisition system (DAQ) is shown in Figure 26.

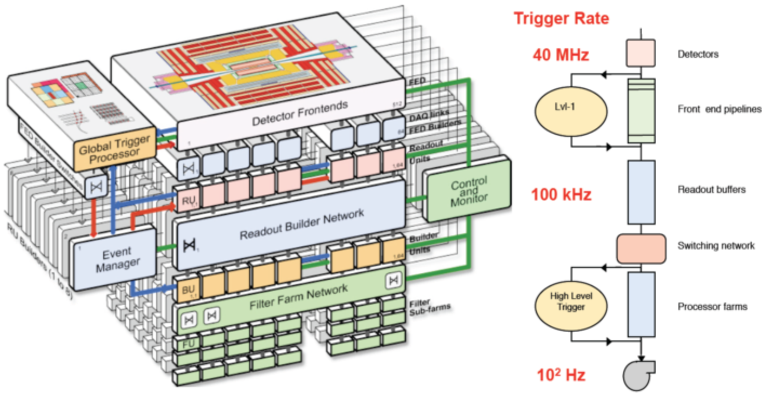


Figure 26: Overview of the CMS trigger and data acquisition system [77]. The DAQ system can be equipped with up to eight slices.

3.5 THE HIGH LUMINOSITY UPGRADE

After having reached an integrated luminosity of 300 fb^{-1} by the end of 2023, an upgrade of the LHC is deemed necessary in order to benefit significantly from recording more data. By this time, the quadrupole magnets that focus the beams at the ATLAS and CMS collision regions are close to the end of their lifetime due to radiation exposure and will be replaced. Together with crab-cavities that will be added to optimize the bunch overlap at the interaction region, they will produce an increase of the luminosity with a factor of 5 to 7. An overview of the conditions after the High Luminosity LHC upgrade (HL-LHC) are shown in Table 1.

The high amount of pileup that results from the increase in luminosity, will also set very challenging conditions for the CMS detector. Additional pileup events in the detector complicate the triggering, offline reconstruction and interpretation of the events, and this on top of the high radiation dose

	2016 Conditions	HL Baseline Scenario	HL Upper limit
$\sqrt{s}$ (TeV)	13	14	14
Instantaneous Lumi. ($cm^{-2}s^{-1}$)	10^{34}	5×10^{34}	7.5×10^{34}
Integrated Luminosity (fb^{-1})	40 (1 year)	3000 (10 year)	4000 (10 year)
Average pileup	27	140	200

Table 1: The main LHC characteristics at the CMS detector during the 2016 data-taking compared to 10 years of HL-LHC operation [78]

that will accumulate during the 10 years of high luminosity operation. Simulations and measurements of the effects of radiation on the CMS detector and test components show that the tracker and endcap calorimeters will have to be replaced entirely [2]. The full set of upgrades that will be performed on the CMS detector is called the Phase II upgrade. The upgrade is necessary to deal with the high luminosity conditions and aims to recover the current performance in the capability to trigger and reconstruct events. The main upgrades to the detector are conducted on the following systems [78]:

- Tracker: A replacement of the full tracker will be carried out with a granularity increase of a factor 6 for the pixel tracker and factor 4 for the outer tracker. In addition, the tracker will be extended up to $|\eta| \approx 3.8$.
  On-detector readout electronics will make it possible to reconstruct tracks at the hardware trigger level.

- Calorimeter endcaps: The electromagnetic and hadronic sections will be replaced by a high granularity calorimeter, with an improved transverse and longitudinal segmentation while maintaining the η coverage.

- Muon detectors: Additional muon chambers will be installed, increasing the redundancy and extending the muon coverage up to $|\eta| \approx 3.0$.

- Trigger: The latency of the L1 trigger will be increased to 12.5 μs and the readout electronics of the electromagnetic calorimeter in the barrel will be upgraded. This will allow the use of information from the tracker in combination with information from the other subdetectors.

Another promising option, currently under study, is to use the improved time resolution of the calorimeters to connect neutral energy deposits to their specific production vertex, something which is only possible for charged particles at the moment from the geometric matching of their reconstructed tracks.

4 THEORY TO DATA COMPARISON

By comparing experimental observations to the standard model predictions we can test our understanding of fundamental particle interactions and search for signs of new physics. In order to compare the theory to experiment, it is crucial to have an accurate description of the SM prediction at the level of the experimental observations. The methods that are used to make this comparison are described in this chapter.
Going from the SM theory to a prediction of the detector readout requires the simulation of multiple intermediate steps. This includes the simulation of particle interactions as they occur at the LHC, a simulation of the interaction of the final state particles with the detector and a modeling of the electronic signals. In addition, the signals that are read out from the CMS detector are not directly interpretable. They are processed in reconstruction algorithms to reproduce the final state objects (Section 3.3.1) in the event. At analysis level, these objects can be used to select a region of phase space that is enhanced in a specific interaction process.
The event simulation and processing of the experimental data requires a large computing infrastructure and software framework. Finally, a description is given of the general procedures within the CMS collaboration for processing and analyzing the data, including the statistical analysis.

4.1 EVENT SIMULATION

The standard model gives accurate predictions for particle interactions involving large momentum transfer through perturbative calculations as described in Section 1.6. To compare the SM predictions to the data, we need not only the scattering cross section of the partons, but also a description of what happens before and after the hard interaction. This regime involves lower energies where the strong coupling of QCD becomes large and perturbation theory breaks down. Most phenomena of low energy QCD can not be calculated, and have to be modeled to the data.

A schematic overview of the different stages of a proton-proton collision is shown in Figure 27. As the length scale decreases the strong interaction between particles becomes asymptotically weaker, causing the quarks and gluons inside the proton to behave as approximately free particles. This is a property of QCD known as asymptotic freedom, and most of the proton-proton interactions happen between one constituent of each proton. A hard scattering occurring between two partons is shown in red in Figure 27. The remaining content of the proton is unaffected at the time scale of the hard in-

teraction and the parton remnants (cyan) travel along the beamline without entering the detector. In rare occasions, a secondary interaction (purple) can take place between a second parton from each proton.
High energetic partons reduce their energy by radiating a gluon, known as QCD radiation (blue), or by a gluon splitting into two quarks. Also QED radiation (yellow) occurs through the emission of soft photons. The remaining low energetic partons form hadrons (light green) by the phenomenon of confinement (Section 1.1.1). The heavy hadrons that are formed finally decay to stable hadrons (dark green).

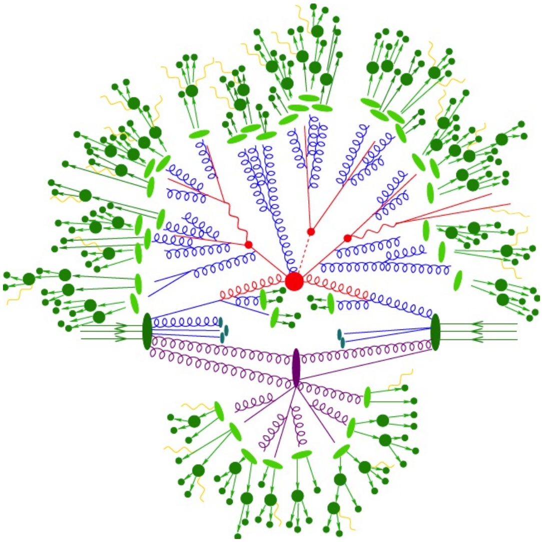


Figure 27: Schematic overview of a simulated $t\bar{t}H$ event [79].

4.1.1 *Collinear Factorization*

Simulation programs rely on factorization to separate the different stages of the event generation. In the calculation of the inclusive cross section σ_n (the cross section to go from pp to the state n described by the outgoing particles after the hard interaction), collinear factorization can be used to separate out the soft, non-perturbative structure of the proton [80]:

$$\sigma_n = \sum_{a,b} \int_0^1 dx_a dx_b \int f_a(x_a, \mu_F) f_b(x_b, \mu_F) \, d\hat{\sigma}_{ab\to n}(\mu_F, \mu_R), \quad (39)$$

with $f_a(x, \mu_F)$ the process-independent parton distribution function (PDF) which gives the probability to find a parton of type a with momentum fraction x_a inside the proton when probing it with an energy scale μ_F, called

the factorization scale. The PDFs can not be calculated perturbatively and are modeled and fitted to data at a fixed scale. Their evolution between different scales is described by evolution equations, known as the DGLAP equations [81–83] in the case of collinear factorization.
The perturbative part of the cross section is given by the parton level cross section:

$$d\hat{\sigma}_{ab\to n} = \int d\Phi_n \, \frac{1}{2\hat{s}} \, |\mathcal{M}_{ab\to n}|^2(\Phi_n; \mu_F, \mu_R), \tag{40}$$

with the squared center-of-mass energy $\hat{s} = x_a x_b s$ equal to the centre-of-mass energy squared s of the pp interaction multiplied by the parton momentum fractions, and $\mathcal{M}_{ab\to n}$ the matrix element corresponding to the sum of Feynman diagrams (Section 1.6). The matrix element depends on the final state phase space Φ_n, the factorization scale (μ_F) and the renormalization scale (μ_R). The factorization scale separates the collinear and soft divergences, while μ_R is introduced to deal with the divergences at high energy, or short distances. The cross section calculated to all orders is independent of these scales. For fixed order calculations, they are usually set to the hard scale in the process and higher order corrections are estimated from varying the scales by a factor 2.

4.1.2 *Parton Showers and Hadronization*

Perturbative calculations can be made for the hard scattering, corresponding to short distance scales, up to the typical scale of hadron formation and decay at the order of 1 GeV. The perturbative calculation is however only feasible up to a limited order in QCD, so that the transition from short to long distances is described by parton shower simulation programs. The initial and final state radiation is simulated using *Monte Carlo* techniques. These techniques rely on the generation of pseudo-random numbers, a deterministically generated set of numbers that exhibit the statistical fluctuations of true random numbers, to obtain numerical results. The intrinsically probabilistic process of particle showers are simulated using the probability for no resolvable QCD radiation to take place when evolving between scales, known as the Sudakov form factor [84]. This method is well suited to describe the radiation of soft and collinear partons and is used to generate the initial and final state radiation.
The formation of hadrons and their decay relies on phenomenological models [85, 86] fitted to data. Also the description of the beam remnants and possible secondary interactions makes use of phenomenological models for their simulation.

4.1.3 *Event generators*

Many event generators have been developed that simulate some or all of the stages of a proton-proton collision. General purpose generators include all of these steps, while specialized generators usually perform one of the tasks e. g. the matrix element generation, or the decay of heavy particles. These are interfaced with general purpose generators to simulate the full event.
Multiple event generators are used in this thesis to simulate events for the signal and background processes. The PYTHIA generator [87, 88] is a general purpose generator that is used to simulate the full event description of several processes. For most processes however, only its showering and hadronization capabilities are used in combination with a dedicated matrix element generator as MG5AMC [34], POWHEG [89], or PHANTOM [90]. MG5AMC is used for producing hard event descriptions at LO and NLO. It contains a high level of automation so that the physics process and theoretic model can be defined by the user. POWHEG is named after the method that is used for interfacing parton shower generators with NLO QCD computations, and can be used for processes that are specifically implemented by the authors. PHANTOM is a fast generator specifically designed for the leading order simulation of six parton final states.

4.2 DETECTOR SIMULATION

The interactions of the outgoing particles with the detector material are simulated with dedicated software. The detector simulation in the CMS collaboration relies on GEANT4 [91] to perform a full simulation of the trajectories and interactions of all particles that enter the detector. The output of this step is a collection of time-stamped energy deposits for every sensor. After applying a digitization step that emulates the electronics response, simulated data is obtained with the same structure as in the real experiment [92].

The detector simulation is a computationally intensive part of the full event simulation chain. In case an optimal accuracy is not required, or to test different detector setups in a reasonable time scale as in Chapter 6, an approximate detector simulation can be employed. The Delphes package [93] is a fast multipurpose detector response simulation framework that reduces the simulation time by a factor of 1000 by using a parametrization for the efficiencies and resolutions. The particle energies are calculated by propagating the stable final state particles and applying efficiency, mistagging factors and resolutions. The values of these quantities can for example be measured from a full Geant-based simulation.

4.3 OBJECT RECONSTRUCTION

The usual approach at hadron colliders has been to reconstruct physics objects by relying only on the most relevant part of the detector for each object, without considering the individual particles that make up the object. Using the full detector information is clearly a more ideal situation and this has proven to be feasible at CMS because of the excellent tracking performance, the ability to resolve the energy deposits of nearby particles and an unambiguous matching of tracker and calorimeter information.

4.3.1 *Particle Flow*

The reconstruction of physics objects (Section 3.3.1) in the events recorded with the CMS detector is based on a particle-flow (PF) algorithm [94]. All the particles that interact with the detector are reconstructed individually by combining information from the different subdetectors. The higher-level objects, such as jets, missing transverse momentum and taus are subsequently build from the reconstructed particles.
The advantages of having access to the individual particles include the ability to remove charged hadrons from pileup, and the use of tracker and ECAL information in the jet energy measurement. The particle-flow approach leads to an improved efficiency and purity of the lepton reconstruction and identification, and to a large improvement in the reconstructed energy and direction of jets, missing transverse momentum and taus.

The general procedure for reconstructing the particle candidates is as follows. In the first step of the algorithm, the hits in the tracker - compatible with a single charged particle - are connected in *tracks*. The same is done in the muon chambers, followed by a clustering of the energy deposits in the calorimeters. These PF elements are then combined in a linking algorithm to generate the list of reconstructed particles.

Tracking and Vertices

The hits in the tracker are connected to form tracks compatible with charged-particle trajectories. The large amount of charged particles, from the multitude of proton proton interactions occuring during a bunch crossing, make it challenging to maintain a high track reconstruction efficiency while sufficiently suppressing the backgrounds. The main background consists of tracks reconstructed from unrelated hits in the tracker, called the combinatorial background. The background from charged particle tracks originating from other interaction vertices is small, as the excellent track position resolution allows to resolve the different vertices in each bunch crossing.
The track reconstruction uses an iterative tracking algorithm, targeting a specific class of tracks with each iteration. The hits making up the reconstructed tracks from each iteration are removed in the subsequent steps, allowing for

a fast track reconstruction while maintaining a high purity. The algorithm is based on Kalman Filtering (KF) [95] and consists of three steps. First, seeds are generated with a high purity. The seed definition is altered with each iteration to reconstruct tracks from different particles and energy ranges. This is followed by a trajectory building step, combining the predicted hit position and the measured ones, and a final fitting to determine the charged particle properties and the fit quality of the track. A dedicated algorithm is applied to identify tracks linked to secondary displaced vertices from nuclear interactions in the tracker material.

The vertices are fitted from the collection of reconstructed tracks, with each vertex corresponding to a pp collision from the same bunch crossing. The primary vertex (PV) is the vertex for which the associated tracks have the largest squared sum of transverse momenta ($\sum_i^{\text{tracks}} p_{T,i}^2$). Secondary vertices are reconstructed to identify the decay of long-lived particles, and particles interacting with the tracker material.

Calorimeter Clusters

The clustering of energy deposits is performed in each subdetector of the calorimeters separately. First, cluster seeds are identified as calorimeter cells containing an energy deposit above a certain threshold and larger than the energy deposited in the 4 or 8 closest cells. From these cluster seeds, topological clusters are created by aggregating nearby cells with large energy deposits. The different energy clusters inside a topological cluster are then reconstructed by fitting a sum of Gaussian energy deposits to the reconstructed energy. The number of Gaussians, their position and amplitude are determined from the fit, while the width of the Gaussians is specific to the calorimeter.

Linking and Reconstruction

The PF elements reconstructed in the tracker, calorimeters and muon chambers are linked together in blocks of nearby elements. A reconstruction of the physics object is then performed for each block. First, the muons are identified and reconstructed, and the PF elements corresponding to the muons are removed from the block. Then, the electrons including the energy of bremsstrahlung photons are reconstructed, as well as the isolated photons.
In the last step, the charged hadrons, neutral hadrons and non-isolated photons are reconstructed. Charged hadrons are identified from matching tracks with clusters in the ECAL and HCAL. Clusters within the tracker acceptance ($|\eta| < 2.5|$) that are not linked to any track, are identified as photons in the case of ECAL clusters and neutral hadrons for HCAL clusters. Outside of the tracker acceptance, neutral and charged hadrons can not be distinguished. ECAL clusters linked to HCAL clusters are reconstructed as hadrons, while ECAL clusters without link are reconstructed as photons.

This procedure is performed on all the blocks to form the list of reconstructed particles in the event.

4.3.2 *Muons*

Muons can be reconstructed from hits in the tracker, in the muon chambers or a combination of the two. The hits in the muon chambers have a high purity as most particles are absorbed in the calorimeters, while the inner tracker hits improve the momentum measurement and allows a better rejection of the backgrounds.
A *standalone muon* track is made up of hits in the muon chambers only, while an *inner track* is build from hits in the tracker detector. By matching a standalone muon track with an inner track, a *global muon* is formed. Multiple scattering in the steel of the return yoke can cause muons with a low momentum to fail the global muons reconstruction. These muons can be recovered by the *tracker muon* reconstruction, which matches an inner track with at least one muon segment (a short track stub made of DT or CSC hits) in the muon detectors [96].

Most global muons are also reconstructed as a tracker muon sharing the same inner track. In this case both are merged into a single muon candidate. The particle-flow muons are selected using the tracker and global muon algorithms, with selection criteria depending on the information from other subdetectors. More stringent identification criteria are imposed on non-isolated muons, having nearby additional tracks or energy depositions in the calorimeters, to suppress the misidentification of charged hadrons as muons.

Identification

Additional identification and isolation requirements can be applied to guarantee a good p_T measurement and to suppress backgrounds from cosmic muons, muons from decays in flight, and hadronic punch-through when hadron shower remnants reach the muon system. The identification consists of requirements on the goodness-of-fit parameter, the number of hits in each subdetector, and the impact parameter (distance between the fitted track and the primary vertex). Multiple working points are defined by the CMS muon physics object group (POG) that can be used by physics analyses.
The tight Id working point, used in Chapter 5, is defined in Table 2. In addition to tight muons, a soft muon selection is used to identify events containing b quarks that is based on the tracker muon with at least one tightly matched segment, combined with requirements optimized for low p_T muons (Table 2). The high purity requirement contains a selection on multiple track parameters to reject the small fraction of bad quality tracks [73] and an arbitration process is applied to assign each segment uniquely to a single tracker muon.

Requirement	Tight	Soft
Global muon	true	-
PF muon	true	-
Tracker muon	true	true
Muon arbitration	true	true
χ^2/dof	<10	-
High-purity track	-	true
Pixel hits	>0	>0
Tracker layers with hits	>5	>5
Muon chamber hits in global track fit	>0	-
Matched muon station segments	>1	>0
d_{xy}(track, vertex) (cm)	<0.2	<0.3
d_z(track, vertex) (cm)	<0.5	<20

Table 2: Muon identification requirements.

Muons that are produced inside jets are rejected by requiring that the muon is isolated, i. e. the energy in a cone around the muon is smaller than a certain fraction of the energy of the muon. When using only the tracker information, the energy around the muon is estimated from summing the p_T of tracks from the primary vertex (PV) inside a cone of $\Delta R < 0.3$ around the muons direction, with $(\Delta R)^2 = (\Delta\eta)^2 + (\Delta\phi)^2$. The relative, tracker-based isolation ($\mathrm{Iso}^{\mu}_{\text{track, rel}}$) is defined as:

$$\mathrm{Iso}^{\mu}_{\text{track, rel.}} = \frac{\sum_i^{\text{tracks from PV}} p_{T,i}}{p_T^{\mu}} \,. \tag{41}$$

This quantity is similar to the isolation that is calculated by the high level trigger.

Using particle-flow candidates, a more robust isolation variable can be constructed. The sum over p_T in a cone of $\Delta R = 0.4$ is computed separately for the charged hadrons, neutral hadrons and photon candidates:

$$\mathrm{Iso}_{\,\mathrm{PF}} = \sum p_T^{\text{charged had}} + \max\left(0, \sum p_T^{\text{neutral had}} + \sum p_T^{\gamma} - \mathrm{Iso}^{\mathrm{PU}}\right) \,. \tag{42}$$

The last term $\mathrm{Iso}^{\mathrm{PU}}$ corrects for the PU contribution in the neutral components, as the production vertex of neutral particles can not be determined. For muons, the $\Delta\beta$ correction is used, which subtracts half the sum of the p_T of the charged particles not originating from the primary vertex in the cone:

$$\mathrm{Iso}^{\mathrm{PU}}_{\Delta\beta} = 0.5 \cdot \sum p_T^{\text{charged PU}} \,. \tag{43}$$

The factor 0.5 corresponds to the average ratio of the neutral to charged particle energy in jets.
The recommended tight, PF isolation cut is:

$$\frac{\mathrm{Iso}_{\,\mathrm{PF},\,\Delta\beta}}{p_T^{\mu}} < 0.15\,. \tag{44}$$

Selecting muons with a high purity by appylying the tight identification and isolation requirements gives a selection efficiency that is larger than 85% for muons with $p_T > 20\,\mathrm{GeV}$ and larger than 98% for muons with $p_T > 40\,\mathrm{GeV}$.

4.3.3 *Electrons*

The reconstruction of electrons is based on associating a track with a cluster of energy in the ECAL. This is complicated by bremsstrahlung, which can cause electrons to lose a significant fraction of their energy while passing through the tracker. The average energy fraction the electron radiates before it reaches the ECAL is between 33% and 86% (Figure 22), depending on its pseudorapidity. The track reconstruction is further complicated when the radiated photons in turn lead to pair production ($\gamma \rightarrow e^+e^-$). To accurately reconstruct the electron energy, one needs to measure both the initial electron and the radiated photons. Dedicated algorithms are used to reconstruct the tracker hits and energy deposits of the electrons [97].

Two approaches are followed to generate the seeds in the electron reconstruction. The *ECAL-based* approach uses the energy and position of ECAL clusters to infer the expected hits in the innermost tracker layers, which works well for high p_T and isolated electrons. The *tracker-based* reconstruction starts from the reconstructed tracks and is more suited for low p_T electrons and electrons inside jets. Seeds from both procedures are merged and used in a dedicated electron tracking algorithm.

The large energy loss causes the electron to follow an unusual trajectory with non-constant curvature in the magnetic field. The usual fitting method based on Kalman filtering does not perform well for electrons with large radiative losses as hits are lost when the change is curvature is large, leading to poorly estimated track parameters. To solve this, a more computationally intensive Gaussian Sum Filter (GSF) method [98] is applied which approximates the energy loss in each layer by a mixture of Gaussian distributions. To include the energy deposited in the calorimeters by bremsstrahlung photons, *superclusters* (SC) are formed by extending the ECAL clusters in ϕ, corresponding to the curvature of the electron in the magnetic field.
Finally, the electron candidates in a PF block are generated from matching the GSF tracks, ECAL supercluster and ECAL clusters linked to the GSF track tangents.

Identification

Additional identification criteria are imposed to reduce backgrounds from photon conversions in the tracker material, jets misidentified as electrons and electrons from heavy quark decays. The variables that provide discriminating power quantify the agreement between the measurements in the ECAL and the tracker, the shape of the energy deposit, and the track quality.
The tight and HLT-safe cut-based working points, as recommended by the corresponding CMS POG, are summarized in Table 3. The tight working point is designed to have an average efficiency of about 70%.

Requirement		Barrel $\lvert\eta_{SC}\rvert \leqslant 1.479$		Endcap $1.479 < \lvert\eta_{SC}\rvert < 2.5$	
		HLT-safe	*Tight*	*HLT-safe*	*Tight*
$\Delta\eta$ (SC, track)	$<$	0.004	0.00308	-	0.00605
$\Delta\phi$ (SC, track)	$<$	0.020	0.0816	-	0.0394
$\lvert 1/E_{SC} - 1/p_{track}\rvert$	$<$	0.013	0.0129	0.013	0.0129
$\sigma_{\eta\eta}$	$<$	0.011	0.00998	0.031	0.0292
H/E_{SC}	$<$	0.060	0.0414	0.065	0.0641
Tracker Isolation	$<$	0.08	-	0.08	-
ECAL PF Cluster Isolation	$<$	0.160	-	0.120	-
HCAL PF Cluster Isolation	$<$	0.120	-	0.120	-
I^{e}_{rel}	$<$	-	0.0588	-	0.0571
Missing hits	$\leqslant$	-	1	-	1
Pass conversion veto		-	yes	-	yes
Gsf Track χ^2/dof	$<$	-	-	3.0	-
d_{xy}(track, vertex) (cm)	$<$	-	0.05	-	0.10
d_z(track, vertex) (cm)	$<$	-	0.10	-	0.20

Table 3: Electron identification requirements.

The shape of the energy deposits in the calorimeter is a useful measure for rejecting jets, as hadronic showers are expected to be wider and longer than electromagnetic showers. The shower shape is quantified by the lateral extension of the shower in the η direction ($\sigma_{\eta\eta}$), and the ratio of the energy deposited in the HCAL to the ECAL supercluster (H/E_{SC}).
Also the isolation variable is useful for rejecting electrons from jets. The PF isolation is defined by Equation 42. For electrons, the neutral component is corrected by the median energy density in the central region of the detector (ρ), multiplied with a η-dependent effective area (A_{eff}) correction to make it independent of pileup:

$$\mathrm{Iso}^{\mathrm{PU}}_{\mathrm{eff.\ Area}} = \rho \cdot A_{\mathrm{eff}}\,. \tag{45}$$

To reject photons converted in the tracker material, additional requirements can be placed on the reconstructed tracks. For converted photons the first hit of the electron track is often not located in the innermost layer of the tracker, resulting in missing hits in the reconstructed track. Another variable that is calculated for each electron is the *conversion veto*. It is set to *true* when track pairs are found that have a good fit result to a common vertex consistent with a converted photon.

Charge Identification

Next to backgrounds from other particles, also a misreconstructed charge can introduce backgrounds at the analysis level when selection two leptons. The charge measurement is affected by bremsstrahlung photons that convert to an electron positron pair. Hits from these particles can be wrongly included in the electron track fit.
The electron charge is estimated from three measurements: the sign of the GSF track curvature, the sign of the associated KF track curvature and the sign of the difference in ϕ between the supercluster and the first hit of the GSF track. The charge of the electron is then defined as the sign measured by at least two of the three methods. To increase the purity, the agreement of all three methods can be required.

4.3.4 *Jets*

Jet Clustering

Jets are reconstructed by clustering nearby PF particles. Multiple jet algorithms exist, mainly divided in two categories: cone algorithms and sequential recombination algorithms [99]. The cone algorithms use a distance in angle to define the jet, leading to cone-shaped clusters of particles. The sequential recombination algorithms are slightly more involved; they add particles to the jet one-by-one and make use of a distance in angle and energy or momentum space. These algorithms are by consequence not as simple and fast as cone algorithms, but are more similar in structure to the parton shower development.

Particle-flow jets are reconstructed by clustering PF particles with a sequential recombination algorithm called the anti-k_t algorithm. It uses as distance measure:

$$d_{ij} = \min(p_{T,i}^{-2}, p_{T,j}^{-2}) \frac{(\Delta R_{ij})^2}{R^2}, \qquad (\Delta R_{ij})^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2, \tag{46}$$

with y_i and ϕ_i the rapidity[1] and azimuth of particle i, and the radius parameter R set to 0.4.
This algorithm clusters jets around the hard particles, leading to circular-shaped jets. An important property is that the algorithm is infrared and collinear safe, meaning one finds the same jets when a collinear splitting or soft radiation is added to the event.

Jet Energy Reconstruction

About 65% of the total jet energy is reconstructed from charged hadrons, which can be accurately measured from the tracker hits. Another 25% of the energy is carried by photons, the decay products of the short-lived neutral pions, and can be measured with the ECAL with a resolution of a few percent [100]. The remaining 10% of the jet energy is carried by neutral hadrons and is measured with a lower energy resolution at the order of 10% in the HCAL.

The measured energy of a jet is biased by the particles from pileup in the jet cone, as well as by the non-linear detector response. Corrections are performed to recover the true jet energy from the measured energy.
The out-of-time pileup is reduced by using the signal timing and pulse shape to fit the in-time and out-of-time pulses simultaneously [101]. About 50% of the in-time pileup in the tracker acceptance is removed by rejecting PF candidates associated with pileup vertices, with an algorithm called charged hadron subtraction (CHS) [102].

Jet Energy Corrections

Jet energy corrections (JEC) are applied to correct for the remaining energy bias. This is done with a factorized approach, where each step corrects for a different effect by rescaling the four-vector of the jets. This rescaled four-vector is then used as input for the next step.
In the first step, the remaining energy from out-of-time pileup and neutral particles from in-time pileup is subtracted. This is estimated per event from the median energy density in simulation [103], calibrated with zero-bias data (data triggered from the bunch crossing time). On top of this pileup offset correction, the non-linear response is corrected from simulation as a function of η and p_T. Small residual corrections are calculated from data-based methods. An overview of the JECs that are applied is shown in Figure 28.

1 The rapidity y is defined as $y \equiv \frac{1}{2} \ln \frac{E+p_z c}{E-p_z c}$. It is invariant under boosts along the z-axis, and is equal to the pseudorapidity for massless particles.

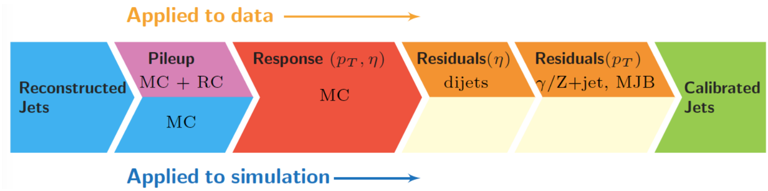


Figure 28: Overview of the jet energy corrections that are applied to simulation and data.

Jet Energy Resolution

To correct for the difference in jet energy resolution between data and simulation, the four-momentum of the reconstructed jets in simulation are rescaled by:

$$1 + (s_{\mathrm{JER}} - 1)\frac{p_T - p_T^{\mathrm{gen}}}{p_T}, \tag{47}$$

with p_T the transverse momentum of the reconstructed jet, p_T^{gen} the transverse momentum of the geometrically matched generator-level jet, and s_{JER} the data-to-simulation resolution scale factor.

Identification

A jet identification is applied to reject noise-contaminated jets originating from calorimeter or readout electronics noise. The loose working point used in Chapter 5 is almost 100% efficient for real jets.

b-tagging

Jets originating from b quarks - also called b-jets - can be distinguished from jets produced by the hadronization of light quarks and gluons by making use of the characteristic properties of b hadrons. Because of the relatively long life-time of b-hadrons, the decay vertex will be displaced compared to the primary vertex, typically by a few millimeters. This secondary vertex has a high mass and high track multiplicity. Another property that can be used is the appearance of leptons in semi-leptonic b decays. An electron or muons is produced in about 40% of the b-jets.
The b-tagging algorithm used in the second part of this thesis is the Combined Secondary Vertex v2 (CSVv2) algorithm [104]. It uses a discriminator from an artificial neural network trained on several variables related to the tracks and secondary vertices.

4.3.5 *Hadronic* τ

The τ is the only lepton heavy enough to decay into hadrons. τ leptons decay hadronically in about $2/3$ of the cases, and are identified and reconstructed by the hadrons-plus-strips (HPS) algorithm [105]. In a first step, the τ candidates are constructed by identifying the particles from one of the main τ decay channels inside a PF jet. Then discriminators are used to separate the τ from gluon and quarks jets, and from electrons and muons. The isolation of the τ is the main handle for reducing the background from jets.

4.3.6 E_T^{miss}

The missing transverse momentum ($\vec{p}_T^{\,\mathrm{miss}}$) is defined as the negative vector sum of of all reconstructed particle-flow particles:

$$\vec{p}_T^{\,\mathrm{miss}}(\mathrm{raw}) = - \sum_{i}^{\text{PF particles}} \vec{p}_{T,i}\,. \tag{48}$$

The jet p_T is usually replaced by the corrected jet p_T (called the type-I MET correction):

$$\vec{p}_T^{\,\mathrm{miss}} = - \sum_{i}^{\text{unclustered PF}} \vec{p}_{T,i} - \sum_{i}^{\text{PF jets}} \vec{p}_{T,i}^{\,\mathrm{JEC}}\,, \tag{49}$$

where the sum is split in particles that are clustered in jets, for which the corrected value p_T^{JEC} is used, and the remaining unclustered PF particles. The size of the missing transverse momentum $\vec{p}_T^{\,\mathrm{miss}}$ is referred to as p_T^{miss} or the missing transverse energy E_T^{miss}.
A post-processing is performed to remove events with an artificially high p_T^{miss}, usually caused by the misidentification or misreconstruction of a muon.

4.4 DATA PROCESSING

In order to process and store the large amount of simulated and experimental data at the LHC, a large computing and data storage infrastructure is required. For this purpose, the LHC Computing Grid has been created, consisting of tens of thousands of PCs distributed in about 50 cluster over the world. The software that is run on the grid is based on the CMSSW framework [106] consisting of many classes written in C++ and python. For statistical analysis and visualization it relies on the ROOT package [107].
The processing of the data is performed in multiple steps, divided over the computing centers. The final datasets that are used for analysis are reduced in size, containing the basic reconstructed objects and event information that is commonly used in analyses. For the analysis in Chapter 5, the total size of the experimental dataset that is processed is about 70 TB.

The quality of the raw data that is recorded with the CMS detector is monitored both during data-taking and offline after the events are reconstructed. A list of runs with high quality data, passing stringent quality requirements, is then passed on to the physics groups.
The datasets are analyzed by Physics Object Groups (POG) to determine optimal reconstruction settings for each physics object and corrections for differences between simulation and data. The extraction of physics results from the data is performed by the Physics Analysis Groups (PAG). It is the responsibility of the analyzers to skim the datasets for information relevant to the specific analysis, and apply the latest reconstruction settings and corrections recommended by the POGs. This first step is performed over the LHC computing grid, while further analysis of the data is usually carried out on a local computer cluster.

4.5 ANALYSIS WORKFLOW

The main components that are needed to compare the experimental data to theory were introduced. The general structure of a physics analysis with data from the CMS detector is as follows:

1. Data and simulation:
   A set of triggers is defined to select the signal events from the pp collisions at the CMS detector. At the same time, the theory prediction is generated in the form of Monte Carlo simulation samples for the signal and background processes that might be present in the signal region. After the data is collected, the pileup distribution and effect of the trigger selection in the simulated samples are matched to the conditions during the data-taking.

2. Objects:
   The reconstruction algorithms and corrections to the physics objects are developed within the respective POGs. Taking into account their recommendations, the optimal selection criteria of each object are defined for the specific analysis. Corrections are applied for the differences in selection efficiency, energy scale and resolution between data and simulation for the various objects.

3. Event selection:
   A phase space is defined that optimizes the significance of the measurement. This can be done by placing selections on the kinematic properties of the physics objects, or on discriminators from supervised machine learning techniques.

4. Background estimation:
   The background processes that are present after the event selection are estimated from the simulation, the data, or a combination of the two. The background estimates are then compared to data in *control regions*,

by selecting events nearby but orthogonal to the event selection, to assess their validity.

5. Statistical analysis:
   The goal of the statistical procedure is to make a quantitative statement on how well certain theoretical models agree with the data, taking into account the statistical and systematic uncertainties. In the simplest case, it is calculated from the event yields of the data and the expected signal and background after the event selection, together with their uncertainties.

4.6 STATISTICAL PROCEDURE

The statistical analysis of the CMS data makes use of a method of statistical inference called hypothesis testing [108]. The method will be discussed here for the simplest case of a signal process without free parameters.
The first step consist in defining two hypotheses: a null or background hypothesis and an alternative or background plus signal hypothesis. The null hypothesis is usually the SM without the signal process, while the alternative model includes both. The expected yield can be expressed as:

$$\mu \cdot s + b\,, \tag{50}$$

with s, b the expected signal and background yields and μ the signal strength modifier. For $\mu = 0$ the expected yield reduces to the expectation under the null hypothesis and $\mu = 1$ gives the expectated yield under the alternative model.
When working with histograms, the expected yield in each bin is distributed according to a Poisson distribution. The total probability to observe n_i events for each bin i of a histogram is the product of those Poisson probabilities:

$$\text{Poisson}(n_i \mid \mu \cdot s + b) = \prod_i \frac{(\mu \cdot s_i + b_i)^{n_i}}{n_i!} e^{-\mu \cdot s_i - b_i}\,. \tag{51}$$

In general, s and b depend on nuisance parameters θ that represent the systematic uncertainties. The degree of belief on what the true value of θ might be, is described by the systematic error probability density function (pdf) $\rho(\theta|\widetilde{\theta})$, with $\widetilde{\theta}$ the default value of the nuisance parameter.
For an observed dataset, n_i are fixed and the probability can be interpreted as a likelihood function with μ and θ as variables. The systematic uncertainty pdfs $\rho(\theta|\widetilde{\theta})$ are re-interpreted as posteriors of some real or imaginary measurement $\widetilde{\theta}$. The pdf for the auxiliary measurement $p(\widetilde{\theta}|\theta)$ can be used to constrain the likelihood, allowing for a purely frequentist treatment of the uncertainties:

$$\mathcal{L}(\text{data}|\mu, \theta) = \text{Poisson}(\text{data}|\mu \cdot s(\theta) + b(\theta)) \cdot p(\widetilde{\theta}|\theta)\,. \tag{52}$$

The next step consist in defining a test statistic that can be used to compare the compatibility of the data with the hypotheses. The test statistic q_μ recommended by the LHC Higgs combination group is based on the profile likelihood ratio [109]:

$$q_\mu = -2\log\frac{\mathcal{L}(\text{data}|\mu,\hat{\theta}_\mu)}{\mathcal{L}(\text{data}|\hat{\mu},\hat{\theta})}, \tag{53}$$

with $0 \leqslant \hat{\mu}$ to require a positive signal and $\hat{\mu} \leqslant \mu$ as additional condition when setting limits. The denominator contains the maximum likelihood with parameter values $\hat{\mu}$ and $\hat{\theta}$, while in the numerator the likelihood is profiled, with $\hat{\theta}_\mu$ the fitted value of θ that maximizes the likelihood for a fixed value of μ.
The pdfs of the test statistic under the alternative $f(q_\mu \mid \mu, \hat{\theta}^{obs}_\mu)$ and null $f(q_\mu \mid 0, \hat{\theta}^{obs}_\mu)$ hypothesis can then be constructed. A hypothesis can be rejected at a pre-specified confidence level by placing a cut on the test statistic. The confidence level defines how likely it is to make a false exclusion.

Excluding the background hypothesis

The confidence level that is required to claim a discovery in particle physics is by convention equal to $1 - 2.87 \cdot 10^{-7}$, meaning that the probability to make a false discovery is $2.87 \cdot 10^{-7}$. This probability corresponds to the probability for an upward fluctuation of 5σ in a Gaussian distributed variable.
The measurement is characterized by the p-value which is the probability that, assuming the null hypothesis is true, the test statistic has a magnitude at least as large as the observed result:

$$p_0 = \int_{q^{obs}_\mu}^{\inf} f(q_\mu \mid 0, \hat{\theta}^{obs}_\mu). \tag{54}$$

The p-value can also be written as a significance, equal to the number of standard deviations that corresponds to an upward fluctuation of a Gaussian distributed variable with probability p_0. A discovery can be claimed when the observed significance is larger than 5σ.

Excluding the alternative hypothesis

A similar procedure is used to exclude an alternative hypothesis. The confidence level that is generally used to place limits on μ is 95%, corresponding to a probability of wrongly excluding the alternative hypothesis of 5%. Instead of requiring this confidence level on the test statistic directly as is done for excluding the background hypothesis, both the p-value under the alternative and null hypothesis are used to construct the quantity CL_S:

$$CL_S = \frac{p_\mu}{1-p_0}. \tag{55}$$

This is known as the modified frequentist method. The denominator protects against setting too strong limits caused by downward fluctuations in the background.

Asymptotic limit

Generating the pdfs of the test statistic under the hypotheses is often impractical in terms of the computing power that is required. In this case, one can rely on the asymptotic behavior of the test statistic to obtain a fair estimate for the p-values. From Wilks theorem [110], one finds that the test statistic q_μ is described by a χ^2 distribution in the asymptotic regime. In practice, this approximation has been shown to work very well, even for cases with few expected events.
The statistical results that are derived in this thesis rely on the asymptotic properties of the test statistic. They are calculated using the RooFit [111] and RooStats [112] packages with a statistics tool (Combine) developed by the CMS Higgs physics analysis group.

Part II

DATA ANALYSIS AND FUTURE EXPECTATIONS

Part II of this thesis explores vector boson scattering with the CMS detector. The first destination on this path is the measurement of a process dominated by VBS. As more data is collected, new objectives come within reach; signs of new physics at a high mass scale can be exposed by performing precision measurements of the quartic couplings, while the role of the Higgs boson in the VBS unitarization can be revealed by measuring the longitudinal scattering component.

An overview is presented of the analysis of the electroweak $W^{\pm}W^{\pm}jj$ production in the leptonic decay channel, leading to its discovery in the 35.9 fb^{-1} of data recorded with the CMS detector at the LHC during 2016. This is followed by a feasibility study of VBS with 3000 fb^{-1} of data collected with the upgraded CMS detector at the high luminosity LHC.

5 MEASUREMENT OF THE ELECTROWEAK $W^\pm W^\pm jj$ PRODUCTION

Measuring the properties of vector boson scattering is one of the primary goals of the LHC physics program. Before the discovery of the Higgs boson, the increase in the longitudinally-polarized VBS amplitude with center-of-mass energy guaranteed new physics to enter at the TeV scale (Section 2.3). Now the Higgs boson has been discovered, the VBS process remains interesting for its sensitivity to modifications in the Higgs sector and to additional resonances (Section 2.4).
VBS searches at the LHC are performed by the CMS and ATLAS experiments, for multiple combinations of vector bosons at center-of-mass energies of 8 or 13 TeV. Public results are available for the $W^\pm W^\pm$ [113, 114], WZ [115], ZZ [116], $W\gamma$ [117], $Z\gamma$ [118, 119] and $\gamma\gamma \to W^+W^-$ [120, 121] channels. All of these channels have been measured in the leptonic final state, while the WV channel (with V either a W or Z boson) has also been studied in the semi-leptonic final state [122] for its high sensitivity to the anomalous quartic gauge couplings.

In this chapter, the analysis of the electroweak production of same-sign W boson pairs in the two jet and two lepton final state is described, in data recorded with the CMS detector during 2016 at a center-of-mass energy of 13 TeV and with an integrated luminosity of 35.9 fb^{-1}. The analysis lead to the first observation of a VBS-dominated process [4], marking an important step in determining the properties of vector boson scattering. In addition some of the most stringent constraints on new physics models are set through limits on the dimension eight EFT operator couplings.

5.1 ANALYSIS OVERVIEW

The scattering of same-sign W bosons that decay to leptons is the VBS process that can be measured with the highest sensitivity at the LHC (Section 2.1.2). It has the smallest relative strong production background because the two same sign W bosons can only be produced in combination with two quarks at the LHC, and the presence of the two same-sign leptons in the final state strongly reduces the backgrounds.
Nevertheless, the small cross section of the process (Section 3.3) in combination with the backgrounds from strong diboson production and from misreconstructed jets make this a challenging analysis.

A detailed description of the analysis is given in the next sections following the usual structure of a physics analysis with data from the CMS detector (Section 4.5). The objective of the analysis is twofold. The main result is the measurement of the electroweak $W^{\pm}W^{\pm}jj$ process and its significance. In addition, limits are placed on beyond the standard model theories in a general effective field theory approach (Section 2.4.1.2). This covers theories in which the new physics is characterized by a high energy scale, which manifests itself at lower energies in deviations to the quartic gauge couplings as predicted by the standard model.

5.1.1 *Signal and Backgrounds*

The signal is defined as the electroweak production of same-sign W boson pairs in the two jet and two same-sign lepton final state. It is measured in a phase space optimized for the VBS component (Figure 4), but also contains contributions from other electroweak production diagrams (Figure 5). Non-resonant and triboson electroweak diagrams (Figure 6) are negligible after the event selection.

The background events can be split in two categories:

- *Irreducible backgrounds* have the same final state particles as the signal process and can not be distinguished from VBS on an event-by-event basis. In this case, only kinematic differences can be used to make a statistical separation.
- *Reducible backgrounds* are processes with a final state different from VBS, but can enter the signal region when objects are not correctly reconstructed (e.g. a jet that is misreconstructed as a lepton) or are not within the detector acceptance. These backgrounds would not be present in the case of a perfect detector and can be reduced by improving the detector, reconstruction algorithms and object selections.

The irreducible background in this analysis is the strong same-sign W boson production (Section 2.1.2). After the event selection the expected number of events from the strong production is only 5% of the signal yield. The interference between the electroweak and strong production is about 3% of the expected signal yield after the event selection and is treated as a systematic uncertainty on the signal.

The main reducible backgrounds, in order of decreasing expected event yield after the event selection, are:

- Backgrounds where one of the reconstructed leptons is a misreconstructed jet, called the *non-prompt* background. The dominant contributions to this background originate from $t\bar{t}$ and W+jets events.

Background	Relative contribution
Non-prompt	60%
WZ	22%
$V\gamma$	7%
charge misassignment	4%
strong $W^{\pm}W^{\pm}jj$	3%
VVV	2%
ZZ, WW DPS	$< 1\%$

Table 4: The relative contribution of each background to the total background yield after the event selection. V indicates a weak vector boson and can be either a W boson or a Z boson.

- WZ production in the fully leptonic decay channel with one of the 3 final state leptons outside of the tracker acceptance or not passing the lepton selection. This contains contributions from the electroweak and strong production. Unlike the signal, this process can also be produced from diagrams without two quarks in the initial and final state.
- $W\gamma$ production where the photon is converted to an electron pair in the detector material and reconstructed as a single electron.
- Backgrounds containing an electron with a misidentified charge, e. g. from Drell-Yan (DY, also called Z+jets) events.

The non-prompt, WZ and charge misassignment backgrounds are estimated using both data and simulation, while the other backgrounds and the signal are estimated from simulation only.
Other backgrounds that are considered are the production of three vector bosons (VVV), ZZ, $Z\gamma$, and double parton scattering (DPS) when two independent hard interactions take place between different partons within one proton-proton collision, each producing a single W boson.
The relative contribution of each background to the total expected background yield after the event selection is shown in Table 4. Backgrounds with a Z boson decaying to leptons also includes the contribution where the Z boson is replaced by an off-shell photon (γ^*).

5.1.2 *Analysis strategy*

The VBS topology is characterized by two forward, highly energetic quarks which are measured as jets in the detector (Section 2.2). These jets are called the *tag jets* and have a large transverse momentum (p_T) compared to the jets originating from pile-up. They are identified as the two jets with the highest p_T, which selects the correct jets in more than 90% of the signal events.

In addition, two same-sign leptons (combinations of electrons and muons) are selected, which can either be directly produced from a W boson decay, or from a W boson with an intermediate τ decay. The leptons are generally produced more centrally in the detector than the jets (Section 2.2.1).

The event selection has been optimized starting from the Run I analysis at $\sqrt{s} = 8$ TeV with the CMS detector [113]. The optimization is performed by scanning over several variables and cut values to measure their effect on the expected signal significance. During this procedure all processes were estimated from simulation only, to avoid sensitivity to statistical fluctuations in the data.
The final event selection is enriched in VBS events and suppresses irreducible backgrounds by selections on the invariant mass and pseudorapidity difference of the jet pair, and by requiring the leptons to be in between the jets in pseudorapidity. To reduce the remaining backgrounds, additional event selections are applied, and the lepton selections are optimized to suppress leptons from misidentified jets, charge misassignment and misidentified photons.

In contrast to what could be assumed from the relative background contributions (Table 4), reducing the non-prompt background further does not significantly improve the signal significance. The sensitivity of the analysis is dominated by the high m_{jj} (> 1.5 TeV) region. Although the non-prompt background is still substantial after the event selection, it becomes very small at high m_{jj}, and it is more important to have a precise description of this background than to reduce it further. The VBS process falls off slower with m_{jj} and the main background in this region comes from WZ events with a lepton that does not pass the reconstruction requirements or is outside of the detector acceptance. This leads to a region with a small amount of background that is very pure in VBS events, but also has a limited number of signal events.

5.2 DATA AND SIMULATION SAMPLES

5.2.1 *Data*

In this analysis, data recorded with the CMS detector during 2016 is used. It consists of proton proton collisions with a center-of-mass energy of 13 TeV and a total integrated luminosity of 35.9 $\text{fb}^{-1} \pm 0.9\text{fb}^{-1}$. The events that pass the trigger selections (Section 3.4.6) are stored in *Primary Datatsets* (PD) that are produced centrally in the CMS collaboration (Section 4.4). The primary datasets are based on a list of similar triggers and contain the events that passes at least one of the triggers in this list. The datasets that are used in this analysis are the double lepton PDs (DoubleMuon, DoubleElectron, MuonEG) and the single lepton PDs (SingleMuon, SingleElectron) to recover

Primary Dataset	Eras	HLT Path
MuonEG	B-F	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_v*
		HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_v*
	F-H	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v*
		HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v*
DoubleMuon	B-H	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_v*
		HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ_v*
	H	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_v*
		HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ_v*
DoubleEG	B-H	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v*
SingleMuon	B-H	HLT_IsoTkMu24_v*
	B-H	HLT_IsoMu24_v*
SingleElectron	B-H	HLT_Ele27_WPTight_Gsf_v*
	B-H	HLT_Ele25_eta2p1_WPTight_Gsf_v*

Table 5: The HLT paths used in the analysis.

some of the dilepton events that did not pass any of the double lepton triggers. The PDs that include the latest available calibrations are used, and the data is required to pass the data quality tests of the physics validation group.

At the beginning of the 2016 data taking, a bug in the L1 muon endcap trigger (EMTF) lead to an inefficiency for events that have more than one muon track in the same 60-degree sector of the endcap. To deal with this small inefficiency, events with more than one muon in the same endcap sector are vetoed both in data and simulation for the affected luminosity fraction. The effect on the signal sample is found to be smaller than 1%.

Triggers

The data taking period is split in different eras, indicated by a letter from A to H. The High Level Triggers (HLT) used in the analysis are summarized in Table 5, showing also in which era (B until H) each trigger was deployed. The trigger names indicate what selections the leptons passed at HLT level, with the numbers showing the p_T cuts (in GeV) that are applied. The trigger efficiency of the signal after the event selection is $\sim 99.9\%$, where about 4% is recovered by including the single lepton triggers. The estimation of the non-prompt background makes use of the additional triggers in Table 6.

Primary Datatset	HLT path
SingleMuon	HLT_Mu8_TrkIsoVVL_v*
	HLT_Mu17_TrkIsoVVL_v*
SingleElectron	HLT_Ele8_CaloIdM_TrackIdM_PFJet30_v*
	HLT_Ele12_CaloIdL_TrackIdL_IsoVL_PFJet30_v*
	HLT_Ele17_CaloIdL_TrackIdL_IsoVL_PFJet30_v*

Table 6: The HLT paths used for the non-prompt background estimation.

5.2.2 *Simulation*

Multiple event generators (Section 4.1.3) are used to simulate events for the signal and background processes. For some processes PYTHIA v8 [88] is used to generate the full event, while others rely on MG5AMC [34] or POWHEG [89] to generate the hard process and are interfaced with PYTHIA to obtain the complete event description with parton showering and hadronization. The detector simulation is performed with GEANT4 [91] as part of the usual CMS workflow (Section 4.2).

The samples for the electroweak and strong production of the same-sign W boson pairs were generated specifically for this analysis. They are generated in the central event production on the LHC Computing Grid, after having the configuration files approved within the CMS collaboration, and use MG5AMC at LO interfaced to PYTHIA v8. The same generator and procedure is used to generate the aQGC samples (Section 2.4.1.2) for a range of non-zero values of the dimension eight Wilson coefficients. It uses a reweighting of the events for multiple points in the Wilson coefficients space to generate the events in a reasonable amount of time. The interference between the electroweak and strong production is estimated with the PHANTOM generator [90].
MG5AMC at LO is also used to produce samples for WZ in multiple jet bins and for $W\gamma$ with two jets. It is used at NLO for the $Z\gamma$ and DY samples.
POWHEG at NLO is used in the production of the ZZ sample and the top samples produced for the estimation of the charge misassignment background. The VVV and DPS samples are generated with PYTHIA v8 at NLO and LO respectively.

Data corrections

The simulated samples are generated with a distribution for the number of pileup interactions that is meant to roughly cover the conditions expected for the different data-taking periods. To correct for the exact pileup conditions after the data is taken, the number of pileup interactions from the simulation is reweighted to match the data.

The effect of the trigger selection on data is reproduced in the simulation samples by applying a scale factor. This scale factor estimates the probability for an event to pass the trigger selection and is calculated from the data as a function of the p_T and η of the two leptons.
The calculation makes use of a *tag and probe* method on pairs of electrons and muons from Z boson decays. The leptons are required to pass the tight identification and isolation selections to measure the trigger efficiency independent of the lepton selections. By selecting a *tag* lepton passing the trigger requirements together with a *probe* lepton such that they together have an invariant mass consistent with the mass of the Z boson, one obtains a clean sample of real, *probe* leptons of a specific flavour. With this method, a collection of *probe* leptons is obtained which can then be used to measure the trigger efficiency in an unbiased way as a function of the p_T and η of the lepton.
The purity of the sample of dilepton events is increased further by performing a fit on the invariant mass distribution to subtract the contribution of non-resonant dilepton pairs. An example of this fit to the collection of probe leptons that fail the selection, and to the probe leptons that pass the selection of which one wants to measure the efficiency is shown in Figure 29.

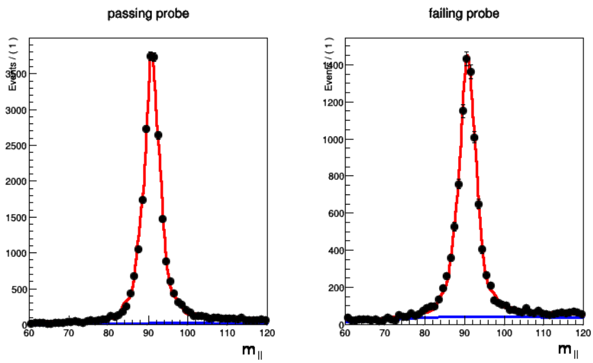


Figure 29: Fit of the m_{ll} distributions to extract the real leptons from the probe lepton collection as part of the tag and probe method for one bin in η and p_T. The real lepton contribution originating from Z boson decays are fitted (in red) together with the background from non-resonant dilepton pairs (in blue).

This procedure is used to calculate the trigger efficiency for each trigger leg in the double and single lepton triggers, from which the efficiency to pass at least one of the triggers in the trigger selection is calculated. Some of the dilepton triggers that are used in the analysis (Table 5) have a selection on

DZ: the distance along the beamline between the track and primary vertex. The efficiency of the DZ selection is measured separately, on events passing the dilepton triggers with the same trigger legs but without the DZ selection. The efficiency is between 95% and 99% depending on the dataset and is applied in the simulation of the relevant trigger paths.

5.3 OBJECT DEFINITIONS

The object definitions that are used in the analysis are summarized here, while a detailed description of the reconstruction and identification algorithms is given in Section 4.3.

5.3.1 *Leptons*

Multiple lepton selections are used to identify electrons and muons:

- Tight lepton:
  Used to select events with two leptons.
- Loose lepton:
  To veto events with extra leptons and used in the non-prompt background estimation.
- Soft muon:
  Used to tag b-jets.

The lepton selections follow the *Physics Object Groups* (POG, Section 4.4) recommendations. Only for the tight electron selection (Table 3) two additional cuts are applied to reduce the backgrounds. The tight electron is required to correspond to a track that is reconstructed without missing hits and the three charge measurements that are performed on the electron need to have the same result, described in more detail in Section 4.3.3.
These cuts reduce the backgrounds from photon conversion by a factor 5, backgrounds from charge misassignment by a factor 20 and backgrounds containing misreconstructed jets with 30%. The lepton selections are summarized in Table 7.

Muons are identified according to the POG recommendations (Table 2) for the tight working point with tighter cuts on impact parameters d_{xy} and d_z to suppress non-prompt muons. The muon isolation criteria follow the recommended tight PF-based combined relative isolation with $\Delta\beta$ correction in a cone with a size of 0.4 in ΔR. The electron ID and isolation is based on the recommended tight working point with additional HLT-safe cuts to make the selection independent of the trigger selection.
The loose lepton selections are optimized for the non-prompt background estimation. It needs to be loose enough to have enough events at high m_{jj} that can be used to estimate the non-prompt background, while at the same

	Electron ($\|\eta\| < 2.5$)	Muon ($\|\eta\| < 2.4$)
Tight	Cut-based Tight + HLT-safe ID + recommended impact parameter cuts + Missing hits = 0 + Triple charge requirement	Tight ID + d_{xy} < 0.02 cm, d_z < 0.1 cm + Tight PF Iso (relative $\Delta\beta$-corrected iso < 0.15)
Loose	HLT-safe ID	Tight ID + d_{xy} < 0.02 cm, d_z < 0.1 cm + rel. PF Iso < 0.4, rel. track Iso < 0.4
Soft	-	Soft muon ID

Table 7: The lepton selections used in the analysis.

time being close enough to the tight selection not to introduce a bias.

Tau leptons that decay leptonically, to a lighter charged lepton and two neutrinos, are reconstructed as an electron or muon. The more frequent hadronic decayed taus are identified with the POG recommended discriminator (decayModeFindingNewDMs) and required to pass a loose isolation selection (LooseCombinedIsolationDeltaBetaCorr3Hits).

Lepton corrections

Corrections are applied for the differences in the lepton energy scale and resolution between simulation and data, following the POG recommendations. For muons, the electron scale correction factors were not available at the time of the analysis which is covered by an increased uncertainty.
Scale factors are derived as a function of the lepton p_T and η with the tag and probe method to correct remaining differences between simulation and data. These are due to differences in the lepton reconstruction efficiency and in the tight lepton selection efficiency. The efficiency and scale factor for the tight electron selection as a function of p_T is shown in Figure 30. For muons, the tight selection efficiency and scale factors are calculated for the identification and isolation selection separately. The tight identification efficiency is larger than 95% for muons with $p_T > 20$ GeV, the tight isolation efficiency is larger than 80% for muons with $p_T > 20$ GeV and larger than 95% for muons with $p_T > 40$ GeV [123].

5.3.2 *Jets, b-tagging and Missing Transverse Momentum*

The jet collection is created by clustering the particle flow candidates with the anti-kt algorithm and a distance parameter of 0.4 (Section 4.3.4). The clustering is performed after applying the Charge Hadron Subtraction (CHS) pileup mitigation. Default jet energy corrections are applied to correct the measured jet energy for variations in the energy response of the detector and other sources of bias. The jet energy resolution is corrected by rescaling

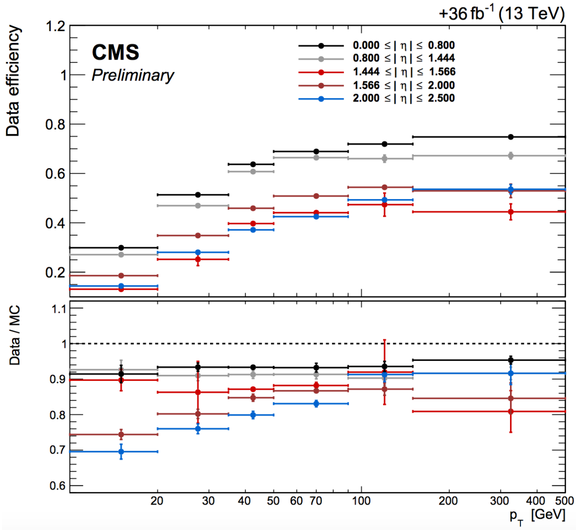


Figure 30: Efficiency of the tight electron selection in data as a function of p_T for several bins in η. The bottom window shows the scale factor that is calculated by dividing the efficiencies in data and simulation (MC).

the four-momentum of the reconstructed jets according to the scale factors provided by the POG. To reject jets originating from calorimeter or read-out electronics noise the loose working point of the PF Jet Identification is applied. Jets that overlap with a loose lepton ($\Delta R(\text{lepton, jet}) < 0.3$, with $(\Delta R)^2 = (\Delta\eta)^2 + (\Delta\phi)^2$) are removed from the jet collection.
B-tagging is performed with the medium working point of the Combined Secondary Vertex v2 (CSVv2) algorithm, corresponding to a b-jet tagging efficiency of about 65% and misidentification rates of 1% for light-flavour (udsg) jets and 15% for c-quark jets [104]. The data/simulation scale factors are computed according to the POG prescription.

The missing transverse momentum ($\vec{p}_T^{\text{miss}}$) is reconstructed as the negative of the vector sum of p_T of all reconstructed particles (Section 4.3.6). The size of the missing transverse momentum (p_T^{miss}) is called the missing transverse energy E_T^{miss}. The Jet energy corrections are propagated in the calculation of the E_T^{miss} and events with large but artificial E_T^{miss} are vetoed by imposing quality filters according to the POG recommendations.

5.4 EVENT SELECTION

The event selection has been optimized on simulated events, starting from the event selection of the analysis at Run I. The final state objects that are selected are:

- two tight, same sign leptons (electron or muon) with $p_T^{lep,1} > 25\,\text{GeV}$ and $p_T^{lep,2} > 20\,\text{GeV}$
- two jets with $p_T > 30$ GeV and $|\eta| < 5.0$
- $E_T^{miss} > 30$ GeV.

Cuts on the m_{jj}, $\Delta\eta_{jj}$ and Zeppenfeld variable for the leptons z_i^{lep} (Equation 32) are applied to extract the VBS topology:

- $m_{jj} > 500$ GeV
- $\Delta\eta_{jj} > 2.5$
- $\max(z_i^{lep}) < 0.5$.

The remaining backgrounds are suppressed by:

- b-jet veto:
  - b-tagged jet veto (CSVv2M, $p_T > 20$ GeV)
  - soft muon veto ($p_T > 3\,\text{GeV}$)
- 3th lepton veto:
  - loose lepton ($p_T > 10$ GeV)
  - hadronic tau ($p_T > 18$ GeV)
- in ee final state:
  $|m_{ll} - m_Z| > 15$ GeV,
  $m_{ll} > 20$ GeV.

The dominant contribution to the non-prompt background comes from $t\bar{t}$ events. It is suppressed by applying a veto on events that contain a b-tagged jet or an additional soft muon. Backgrounds with more than two leptons are suppressed by vetoing events with a third lepton, identified either as an electron or muon passing the loose lepton selection, or as a tau that decayed hadronically. Charge misassignment backgrounds are reduced by vetoing events with two electrons that have an invariant mass within 15 GeV of the Z boson mass, or an invariant mass smaller than 20 GeV to reject backgrounds from low mass resonances.

5.5 BACKGROUND ESTIMATIONS

Most of the backgrounds can be estimated directly from the simulation samples. For backgrounds that involve misreconstructed objects, it is usually necessary to rely on data to make an estimate. Misreconstructions typically happen with a probability at the order of 10^{-4} or lower, meaning a very large amount of simulated events would have to be generated. The amount of computing power that is required to do this, makes that it is usually not feasible to estimate these backgrounds from simulation. In addition, the edge of phase space where the misreconstruction happens is not necessarily well described by the simulation. The non-prompt background for example involves jets that pass the lepton isolation requirement, which is the very corner of its isolation distribution. An inadequate modeling of the underlying processes by which jets are misidentified as leptons or even a small mismodeling of the pile up can introduce large differences between simulation and data.

The non-prompt and charge misassignment backgrounds in this analysis are estimated using a data-driven method. For the WZ background estimate, a normalization to data is used to deal with the large theoretical uncertainties associated with the simulation of VBS processes at leading order, involving electroweak, strong and interference contributions.

5.5.1 *Non-prompt background*

The leptons that are reconstructed with the CMS detector can originate from real leptons, misreconstructed jets, misreconstructed photons or from tracks that are matched to unrelated hits in the ECAL or muon chambers [124].

For estimating the non-prompt background, the sources of leptons are split in three categories. First of all, there are leptons from the leptonic decay of heavy electroweak bosons (W, Z, or an off-shell photon γ^*) produced in the hard interaction that decay immediately at their production vertex and are called *prompt* leptons. The leptons we want to select to identify the W bosons are in this category.

There are also leptons originating from light and heavy flavour hadron decays. These hadrons typically travel some distance before decaying and can produce *non-prompt* leptons. Finally, there are reconstructed leptons called *fake* leptons which do not originate from a real lepton. In most cases they are caused by the misidentification of a jet. Leptons from misreconstructed photons would normally also fall in this last category, but because they are not estimated from data in this analysis but from simulation, misreconstructed photons will be considered as part of the prompt leptons in the rest of the discussion on the non-prompt background.

The name *non-prompt* background is used to indicate backgrounds that contain a lepton that is either non-prompt or fake, or in other words a lepton that does not originate from a heavy boson decay or a misreconstructed photon.

Because electrons and muons are reconstructed differently, also the main sources of the non-prompt background are different in the electron and muon final states. For muons, this background is dominated by non-prompt muons from $t\bar{t}$ events in the semi-leptonic decay channel. An example Feynman diagram is shown in Figure 31. There is one prompt lepton from the W decay, while a second, non-prompt lepton can be produced in the b-quark decay.

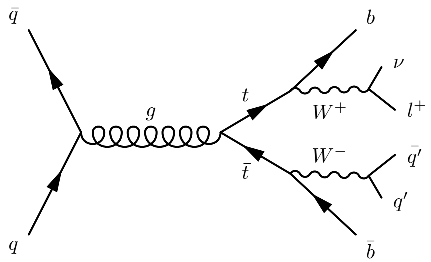


Figure 31: One of the possible Feynman diagrams for semi-leptonic $t\bar{t}$ production.

For electrons also fake leptons from light flavour jets become important. This is mainly from W+jets events where a light flavour jet is misreconstructed as an electron.
Although the probability for the misreconstruction is low, these backgrounds become important because the cross section of the processes involved is up to a factor 10^6 higher than the signal process (Figure 18).

5.5.1.1 *Fakeable object method*

The data-driven method that is used to estimate the non-prompt background is known as the fakeable object method. The first step of the method consists in measuring the probability for a jet that passes a loose lepton selection to also pass a tighter lepton selection. The measurement is performed in a data region that is enhanced in non-prompt and fake leptons, and will be called the *measurement region*. In the second step the measured probability is used to estimate the non-prompt background in the *signal region* (defined by the event selection).
The probability is called the fake rate (ϵ_{FR}) and is measured as a function of the p_T, η and flavour of the lepton:

$$\epsilon_{FR} = \frac{N_{\text{Tight leptons}}(\text{flavour}, p_T, |\eta|)}{N_{\text{Loose leptons}}(\text{flavour}, p_T, |\eta|)}, \tag{56}$$

with N the number of tight or loose leptons. The remaining contamination from prompt lepton is corrected by subtracting their estimate from simulation from the numerator and denominator.

The fake rate is also sensitive to the flavour and p_T of the quark from which the non-prompt or fake lepton originates and the hadronic energy in the event. Differences in these variables between the measurement region and the signal region can introduce a bias in the non-prompt estimate and care has to be taken in defining a loose lepton object so that the fake rate is not strongly dependent on them.

The estimate of the non-prompt background in the signal region relies on two nearby regions in data. The Tight-Loose (TL) region has the same event selection as the signal region, except one of the leptons does not pass the tight lepton selection but instead passes the loose lepton selection. In the Loose-Loose (LL) region both leptons only pass the loose lepton selection.
The ratio of fake and non-prompt leptons passing the tight selection over fake and non-prompt leptons passing only the loose selection is given by:

$$w_L = \frac{\epsilon_{FR}(\text{flavour}, p_T, \eta)}{1 - \epsilon_{FR}(\text{flavour}, p_T, \eta)} . \tag{57}$$

The denominator contains a small correction that reflects the different definition of the loose lepton category in Equation 56; for the efficiency ϵ_{FR} the loose leptons also include leptons passing the tight selection while for the ratio w_L the tight and loose categories are exclusive.
The non-prompt contribution in the signal region can be estimated by weighting the data events in the TL region with the ratio w_L calculated on the loose lepton. For the estimated non-prompt background yield this gives:

$$N^{\text{TL, data}}_{\text{non-prompt}} = \sum^{\text{TL data}} w_L . \tag{58}$$

The TL region however does not only contains events with one prompt and one non-prompt lepton, but also events with two prompt leptons of which one does not pass the tight lepton selection. To correct for these events, the same weights are calculated for simulation events with two prompt leptons in the TL region and subtracted:

$$N^{\text{TL}}_{\text{non-prompt}} = \sum^{\text{TL data}} w_L - \sum^{\text{TL prompt sim}} w_L . \tag{59}$$

We can use simulation events for this subtraction as the prompt leptons are well-described by the simulation.
An additional small correction is applied for the double counting of events with two non-prompt leptons. Their effect is counted twice because they end up in the TL region when either the first or the second lepton passes the tight selection [125]. The expected non-prompt contribution of two non-prompt lepton events in the TT region, is calculated by weighting the events in the LL region by the product of the ratios w_L calculated on both loose

leptons. After correcting for the prompt leptons from simulation, this gives as estimate for the non-prompt background yield:

$$\begin{aligned} N^{\text{non-prompt}} &= \sum^{\text{TL data}} w_L - \sum^{\text{TL prompt sim}} w_L \\ &- \sum^{\text{LL data}} w_{L,1} \cdot w_{L,2} + \sum^{\text{LL prompt sim}} w_{L,1} \cdot w_{L,2} \,. \end{aligned} \tag{60}$$

In summary, the non-prompt background is estimated by weighting data events in the TL and LL regions and corrected for prompt lepton events from simulation according to Equation 60.

5.5.1.2 *Fake Rate Measurement*

The fake rate is measured in a data region enriched in di-jet events. This *measurement region* is defined by the single lepton triggers in Table 6 and an event selection that suppresses real lepton contributions:

- exactly 1 loose lepton (electron or muon) with $p_T > 10\,\text{GeV}$
- $E_T^{\text{miss}} < 20\,\text{GeV}$
- $m_T^{\ell, E_T^{\text{miss}}} < 20\,\text{GeV}$
- jet with $p_T > 35\,\text{GeV}$ and $\Delta R(\text{jet,lepton}) > 1$.

With $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ and $m_T^{\ell, E_T^{\text{miss}}}$ the transverse mass[1] of the lepton and E_T^{miss}. For events with a W boson decaying to leptons, $m_T^{\ell, E_T^{\text{miss}}}$ has a maximum at the W boson mass (80 GeV). The remaining prompt lepton contribution, as estimated from the DY and W+jets simulation samples, is subtracted from the numerator and denominator of the fake rate. It is cross-checked in a control region with inverted MET and $m_T^{\ell, E_T^{\text{miss}}}$ cuts that the prediction of the prompt leptons agrees with data.
The fake rate as a function of the p_T and $|\eta|$ is shown in Figure 32 for electrons and Figure 33 for muons.

5.5.1.3 *Validation and Uncertainty*

Several test are performed to assess the validity of the fakeable object method, with the defined lepton selections and measurement and signal region, and to determine the uncertainty on the non-prompt background estimate.

1 In the case of a decay into two particles of which the full energy can not be reconstructed, the transverse mass is defined as:
$M_T^2 = (E_{T,1} + E_{T,2})^2 - (\vec{p}_{T,1} + \vec{p}_{T,2})^2$
with $E_T^2 = m^2 + \vec{p}_T^{\,2}$.

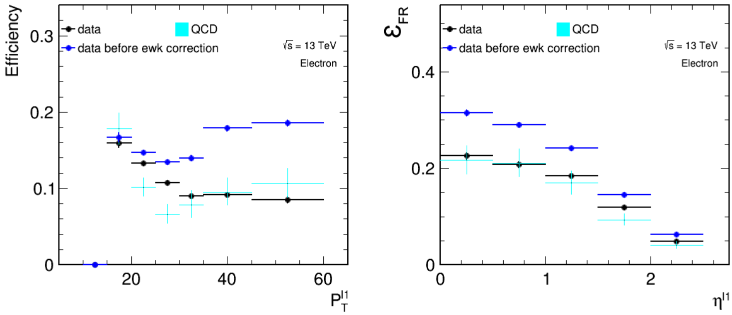


Figure 32: The electron fake rate (ϵ_{FR}) as a function of the p_T and $|\eta|$ of the electron measured on data, after (black) and before the prompt lepton subtraction (blue), and on a simulated QCD sample (light blue).

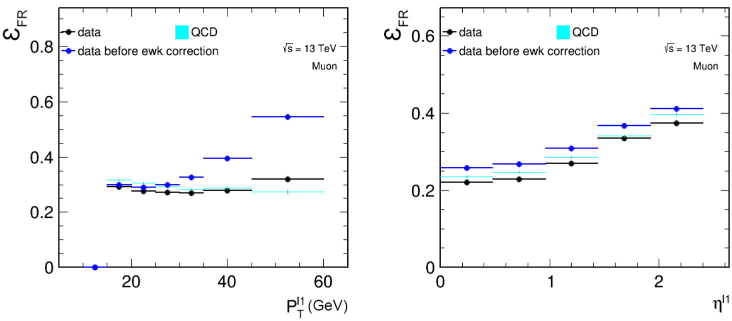


Figure 33: Muon fake rate (ϵ_{FR}) as a function of the p_T and $|\eta|$ of the muon. The fake rate is measured on data, shown after (black) and before the prompt lepton subtraction (blue), and on a simulated QCD sample (light blue).

Closure Test on Simulation

The first test consists in cross-checking the fakeable object method on simulation. Although the non-prompt background can not be estimated accurately from simulation alone, a cross-check that relies only on simulation is useful for testing the method and for investigating the source of possible discrepancies by using the generator level information [126]. The *closure test* consists of comparing the non-prompt background estimate from applying the fakeable object method to simulated events in the TL and LL regions with the direct non-prompt estimate using the simulated events that end up in the TT region.

The fake rate is measured in the QCD region, as is done when using the fakeable object method on data, but here it is calculated on simulated QCD samples. The event weights (Equation 60) are applied in the TL and LL signal region on $t\bar{t}$ and W+jets simulation samples. A global closure is found within 20%, which is shown for the $t\bar{t}$ sample in Figure 34 with the estimate from the TT region (simulated) in blue and the estimate from weighting the LL and TL events (estimated) in green. The amount of non-closure gives an indication of the remaining sensitivity to differences in flavour composition and the mother parton p_T between the measurement and signal region.

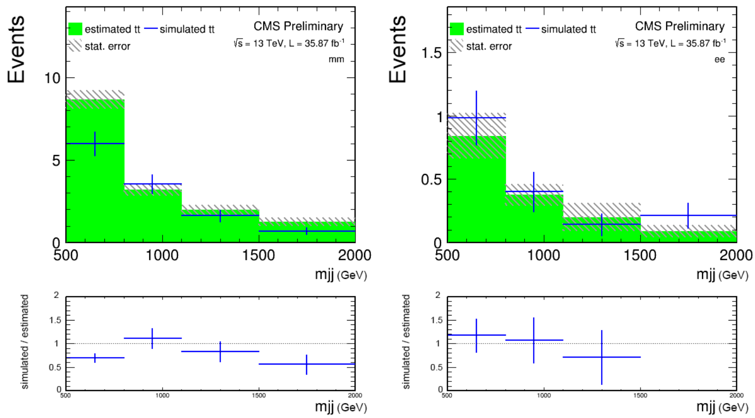


Figure 34: The m_{jj} distribution in the signal region for the $t\bar{t}$ sample directly from simulation (simulated, in blue) and from applying the fakeable object method to simulation (estimated, in green). The $\mu^{\pm}\mu^{\pm}$ final state is shown on the left and $e^{\pm}e^{\pm}$ final state on the right.

Control regions

Another important test is the comparison of the non-prompt estimate with the data in control regions. Two control regions are defined:

- low $m_{jj} < 500\,\text{GeV}$ region:
  with a VBS selection that is inverted compared to the signal region
- b-tag region: at least one b-tagged jet is required, instead of the b veto.

The m_{jj} distribution of both control regions is shown in Figure 35, combining all lepton flavour final states. The m_{ll} distributions for the low m_{jj}

region, split in lepton flavour final states, is shown in Figure 36. The relatively good closure in these regions shows that the non-prompt estimate is quite robust against changes in the jet flavour and energy.

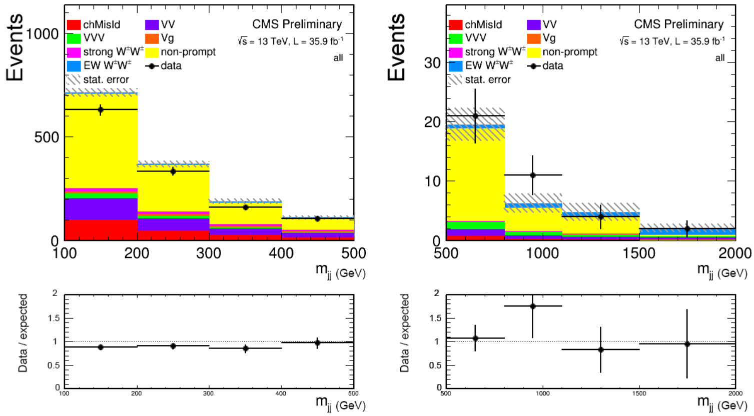


Figure 35: The m_{jj} distribution in the low m_{jj} (left) and b-tag (right) control regions.

Varying the measurement region

To estimate the dependence on the p_T of the parton that the jet originates from, the jet p_T cut in the measurement region is varied within 15 GeV of its default cut value. This leads to a difference of $10\% - 15\%$ in the non-prompt estimate.

Alternative prompt lepton correction

The uncertainty on the subtraction of the prompt leptons from simulation in the signal region is estimated by using an alternative method that relies on data to subtract the prompt lepton contribution. Next to the fake rate ϵ_{FR}, it also uses the prompt rate ϵ_{PR}: the probability for a prompt lepton that passes the loose selection to also pass the tight selection. The prompt rate is defined in the same way as the fake rate in Equation 56, but is measured on prompt leptons in a data region enhanced in Drell-Yan events. The prompt rate as a function of the lepton p_T and $|\eta|$ is shown for muons in Figure 37. The prompt rate for electrons is about 20% lower because of the stronger requirements in the tight electron selection (Section 5.3.1).

To simplify the notation, we will use $f \equiv \epsilon_{FR}$ and $p \equiv \epsilon_{PR}$. The non-prompt background yield can be estimated as [125]:

$$N^{\text{non-prompt}} = f^2 \cdot N_{ff} + f \cdot p \cdot N_{pf}, \tag{61}$$

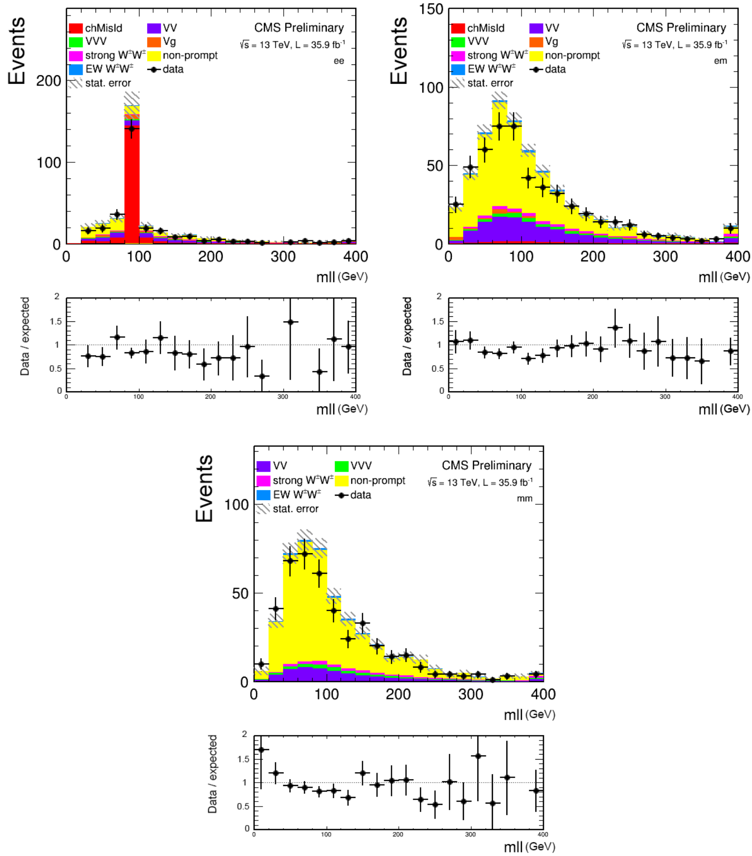


Figure 36: The m_{ll} distribution in the low m_{jj} control region, split in the ee (left), $e\mu$ (right) and $\mu\mu$ (bottom) final states.

with N_{ff} the number of events with two non-prompt or fake leptons, and N_{pf} the number of events with one prompt and one non-prompt or fake lepton.

The quantities N_{ff} and N_{pf} are not directly measurable in data, we can only determine the number of leptons that pass the tight selection. The number of events with two leptons that pass the tight selection N_{TT} for example, can be written as:

$$N_{TT} = f^2 \cdot N_{ff} + f \cdot p \cdot N_{pf} + p^2 \cdot N_{pp}\,. \tag{62}$$

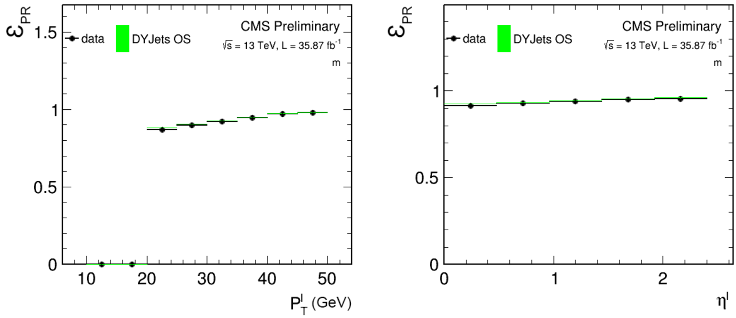


Figure 37: The muon prompt rate (ϵ_{PR}) as a function of the p_T and $|\eta|$ of the muon measured on data (black) and on a simulated DY sample (green).

The same can be done for the number of events with one tight and one loose lepton N_{TL} and two loose leptons N_{LL}. By inverting these equations, one finds the expressions for N_{ff} and N_{pf}:

$$\begin{pmatrix} N_{pp} \\ N_{pf} \\ N_{ff} \end{pmatrix} = \frac{1}{(p-f)^2} \begin{pmatrix} f^2 & -f(1-f) & (1-f)^2 \\ -2fp & p(1-f)+f(1-p) & -2(1-p)(1-f) \\ p^2 & -p(1-p) & (1-p)^2 \end{pmatrix} \begin{pmatrix} N_{LL} \\ N_{TL} \\ N_{TT} \end{pmatrix} . \tag{63}$$

This leads to $N^{\text{non-prompt}}$ as a function of N_{LL}, N_{TL} and N_{TT}. The non-prompt background is estimated by weighting events in the LL, TL and TT regions in data according to Equation 61, where the prompt and fake rate are a function of the p_T and $|\eta|$ of the leptons. By setting $p = 1$ one recovers the data weights that are used in the default method (Equation 60). The non-prompt background estimate using the alternative method is found to agree with the estimate from the default method within 1%.

For the alternative method a distinction needs to be made between the prompt leptons and the leptons from photon misreconstruction, which were considered as part of the prompt leptons in this discussion (Section 5.5.1). The misreconstructed photons are not well described by the prompt or fake rate, but have a rate that lies somewhere in between. The contribution from the $V\gamma$ background where a photon is misreconstructed as an electron is still subtracted from simulation, as is done in the default method.

Total uncertainty

The maximum discrepancy that is observed in each test is summarized in Table 8.

Test	Maximum discrepancy
Control regions	20%
Closure test on simulation	20%
Varying measurement region	15%
Alternative prompt lepton correction	1%

Table 8: The measured discrepancy in the non-prompt background tests.

Two systematic uncertainties are applied on the non-prompt estimate: a bin-by-bin statistical uncertainty and a global systematical uncertainty of 30%, which is a conservative estimate motivated by the measured discrepancies in Table 8.
Another cross-check was performed to examine if the shape uncertainty on the fake rate is covered by the combination of the statistical and global systematic uncertainty. This is done by constructing a shape systematic uncertainty from the variation in ϵ_{FR} that happens when changing the jet p_T cut in the measurement region. The effect of the shape uncertainty on the final results are within 1%. This can be expected, as the most sensitive data region in this analysis is dominated by the statistical uncertainties.

5.5.2 *Charge Misassignment background*

Final states with two oppositely charged leptons are much more common than with two same-sign leptons at proton proton colliders. Oppositely charged leptons are produced in quark or gluon annihilation interactions with for example an intermediate Z or $t\bar{t}$ state, while higher order interactions are needed to create two leptons with the same charge.

When the charge of one the opposite-sign leptons is wrongly identified, the event can enter the signal region of this analysis. The misreconstruction of the muons charge is negligible due to the large distance the muon propagates through the detector, allowing an accurate measurement of its curvature. For electrons the probability of charge misreconstruction is between 10^{-2} and 10^{-5} depending on the p_T and η of the electron. This probability is already strongly reduced by the extra requirements in the tight electron identification (Section 5.3.1).
As there are not enough simulated events with a misreconstructed charge in the signal region to directly estimate this background from simulation, the charge misassignment probability is measured in simulation and corrected for the difference with data in a region enhanced in Drell-Yan events. This probability is then used to estimate the charge misassignment background from simulated events passing the event selection with an inverted requirement on the lepton charge, i. e. from events with opposite-sign leptons. Simulated opposite-sign lepton events from DY, $t\bar{t}$, tW, and W^+W^- processes

are weighted to estimate the charge misassignment background.

The ratio of electrons with a misreconstructed charge over the electrons with a correctly reconstructed charge (R_{ch}) is measured in a simulated Drell-Yan sample as a function of p_T and η:

$$R_{ch}(p_T, |\eta|) = \frac{N_{\text{wrong charge}}(p_T, |\eta|)}{N_{\text{correct charge}}(p_T, |\eta|)}. \tag{64}$$

Whether the charge of a lepton is misreconstructed is determined by comparing the reconstructed charge to the charge at the generator level.

To correct for the difference with data, a scale factor is derived by dividing R_{ch} calculated on data with R_{ch} calculated on simulation. The scale factor is measured as function of the electron $|\eta|$, while no significant dependence on the p_T of the electron was found.
Measuring R_{ch} in data requires a different approach, as it is not known which one of leptons has a misreconstructed charge. Two regions dominated in Drell-Yan events are used for the measurement. They have the common event selection listed in Table 9, with the only difference that one region has two same-sign (SS) reconstructed electrons and the other has two opposite-sign (OS) reconstructed electrons. The pseudorapidity distribution of the highest p_T lepton in both regions is shown in Figure 38. Contributions from other processes are very small in both regions and are neglected.

Event selection
Exactly 2 electrons with $p_T^{\text{lep},1} > 25$ GeV and $p_T^{\text{lep},2} > 20$ GeV
$\lvert m_{ll} - m_Z \rvert < 15$ GeV
charge$^{\text{lep},1}$ · charge$^{\text{lep},2}$ = -1 (OS) or = 1 (SS)

Table 9: The event selection used to identify the Drell-Yan events in the opposite-sign (OS) and same-sign (SS) regions for measuring R_{ch} in data.

The probability for an opposite-sign lepton pair to be reconstructed as same-sign is the sum of the probabilities for either lepton to have a misreconstructed charge. For the ratio of same-sign events over opposite-sign events this yields:

$$\frac{N_{\text{same sign}}(|\eta_{l,1}|, |\eta_{l,2}|)}{N_{\text{opposite sign}}(|\eta_{l,1}|, |\eta_{l,2}|)} = \frac{R_{ch}(|\eta_{l,1}|) + R_{ch}(|\eta_{l,2}|)}{1 + R_{ch}(|\eta_{l,1}|) \cdot R_{ch}(|\eta_{l,2}|)}. \tag{65}$$

Five bins in $|\eta|$ are used, which means there are five unknowns: $R_{ch}(|\eta|)$ in each $|\eta|$ bin. At the same time, there are 25 equations corresponding to Equation 65 for each combination of $|\eta_{l,1}|$ and $|\eta_{l,2}|$.
The overdetermined system is solved by performing a least squares fit to $N_{\text{same sign}}/N_{\text{opposite sign}}$ measured in data. Dividing $R_{ch}(|\eta|)$ in data and

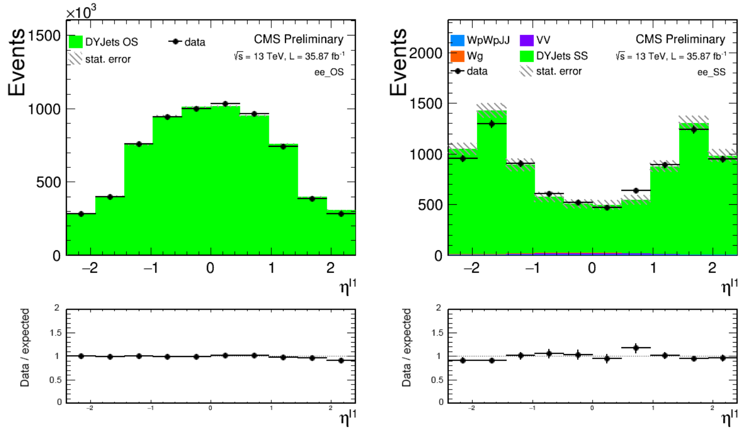


Figure 38: The pseudorapidity distribution of the highest p_T electron in the opposite-sign (OS, left) and same sign (SS, right) DY region defined in Table 9.

$\|\eta\|$ range	R_{ch}^{data}	R_{ch}^{sim}	Scale factor ($R_{ch}^{data}/R_{ch}^{sim}$)
0.0-0.5	$(3.8 \pm 0.5) \cdot 10^{-5}$	$(3.0 \pm 0.3) \cdot 10^{-5}$	1.27 ± 0.22
0.5-1.0	$(1.0 \pm 0.1) \cdot 10^{-4}$	$(0.7 \pm 0.05) \cdot 10^{-4}$	1.43 ± 0.14
1.0-1.5	$(5.7 \pm 0.2) \cdot 10^{-4}$	$(6.0 \pm 0.2) \cdot 10^{-4}$	0.94 ± 0.04
1.5-2.0	$(2.4 \pm 0.05) \cdot 10^{-3}$	$(2.6 \pm 0.05) \cdot 10^{-3}$	0.93 ± 0.02
2.0-2.5	$(2.4 \pm 0.05) \cdot 10^{-3}$	$(2.4 \pm 0.05) \cdot 10^{-3}$	0.98 ± 0.03

Table 10: The ratio of electrons with a misreconstructed charge over the electrons with a correct reconstructed charge in data and simulation, and the resulting scale factor as a function of the electron $|\eta|$.

simulation leads to the scale factors in Table 10.

The estimate of the charge misassignment background is in agreement with data, as can be seen from the m_{ll} distributions in the ee final state of the low m_{jj} control region (Figure 36). The method was also verified in the same-sign DY measurement region, by comparing the estimate from weighted opposite-sign simulation events to data. A relatively good closure is found in the DY region, shown for the m_{jj} and $\eta^{lep,1}$ distributions in Figure 39. The uncertainty on the measurement of R_{ch}^{data} covers the difference and is propagated in the statistical procedure of the analysis.

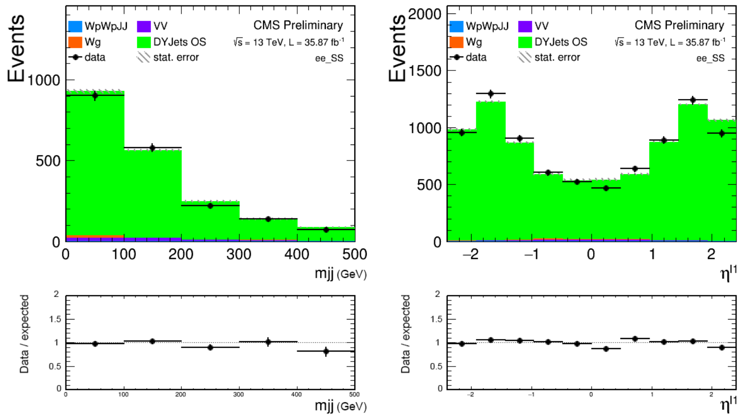


Figure 39: The pseudorapidity of the highest p_T electron (left) and the m_{jj} distribution (right) in the same-sign DY region defined in Table 9 with the charge missasignment background estimated from weighted opposite-sign simulation event.

5.5.3 WZ

Simulated samples are produced to describe the electroweak and strong WZ production in multiple jet bins. The sum of the WZ contributions is fitted to data as part of the statistical analysis. The details of this binned maximum likelihood fit that is performed simultaneously in the signal region and a WZ control region will be described in Section 5.6.1. The WZ region is defined by modifying the following cuts on top of the signal event selection (Section 5.4):

- Exactly 3 leptons with total charge equal to one,
  $p_T^{\text{lep},1} > 25\,\text{GeV}$, $p_T^{\text{lep},2} > 20\,\text{GeV}$ and $p_T^{\text{lep},3} > 10\,\text{GeV}$
- Two opposite-sign same-flavor leptons
  with $|m_{ll} - m_Z| < 15$ GeV.

The m_{jj} and m_{ll} distributions in the WZ control region before the fit to data in the statistical analysis are shown in Figure 40.

5.6 SIGNAL REGION

The signal and background contributions after the event selection (Section 5.4) are estimated from the simulated samples and the background estimation methods described in the previous sections. Before the statistical analysis, only the non-prompt and charge misassignment backgrounds are estimated with data-driven method, while the other backgrounds and the signal process are estimated from simulation. At this point, also the WZ background

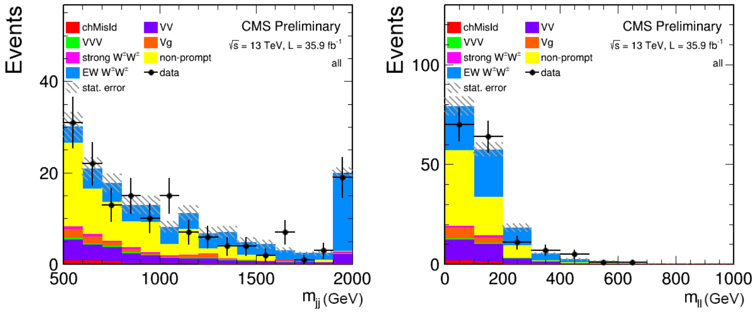


Figure 40: The pre-fit m_{jj} (left) and m_{ll} (right) distributions in the WZ control region.

is still fully estimated from simulation.

The pre-fit distributions in several variables for the data and the estimated signal and backgrounds are shown in Figure 41. It shows the distributions in the $p_T^{lep,1}$, $p_T^{lep,2}$, E_T^{miss} and z_1^{lep} (Equation 32) variables that are used in the event selection, as well as the m_{jj} and m_{ll} variables from which the measurements are performed in the statistical analysis. It is clear that the electroweak $W^{\pm}W^{\pm}jj$ process makes up a significant fraction of the events that pass the event selection, and that a high purity is obtained in the high m_{jj} region. The event yields are listed in Table 11 after an accumulative application of the cuts in the event selection.

5.6.1 *Statistical Procedure*

The two measurements that are performed in this analysis, of the electroweak $W^{\pm}W^{\pm}jj$ signal process and of the anomalous quartic gauge couplings (aQGC) (Section 5.1), make use of slightly different methods. For both, quantitative statements on the agreement of the data with the theory predictions are made from the asymptotic behavior of the test statistic as defined in Section 4.6. The test statistic is based on a ratio of likelihoods and is a function of the parameter of interest μ. For the SM measurement, this parameter is a global scale factor to the expected signal process yield, while for the aQGC measurement it corresponds to the Wilson coefficients (Section 2.4.1.2) that are part of the signal model.

Evaluating the test statistic for a specific value of μ is done by performing a binned maximum likelihood fit to the data. The statistical analysis relies on a simultaneous fit in the signal region and WZ control region, which allows

	data	non-prompt	VV	$V\gamma$	chMisId	strong $W^\pm W^\pm jj$	VVV	Sum Backgrounds	electroweak $W^\pm W^\pm jj$
2 same-sign lep, >= 2 jets, p_T, m_{jj}, $\Delta\eta_{jj}$	569	327.3	70.8	27.5	72.1	8.2	17.8	523.7	91.1
E_T^{miss}	457	227.5	59.0	19.1	36.7	7.4	16.3	366	82.1
$\max(z_i^{lep})$	318	142.2	41.5	11.2	22.5	3.5	9.8	230.7	72.1
b-jet veto	227	98.4	36.4	9.2	18.0	3.2	3.9	169.1	68.1
soft muon veto	196	91.5	32.1	8.7	17.0	3.1	3.1	155.5	65.3
hadronic tau veto	180	83.1	24.6	8.4	16.4	2.9	2.4	137.8	63.0
$\|m_{ee} - m_Z\|$	160	64.1	23.6	7.5	4.0	2.9	2.3	104.4	61.4
m_{ee}	159	64.0	23.5	7.5	3.8	2.9	2.3	104	61.4

Table 11: Data and expected event yields after successive application of the event selections. The sum of the background yields is shown in the second last column.

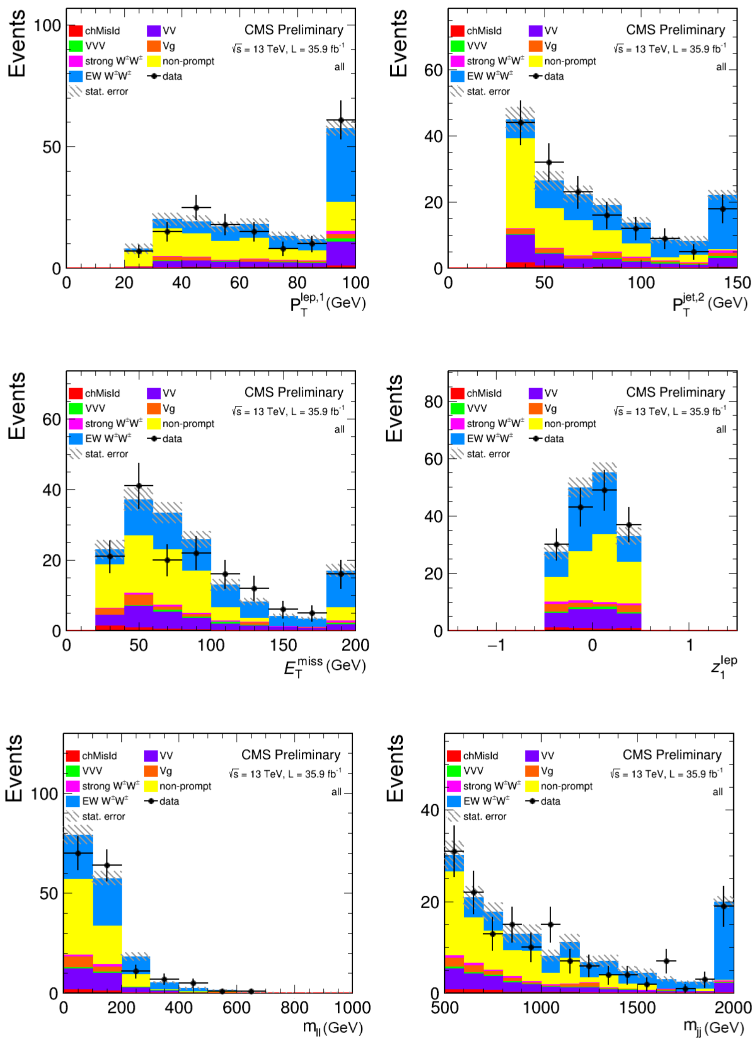


Figure 41: The expected signal and background distributions in the signal region before performing the statistical analysis (pre-fit). The signal contribution (EW $W^{\pm}$ $W^{\pm}$) in blue is stacked on top of the background contributions.

to correlate the processes and systematic uncertainties in both regions. The free parameters in the fit are the parameter of interest μ and the normalization of the WZ contribution. The other nuisance parameters are described by a log-normal probability distribution.

The statistical analysis of the event yields is performed on histograms in the m_{jj} and m_{ll} variables. The variables and bins that are used in the statistical analysis for each measurement are summarized in Table 12, and described in more detail below.

	EW $W^\pm W^\pm$ measurement	Anomalous coupling limits
Signal region	2D histogram in m_{jj} and m_{ll}	1D histogram in m_{ll}
WZ region	1D histogram in m_{jj}	1D histogram in m_{ll}
Binning	m_{jj}: (500, 800, 1100, 1500, ∞)	
	m_{ll}: (20, 100, 180, 300, ∞)	m_{ll}: (20,100,180,300,400, ∞)

Table 12: Overview of the statistical analysis.

Electroweak $W^\pm W^\pm jj$

The statistical analysis of the event yields relies on a two-dimensional simultaneous fit of the m_{jj} and m_{ll} distributions in the signal region, and the one-dimensional m_{jj} distributions in the WZ region. The result is a measurement of the signal strength and its significance, quantified by the probability that the observed data is produced by the standard model without the signal process.
The free parameters in the fit are the signal strength modifier μ equal to $N^{\text{signal}}/N^{\text{signal}}_{\text{SM expectation}}$, and the WZ normalization in each bin of m_{jj}. As post-fit the WZ background estimate in m_{jj} is determined completely from the fit, the WZ simulation only effects the final result through the WZ background shape in m_{ll}.

Anomalous Quartic Gauge Couplings

The statistical procedure to quantify the sensitivity to the dimension eight EFT operators uses a simultaneous fit to the m_{ll} distributions in the signal region and the WZ control region. The m_{ll} variable is sensitive to the new physics characterized by the high energy scale Λ described with the effective field theory, as the higher order operators enhance the cross section at high m_{WW}. The effect of a non-zero aQGC coefficient on the m_{ll} distribution is shown in Figure 42. The coefficient associated with the M_0 operator is chosen, although a very similar effect is found when modifying the other anomalous quartic gauge couplings. As can be seen, the effect of a non-zero aQGC depends strongly on m_{ll} and the sensitivity comes almost exclusively from the last bin in m_{ll} (which includes overflow).
As the data shows no significant deviations from the SM expectation, lim-

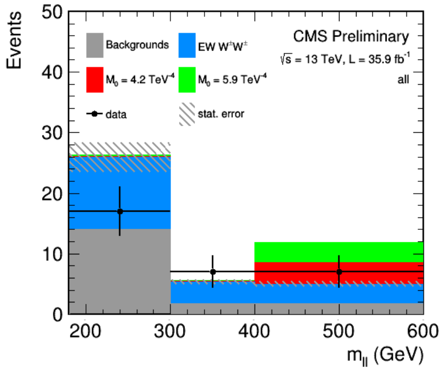


Figure 42: The effect of a non-zero coefficient f_{M0}/Λ^4 (associated with the M_0 operator) on the electroweak $W^{\pm}W^{\pm}jj$ production as a function of m_{ll}. The chosen values of f_{M0} correspond to the measured exclusion limits at 68% (red) and 95% (green) confidence level. Their effect is shown on top of the backgrounds (grey) and electroweak $W^{\pm}W^{\pm}jj$ as predicted by the SM (blue).

its are placed on the Wilson coefficients. A limit is placed on each Wilson coefficient, with the parameter of interest μ equal to the coefficient under study and the other Wilson coefficients set to zero (corresponding to their SM value).

The signal in this measurement is the effect of the dimension eight operators on the electroweak $W^{\pm}W^{\pm}jj$ production, while the electroweak $W^{\pm}W^{\pm}jj$ production as predicted by the standard model is considered as a background.
The limits are insensitive to a possible effect of the aQGCs on the WZ process. As this background is determined completely from the data, any effect of new physics is absorbed in its estimate.

5.7 SYSTEMATIC UNCERTAINTIES

Next to the uncertainties due to the statistical fluctuations in simulation and data, also systematic uncertainties have to be taken into account. The sources of systematic uncertainties that are considered in the analysis can be split in four categories: uncertainties arising from the collider setup, from the detector simulation and reconstruction, from the event simulation called the theory uncertainties, and from the background estimation methods.

5.7.1 *Collider uncertainties*

Luminosity

The luminosity is measured by counting the mean number of pixel clusters per bunch crossing, calibrated with Van der Meer scans. These are performed with a dedicated LHC machine set up, by scanning the two beams through one another in the transverse plane of the detector from which the size of the beams at the collision point can be determined. In combination with information on the number of circulating protons, this allows the determination of an absolute luminosity scale [127].
The uncertainty on the luminosity is 2.5%, giving a total integrated luminosity of 35.9 ± 0.9 fb^{-1}[128].

Pileup

The pileup in data is calculated by multiplying the instantaneous luminosity with the total inelastic cross section. The uncertainty is estimated by varying the cross section by 5% in both directions and using the associated variations in the pileup distribution as shape systematic uncertainties. Their effect on the event yields in each bin after applying the event selection is smaller than 2%.

5.7.2 *Detector uncertainties*

Jet Energy

One of the main systematic uncertainties in the analysis comes from the uncertainty on the jet energy reconstruction. It is estimated by rescaling the jet four-momenta (up and down) with the jet energy uncertainty as measured by the corresponding POG. With this procedure, both the acceptance after the event selection and the shape difference in the jet distributions are propagated. The jet energy scale uncertainty has an effect between 1% and 7% on the final event yields over all bins.
The uncertainty on the jet energy resolution is also considered as a shape systematic uncertainty with a very small effect below 1%.

Lepton and Trigger

As for the jets, the uncertainty on the lepton momentum scale of about 1% is propagated by varying the momentum of the leptons by their uncertainties. Other systematic uncertainties involving the leptons are related to the identification and isolation scale factors and the trigger scale factor. These have a total effect of about 2% per lepton.

Missing transverse energy

The uncertainty on the E_T^{miss} is computed from the energy scale uncertainties of its components. It is propagated as a shape uncertainty and has an effect on the event yields of about 1%.

B-tag efficiency

The uncertainties on the b-tag scale factor for light flavour jets and for heavy flavour jets are included as two shape systematic uncertainties. Their total effect is within 3%.

5.7.3 *Theory uncertainties*

Interference

The interference between the electroweak and strong production of same-sign W bosons is estimated with the PHANTOM generator. Large simulation samples are generated for the total electroweak $W^{\pm}W^{\pm}jj$ production, and for the electroweak and strong production components separately. The interference is calculated in the relevant variables by subtracting the electroweak and strong contributions from the total electroweak $W^{\pm}W^{\pm}jj$ production. The study is performed at generator level, with an event selection applied to the generator level objects approximating the event selection used at reconstruction level in the analysis. The estimated interference is applied as a shape systematic uncertainty on the signal and is found to be smaller than 4% in all the bins used in the statistical analysis.

QCD Scales and PDFs

The uncertainty on the signal simulation leads to the main systematic uncertainty in this analysis. All processes that are estimated from simulation dependent on the Parton Distribution Functions (PDF), the strong coupling α_s, and the factorization (μ_f) and renormalization (μ_r) scales that are used for the event generation. The information that is needed to estimate the uncertainties is produced during the event generation as event weights. The scales μ_f and μ_r are varied by a factor 2 to estimate their uncertainty. The uncertainty on the signal yield is between 3% and 12% over all bins, and it is largest for the triboson background where is is between 10% and 25%. The procedure for estimating uncertainties from the PDF and α_s follows the recommendations from the PDF4LHC group [129]. A shape systematic is derived by reweighting the events according to a predefined set of PDFs, with the uncertainty on α_s added in quadrature. This leads to another uncertainty up to 5% of the event yields.

5.7.4 *Background estimation*

The systematic uncertainties described so far are all uncertainties related to the simulation. Different systematic uncertainties are applied for backgrounds that are estimated using data.

Non-prompt

The systematic uncertainty on the non-prompt background estimate is quantified by a global uncertainty of 30% as motivated in Section 5.5.1.3. The non-prompt estimate relies on simulated samples to remove the contribution from prompt lepton, which are subject to the usual simulation systematic uncertainties. The effect of these systematic uncertainties on the non-prompt estimate was measured and has been found to be covered by the global uncertainty.

WZ

An unconstrained uniform uncertainty is used for the WZ background normalization in each bin in m_{jj} for the electroweak $W^{\pm}W^{\pm}jj$ measurement and in m_{ll} for the anomalous coupling measurement. The systematic uncertainties on the simulation also influence the WZ estimate in the electroweak $W^{\pm}W^{\pm}jj$ measurement through its effect on the m_{ll} distribution. Post-fit the dominant uncertainty comes from the fit in the WZ control region that has limited statistics, giving a total uncertainty between 20% and 60% over the different bins. The post-fit WZ uncertainty in the most sensitive m_{jj} bin of the electroweak $W^{\pm}W^{\pm}jj$ measurement is 37%.

Charge misassignment

The charge misassignment background is estimated from weighted opposite-sign simulation events. As a consequence the estimate is subject to the systematic uncertainties on the simulation. The weights are derived from simulation as well, and are corrected with a scale factor to data. A global systematic uncertainty of 7% is applied that covers the uncertainty from the scale factor measurement.

5.7.5 *Uncertainty impacts*

The most important systematic uncertainties in the fit to data are shown in Figure 43 ranked by decreasing impact on the parameter of interest. The impact of a systematic uncertainty is defined as the change in the measured signal strength $\hat{\mu}$ when varying the post-fit value of the uncertainty by $\pm 1\,\sigma$. The figure also shows the pull (and its post-fit uncertainty) for each nuisance θ, defined as:

$$\text{pull} = \frac{\hat{\theta} - \theta_0}{\Delta\theta}\,, \tag{66}$$

with $\hat{\theta}$ the post-fit value of θ, and θ_0 and $\Delta\theta$ the initial (pre-fit) value and uncertainty.

Three of the five systematic uncertainties with the largest impact are related to the theory. These are the systematic uncertainties related to the QCD scales on the signal process (QCDscale_EWK), the quark PDFs (pdf_qqbar), and the interference between the electroweak and strong production (EWK_Intf). The uncertainty on the QCD scale in the leading order generation of the signal process corresponds to an uncertainty of about 10% on the measured signal strength μ. Other important systematic uncertainties come from the estimate of the non-prompt lepton background (fake_syst) and the energy scale uncertainties of jets (CMS_scale_j) and missing transverse energy (CMS_scale_met).

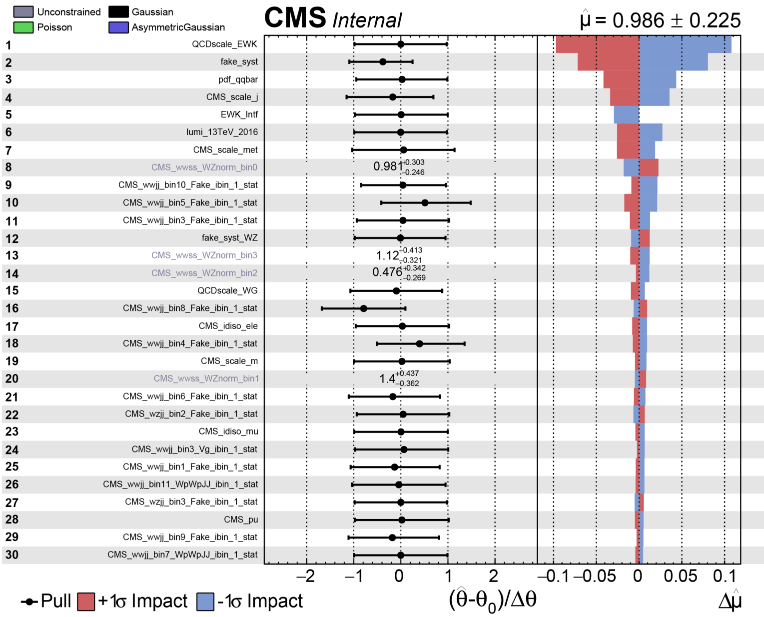


Figure 43: The main systematic uncertainties ranked by decreasing impact on the post-fit value of the parameter of interest $\hat{\mu}$ (right), and the mean and standard deviation of the pull (middle).

5.8 RESULTS

5.8.1 *Electroweak* $W^\pm W^\pm jj$

The m_{jj} and m_{ll} distributions in the signal region after the maximum likelihood fit are shown in Figure 44. The post-fit distribution of the m_{jj} variable in the WZ region is shown in Figure 45.
The observed signal strength relative to the SM expectation μ is $0.99^{+0.26}_{-0.22}$, while the value that is expected from the estimated signal and backgrounds is $1.00^{+0.26}_{-0.22}$. The observed significance of the measurement is $6.1\,\sigma$. The expected significance is calculated with the background estimates from data by performing the fit with $\mu = 1$ (a posteriori), and is equal to $6.1\,\sigma$.

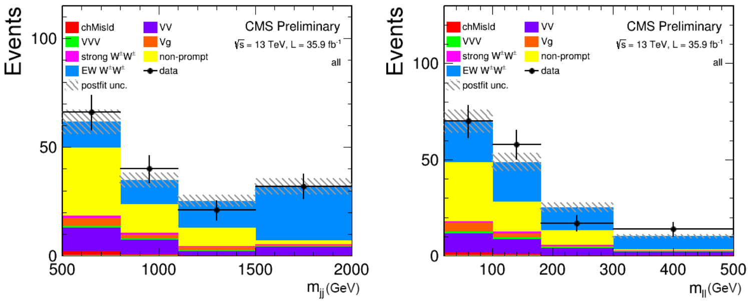


Figure 44: The m_{jj} (left) and m_{ll} (right) distributions in the signal region after performing the statistical analysis.

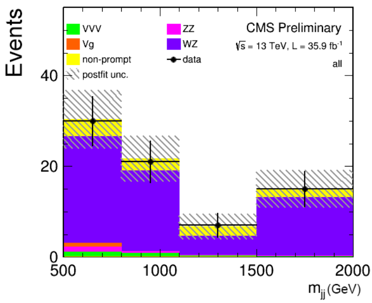


Figure 45: The m_{jj} distribution in the WZ region after performing the statistical analysis.

The version of the analysis that is submitted for publication [4] contains small differences to the version discussed in this thesis. A first difference is that the tight electron selection does not include the HLT safe selection. In addition, small differences are present in the event selection, with the main difference the slightly tighter VBS selection in this thesis, while the MET selection is looser. The exact cut values that are used in the two version of the analysis are listed in Table 13. Both versions of the analysis use a completely independent framework, but have been synchronized at the percent level in the muon final state after applying the event selection used in the publication version. The expected and observed significances with the selections from the paper are respectively 5.5σ and 5.7σ. These results are the first observation of the electroweak $W^{\pm}W^{\pm}jj$ production with a significance greater than 5σ.

Differences	Thesis version	Publication version
Tight electron ID	requires HLT safe	not required
$\max(z_i^{lep})$	< 0.5	< 0.75
E_T^{miss}	> 30 GeV	> 40 GeV

Table 13: The main differences between the analysis version in this thesis and the version used in the publication.

The measured signal cross-section in a fiducial region is reported in the publication [4]. The cross section is measured in a fiducial region to obtain a maximally model-independent result, that allows a straightforward comparison to theory predictions. The fiducial region is defined on generator objects with the selections in Table 14, using MG5aMC and excluding the $W \to \tau\nu \to l\nu\nu\nu$ decays. The fiducial cross section is measured to be 3.83 ± 0.66 (stat) ± 0.35 (syst) fb. The predicted theoretical cross section at LO is 4.25 ± 0.21fb in the fiducial region, in agreement with the measurement.

Event selection
two same sign leptons (electron or muon) with $p_T > 20$ GeV and $\|\eta\| < 2.5$
two jets with $p_T > 30$ GeV and $\|\eta\| < 5.0$
$m_{jj} > 500$ GeV
$\Delta\eta_{jj} > 2.5$

Table 14: The event selection of the fiducial region.

5.8.2 *Anomalous Quartic Gauge Couplings*

The m_{ll} distributions in the signal and WZ region are fitted simultaneously to extract limits on the dimension eight operator coefficients. The limits are calculated on top of the analysis version that is submitted for publication [4]. The expected and observed 95% confidence level (CL) limits are given in Table 15. The observed limits are up to a factor 6 stronger than the limits observed at Run I in CMS analyses, and the strongest limits that have been set on some of the anomalous quartic gauge couplings. This is shown for the M-operators in Figure 46, which gives an overview of the aQGC limits obtained by CMS and ATLAS. The strongest limits on the aQGCs associated to the M_1, M_6 and M_7 operators were set in this analysis.

	Observed limits (TeV^{-4})	Expected limits (TeV^{-4})	Run-I limits (TeV^{-4})	Run-I analysis
f_{S0}/Λ^4	[-7.7, 7.7]	[-7.0, 7.2]	[-38 , 40]	ss WW
f_{S1}/Λ^4	[-21.6,21.8]	[-19.9,20.2]	[-118 , 120]	ss WW
f_{M0}/Λ^4	[-6.0, 5.9]	[-5.6, 5.5]	[-4.6 , 4.6]	$\gamma\gamma \to$WW
f_{M1}/Λ^4	[-8.7 ,9.1]	[-7.9, 8.5]	[-17 , 17]	$\gamma\gamma \to$WW
f_{M6}/Λ^4	[-11.9,11.8]	[-11.1,11.0]	[-65 , 63]	ss WW
f_{M7}/Λ^4	[-13.3,12.9]	[-12.4,11.8]	[-70 , 66]	ss WW
f_{T0}/Λ^4	[-0.62,0.65]	[-0.58,0.61]	[-3.8 , 3.4]	$Z\gamma$
f_{T1}/Λ^4	[-0.28,0.31]	[-0.26,0.29]	[-1.9 , 2.2]	ss WW
f_{T2}/Λ^4	[-0.89,1.02]	[-0.80,0.95]	[-5.2 , 6.4]	ss WW

Table 15: Observed and expected 95% CL limits on the coefficients of dimension-eight operators in the EFT Lagrangian. The last two columns show the strongest limits obtained by CMS with Run-I data.

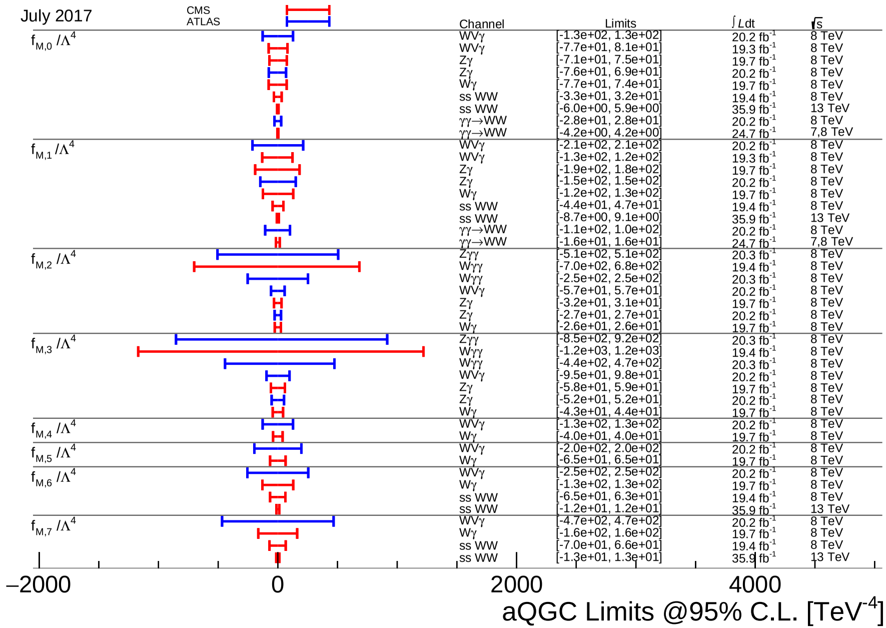


Figure 46: Comparison of limits on the Wilson coefficients for the dimension eight M-operators by CMS and ATLAS analyses [130]. The aQGC limits that were set in this analysis (ss WW) are the strongest limits on the f_{M1}/Λ^4, f_{M6}/Λ^4, f_{M7}/Λ^4 coefficients.

6 VBS AT THE HL-LHC

The high luminosity (HL) upgrade of the LHC (Section 3.5), planned for 2024, will provide the opportunity to study many features of the SM and beyond in detail. It will be combined with upgrades to the CMS experiment to deal with the high luminosity environment. These upgrades to the detector are necessary to make optimal use of the 3000 fb^{-1} of data that will be recorded in the decade after the HL upgrade, which is a tenfold of the amount of data that will be collected before 2024.
The study of the Higgs boson will be central to the HL-LHC program [2]; the large dataset allows precise measurements of the Higgs couplings, including the Higgs self-coupling through the di-Higgs production, and searches for rare SM and BSM decays. This will be complemented by studies of vector boson scattering to test the role of the Higgs boson in the electroweak symmetry breaking.
The other main component of the program consists of searches for new physics, e. g. supersymmetry, new heavy gauge bosons, extra dimensions and dark matter. The upgrades also ensure the continued study of SM phenomena with high statistics, which is useful to test our current knowledge of particle physics and to have well-modeled backgrounds for the discovery portion of the program.

Even after the first discovery of the electroweak production of di-bosons pairs in combination with two jets, the study of VBS remains important as many of the interesting properties of VBS are still unexplored by the LHC experiments. Measurement are needed to precisely determine the role of the Higgs boson in the normalization of the VBS process, the longitudinal scattering cross section and the quartic gauge couplings.
The prospects for the electroweak $W^{\pm}W^{\pm}jj$ production at the HL-LHC are explored in this chapter. The results have been used to support the technical proposal for the Phase II CMS detector [2] and were later updated and summarized in a public analysis summary [3].

6.1 ANALYSIS OVERVIEW

Because of the large amount of data that will be recorded with the phase II CMS detector at the HL-LHC, not only is it possible to measure the electroweak $W^{\pm}W^{\pm}jj$ production cross section, but also measurements of the VBS properties can be performed. In total, four separate measurements are chosen for which the sensitivity is estimated with 3 ab^{-1}.

First of all, a precision measurement of the electroweak $W^{\pm}W^{\pm}jj$ process can be performed. The sensitivity of the measurement is quantified by the uncertainty on the cross section.

Information on the role of the Higgs boson in VBS can be obtained in several ways. One way is by measuring the longitudinal component of the process, for which the incoming and outgoing W bosons are longitudinally polarized ($W^{\pm}_{L}W^{\pm}_{L} \rightarrow W^{\pm}_{L}W^{\pm}_{L}$). The different polarization components can be separated using the kinematics of the outgoing particles, which depend on the polarization of the W bosons. By measuring the longitudinal component, one can test if the longitudinal scattering process is unitarized through the presence of the Higgs boson (Section 2.3) as predicted by the standard model.
Another possibility is to measure the effect of partial unitarization on the *total* electroweak $W^{\pm}W^{\pm}jj$ process. In theories involving more than a single Higgs boson, e. g. a two Higgs doublet model introducing additional heavy bosons, the VBS process is only partially unitarized by the Higgs boson (Section 2.4). Models that are nearly degenerate at the Higgs resonance could be resolved from their different behaviors at higher energies [131]. The sensitivity to partially unitarized theories is estimated both in a qualitative and quantitative way.

Finally, the sensitivity to new physics in the electroweak sector at a high mass scale is estimated. The new physics could lead to strong enhancements in the cross section at high energy, which can be approximated through the effect of higher dimensional operators in a effective field theory (Section 2.4.1.2). As before, only dimension eight operators are considered.

6.1.1 *Detector scenarios*

As the study was initiated for the technical proposal of the CMS upgrade, its main goal has been to assess the performance of the upgraded detector in the high luminosity environment. For VBS studies specifically, the tracker and calorimeter extensions and improved granularity are crucial for the tagging of VBS events as they allow for an improved reconstruction of the forward jets.
The main detector scenario that is used in the analysis corresponds to the upgraded Phase II detector, described in Section 3.5, in the high luminosity conditions with on average 140 pileup interactions per bunch crossing.

To compare the performance of the upgraded detector, two other detector scenarios are considered.
The *Phase I, 50 PU* detector scenario corresponds to the current Phase I detector and data conditions as they are before the HL upgrade, with on average 50 pileup events per bunch crossing and without assuming degradations due

to radiation.
The *Phase I aged, 140 PU* detector scenario describes an *aged* phase I detector, including radiation damage after 1 ab^{-1}, in the HL conditions. At this point, a large fraction of pixels in the pixel tracker lose functionality and the ECAL and HCAL scintillating elements become opaque. As both the tracker and forward calorimeters are close to becoming completely dysfunctional with this amount of radiation damage, this is not a detector scenario that could be used throughout the HL operation, but is used to identify the key areas to be addressed by the upgrade.

The target of the CMS phase II upgrade is to recover the current performance of the detector, described by the *Phase I, 50 PU* scenario, in the challenging conditions after the HL upgrade characterized by an increased luminosity and pileup, described by the *Phase II, 140 PU* scenario.
An overview of the detector scenarios and analysis benchmarks that are used in the analysis is summarized in Table 16. In the following, the three detector scenarios will be denoted by *Phase II*, *Phase I* and *Aged*.

Analysis benchmarks	Detector scenarios	
Electroweak $W^{\pm}W^{\pm}jj$ cross section	PHASE II	Phase II, 140 PU
Longitudinal scattering cross section	PHASE I	Phase I, 50 PU
Partial unitarization	AGED	Phase I, 140 PU + radiation damage
Anomalous quartic gauge couplings		

Table 16: Overview of the analysis benchmarks and detector scenarios.

6.1.2 *Analysis strategy*

The signal is defined as the electroweak production of same-sign W boson pairs in the two jet and two same-sign lepton final state, as in the Run II data analysis Section 5.1.1. The interference between the electroweak and strong production is found to be at the order of a few percent after the event selection and its effect is neglected in the analysis. The analysis strategy is also similar to the one described in the previous chapter, as the VBS topology is exploited to select a phase space enriched in VBS events. An additional handle that is used to select the electroweak events is the amount of soft QCD radiation. Backgrounds that are initiated by the strong interaction are expected to contain a more QCD radiation in the central region of the detector than the signal process, for which the QCD radiation is suppressed in the rapidity gap (Figure 10) between the tag jets [32].
The main difference with the analysis strategy described in the previous chapter is that the backgrounds can not be estimated from data but have to be estimated completely from simulation. This is not straightforward for the estimation of the non-prompt background, as it is known to be not well

described by simulation. To deal with this limitation, the final results are produced not only as a function of the integrated luminosity, but also as a function of a global scale factor to the non-prompt background estimation.

Because of the large integrated luminosity of 3000 fb^{-1}, stronger event selections can be applied than in the Run 2 analysis. This leads to a very pure signal region, with only small background contributions, of which the non-prompt background is the dominant one. The large amount of data also allows to perform the statistical analysis with a 2-dimensional binned fit in multiple final state categories.

6.2 SIMULATION SAMPLES

Simulation samples for all the backgrounds and signal, with three different detector scenarios, are needed to perform this study. As it is not feasible to produce all these samples with a detailed detector simulation, the detector simulation is performed using a parametric implementation of the detector response based on the Delphes package, described in Section 4.2. In this way, the amount of time and computing resources that are needed is strongly reduced.
In the signal and background samples, only decays of the vector bosons to electrons and muons are considered, while decays to taus are excluded in the simulation. This reduces the recources needed to produce the simulation samples with only a small effect on the final results.

No additional corrections are applied to the simulation as is done in a data analysis. Also no trigger inefficiency correction needs to be applied as indicated by the trigger studies performed for the HL-LHC [2]. The studies show that the addition of tracker information in the L1 trigger allows for dilepton triggers with an efficiency of $\sim 100\%$, for p_T thresholds of 20 GeV for the first and 10 GeV for the second lepton.

6.2.1 *Event Generators*

Multiple matrix element generators (Section 4.1.3) are used to simulate the hard interactions, which are interfaced with PYTHIA v8 for the parton shower and hadronization steps.
The samples for the $W^{\pm}W^{\pm}jj$ signal, irreducible background and their interference are produced at leading order with the PHANTOM 1.2.6 generator. This event generator is also used to produce the WZ background samples, samples with opposite sign leptons and two jets (used to estimate the charge misassignment background as well as a small fraction of the non-prompt background), and signal samples for the partial unitarization benchmark with rescaled Higgs to vector boson couplings.
For the reducible non-prompt background estimate, a semi-leptonic $t\bar{t}$ sam-

ple is generated with POWHEG v2 at NLO QCD, and samples with a single vector boson in association with jets are generated with MadGraph v5.2.2.3 [132] at LO. The MadGraph generator is also used to produce the signal samples for the anomalous couplings study and the longitudinal scattering signal samples with the DECAY v4 package (which is part of the MadGraph framework) to preserve information on the polarization.
The CTEQ6L [133] parton density functions (PDF) are used for MadGraph and PHANTOM, while POWHEG relies on the CT10 [134] PDF.

6.2.2 *Detector Simulation*

The detector simulation for all the samples is performed with Delphes. A limited set of samples has been processed as well with a detailed simulation performed with Geant4 (Section 4.2), and are used to tune the Delphes setup. The efficiencies and resolutions of the reconstructed leptons, jets, and missing energy, as well as the b-quark tagging efficiency and the particle misidentification probability are measured in the detailed simulation samples, and consequently implemented in Delphes to perform the fast detector simulation.

For the Phase II and Phase I detector scenarios, a dedicated Delphes configuration was made for the technical proposal studies, implementing the detector geometry and performance as observed in detailed simulation. For the aged scenario, no Delphes tune had been studied and the phase II configuration was modified according to the POG recommendations and from comparing the corresponding detailed simulation samples. Several systematic uncertainties were downgraded and the signal yield was scaled down to 70%, corresponding to the loss in signal efficiency as measured in detailed simulation.

The reliability of the Delphes simulation to reproduce the kinematics of VBS events is validated by comparing the main analysis variables after a typical VBS event selection in signal samples with Delphes and detailed simulation. A good agreement in the main variables is observed between the samples, as can be seen in Figure 47 showing a few of these variables for the phase I detector scenario.
Another validation of the Delphes simulation is performed by comparing the event selection efficiencies, shown in Figure 48. The largest difference is observed for the matching of reconstructed to generator level jets. This matching is not used in the analysis and the discrepancy is caused by small differences in the jet resolution and in the parton showering, hadronization and PU description between Delphes samples using PYTHIA v8 and detailed simulation using PYTHIA v6.

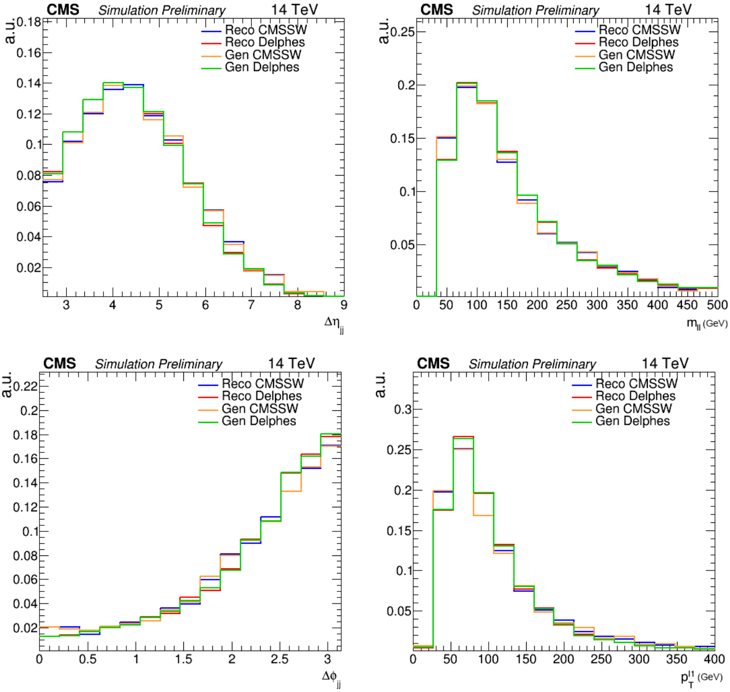


Figure 47: Distributions of some of the main analysis variables after a typical VBS selection for the signal sample with dedicated simulation (CMSSW) and Delphes simulation, and at generator (Gen) level and after reconstruction (Reco), in the phase I detector scenario.

6.3 OBJECT RECONSTRUCTION

To make use of the improved phase II detector design in the object reconstruction (Section 4.3), reconstruction algorithms have been designed for the new sub-detectors as part of the detailed simulation. In the Delphes simulation, the reconstruction of physics objects is performed with a simplified particle flow method that combines information from the subdetectors to perform particle identification [93].

The efficiencies and resolutions of the reconstructed objects are implemented as measured from detailed simulation. In addition, the leptons are required to be isolated. The isolation variable, defined as in Equations 42 and 45 with a ρ-based PU correction, needs to be smaller than 0.25 and 0.6 in the 50 PU and 140 PU scenario respectively. Loose leptons are defined using the same identification selections and a relaxed isolation requirement

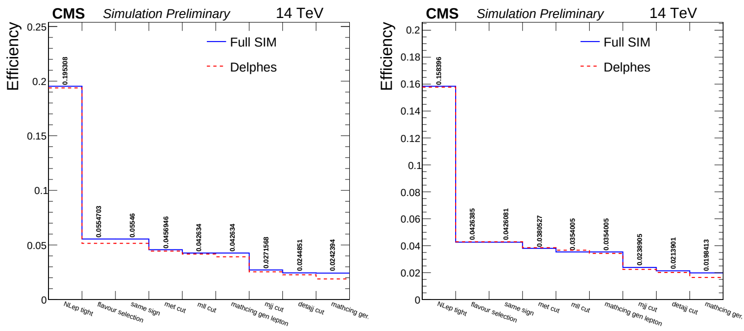


Figure 48: Efficiency of a typical VBS selection for the signal sample with dedicated simulation (Full SIM) and Delphes simulation in the phase I (left) and phase II (right) detector scenarios.

< 0.35 in the 50 PU and < 0.75 in the 140 PU scenario.
The identification and isolation efficiency for tight leptons with $p_T > 20$ GeV and $|\eta| < 1.5$ is 75% for electrons and 80% for muons. The efficiencies reduces to respectively 55% and 65% at $|\eta| \sim 2.5$ and remains constant up to $|\eta| \sim 4.0$ for the Phase II detector.

Jets are clustered from the PF particles that have been cleaned for the contribution of charged particles from pileup with charged hadron subtraction (Section 4.3.4). They are reconstructed with the anti-k_t algorithm with cone radius $\Delta R = 0.4$. A median energy subtraction in bins of $|\eta|$ is applied to mitigate residual pileup effects.
Jets containing b quarks are identified with the CSV algorithm which provides an average efficiency of 70% with a mistag rate of 1% for light flavour and gluon jets. The b-tagging algorithm is used only up to $|\eta| < 2.5$ for the Phase II detector scenario. This conservative approach is used as the b-tagging algorithm requires further study to be extended to the full tracker acceptance.
To estimate the amount of hadronic energy in the event, *track jets* are used because of their low dependence on pileup. They are reconstructed with the anti-k_t algorithm with $\Delta R = 0.4$, by clustering tracks (Section 4.3) with a $p_T > 0.5$ GeV originating from the primary vertex. The estimator of the total hadronic activity in the event H_T is defined as the sum of the transverse momenta of the track jets with a $p_T > 2$ GeV:

$$H_T = \sum_{i}^{\text{track jets}} p_T^i , \tag{67}$$

excluding the track jets that are geometrically matched to one of the tag jets or leptons.

6.4 EVENT SELECTION

The cuts to select the final state objects is the same as in the Run I analysis:

- two tight, same sign leptons (electron or muon) with $p_T > 20\,\text{GeV}$ and $|\eta| < 2.4$
- two jets with $p_T > 30$ GeV and $|\eta| < 4.7$
- $E_T^{miss} > 40$ GeV.

The leptons are selected only up to $|\eta| = 2.4$, even for the Phase II detector scenario with a tracker up to $|\eta| = 4.0$. The reason is that less than 10% of the signal events have a lepton at a pseudorapidity larger than 2.4, while the backgrounds remain significant in this region. In addition, we can not accurately estimate the non-prompt background at high pseudorapidity.
The VBS topology is extracted with the following selections:

- $m_{jj} > 850$ GeV
- $\Delta\eta_{jj} > 4.0$
- $\max(z_i^{lep}) < 0.5$,
- $H_T < 125$ GeV (< 150 GeV for phase II) .

The remaining backgrounds are suppressed by:

- b-jet veto:
  - loose b-tagged jet veto ($p_T > 30$ GeV)
  - soft muon veto ($p_T > 5$ GeV, inside a jet with $p_T > 20$ GeV)
- 3th loose lepton veto ($p_T > 20$ GeV, $|\eta| < 2.4$ (phase I) / 4.0 (phase II))
- Z mass veto: $|m_{ll} - m_Z| > 20$ GeV
- $m_{ll} > 40$ GeV
- $\Delta R(jj, ll) < 6$
- $\Delta\eta_{ll} < 2$.

Tight selections are placed on the m_{jj} and $\Delta\eta_{jj}$ variables. This is complemented with a cut on the Zeppenfeld variable for the leptons z_i^{lep} (Equation 32) and on H_T (Equation 67). For the phase II scenario, a tighter selection on H_T is applied because of the extended tracker pseudorapidity coverage of the Phase II detector.

Variables that have not yet been used in the event selection of the run I and II data analyses are the distance in R between the di-jet and di-lepton system:

$$\Delta R(jj, ll) = \sqrt{(\Delta\eta_{jj,ll})^2 + (\Delta\phi_{jj,ll})^2},$$

and the η-separation between the leptons.

6.5 BACKGROUND ESTIMATIONS

Most of the backgrounds (and the signal) are estimated directly from the simulation samples. Also the WZ background is estimated directly from simulation, which is different from the approach that was followed for the data analysis in Chapter 5.
Because of the tighter event selection that is used in this study, some of the smaller backgrounds in the Run II data analysis, as ZZ and VVV, become negligible. These backgrounds are therefore not considered in this study. Also the $V\gamma$ background has been neglected.

For the backgrounds that involve a misreconstruction, dedicated methods have to be developed. To estimate the background containing electrons with a misreconstructed charge, the charge misreconstruction probability is used as an event weight on opposite-sign lepton events passing all other event selections. The probability measured on data at $\sqrt{s} = 8$ TeV is used as function of p_T and $|\eta|$ [135]. The estimation of the background from jets that are misreconstructed as leptons is generally performed on data. It will be referred to as the non-prompt or fake lepton background, which includes all backgrounds with a reconstructed lepton that originates from a jet as in Section 5.5.1. To provide a preliminary estimate for this study, an extrapolation procedure using only simulation samples has been developed.

6.5.1 *Non-prompt background*

Similar to the fakeable object method used in Section 5.5.1.1, the probability for a lepton-like object to be reconstructed as a tight lepton is measured. The measurement is performed in a large detailed simulation sample and the probability is then applied to Delphes samples to estimate the non-prompt background.
Instead of defining a loosely identified lepton, the jet definition is used as the denominator object. As there is no misreconstruction simulated in Delphes, using loose leptons is not possible. The jet on the other hand, is a robust object with a very similar reconstruction in the Delphes and detailed simulation samples. The measured probability on the jets of the detailed simulation sample can therefore be applied on the jets reconstructed in the Delphes samples.

The small misreconstruction probability is susceptible to bias introduced by differences between the measurement and application region. To control this potential bias, the probability is measured on a $t\bar{t}$ sample which is the dominant contribution to the non-prompt background and the measurement is performed for b jets and non-b jets separately.
A detailed simulation $t\bar{t}$ sample with $\sqrt{s} = 13$ TeV and on average 20 pileup events per bunch crossing is used for the measurement. This samples was chosen because of its large amount of events (25M), needed to precisely measure the probability at the order of $5 \cdot 10^{-4}$ for a jet to be misreconstructed as a lepton.

The fake rate (ϵ_{FR}) is defined as:

$$\epsilon_{FR} = \frac{N_{\text{Tight leptons}}(\text{flavour}, p_T, |\eta|)}{N_{\text{jets}}(\text{flavour}, p_T, |\eta|)}\,. \tag{68}$$

ϵ_{FR} is measured as a function of p_T and $|\eta|$ of the jet, separately for electrons and muons, and for b-jets and non-b jets.
To make sure that the jets and leptons that are used in the measurement are not prompt leptons, all the generator level leptons from W boson decays in the $t\bar{t}$ event (Figure 31) that are within the tracker acceptance are required to be geometrically matched to a reconstructed lepton. Also the jets are cleaned for prompt leptons and are required to originate from the primary vertex so that generator level information can be used to associate a flavour to them. In addition, for the jet to electron measurements the reconstructed jets are required not to pass the b-tag working point used in the event selection. For the jet to muon measurements this is not required as no dependence on the b-tag selection was found.
The denominator of the fake rate is filled with the cleaned jets, while the numerator is filled with cleaned jets that are matched to a lepton with $p_T > 20$ GeV passing the tight lepton identification and isolation.

Because the method uses two different physics objects, an additional step has to be implemented to account for the difference in their kinematics. The transverse momentum of the non-prompt lepton will in general be smaller than the p_T of its originating jet. To transform the jets to leptons in the estimation of the non-prompt background, the p_T has to be changed as well:

$$M_{p_T} = \left[P_T^{\text{jet}} - P_T^{\text{lepton}}\right](\text{flavour}, p_T, |\eta|)\,. \tag{69}$$

The p_T migration M_{p_T} is measured as a function of jet p_T and $|\eta|$ for the different flavours of jets and leptons.

Both measured quantities are then used to estimate the non-prompt background from simulated events with at least 1 prompt lepton and at least 3 jets. For each jet in the event, two new event are created in which the jet is transformed to an electron and to a muon with a p_T lowered by the p_T migration matrix, and weighted with the corresponding fake rate.

Fake rate and p_T migration

The measured fake rate as a function of the jet p_T is shown in Figure 49 and as a function of the jet $|\eta|$ for electrons in Figure 50. Differences in the fake rate are observed for b jets and non-b jets, and for electrons and muons. To understand these differences, it is important to realize that the decay chain of heavy quarks often contains leptons. Non-prompt leptons that are present in the b jets can pass the lepton selections, while for light flavour jets a misreconstruction needs to take place for it to be reconstructed as a lepton.

The b jets are almost two times as likely to be reconstructed as a muon than as an electron, which can be explained by the tight selections on electrons e. g. on the ECAL cluster shape and H/E requirements (Section 4.3.3). Light jets on the other hand, have a very small probability to be reconstructed as a muon as they are unlikely to produce a consistent lepton signature up to the muon chambers.
The importance of separating b jets and non-b jets also shows in the different shape of the non-b and b-jet fake rate as a function of p_T. For b jets, a higher p_T causes the jets to be more collimated which reduces the probability for the lepton to pass the isolation requirement. For light flavour jets, the behavior is very different; the fake rate increases with p_T, indicating the dominant contribution comes from the probability for the jet to pass the lepton identification requirements.

Also small differences in the p_T migration are observed. The p_T migration for jets that are misreconstructed as electrons is shown in Figure 51.

Validation

To check the dependance of the measured quantities on the center-of mass energy, the fake rate and the p_T migration factors are also determined using a $t\bar{t}$ sample at $\sqrt{s} = 8$ TeV with $7M$ events and similar pileup conditions. The p_T migration factors are found to be in good agreement between both samples, the fake rate shapes are in agreement as well while the normalizations are within a factor 2. The fake rate as a function of p_T is shown for b jets to muons in Figure 52. The large difference in normalization is expected as different reconstructions are applied in the samples while the same lepton identification and isolation requirements were used. This suggest that the fake rate does not depend strongly on the center-of-mass energy and the measured quantities at $\sqrt{s} = 13$ TeV can be used for this study at $\sqrt{s} = 14$ TeV as well.

To validate if the fake rate estimate can be used for the different upgrade scenarios, the same quantities were calculated on the $t\bar{t}$ detailed simulation

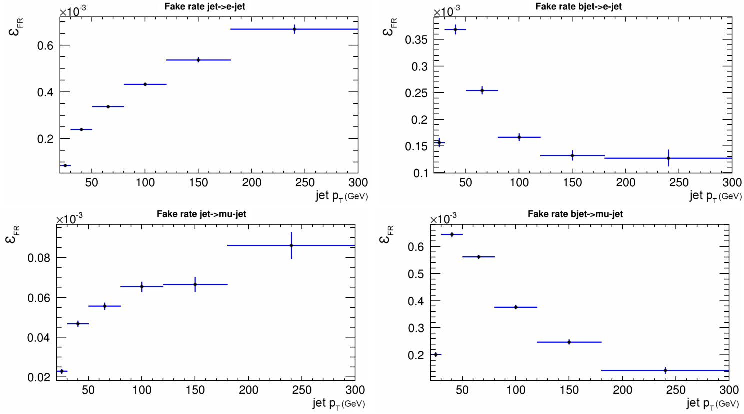


Figure 49: The fake rate as a function of the p_T of non-b jets (left) and b jets (right) for electrons (top) and muons (bottom).

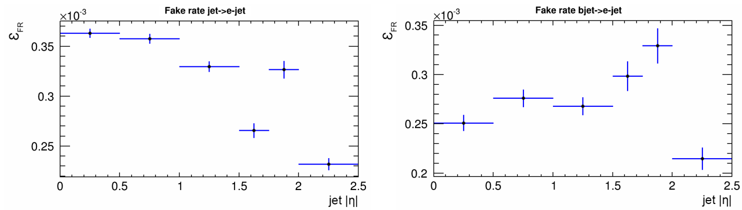


Figure 50: The fake rate as a function of the p_T of non-b jets (left) and b jets (right) for electrons.

samples produced for the HL studies with $0.5M$ events. Again an agreement of the fake rate shapes are found within the uncertainties.
It is therefore reasonable to use the fake rate derived from the $\sqrt{s} = 13$ TeV sample with small statistical uncertainties for all the detector scenarios. Because the normalization of the fake rate determined from simulation has been found to be unreliable in previous studies on data, the final results will be stated as a function of a global fake rate scale factor to account for the unknown data-to-simulation ratio.

6.6 SIGNAL REGION

The expected signal and background distributions in several variables after the event selection are shown in Figure 53, for the Phase II detector configuration and an integrated luminosity of 3 ab^{-1}. Two signal models stacked on the backgrounds are shown in this figure: the signal as predicted by the standard model (blue continuous line) and the signal process in absence of

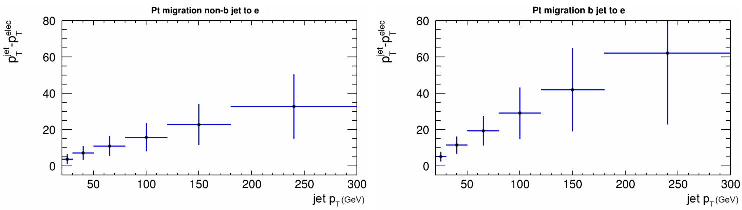


Figure 51: The jet to electron p_T migration as a function of the p_T of non-b jets (left) and b jets (right). The error bar corresponds to the root mean square of the p_T migration distribution in each bin.

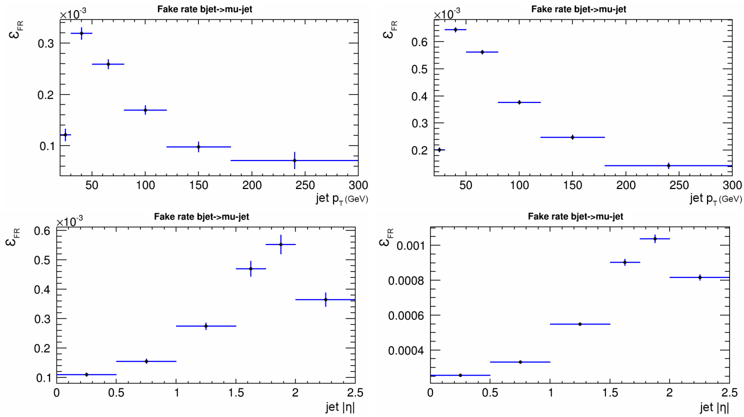


Figure 52: The fake rate of b jets to muons as a function of P_T (top) and $|\eta|$ (bottom) at 8 TeV (left) and 13 TeV (right).

the Higgs boson (dashed pink line). The difference between the two is shown with the continuous red line.

6.6.1 *Statistical procedure*

The final state is categorized according to the lepton flavour (ee, $e\mu$, μe, $\mu\mu$, where the leptons are sorted in p_T) and the lepton charge ($++$, $--$), yielding a total of eight categories. The final state with two positive charged leptons has a larger signal contribution due to the positive charge of the two initial protons. Different backgrounds are present in the lepton flavour categories e. g. the non-prompt background depends on the lepton flavour and p_T.

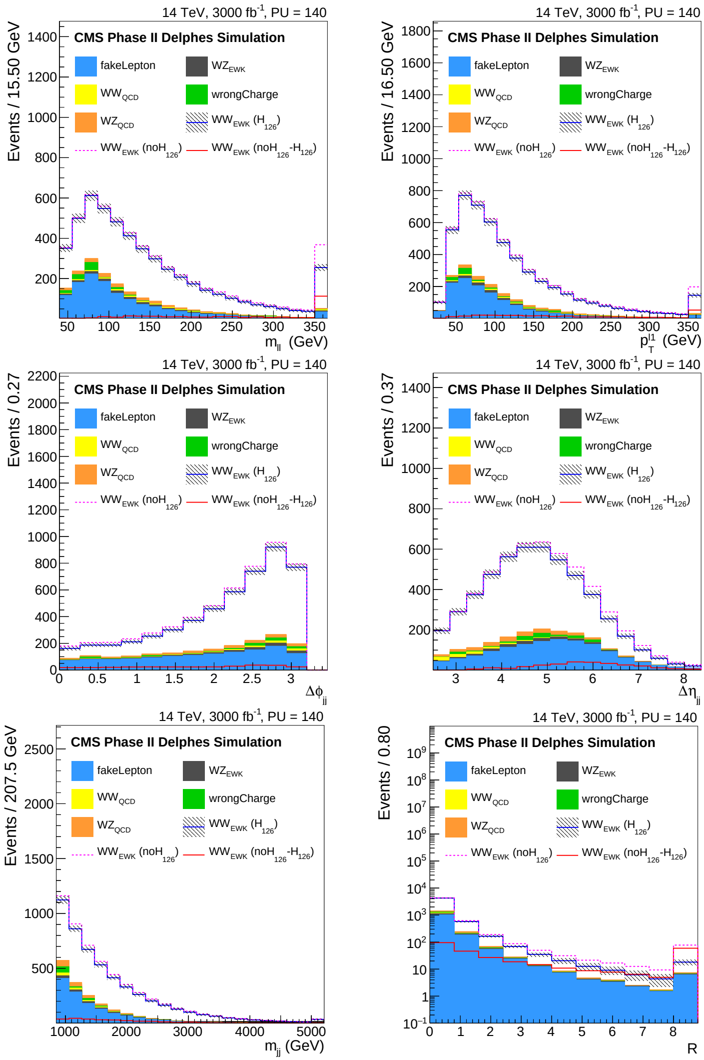


Figure 53: Distributions after the event selection with an integrated luminosity of 3 ab^{-1} and the Phase II detector configuration. The blue continuous line corresponds to the signal as predicted by the standard model, the dashed pink line corresponds to the signal process in absence of the Higgs boson, and the continuous red line is equal to the difference between the two.

The results are obtained by a maximum likelihood fit with the same methods as described in Section 5.6.1, relying on the asymptotic behavior of the test statistic defined in Section 4.6.
For each benchmark the fit is performed simultaneous in all the final state categories. A one-dimensional fit is performed on m_{ll} for the anomalous coupling measurement, while the results for the other benchmarks are obtained by a two-dimensional fit using the pair of variables that maximizes the analysis sensitivity. A large set of variables are tested, with the main ones listed in Table 17.

An example of the two-dimensional scan to determine the pair of variables that gives the highest significance for the longitudinal scattering measurement is shown in Figure 54. This procedure is repeated for the other benchmarks and leads to the pair of variables that maximizes the sensitivity for each benchmark, listed in Table 18.
The final results are presented as a function of the integrated luminosity with the non-prompt scale factor equal to one, and as function of a global scale factor to the non-prompt background estimate with the integrated luminosity set to 3 ab^{-1}.

Description	Definition
jet p_T	$p_T^{j_1}, p_T^{j_2}$
lepton p_T	$p_T^{l_1}, p_T^{l_2}$
R-variable	$R = (p_T^{l_1} \cdot p_T^{l_2})/(p_T^{j_1} \cdot p_T^{j_2})$
jet asymmetry	$asim_j = (p_T^{j_1} - p_T^{j_2})/(p_T^{j_1} + p_T^{j_2})$
lepton asymmetry	$asim_l = (p_T^{l_1} - p_T^{l_2})/(p_T^{l_1} + p_T^{l_2})$
VBS variables	$\Delta\eta_{jj}, m_{jj}$
missing energy value	E_T^{miss}
ϕ separation	$\Delta\phi_{ll}, \Delta\phi_{jj}, \Delta\phi_{ll,E_T^{miss}}, \Delta\phi_{ll,jj}$
Razor mass	$m_R = \sqrt{(E^{l_1} + E^{l_2})^2 - (p_z^{l_1} + p_z^{l_2})^2}$
Transverse Razor mass	$m_T^R = \sqrt{[E_T^{miss}(p_T^{l_1} + p_T^{l_2}) - \vec{E}_T^{miss} \cdot (\vec{p}_T^{l_1} + \vec{p}_T^{l_2})]/2}$

Table 17: The main variables considered for the 2-dimensional maximum likelihood fit in order to optimize the analysis sensitivity.
The *Razor* variables contain information on the mass of pair-produced heavy particles that decay to a detectable and undetectable part [136, 137].

6.7 SYSTEMATIC UNCERTAINTIES

The statistical procedure takes into account the systematic uncertainties in the analysis (Section 4.6). Shape systematic uncertainties modify each bin used in the fit, while global systematic uncertainties affect only the normalizations.

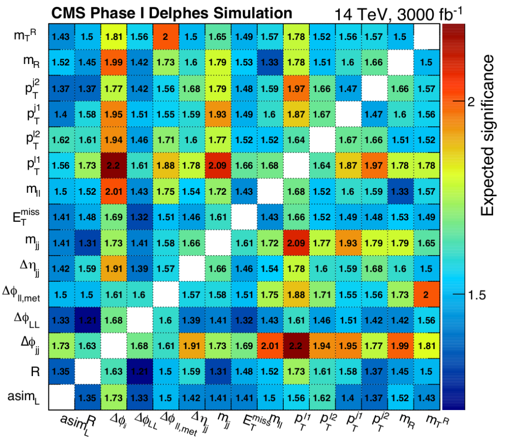


Figure 54: The expected significance of the longitudinal scattering measurement in the phase I scenario, using the pairs of the variables defined in Table 17. The highest significance is obtained when combining the $\Delta\phi_{jj}$ and $p_T^{l_1}$ variables.

Analysis benchmarks	Fit variables
Electroweak $W^{\pm}W^{\pm}jj$ cross section	R, m_{ll}
Longitudinal scattering cross section	$\Delta\phi_{jj}$, $p_T^{l_1}$
Partial unitarization	R, m_{ll}
Anomalous quartic gauge couplings	m_{ll}

Table 18: Overview of the variables used for each analysis benchmarks in the maximum likelihood fit.

The systematic uncertainties estimated for the Phase I detector are assumed to be identical for the Phase II scenario, as the performance of the object reconstruction is expected to be similar [2].
For the aged scenario, several systematic uncertainties are downgraded (related to the b-tagging, energy scale and energy resolution) according to the POG recommendations and from comparing the Phase II and aged detailed simulation samples.
The full list of systematic uncertainties that are considered and their relative effect on the event yields are summarized in Table 19.

The systematic uncertainties that rely on data for their estimation are taken from the analysis at Run I [113]. These are the uncertainty on the luminosity and on the data-driven estimate of the non-prompt and charge misassignment backgrounds. Also the uncertainty on the trigger, lepton reconstruction

and selection efficiencies and on the b-tag efficiency are taken from the Run I analysis.
The uncertainties on the energy scale and resolution of the objects are estimated with the same methods as described in Section 5.7 and applied as shape systematic uncertainties.
The QCD scale uncertainties are estimated by varying the renormalization and factorization scale by a factor two from their nominal value used in the simulation, while PDF uncertainties are taken from the Run I analysis. The effect of the theory uncertainties on the event selection efficiency is separated out for the signal process (called the signal acceptance uncertainty in Table 19).

Systematic sources	Detector scenarios		
	Phase I	Aged	Phase II
LHC luminosity	2.6%	2.6%	2.6%
jet energy scale	1–3%	1.5–4%	1–3%
jet energy resol.	5%	6.5%	5%
muon energy scale	1%	2%	1%
muon energy resol.	1%	2%	1%
electron energy scale	2%	4%	2%
electron energy resol.	2%	4%	2%
lepton efficiency	2%	2%	2%
b-tag efficiency	4%	5.5%	4%
lepton fake rate	30%	30%	30%
lepton wrong charge	30%	30%	30%
signal acceptance	2%	2%	2%
QCD scale	3-5%	3-5%	3-5%
PDF	5-7%	5-7%	5-7%

Table 19: The relative effect of the systematic uncertainties on the event yields for each of the detector scenarios.

6.8 RESULTS

6.8.1 *Electroweak* $W^{\pm}W^{\pm}jj$

The measurement of the electroweak $W^{\pm}W^{\pm}jj$ cross section with the largest expected significance is obtained with a two-dimensional fit on R and m_{ll}, with the R-variable defined as the product of the transverse momenta of the leptons divided by the product of the transverse momenta of the tag jets (Table 17). The expected uncertainty on the cross section measurement as a function of the fake rate scale factor and integrated luminosity is shown

in Figure 55. With 3 ab^{-1} of data, the expected uncertainty is about 6% for the Phase II and Phase I detector scenarios, and 8% for the aged detector scenario. This means the Phase II detector in the high luminosity environment recovers the performance of the Phase I detector at the current pileup conditions. As explained in Section 6.1.1, the Aged detector scenario is not a realistic detector scenario beyond 1000 fb^{-1} of data, but is used to identify the key areas to be addressed by the upgrade.
Note that the uncertainty as a function of luminosity is shown up to 6 ab^{-1} of collected data, while the expected luminosity that will be collected at the HL-LHC is between 3 and 4 ab^{-1} (Section 3.5). At this point, the uncertainty on the cross section measurement is dominated by the systematic uncertainties.

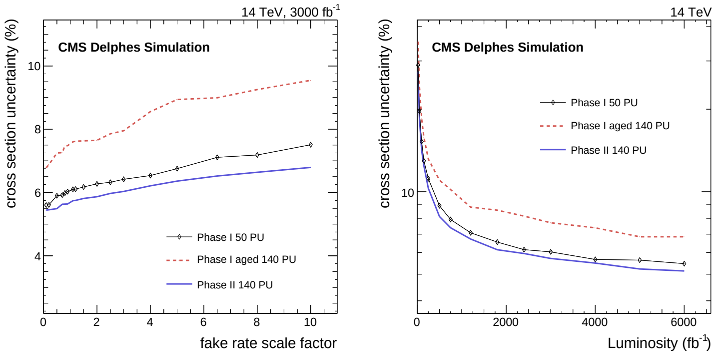


Figure 55: The expected uncertainty on the cross section measurement for the various detector scenarios, as a function of the fake rate scale factor with 3 ab^{-1} of data (left) and as a function of the integrated luminosity with unity scale factor (right).

6.8.2 *Longitudinal scattering*

Even with the large amount of data that will be collected at the HL-LHC, measuring the $W^\pm_L W^\pm_L \to W^\pm_L W^\pm_L$ process remains challenging. The component with two longitudinally polarized W bosons makes up only about 5% of the total electroweak $W^\pm W^\pm jj$ production (Section 2.3). The largest significance for the measurement of the longitudinal component was found with the two-dimensional fit on $\Delta\phi_{jj}$ and $p_T^{l_1}$. The normalized distributions in these variables for the longitudinal polarized (LL), transverse polarized (TT) and mixed state (TL) contributions are shown in Figure 56.
The significance of the measurement as a function of the fake rate scale factor and integrated luminosity is shown in Figure 57. A significance between 2σ and 3σ is expected after 3 ab^{-1} of data collected with the Phase II CMS

detector, which recovers the performance of the current Phase I detector and luminosity conditions.

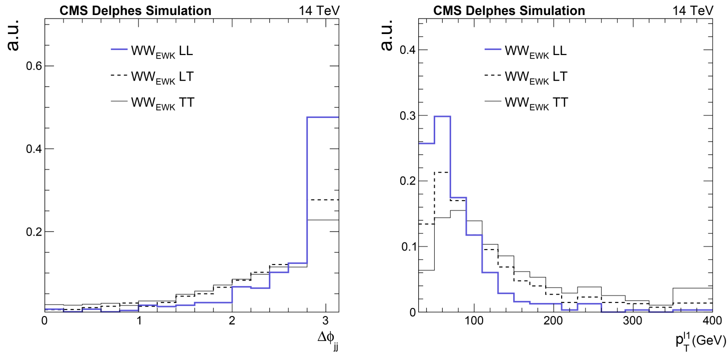


Figure 56: Normalized distributions of the LL, TL, and TT components of the electroweak $W^{\pm}W^{\pm}jj$ production in $\Delta\phi_{jj}$ and p_T^{l1} with the Phase II detector after the event selection.

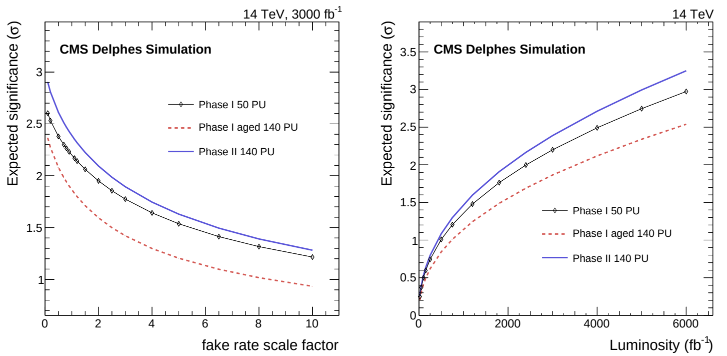
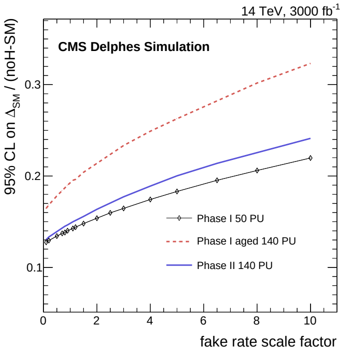


Figure 57: The expected significance for the $W^{\pm}_{L}W^{\pm}_{L} \rightarrow W^{\pm}_{L}W^{\pm}_{L}$ process in association with two jets for the various detector scenarios, as a function of the fake rate scale factor with 3 ab^{-1} of data (left) and as a function of the integrated luminosity with unity scale factor (right).

6.8.3 *Partial unitarization*

The sensitivity to models in which the Higgs boson only partially unitarizes the VBS process is studied. Instead of testing specific BSM models, a more general approach is chosen.
As a first step, we can exclude the model in which the Higgs boson does

not take part in the unitarization of the VBS process. The distributions of the electroweak $W^\pm W^\pm jj$ production in the no-Higgs model after the event selection are shown in Figure 53 with the pink dotted line. To test the sensitivity to partial unitarization, we define the difference between the SM electroweak $W^\pm W^\pm jj$ production and the same process in absence of the Higgs boson as the signal, and the SM process as background. An exclusion limit can be placed on the fraction of the difference between the no-Higgs and SM process (shown as the red line in Figure 53). The strongest limits are found with a two-dimensional fit on R and m_{ll}.
When an exclusion limit at 95% confidence level (CL) can be placed at 1 the no-Higgs model is excluded, while stronger limits provide a qualitative assessment of the potential to separate partially unitarized theories from the SM. The exclusion limits at 95% CL as a function of the fake rate scale factor and integrated luminosity are shown in Figure 58. The no-Higgs model is expected to be excluded already with the 300 fb^{-1} of data collected before the high luminosity upgrade. With 3 ab^{-1} of data, a sensitivity of about 0.15 of the difference between no-Higgs and SM process is expected with the Phase II detector scenario.



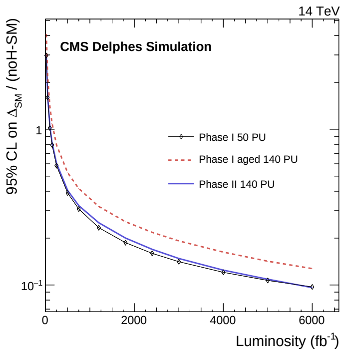


Figure 58: The expected limits at 95% CL on the fraction of the difference between the no-Higgs model and the SM, as a function of the fake rate scale factor with 3 ab^{-1} of data (left) and as a function of the integrated luminosity with unity scale factor (right).

A second quantitative study was added after the technical proposal, by rescaling the coupling between the Higgs boson and vector bosons κ_V. Similar to the previous study, we can then calculate the expected significance for measuring the electroweak $W^\pm W^\pm jj$ process with rescaled κ_V, with the SM process as background. As the signal diagrams containing an intermediate Higgs boson have an amplitude proportional to κ_V^2, the results are presented as a function of the squared coupling. The expected significance as a function of the rescaled squared coupling for the different detector scenarios is shown in Figure 59, together with the probability density of the test statistic for $\kappa_V^2 = 1$ and $\kappa_V^2 = 0.5$. Partially unitarization models with $\kappa_V^2 < 0.75$ are expected to be excluded at 95% CL for the Phase II detector scenario.

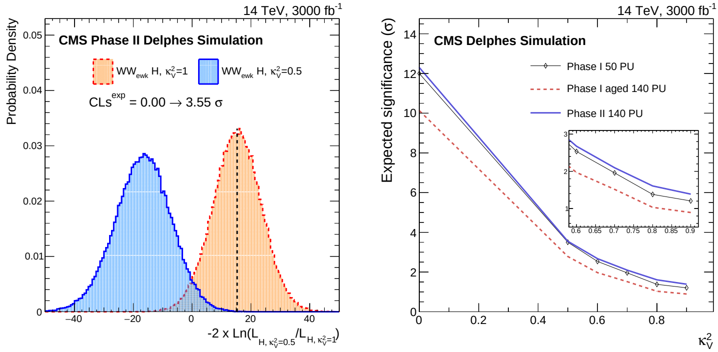


Figure 59: The test statistic for the partial unitarization hypothesis with $\kappa_V^2 = 0.5$ and the SM ($\kappa_V^2 = 1$) for the Phase II detector (left). The expected significance for the partial unitarization scenarios as a function of κ_V^2 for the three detector scenarios (right). Both with an integrated luminosity of 3 ab^{-1} and unity fake rate scale factor.

6.8.4 *Anomalous Quartic Gauge Couplings*

Exclusion limits are placed on the Wilson coefficients associated to the dimension eight operators in the SM EFT (Section 2.4.1.2) that modify the WW scattering cross section. The coefficients with dimension TeV^{-4} are scanned independently, assuming the other operators to be zero as in the SM. For the S_0 and S_1 operators, a two-dimensional scan is performed, while setting the other seven coefficients to zero.

The expected limits are determined from a fit to the m_{ll} distribution. The expected m_{ll} distributions for the SM signal and its modifications for two values of the coefficients associated with the S_0 and T_0 operators are shown in Figure 60 for the Phase II detector scenario.
The expected upper limits are shown in Table 20 for the different detector scenarios with an integrated luminosity of 3 ab^{-1} and unity scale factor of the fake rate. The evolution of the limit with the integrated luminosity is shown in Figure 61 for the S_1 operator, together with the two-dimensional limit on the S_0 and S_1 Wilson coefficients.
The limits with the Phase II detector scenario are 10% to 15% better than the limits obtained with the aged detector scenario.

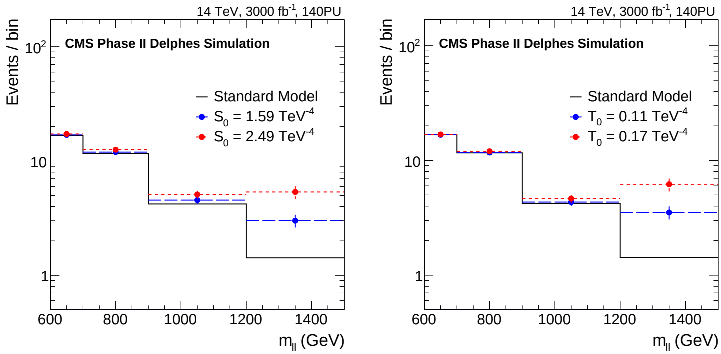


Figure 60: The m_{ll} distributions (with $m_{ll} > 600\,\text{GeV}$) for the electroweak $W^{\pm}W^{\pm}jj$ production and its modifications for two values of the coefficient associated with the S_0 (left) and T_0 (right) operators with the Phase II detector scenario.

	Phase I (TeV^{-4})	Phase II (TeV^{-4})	Aged (TeV^{-4})
S_0	2.47	2.49	2.85
S_1	8.19	8.25	9.45
M_0	1.88	1.76	2.03
M_1	2.54	2.38	2.72
M_6	3.78	3.54	4.05
M_7	3.42	3.24	3.75
T_0	0.17	0.17	0.19
T_1	0.078	0.070	0.080
T_2	0.25	0.23	0.25

Table 20: Expected 95% CL limits on the dimension-eight operator coefficients in the SM EFT Lagrangian for 3 ab^{-1} of data and unity scale factor of the fake rate for the three detector scenarios.

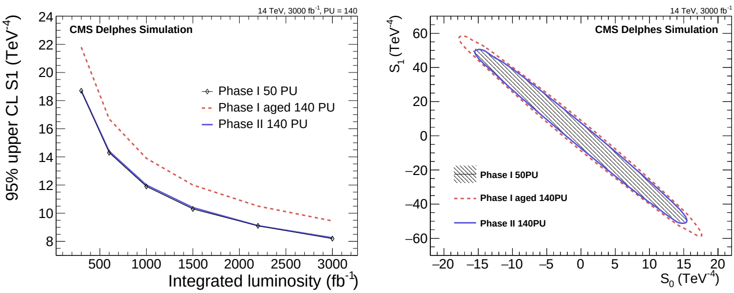


Figure 61: The expected 95% CL limits on the S_1 Wilson coefficient as a function of the integrated luminosity (left) and the 2D contour 95% CL limits on the coefficients associated to the S_0 and S_1 operators for 3 ab^{-1} of data (right), for the three detector scenarios and with unity scale factor of the fake rate.

7 CONCLUSIONS AND OUTLOOK

A measurement of the electroweak $W^{\pm}W^{\pm}jj$ production in the leptonic decay channel was described in Chapter 5, in data collected with the CMS detector throughout 2016. This was followed by a feasibility study of the same process at the High Luminosity LHC with the Phase II CMS detector in Chapter 6. Measurements of the VBS process, like the ones that are described in this thesis, are one of the primary goals of the LHC physics program, as they play an important role in determining the nature of the electroweak symmetry breaking mechanism and the self-interactions of the electroweak bosons.

The electroweak $W^{\pm}W^{\pm}jj$ production was measured in Run II CMS data at a center-of-mass energy of 13 TeV and with an integrated luminosity of 35.9 fb^{-1}. This measurement corresponds to the first discovery of a VBS-dominated process, with an observed (and expected) significance of $6.1\,\sigma$. Limits were set on anomalous quartic gauge couplings in an EFT framework. These measured limits are a factor 5 to 6 stronger than the limits obtained in the Run I measurement of the same process, and are the strongest limits obtained so far for some of these couplings. The large improvement in these limits is due to the increase in center-of-mass energy from 8 TeV to 13 TeV in 2015, which allows to probe a region that is more sensitive to new physics at a high energy scale.

In the near future, the analysis of the same-sign WW boson scattering in 2017 and 2018 Run II data could benefit from several improvements. The main one is the simulation of the signal with the full NLO corrections that were calculated recently. The dominant systematic uncertainties in the analysis originate from the theory uncertainties in the signal estimate, which shows the importance of including the NLO corrections in the theory calculations.

As more data is collected, splitting the events in charge and flavour final state categories could also improve the sensitivity. This has also been studied for the current analysis but did not improve the results due to limited statistics in the already pure high-m_{jj} region.

During the Run II and Run III LHC operation, it will also become possible to discover other electroweak $VVjj$ processes, with the most obvious candidates the $Z\gamma$ and WZ scattering processes that were measured with the highest expected significance at Run I. Public results with Run II data are available only for the ZZ scattering processes [116] so far, while measuments of other VBS processes are under study, including the first measurement of the W^+W^- scattering process.

The study of VBS after the high luminosity upgrade of the LHC is explored in the last chapter, with the Phase II CMS detector. It was found that the phase II detector, in the higher luminosity environment present during the HL-LHC operation, reaches the target of recovering the performance of the current detector (Phase I 50 PU). The tracker extension up to $\eta = 4.0$ is useful for VBS studies as it improves the selection and reconstruction of the tag jets that characterize the signal events, and can be used to reject backgrounds from WZ and misreconstructed leptons. Sensitivity was found to models in which the Higgs boson only partially unitarizes the vector boson scattering, and improvements are expected to the anomalous quartic gauge coupling limits. Detecting the longitudinal component of the electroweak $W^{\pm}W^{\pm}jj$ production in the leptonic decay channel remains challenging, with an expected significance of 2.4 σ. Another possibility, which was not mentioned in this study, would be to measure differential cross sections.

The results of the VBS study at the HL-LHC are a first estimate, and some approximations that were used, will have to be addressed when performing a data analysis. Several backgrounds will need to be estimated with a data-driven approach, an estimate for the $V\gamma$ background should be added, and the tau decays of W bosons need to be included in the simulation which can increase the expected signal yield by about 15%.
In general, one could expect that better results might be attained in future measurements at the HL-LHC. First of all, conservative estimates were used for the pixel detector performance. Also in the event reconstruction there is room for improvement; the object identification and reconstruction algorithms will be optimized in the next decade and the b-tagging algorithm can be extended up to $|\eta| = 4.0$, which was only applied up to $|\eta| < 2.5$ in these studies. Another improvement which is currently under study, is to use the enhanced time resolution of the upgraded detector to associate the neutral energy with specific vertices in the event. This could greatly reduce the impact of pileup on the object reconstruction, improve the identification of leptons and jets, and the di-jet invariant mass reconstruction.
Next to the experimental advancements, also improvements from the theory side can be expected; higher order calculations of the VBS processes can reduce some of the main systematic uncertainties.
Finally, studying the so far unexplored semi-leptonic decay channel of the electroweak $W^{\pm}W^{\pm}jj$ production is another promising opportunity to enhance the analysis significance. Although the signal over background ratio is smaller in this final state, recent developments in the study of boosted topologies [138, 139] and the possibility to reconstruct the complete final state add separation power to the measurement of the longitudinal scattering and anomalous couplings.

Part III

APPENDIX

A

VBF HIGGS TO B QUARKS

The event topology of Higgs production through vector boson fusion (VBF) is very similar to that of vector boson scattering. Instead of producing two outgoing vector bosons, the incoming bosons merge to a Higgs boson, which subsequently decays to two b quarks in about 60% of the decays. The main Feynman diagrams for this process are shown in Figure 62, with the final state consisting of four jets: two tag jets and two b jets from the Higgs decay.

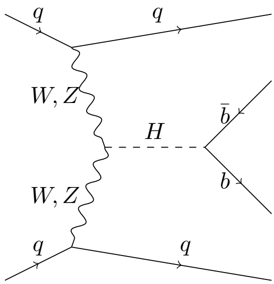


Figure 62: Feynman diagrams of the VBF-produced Higgs boson process decaying to b quarks, with the Higgs boson produced from two Z bosons or two oppositely charged W bosons.

The main challenge for measuring this process is the overwhelming QCD background that can not be estimated accurately from simulation. A first measurement of this process was performed in Run I data recorded with the CMS detector [1, 140], at a center-of-mass energy of 8 TeV and with an integrated luminosity of 19.8 fb^{-1}.
A brief overview of the analysis is given, with a focus on the statistical procedure and its validation method. The analysis is described here for the largest of two datasets that are used in the analysis. Both datasets are recorded during the same data-taking period, but rely on a different set of triggers.

ANALYSIS OVERVIEW

Measuring the small signal is only possible with a dedicated approach at all levels of the analysis. A topological trigger on four-jet events was developed to select the signal events. The trigger requires four jets passing p_T and b-tagging requirements, as well as selections on the invariant mass and the pseudorapidity difference of the tag jets. To identify the jets originating from the Higgs boson in the offline analysis, a boosted decision tree (BDT) [141] is trained on the η and b-tagging values of the jets. This is followed by the event selection listed in Table 21. The pair of most probable jets from the Higgs decay will be denoted by *bb* and the pair of most probable tag jets by *qq*.

Event selection
4 jets with $p_T^{1,2,3,4} > 80, 70, 50, 40\,\text{GeV}$ and $\lvert\eta\rvert < 4.5$
2 jets that pass a loose b-tagging requirement (CSV algorithm (Section 4.3.4))
$\Delta\phi_{bb} < 2.0$
$m_{qq} > 250$ GeV and $\Delta\eta_{qq} > 2.5$
lepton veto

Table 21: The event selection, with *bb* the pair of jets identified as originating from the Higgs decay and *qq* the identified pair of tag jets.

The event selection consists only of loose selections because the main selection of the signal region is performed with a boosted decision tree. This multivariate discriminant makes use of variables related to the kinematics of the tag jets, the angle between the qq and bb planes, b-tagging values, a quark-gluon discriminant and the hadronic activity in the event. The output of this BDT for the signal, backgrounds and data is shown in Figure 63.

The events are split in multiple categories according to the BDT output. A signal extraction method is developed for this analysis that is based on the search for a signal "bump" on top of the smoother QCD background shape in the m_{bb} distribution in each category. The BDT output values at which the boundaries are set, are optimized to maximize the expected significance while keeping sufficient statistics to fit the signal and background templates in each category. The categorization in the BDT output is shown in Figure 64.

The lowest BDT category (*CAT-1*) is not used in the statistical procedure because the signal extraction method requires the m_{bb} distribution to not depend strongly on the BDT category, which is not fulfilled for this category. This can be seen in Figure 65 that contains the average m_{bb} as a function of the BDT output.

The signal discrimination is enhanced further by a jet regression technique developed for this analysis that improves the energy scale and resolution of the invariant $b\bar{b}$ mass.

STATISTICAL PROCEDURE

The signal is measured with a parametric fit to the binned invariant mass distribution of the $b\bar{b}$ system in multiple categories. In the signal distribution

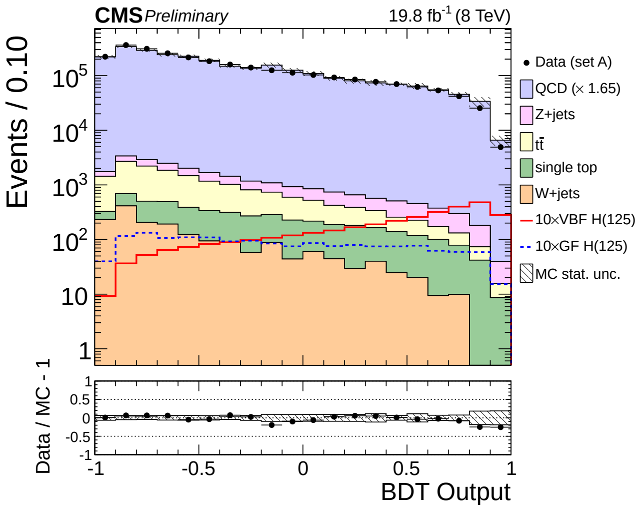


Figure 63: The distributions of the BDT output used to separate the signal and background regions. The VBF and gluon fusion (GF) Higgs boson signal have been multiplied by a factor 10 for a better visibility, the QCD simulation estimate is scaled to the data.

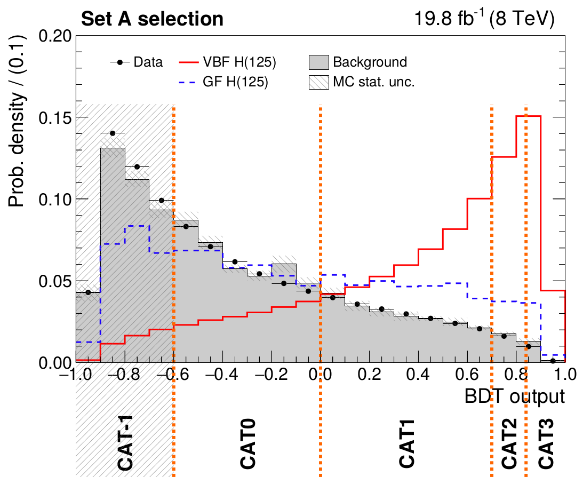


Figure 64: The normalized data, signal and background distributions in the BDT output, with the categories indicated. The plot is taken from [140].

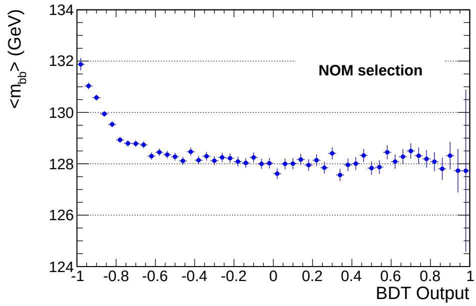


Figure 65: The average m_{bb} as a function of the BDT output. The region with a BDT output smaller than -0.6 is not used in the statistical procedure.

a bump at the Higgs mass is expected, while the large QCD background has a falling spectrum as a function of m_{bb}. The invariant mass distributions of the data and simulation, summed over all categories, are shown in Figure 66.

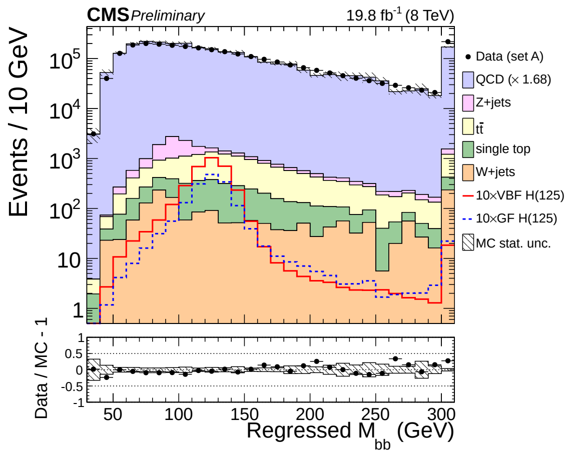


Figure 66: The m_{bb} distribution after the jet regression. The VBF and gluon fusion (GF) Higgs boson signal have been multiplied by a factor 10 for a better visibility, the QCD simulation estimate is scaled to the data.

In the fit model, the signal and backgrounds from the production of a Z or W boson (called the Z background) and from single top and $t\bar{t}$ production (called the top background) are modeled from simulation. The dominant QCD background can not be accurately determined from theory and has to be estimated from data. The QCD background model consists of a function that describes the global shape of the background, multiplied by a transfer function that is category dependent. As the m_{bb} distribution does not depend strongly on the BDT output, factorizing out this contribution strongly reduces the degrees of freedom needed to describe the QCD background in all categories. The fit model f_i for category i can be written as:

$$\begin{aligned} f_i(m_{bb}) = \mu \cdot N_{i,H} \cdot H_i(m_{bb}) + N_{i,Z} \cdot Z_i(m_{bb}) + \\ N_{i,Top} \cdot T_i(m_{bb}) + N_{i,QCD} \cdot K_i(m_{bb}) \cdot B(m_{bb}), \end{aligned} \tag{70}$$

with the signal strength μ the parameter of interest, and H, Z and T the parametric models that describe the signal, Z and top components with normalization N. The global QCD shape $B(m_{bb})$ is measured in the lowest BDT output category (CAT0) where the signal is negligible. The measured $B(m_{bb})$ template is identical in all categories of the fit model. The dependency of the QCD shape on the BDT discriminant is contained in the transfer function $K_i(m_{bb})$ which is measured from the m_{bb} ratio between the categories, blinded in the signal region. The measured transfer functions are shown in Figure 67. The functional form of the QCD shape $B(m_{bb})$ and transfer functions $K_i(m_{bb})$ are determined from a bias study, which is explained in the next section.

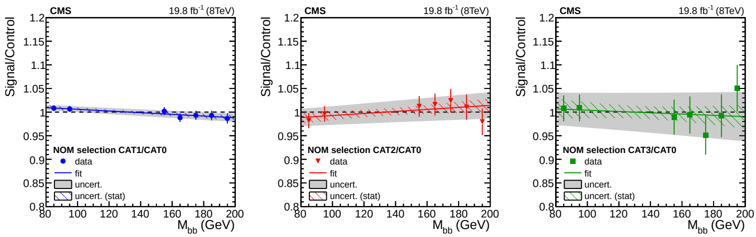


Figure 67: The transfer functions fitted to the ratio of the normalized m_{bb} distributions, for CAT1 to CAT3 divided by CAT0.

The QCD normalization $N_{i,QCD}$ is unconstrained in the fit, while the top and Z normalizations ($N_{i,Top}$ and $N_{i,Z}$) are allowed to float within 30% of their expected values. The signal normalization is modified by the unconstrained signal strength μ, while $N_{i,H}$ is fixed to its SM prediction. The models determined from simulation include parameters that describe their variation with the jet energy scale and resolution, which are free in the fit. The parameters of the QCD shape $B(m_{bb})$ are allowed to float according their uncertainties, while the transfer functions $K_i(m_{bb})$ parameters are unconstrained in the fit. The fit results to data in all categories are shown in Figure 68.

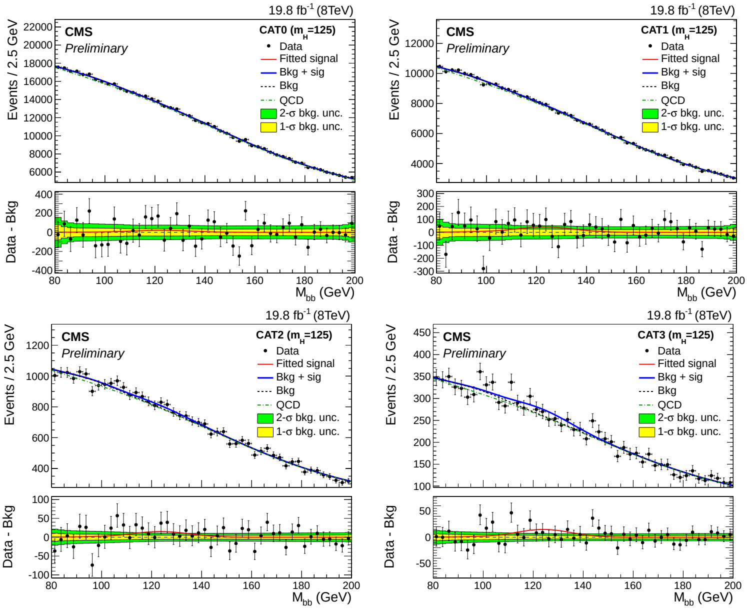


Figure 68: The total fit model (in blue) and the separate components after fitting to the m_{bb} distribution in all categories, with CAT0 the least sensitive and CAT3 the most sensitive category. In the bottom panel the background subtracted distribution is shown, including the fitted signal and the 1σ and 2σ background uncertainty bands.

BIAS STUDY

As the underlying function that describes the QCD background can not be determined from theory arguments, a functional form has to be chosen. Ideally, the background model would be general enough to accurately describe the true underlying model (which is unknown), while at the same time not being able to absorb the signal contribution as well. Also for the transfer functions, a parametric model needs to be chosen. The optimal functions are determined from data with a bias study, meaning the functional form is chosen that minimizes the potential bias in the fit.
In order to estimate the potential bias on the parameter of interest μ, different classes of functional forms that can describe the QCD background shape $B(m_{bb})$ are tested, which are listed in Table 22. For each possible $B(m_{bb})$ model 1000 pseudo-dataset are generated with the full model (Equation 70). The same fit model, but with a different function for $B(m_{bb})$, is then fitted to each of the pseudo-datasets with a simultaneous maximum likelihood fit in all categories.

The signal bias is defined as the difference between the measured and in-

Function name	Parametrization
expPow(x)	$(x-30)^a \cdot \exp(-bx^c)$
modG(x)	$\exp(-ax) \cdot erfc(b-cx)$
tanh(x)	$a - tanh(bx-c)$
sine(x)	$1 + a \cdot sin(bx-c)$
Bernstein4-6(x)	$\sum_{\nu=0}^{n=4-6} \beta_\nu b_{\nu,n}(x)$, with $b_{\nu,n}(x) = \binom{n}{\nu} x^\nu (1-x)^{n-\nu}$

Table 22: The functional forms that are tested in the bias study of the QCD background shape $B(m_{bb})$. *Bernstein4-6* corresponds to polynomial functions of order 4 to 6, with the Bernstein form chosen for numerical stability.

jected signal.The signal bias is considered as acceptable when it is smaller than 20% of the uncertainty on the measured signal, quantified by the root mean square of the measured signal strength distribution.

A polynomial function of order 5 as $B(m_{bb})$ model gives the best results, with an average bias within 20% of the uncertainty on μ for all the $B(m_{bb})$ models (Table 22) used in the pseudo-dataset generation. The measured bias with the Bernstein5 QCD function in the fit model as a function of the injected signal strength is shown in Figure 69.
To optimize the fit window and bin size in m_{bb}, the bias study is repeated with different options for the fit window and bin size. The optimal settings are a fit window between 80 GeV and 200 GeV with a bin size of 0.5 GeV in m_{bb}. The functional form of the transfer functions $K_i(m_{bb})$ is decided on with a second bias study in which linear, quadratic polynomial and exponential functions are tested for $K_i(m_{bb})$. It is found that the linear transfer function with free parameters that is used in the fit model has an expected bias well within 20% of the total uncerainty.

RESULTS

The analysis is performed on two datasets that rely on a different set of triggers that were active during the same data-taking period. On the second, smaller dataset, the same analysis strategy is applied, but with small differences in the event selection and statistical procedure. The final results are derived from combining the four categories described in this chapter, and three additional categories from the second dataset.

The observed significance of the signal measurement, for a Higgs boson mass of 125 GeV, is 2.2 standard deviations, while the expected significance is 0.8σ. The fitted signal strength is $\mu = \sigma/\sigma_{SM} = 2.8^{+1.6}_{-1.4}$.

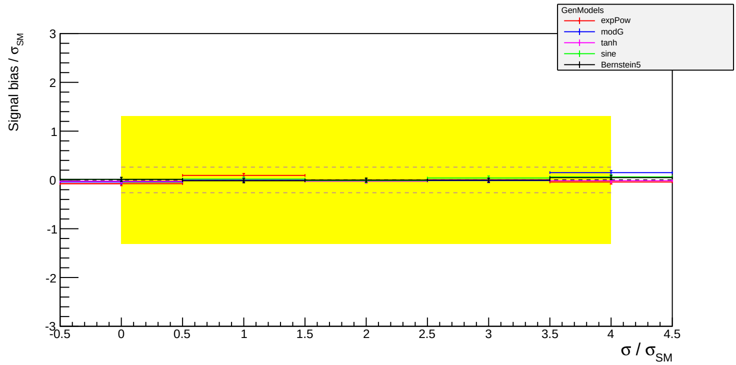


Figure 69: The average bias as a function of the injected signal strength (σ/σ_{SM}) for a fit model with $B(m_{bb})$ a 5th order Bernstein polynomial. The legend shows the functions that were used for $B(m_{bb})$ while generating the pseudo-datasets. The root mean square of the measured signal distribution is shown by the yellow band, and σ_{SM} denotes the signal cross section as predicted by the standard model.

In addition, upper limits are set at 95% confidence level on the signal cross section, for five different Higgs boson mass hypotheses (between 115 and 135 GeV). The expected and observed limits are shown in Figure 70. The expected limits are shown in the case there is no Higgs boson and in the case of a Higgs boson with a mass of 125 GeV.

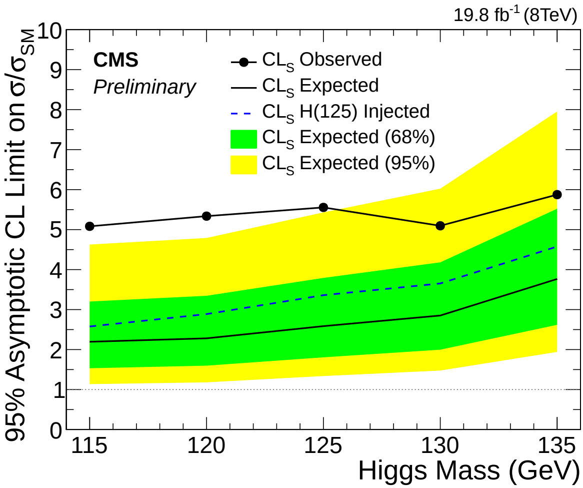


Figure 70: The observed and expected 95% confidence level limits on the signal cross section in units of the SM expected cross section, as a function of the Higgs boson mass. The limits expected in the presence of a SM Higgs boson with a mass of 125 GeV are indicated by the dotted curve.

BIBLIOGRAPHY

[1] Vardan Khachatryan et al. "Search for the standard model Higgs boson produced through vector boson fusion and decaying to $b\bar{b}$." In: *Phys. Rev.* D92.3 (2015), p. 032008. DOI: 10.1103/PhysRevD.92.032008. arXiv: 1506.01010 [hep-ex] (cit. on pp. v, 131).

[2] D Contardo et al. *Technical Proposal for the Phase-II Upgrade of the CMS Detector*. Tech. rep. CERN-LHCC-2015-010. LHCC-P-008. CMS-TDR-15-02. Geneva, 2015. URL: https://cds.cern.ch/record/2020886 (cit. on pp. v, 43, 103, 106, 118).

[3] CMS Collaboration. *Prospects for the study of vector boson scattering in same sign WW and WZ interactions at the HL-LHC with the upgraded CMS detector*. Tech. rep. CMS-PAS-SMP-14-008. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2220831 (cit. on pp. vi, 103).

[4] Albert M Sirunyan et al. "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 $\sqrt{s} = 13$ TeV." In: (2017). Submitted to Phys. Rev. Lett. arXiv: 1709.05822 [hep-ex] (cit. on pp. vi, 65, 99, 100).

[5] Serguei Chatrchyan et al. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC." In: *Phys. Lett.* B716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex] (cit. on p. 3).

[6] Georges Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC." In: *Phys. Lett.* B716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex] (cit. on p. 3).

[7] S. L. Glashow. "Partial Symmetries of Weak Interactions." In: *Nucl. Phys.* 22 (1961), pp. 579–588. DOI: 10.1016/0029-5582(61)90469-2 (cit. on p. 5).

[8] Steven Weinberg. "A Model of Leptons." In: *Phys. Rev. Lett.* 19 (1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264 (cit. on p. 5).

[9] Abdus Salam. "Weak and Electromagnetic Interactions." In: *Conf. Proc.* C680519 (1968), pp. 367–377 (cit. on p. 5).

[10] W. Greiner and J. Reinhardt. *Field Quantization* (cit. on p. 5).

[11] F. Halzen en A.D. Martin. *Quarks and Leptons* (cit. on pp. 5, 6).

[12] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons." In: *Phys. Rev. Lett.* 13 (1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321 (cit. on p. 5).

[13] Peter W. Higgs. “Broken symmetries, massless particles and gauge fields.” In: *Phys. Lett.* 12 (1964), pp. 132–133. DOI: 10.1016/0031-9163(64)91136-9 (cit. on p. 5).

[14] Peter W. Higgs. “Broken Symmetries and the Masses of Gauge Bosons.” In: *Phys. Rev. Lett.* 13 (1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508 (cit. on p. 5).

[15] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. “Global Conservation Laws and Massless Particles.” In: *Phys. Rev. Lett.* 13 (1964), pp. 585–587. DOI: 10.1103/PhysRevLett.13.585 (cit. on p. 5).

[16] D. V. Schroeder M. E. Peskin. *An introduction to quantum field theory* (cit. on p. 6).

[17] F. Mandl and G. Shaw. *Quantum field theory*. 2nd ed. Wiley, 2010 (cit. on p. 6).

[18] C. S. Wu et al. “Experimental Test of Parity Conservation in Beta Decay.” In: *Phys. Rev.* 105 (4 1957), pp. 1413–1415. DOI: 10.1103/PhysRev.105.1413. URL: https://link.aps.org/doi/10.1103/PhysRev.105.1413 (cit. on p. 7).

[19] Gerard ’t Hooft and M.J.G. Veltman. “Regularization and Renormalization of Gauge Fields.” In: *Nucl.Phys.* B44 (1972), pp. 189–213. DOI: 10.1016/0550-3213(72)90279-9 (cit. on p. 8).

[20] Y. Fukuda et al. “Evidence for oscillation of atmospheric neutrinos.” In: *Phys. Rev. Lett.* 81 (1998), pp. 1562–1567. DOI: 10.1103/PhysRevLett.81.1562. arXiv: hep-ex/9807003 [hep-ex] (cit. on p. 10).

[21] Y. Fukuda et al. “Measurement of the flux and zenith angle distribution of upward through going muons by Super-Kamiokande.” In: *Phys. Rev. Lett.* 82 (1999), pp. 2644–2648. DOI: 10.1103/PhysRevLett.82.2644. arXiv: hep-ex/9812014 [hep-ex] (cit. on p. 10).

[22] Q. R. Ahmad et al. “Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory.” In: *Phys. Rev. Lett.* 89 (2002), p. 011301. DOI: 10.1103/PhysRevLett.89.011301. arXiv: nucl-ex/0204008 [nucl-ex] (cit. on p. 10).

[23] Michael Rauch. “Vector-Boson Fusion and Vector-Boson Scattering.” In: (2016). arXiv: 1610.08420 [hep-ph] (cit. on pp. 15, 19).

[24] Philipp Anger, Michael Kobel, and Sabine Lammers. “Probing Electroweak Gauge Boson Scattering with the ATLAS Detector at the Large Hadron Collider.” Presented 01 Sep 2014. PhD thesis. Dresden, Tech. U., 2014. URL: https://cds.cern.ch/record/1753849 (cit. on p. 17).

[25] Francisco Campanario et al. “Next-to-leading order QCD corrections to W^+W^+ and W^-W^- production in association with two jets.” In: *Phys. Rev.* D89.5 (2014), p. 054009. DOI: 10.1103/PhysRevD.89.054009. arXiv: 1311.6738 [hep-ph] (cit. on pp. 17, 18).

[26] Barbara Jager and Giulia Zanderighi. "NLO corrections to electroweak and QCD production of W+W+ plus two jets in the POWHEGBOX." In: *JHEP* 11 (2011), p. 055. DOI: 10.1007/JHEP11(2011)055. arXiv: 1108.0864 [hep-ph] (cit. on p. 17).

[27] Tom Melia et al. "W^+W^+ plus dijet production in the POWHEGBOX." In: *Eur. Phys. J.* C71 (2011), p. 1670. DOI: 10.1140/epjc/s10052-011-1670-x. arXiv: 1102.4846 [hep-ph] (cit. on p. 17).

[28] A. Denner, L. Hosekova, and S. Kallweit. "NLO QCD corrections to W+ W+ jj production in vector-boson fusion at the LHC." In: *Phys. Rev.* D86 (2012), p. 114014. DOI: 10.1103/PhysRevD.86.114014. arXiv: 1209.2389 [hep-ph] (cit. on p. 17).

[29] Benedikt Biedermann, Ansgar Denner, and Mathieu Pellen. "Large electroweak corrections to vector-boson scattering at the Large Hadron Collider." In: (2016). arXiv: 1611.02951 [hep-ph] (cit. on p. 18).

[30] Benedikt Biedermann et al. "Automation of NLO QCD and EW corrections with Sherpa and Recola." In: (2017). arXiv: 1704.05783 [hep-ph] (cit. on p. 18).

[31] Benedikt Biedermann, Ansgar Denner, and Mathieu Pellen. "Complete NLO corrections to W^+W^+ scattering and its irreducible background at the LHC." In: (2017). arXiv: 1708.00268 [hep-ph] (cit. on p. 18).

[32] Dieter Zeppenfeld. "Collider physics." In: *Neutrinos in physics and astrophysics from 10**(-33) to 10**28 CM. Proceedings, Conference, TASI'98, Boulder, USA, June 1-26, 1998*. 1999, pp. 303–350. arXiv: hep-ph/9902307 [hep-ph]. URL: http://alice.cern.ch/format/showfull?sysnb=0304571 (cit. on pp. 19, 105).

[33] David L. Rainwater, R. Szalapski, and D. Zeppenfeld. "Probing color singlet exchange in Z + two jet events at the CERN LHC." In: *Phys. Rev.* D54 (1996), pp. 6680–6689. DOI: 10.1103/PhysRevD.54.6680. arXiv: hep-ph/9605444 [hep-ph] (cit. on p. 20).

[34] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations." In: *JHEP* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph] (cit. on pp. 20, 48, 70).

[35] Duane A. Dicus and Vishnu S. Mathur. "Upper Bounds on the Values of Masses in Unified Gauge Theories." In: *Phys. Rev. D* 7 (10 1973), pp. 3111–3114. DOI: 10.1103/PhysRevD.7.3111. URL: https://link.aps.org/doi/10.1103/PhysRevD.7.3111 (cit. on p. 21).

[36] Benjamin W. Lee, C. Quigg, and H. B. Thacker. "Weak Interactions at Very High-Energies: The Role of the Higgs Boson Mass." In: *Phys. Rev.* D16 (1977), p. 1519. DOI: 10.1103/PhysRevD.16.1519 (cit. on p. 21).

[37] Vernon D. Barger et al. "Strong W+ W+ scattering signals at p p super-colliders." In: *1990 DPF Summer Study on High-energy Physics: Research Directions for the Decade (Snowmass 90) Snowmass, Colorado, June 25-July 13, 1990*. 1990 (cit. on p. 21).

[38] Michał Szleper. "The Higgs boson and the physics of WW scattering before and after Higgs discovery." In: (2014). arXiv: 1412.8367 [hep-ph] (cit. on pp. 21, 23).

[39] S. D. Rindani. "Strong gauge boson scattering at the LHC." In: *Physics at the Large Hadron Collider*. Ed. by Amitava Datta et al. 2009, pp. 145–155. DOI: 10.1007/978-81-8489-295-6_10. arXiv: 0910.5068 [hep-ph]. URL: https://inspirehep.net/record/835249/files/arXiv:0910.5068.pdf (cit. on p. 21).

[40] Michael S. Chanowitz and Mary K. Gaillard. "The TeV physics of strongly interacting W's and Z's." In: *Nuclear Physics B* 261 (1985), pp. 379 –431. ISSN: 0550-3213. DOI: http://dx.doi.org/10.1016/0550-3213(85)90580-2. URL: http://www.sciencedirect.com/science/article/pii/0550321385905802 (cit. on p. 21).

[41] Kingman Cheung, Cheng-Wei Chiang, and Tzu-Chiang Yuan. "Partially Strong WW Scattering." In: *Phys. Rev.* D78 (2008), p. 051701. DOI: 10.1103/PhysRevD.78.051701. arXiv: 0803.2661 [hep-ph] (cit. on p. 23).

[42] G. F. Giudice et al. "The Strongly-Interacting Light Higgs." In: *JHEP* 06 (2007), p. 045. DOI: 10.1088/1126-6708/2007/06/045. arXiv: hep-ph/0703164 [hep-ph] (cit. on p. 24).

[43] Marcela Carena, Leandro Da Rold, and Eduardo Pontón. "Minimal Composite Higgs Models at the LHC." In: *JHEP* 06 (2014), p. 159. DOI: 10.1007/JHEP06(2014)159. arXiv: 1402.2987 [hep-ph] (cit. on p. 24).

[44] B. Holdom. "The accidental Higgs boson." In: *Phys. Rev.* D90.1 (2014), p. 015004. DOI: 10.1103/PhysRevD.90.015004. arXiv: 1404.6229 [hep-ph] (cit. on p. 24).

[45] Vernon Barger et al. "CERN LHC phenomenology of an extended standard model with a real scalar singlet." In: *Phys. Rev. D* 77 (3 2008), p. 035005. DOI: 10.1103/PhysRevD.77.035005. URL: https://link.aps.org/doi/10.1103/PhysRevD.77.035005 (cit. on p. 24).

[46] G.C. Branco et al. "Theory and phenomenology of two-Higgs-doublet models." In: *Physics Reports* 516.1 (2012). Theory and phenomenology of two-Higgs-doublet models, pp. 1 –102. ISSN: 0370-1573. DOI: http://dx.doi.org/10.1016/j.physrep.2012.02.002. URL: http://www.sciencedirect.com/science/article/pii/S0370157312000695 (cit. on p. 24).

[47] G. Gounaris et al. "Triple gauge boson couplings." In: *AGS / RHIC Users Annual Meeting Upton, New York, June 15-16, 1995*. [,525(1996)]. 1996, pp. 525–576. arXiv: hep-ph/9601233 [hep-ph] (cit. on p. 24).

[48] S. Schael et al. "Electroweak Measurements in Electron-Positron Collisions at W-Boson-Pair Energies at LEP." In: *Phys. Rept.* 532 (2013), pp. 119–244. DOI: 10.1016/j.physrep.2013.07.004. arXiv: 1302.3415 [hep-ex] (cit. on p. 24).

[49] Ashutosh V. Kotwal, Heidi Schellman, and Jadranka Sekaric. "Review of Physics Results from the Tevatron: Electroweak Physics." In: *Int. J. Mod. Phys.* A30.06 (2015), p. 1541004. DOI: 10.1142/S0217751X15410043. arXiv: 1409.5163 [hep-ex] (cit. on p. 24).

[50] Daniel R. Green, Patrick Meade, and Marc-Andre Pleier. "Multi-Boson Interactions at the Run 1 LHC." In: *Submitted to: Rev. Mod. Phys.* (2016). arXiv: 1610.07572 [hep-ex] (cit. on p. 24).

[51] Chen Ning Yang. "Fermi's β-decay Theory." In: *Asia Pacific Physics Newsletter* 01.01 (2012), pp. 27–30. DOI: 10.1142/S2251158X12000045. eprint: http://www.worldscientific.com/doi/pdf/10.1142/S2251158X12000045. URL: http://www.worldscientific.com/doi/abs/10.1142/S2251158X12000045 (cit. on p. 24).

[52] Landon Lehman. "Extending the Standard Model Effective Field Theory with the Complete Set of Dimension-7 Operators." In: *Phys. Rev.* D90.12 (2014), p. 125023. DOI: 10.1103/PhysRevD.90.125023. arXiv: 1410.4193 [hep-ph] (cit. on p. 25).

[53] M. Baak et al. "Working Group Report: Precision Study of Electroweak Interactions." In: *Proceedings, 2013 Community Summer Study on the Future of U.S. Particle Physics: Snowmass on the Mississippi (CSS2013): Minneapolis, MN, USA, July 29-August 6, 2013*. 2013. arXiv: 1310.6708 [hep-ph]. URL: http://www.slac.stanford.edu/econf/C1307292/docs/EnergyFrontier/Electroweak-19.pdf (cit. on p. 26).

[54] O. J. P. Éboli, M. C. Gonzalez-Garcia, and J. K. Mizukoshi. "$pp \to jje^{\pm}\mu^{\pm}\nu\nu$." In: *Phys. Rev. D* 74 (7 2006), p. 073005. DOI: 10.1103/PhysRevD.74.073005. URL: https://link.aps.org/doi/10.1103/PhysRevD.74.073005 (cit. on p. 26).

[55] Ana Alboteanu, Wolfgang Kilian, and Juergen Reuter. "Resonances and Unitarity in Weak Boson Scattering at the LHC." In: *JHEP* 11 (2008), p. 010. DOI: 10.1088/1126-6708/2008/11/010. arXiv: 0806.4145 [hep-ph] (cit. on p. 27).

[56] Celine Degrande et al. "Monte Carlo tools for studies of non-standard electroweak gauge boson interactions in multi-boson processes: A Snowmass White Paper." In: *Proceedings, 2013 Community Summer Study on the Future of U.S. Particle Physics: Snowmass on the Mississippi (CSS2013): Minneapolis, MN, USA, July 29-August 6, 2013*. 2013. arXiv: 1309.7890 [hep-ph]. URL: https://inspirehep.net/record/1256129/files/arXiv:1309.7890.pdf (cit. on p. 27).

[57] Celine Degrande et al. “Effective Field Theory: A Modern Approach to Anomalous Couplings.” In: *Annals Phys.* 335 (2013), pp. 21–32. DOI: 10.1016/j.aop.2013.04.016. arXiv: 1205.4231 [hep-ph] (cit. on p. 27).

[58] Thomas Schörner-Sadenius. “The Large Hadron Collider - Background and History.” In: *The Large Hadron Collider: Harvest of Run 1*. Ed. by Thomas Schörner-Sadenius. 2015, pp. 1–26. DOI: 10.1007/978-3-319-15001-7_1 (cit. on p. 30).

[59] Lyndon Evans and Philip Bryant. “LHC Machine.” In: *JINST* 3 (2008), S08001. DOI: 10.1088/1748-0221/3/08/S08001 (cit. on p. 30).

[60] Esma Mobs. “The CERN accelerator complex. Complexe des accélérateurs du CERN.” In: (2016). General Photo. URL: https://cds.cern.ch/record/2197559 (cit. on p. 31).

[61] Andreas Hoecker. *Physics at the LHC Run-2 and Beyond*. Tech. rep. arXiv:1611.07864. Comments: Lecture notes from the 2016 European School of High-Energy Physics, 15-28 June 2016, Skeikampen, Norway (61 pages, 56 figures). 2016. URL: http://cds.cern.ch/record/2236645 (cit. on p. 32).

[62] G. Aad et al. “The ATLAS Experiment at the CERN Large Hadron Collider.” In: *JINST* 3 (2008), S08003. DOI: 10.1088/1748-0221/3/08/S08003 (cit. on p. 32).

[63] S. Chatrchyan et al. “The CMS Experiment at the CERN LHC.” In: *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004 (cit. on pp. 32, 40).

[64] K. Aamodt et al. “The ALICE experiment at the CERN LHC.” In: *JINST* 3 (2008), S08002. DOI: 10.1088/1748-0221/3/08/S08002 (cit. on p. 32).

[65] A. Augusto Alves Jr. et al. “The LHCb Detector at the LHC.” In: *JINST* 3 (2008), S08005. DOI: 10.1088/1748-0221/3/08/S08005 (cit. on p. 32).

[66] G. Anelli et al. “The TOTEM experiment at the CERN Large Hadron Collider.” In: *JINST* 3 (2008), S08007. DOI: 10.1088/1748-0221/3/08/S08007 (cit. on p. 32).

[67] O. Adriani et al. “The LHCf detector at the CERN Large Hadron Collider.” In: *JINST* 3 (2008), S08006. DOI: 10.1088/1748-0221/3/08/S08006 (cit. on p. 32).

[68] James Pinfold et al. *Technical Design Report of the MoEDAL Experiment*. Tech. rep. CERN-LHCC-2009-006. MoEDAL-TDR-001. 2009. URL: http://cds.cern.ch/record/1181486 (cit. on p. 32).

[69] G. Antchev et al. “Luminosity-independent measurements of total, elastic and inelastic cross-sections at $\sqrt{s} = 7$ TeV.” In: *Europhys. Lett.* 101 (2013), p. 21004. DOI: 10.1209/0295-5075/101/21004 (cit. on p. 32).

[70] W.J. Stirling. *private communication*. URL: http://www.hep.ph.ic.ac.uk/~wstirlin/plots/plots.html (cit. on p. 33).

[71] Juska Tomi Kaleva Pekkanen. *Jet Particology - Studying the Structure of Jets with the CMS Particle Flow Algorithm*. Tech. rep. CMS-CR-2017-139. Geneva: CERN, 2017. URL: http://cds.cern.ch/record/2264517 (cit. on p. 34).

[72] C. Patrignani et al. "Review of Particle Physics." In: *Chin. Phys.* C40.10 (2016), p. 100001. DOI: 10.1088/1674-1137/40/10/100001 (cit. on p. 36).

[73] Serguei Chatrchyan et al. "Description and performance of track and primary-vertex reconstruction with the CMS tracker." In: *JINST* 9.10 (2014), P10009. DOI: 10.1088/1748-0221/9/10/P10009. arXiv: 1405.6569 [physics.ins-det] (cit. on pp. 37, 38, 51).

[74] V Karimäki et al. *The CMS tracker system project: Technical Design Report*. Technical Design Report CMS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/368412 (cit. on p. 37).

[75] A Benaglia. "The CMS ECAL performance with examples." In: *Journal of Instrumentation* 9.02 (2014), p. C02008. URL: http://stacks.iop.org/1748-0221/9/i=02/a=C02008 (cit. on p. 39).

[76] M. I. Pedraza-Morales, M. A. Shah, and M. Shopova. "First results of CMS RPC performance at 13 TeV." In: *JINST* 11.12 (2016), p. C12003. DOI: 10.1088/1748-0221/11/12/C12003. arXiv: 1605.09521 [physics.ins-det] (cit. on p. 41).

[77] V. Halyo et al. "GPU Enhancement of the Trigger to Extend Physics Reach at the LHC." In: *JINST* 8 (2013), P10005. DOI: 10.1088/1748-0221/8/10/P10005. arXiv: 1305.4855 [physics.ins-det] (cit. on p. 42).

[78] J Butler et al. *CMS Phase II Upgrade Scope Document*. Tech. rep. CERN-LHCC-2015-019. LHCC-G-165. Geneva: CERN, 2015. URL: https://cds.cern.ch/record/2055167 (cit. on p. 43).

[79] T. Gleisberg et al. "Event generation with SHERPA 1.1." In: *JHEP* 02 (2009), p. 007. DOI: 10.1088/1126-6708/2009/02/007. arXiv: 0811.4622 [hep-ph] (cit. on p. 46).

[80] Andy Buckley et al. "General-purpose event generators for LHC physics." In: *Phys. Rept.* 504 (2011), pp. 145–233. DOI: 10.1016/j.physrep.2011.03.005. arXiv: 1101.2599 [hep-ph] (cit. on p. 46).

[81] V. N. Gribov and L. N. Lipatov. "Deep inelastic e p scattering in perturbation theory." In: *Sov. J. Nucl. Phys.* 15 (1972). [Yad. Fiz.15,781(1972)], pp. 438–450 (cit. on p. 47).

[82] Yuri L. Dokshitzer. "Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics." In: *Sov. Phys. JETP* 46 (1977). [Zh. Eksp. Teor. Fiz.73,1216(1977)], pp. 641–653 (cit. on p. 47).

[83] Guido Altarelli and G. Parisi. “Asymptotic Freedom in Parton Language.” In: *Nucl. Phys.* B126 (1977), pp. 298–318. DOI: 10.1016/0550-3213(77)90384-4 (cit. on p. 47).

[84] V. V. Sudakov. “Vertex parts at very high-energies in quantum electrodynamics.” In: *Sov. Phys. JETP* 3 (1956). [Zh. Eksp. Teor. Fiz.30,87(1956)], pp. 65–71 (cit. on p. 47).

[85] Bo Andersson et al. “Parton Fragmentation and String Dynamics.” In: *Phys. Rept.* 97 (1983), pp. 31–145. DOI: 10.1016/0370-1573(83)90080-7 (cit. on p. 47).

[86] B. R. Webber. “A QCD Model for Jet Fragmentation Including Soft Gluon Interference.” In: *Nucl. Phys.* B238 (1984), pp. 492–528. DOI: 10.1016/0550-3213(84)90333-X (cit. on p. 47).

[87] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. “PYTHIA 6.4 Physics and Manual.” In: *JHEP* 05 (2006), p. 026. DOI: 10.1088/1126-6708/2006/05/026. arXiv: hep-ph/0603175 [hep-ph] (cit. on p. 48).

[88] Torbjörn Sjöstrand et al. “An Introduction to PYTHIA 8.2.” In: *Comput. Phys. Commun.* 191 (2015), pp. 159–177. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph] (cit. on pp. 48, 70).

[89] Stefano Frixione, Paolo Nason, and Carlo Oleari. “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method.” In: *JHEP* 11 (2007), p. 070. DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph] (cit. on pp. 48, 70).

[90] Alessandro Ballestrero et al. “PHANTOM: A Monte Carlo event generator for six parton final states at high energy colliders.” In: *Comput. Phys. Commun.* 180 (2009), pp. 401–417. DOI: 10.1016/j.cpc.2008.10.005. arXiv: 0801.3359 [hep-ph] (cit. on pp. 48, 70).

[91] S. Agostinelli et al. “GEANT4: A Simulation toolkit.” In: *Nucl. Instrum. Meth.* A506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8 (cit. on pp. 48, 70).

[92] G. Boudoul et al. “Monte Carlo Production Management at CMS.” In: *J. Phys. Conf. Ser.* 664.7 (2015), p. 072018. DOI: 10.1088/1742-6596/664/7/072018 (cit. on p. 48).

[93] J. de Favereau et al. “DELPHES 3, A modular framework for fast simulation of a generic collider experiment.” In: *JHEP* 02 (2014), p. 057. DOI: 10.1007/JHEP02(2014)057. arXiv: 1307.6346 [hep-ex] (cit. on pp. 48, 108).

[94] A. M. Sirunyan et al. “Particle-flow reconstruction and global event description with the CMS detector.” In: *JINST* 12 (2017), P10003. DOI: 10.1088/1748-0221/12/10/P10003. arXiv: 1706.04965 [physics.ins-det] (cit. on p. 49).

[95] Wolfgang Adam et al. *Track Reconstruction in the CMS tracker*. Tech. rep. CMS-NOTE-2006-041. Geneva: CERN, 2006. URL: https://cds.cern.ch/record/934067 (cit. on p. 50).

[96] Serguei Chatrchyan et al. “Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV.” In: *JINST* 7 (2012), P10002. DOI: 10.1088/1748-0221/7/10/P10002. arXiv: 1206.4071 [physics.ins-det] (cit. on p. 51).

[97] Vardan Khachatryan et al. “Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV.” In: *JINST* 10.06 (2015), P06005. DOI: 10.1088/1748-0221/10/06/P06005. arXiv: 1502.02701 [physics.ins-det] (cit. on p. 53).

[98] Wolfgang Adam et al. *Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC*. Tech. rep. CMS-NOTE-2005-001. Geneva: CERN, 2005. URL: https://cds.cern.ch/record/815410 (cit. on p. 53).

[99] Gavin P. Salam. “Towards Jetography.” In: *Eur. Phys. J.* C67 (2010), pp. 637–686. DOI: 10.1140/epjc/s10052-010-1314-6. arXiv: 0906.1833 [hep-ph] (cit. on p. 55).

[100] “2015 ECAL detector performance plots.” In: (2015). URL: https://cds.cern.ch/record/2114735 (cit. on p. 56).

[101] Vardan Khachatryan et al. “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV.” In: *JINST* 12.02 (2017), P02014. DOI: 10.1088/1748-0221/12/02/P02014. arXiv: 1607.03663 [hep-ex] (cit. on p. 56).

[102] Henning Kirschenmann. *Jet performance in CMS*. Tech. rep. CMS-CR-2013-325. Geneva: CERN, 2013. URL: https://cds.cern.ch/record/1627818 (cit. on p. 56).

[103] Matteo Cacciari and Gavin P. Salam. “Pileup subtraction using jet areas.” In: *Phys. Lett.* B659 (2008), pp. 119–126. DOI: 10.1016/j.physletb.2007.09.077. arXiv: 0707.1378 [hep-ph] (cit. on p. 56).

[104] Serguei Chatrchyan et al. “Identification of b-quark jets with the CMS experiment.” In: *JINST* 8 (2013), P04013. DOI: 10.1088/1748-0221/8/04/P04013. arXiv: 1211.4462 [hep-ex] (cit. on pp. 57, 74).

[105] Vardan Khachatryan et al. “Reconstruction and identification of τ lepton decays to hadrons and ν_τ at CMS.” In: *JINST* 11.01 (2016), P01019. DOI: 10.1088/1748-0221/11/01/P01019. arXiv: 1510.07488 [physics.ins-det] (cit. on p. 58).

[106] Kenneth Bloom. “CMS software and computing for LHC Run 2.” In: *PoS* ICHEP2016 (2016), p. 185. arXiv: 1611.03215 [physics.ins-det] (cit. on p. 58).

[107] R. Brun and F. Rademakers. "ROOT: An object oriented data analysis framework." In: *Nucl. Instrum. Meth.* A389 (1997), pp. 81–86. DOI: 10.1016/S0168-9002(97)00048-X (cit. on p. 58).

[108] Olaf Behnke et al. *Data analysis in high energy physics: a practical guide to statistical methods*. Weinheim: Wiley-VCH, 2013. URL: https://cds.cern.ch/record/1517556 (cit. on p. 60).

[109] *Procedure for the LHC Higgs boson search combination in Summer 2011*. Tech. rep. CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11. Geneva: CERN, 2011. URL: http://cds.cern.ch/record/1379837 (cit. on p. 61).

[110] S. S. Wilks. "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses." In: *Annals Math. Statist.* 9.1 (1938), pp. 60–62. DOI: 10.1214/aoms/1177732360 (cit. on p. 62).

[111] Wouter Verkerke and David P. Kirkby. "The RooFit toolkit for data modeling." In: *eConf* C0303241 (2003). [,186(2003)], MOLT007. arXiv: physics/0306116 [physics] (cit. on p. 62).

[112] Lorenzo Moneta et al. "The RooStats Project." In: *PoS* ACAT2010 (2010), p. 057. arXiv: 1009.1003 [physics.data-an] (cit. on p. 62).

[113] Vardan Khachatryan et al. "Study of vector boson scattering and search for new physics in events with two same-sign leptons and two jets." In: *Phys. Rev. Lett.* 114.5 (2015), p. 051801. DOI: 10.1103/PhysRevLett.114.051801. arXiv: 1410.6315 [hep-ex] (cit. on pp. 65, 68, 118).

[114] Georges Aad et al. "Evidence for Electroweak Production of $W^{\pm}W^{\pm}jj$ in pp Collisions at $\sqrt{s} = 8$ TeV with the ATLAS Detector." In: *Phys. Rev. Lett.* 113.14 (2014), p. 141803. DOI: 10.1103/PhysRevLett.113.141803. arXiv: 1405.6241 [hep-ex] (cit. on p. 65).

[115] Georges Aad et al. "Measurements of $W^{\pm}Z$ production cross sections in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector and limits on anomalous gauge boson self-couplings." In: *Phys. Rev.* D93.9 (2016), p. 092004. DOI: 10.1103/PhysRevD.93.092004. arXiv: 1603.02151 [hep-ex] (cit. on p. 65).

[116] Albert M Sirunyan et al. "Measurement of vector boson scattering and constraints on anomalous quartic couplings from events with four leptons and two jets in proton-proton collisions at $\sqrt{s} = 13$ TeV." In: (2017). arXiv: 1708.02812 [hep-ex] (cit. on pp. 65, 127).

[117] Vardan Khachatryan et al. "Measurement of electroweak-induced production of Wγ with two jets in pp collisions at $\sqrt{s} = 8$ TeV and constraints on anomalous quartic gauge couplings." In: *JHEP* 06 (2017), p. 106. DOI: 10.1007/JHEP06(2017)106. arXiv: 1612.09256 [hep-ex] (cit. on p. 65).

[118] Vardan Khachatryan et al. "Measurement of the cross section for electroweak production of $Z\gamma$ in association with two jets and constraints on anomalous quartic gauge couplings in proton–proton collisions at $\sqrt{s} = 8$ TeV." In: *Phys. Lett.* B770 (2017), pp. 380–402. DOI: 10.1016/j.physletb.2017.04.071. arXiv: 1702.03025 [hep-ex] (cit. on p. 65).

[119] Morad Aaboud et al. "Studies of $Z\gamma$ production in association with a high-mass dijet system in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector." In: *JHEP* 07 (2017), p. 107. DOI: 10.1007/JHEP07(2017)107. arXiv: 1705.01966 [hep-ex] (cit. on p. 65).

[120] Vardan Khachatryan et al. "Evidence for exclusive $\gamma\gamma \to W^+W^-$ production and constraints on anomalous quartic gauge couplings in pp collisions at $\sqrt{s} = 7$ and 8 TeV." In: *JHEP* 08 (2016), p. 119. DOI: 10.1007/JHEP08(2016)119. arXiv: 1604.04464 [hep-ex] (cit. on p. 65).

[121] Morad Aaboud et al. "Measurement of exclusive $\gamma\gamma \to W^+W^-$ production and search for exclusive Higgs boson production in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector." In: *Phys. Rev.* D94.3 (2016), p. 032011. DOI: 10.1103/PhysRevD.94.032011. arXiv: 1607.03745 [hep-ex] (cit. on p. 65).

[122] Morad Aaboud et al. "Search for anomalous electroweak production of WW/WZ in association with a high-mass dijet system in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector." In: *Phys. Rev.* D95.3 (2017), p. 032001. DOI: 10.1103/PhysRevD.95.032001. arXiv: 1609.05122 [hep-ex] (cit. on p. 65).

[123] S. Lee et al. "Common analysis object definitions and trigger efficiencies for the H $\to$ WW analysis with 2016 full data." CMS AN-17-082 (cit. on p. 73).

[124] G. Bauer et al. "Study of data-driven methods for estimation of fake lepton backgrounds." CMS AN-2009/120 (cit. on p. 76).

[125] H. Bakhshian et al. "Computing the contamination from fakes in leptonic final states." CMS AN-2010/261 (cit. on pp. 78, 82).

[126] F. Golf et al. "Studies of methods to estimate the non-prompt lepton background to searches for new physics." CMS AN-2014/261 (cit. on p. 80).

[127] S van der Meer. *Calibration of the effective beam height in the ISR*. Tech. rep. CERN-ISR-PO-68-31. ISR-PO-68-31. Geneva: CERN, 1968. URL: https://cds.cern.ch/record/296752 (cit. on p. 94).

[128] *CMS Luminosity Measurements for the 2016 Data Taking Period*. Tech. rep. CMS-PAS-LUM-17-001. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2257069 (cit. on p. 94).

[129] Jon Butterworth et al. "PDF4LHC recommendations for LHC Run II." In: *J. Phys.* G43 (2016), p. 023001. DOI: 10.1088/0954-3899/43/2/023001. arXiv: 1510.03865 [hep-ph] (cit. on p. 95).

[130] CMS collaboration. *Limits on anomalous triple and quartic gauge couplings*. (non-public page). December 2017. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSMPaTGC (cit. on p. 101).

[131] Yanou Cui and Zhenyu Han. "New light on WW scattering at the LHC with W jet tagging." In: *Phys. Rev.* D89.1 (2014), p. 015019. DOI: 10.1103/PhysRevD.89.015019. arXiv: 1304.4599 [hep-ph] (cit. on p. 104).

[132] Johan Alwall et al. "MadGraph 5 : Going Beyond." In: *JHEP* 06 (2011), p. 128. DOI: 10.1007/JHEP06(2011)128. arXiv: 1106.0522 [hep-ph] (cit. on p. 107).

[133] J. Pumplin et al. "New generation of parton distributions with uncertainties from global QCD analysis." In: *JHEP* 07 (2002), p. 012. DOI: 10.1088/1126-6708/2002/07/012. arXiv: hep-ph/0201195 [hep-ph] (cit. on p. 107).

[134] Jun Gao et al. "CT10 next-to-next-to-leading order global analysis of QCD." In: *Phys. Rev.* D89.3 (2014), p. 033009. DOI: 10.1103/PhysRevD.89.033009. arXiv: 1302.6246 [hep-ph] (cit. on p. 107).

[135] Vardan Khachatryan et al. "Search for the associated production of the Higgs boson with a top-quark pair." In: *JHEP* 09 (2014). [Erratum: JHEP10,106(2014)], p. 087. DOI: 10.1007/JHEP09(2014)087, 10.1007/JHEP10(2014)106. arXiv: 1408.1682 [hep-ex] (cit. on p. 111).

[136] Christopher Rogan. "Kinematical variables towards new dynamics at the LHC." In: (2010). arXiv: 1006.2727 [hep-ph] (cit. on p. 117).

[137] *Search for supersymmetry with the razor variables at CMS*. Tech. rep. CMS-PAS-SUS-11-008. Geneva: CERN, 2011. URL: https://cds.cern.ch/record/1404167 (cit. on p. 117).

[138] Vardan Khachatryan et al. "Identification techniques for highly boosted W bosons that decay into hadrons." In: *JHEP* 12 (2014), p. 017. DOI: 10.1007/JHEP12(2014)017. arXiv: 1410.4227 [hep-ex] (cit. on p. 128).

[139] John Stupak. *ATLAS+CMS Boosted topologies (Run1 results, Run2 potential)*. Tech. rep. CMS-NOTE-2015-005. CERN-CMS-NOTE-2015-005. Geneva: CERN, 2015. URL: https://cds.cern.ch/record/2043172 (cit. on p. 128).

[140] Sara Alderweireldt. "Search for the Standard Model Higgs boson produced via vector boson fusion and decaying to bottom quarks with the CMS detector at the Large Hadron Collider." PhD thesis. 2016. URL: http://hdl.handle.net/10067/1321240151162165141 (cit. on pp. 131, 133).

[141] Andreas Hocker et al. "TMVA - Toolkit for Multivariate Data Analysis." In: *PoS* ACAT (2007), p. 040. arXiv: physics/0703039 [PHYSICS] (cit. on p. 132).

ACKNOWLEDGEMENTS

I would like to express my gratitude to the people that I worked with and that supported me during the last years.

First of all, thanks to Xavier, the person I worked most closely together with during my PhD, and who I learned a lot from. Also thanks to Nick for introducing me to Higgs physics and for providing guidance.

Thank you to all the people I worked together with throughout the last years. It has always surprised me how friendly and helpful people have been. Thanks to the people that shared an office with me (most of the time Sara, Davide and Xavier), and also to all other colleges, who made the time spend at the University not feel like work.

Last but not least, thanks to my family and friends for their support. In particular I want to thank Deniz, who was always there for me, and with whom I hope to share many more happy days.

"Alas", said the mouse, "the whole world is growing smaller every day. At the beginning it was so big that I was afraid, I kept running and running, and I was glad when I saw walls far away to the right and left, but these long walls have narrowed so quickly that I am in the last chamber already, and there in the corner stands the trap that I am running into."
"You only need to change your direction," said the cat, and ate it up.

— Franz Kafka

BRIEF ACADEMIC CV

PUBLICATIONS

- On a Wilson lines approach to the study of jet quenching, I.O. Cherednikov, J. Lauwers, P. Taels, THE EUROPEAN PHYSICAL JOURNAL C (2014) 74:2721
- Search for the standard model Higgs boson produced through vector boson fusion and decaying to b bbar, CMS Collaboration, arXiv:1506.01010, Phys.Rev. D92 (2015) no.3, 032008
- 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 $\sqrt{s} = 13$ TeV, CMS Collaboration, arXiv:1709.05822, Submitted to Phys. Rev. Lett. (2017)

TEACHING

- Programming for physicist (3rd Bachelor), Exercises, 2nd semester 2017 and 2016
- Introduction to programming (2nd Bachelor), Exercises, 2nd semester 2015 and 2014
- Introduction to ROOT, 6/10 and 23/10 2014
- Masterclass particle physics: Particle physics introduction for high school students, UA, 12/03/2016, 4/11/2015, 22/03/2014

CMS RESPONSIBILITIES

Analyses:

- Measurement of VBF H$\rightarrow$ $b\overline{b}$ in Run I data,
  "Analysis modifications after the ARC review" (Nov 2013)
- Feasibility study of VBS at the HL-LHC,
  Pre-approval talk for TP (Apr 2015),
  Pre-approval (Feb 2016) and Approval talk for PAS (Sep 2016)
- Measurement of electroweak $W^{\pm}W^{\pm}jj$ in Run II data,
  Approval talk (May 2017)

Service tasks:

- Data quality monitoring contact for the Higgs triggers, 2015-2017
- Computing Operations Workflow shifts, 2014
- Development of high level trigger monitoring for the VBFHbb analysis, 2014
- Central Detector Control System shifts, 2014 - 2016

EXTENDED RESEARCH STAYS

- Three month research stay at CERN (Geneva):
  Vector boson scattering studies in view of the high luminosity LHC upgrade, 19/10/2014 - 20/12/2014, 4/1/2015 - 6/2/2015

CONFERENCES AND SCHOOLS

- VBSCan meeting: EU programme (COST) on VBS, Split, Croatia, 28-30/06/2017
- Presentation at conference: "CMS HL-LHC Multi-boson Physics Prospect", Multi-Boson Interactions 2016, Wisconsin, Madison, United States, 24-26/08/2016,
- Poster presentation at colloquium "Particle physics after the 2013 Nobel prize", $H \rightarrow WW^* \rightarrow 2l\,2\nu$ with the CMS experiment at the LHC, Brussels, 28/03/2015
- BND school: Lectures on subjects within experimental and theoretical particle physics, exercises and presentations, Kerkrade (NL) Nikhef, 25/08/2014 - 5/09/2014
- CMSDAS: Exercise sessions on the CMS software, CERN (Genève), 13-18/01/2014
- MPI @ LHC 2013, Workshop on Multi-Parton Interactions at the LHC, Antwerp, 2-6/12/2013
- BND school, Brussel (ULB), 25/08/2013 - 4/09/2013