UNIVERSITÀ DEGLI STUDI DI TRIESTE

XXXIII CICLO DEL DOTTORATO DI RICERCA IN

FISICA

MEASUREMENT OF THE Wγγ PRODUCTION CROSS SECTION IN pp COLLISIONS AT 13 TEV WITH FULL RUN 2 DATA AND LIMITS ON ANOMALOUS QUARTIC GAUGE COUPLINGS

Settore scientifico-disciplinare: FIS/01 FISICA SPERIMENTALE

DOTTORANDO
ALESSANDRO DA ROLD

COORDINATORE
PROF. FRANCESCO LONGO

SUPERVISORE DI TESI
PROF. GIUSEPPE DELLA RICCA

ANNO ACCADEMICO 2019/2020

Università degli Studi di Trieste

Dipartimento di Fisica

PhD programme in Physics - XXXIII Cycle

PhD Thesis

Measurement of the $W\gamma\gamma$ production cross section in pp collisions at 13 TeV with full Run2 data and limits on anomalous quartic gauge couplings

Candidate:
Alessandro Da Rold

Supervisor:
Prof. Giuseppe Della Ricca

Chair of the PhD Programme:
Prof. Francesco Longo

ACADEMIC YEAR 2019–2020

INTRODUCTION

The importance of the study of the $W\gamma\gamma$ production in proton-proton collisions is manifold. The measurement of the production cross section constitutes a precision test of the Standard Model since it is sensitive to the self interaction of electroweak bosons. This final state is also one of the possible background sources in the searches for new particles that can mediate multi-boson interactions and hence deserves to be studied in detail. Moreover, the process can be interpreted in the framework of the Effective Field Theories as it is sensitive to the presence of anomalous gauge couplings and can probe the presence of effects that are beyond the Standard Model. Being the $W\gamma\gamma$ production a rare process, a dataset with high integrated luminosity is needed.

In the past, searches for anomalous quartic gauge couplings in multi-boson final states have been performed at LEP, for example by the ALEPH [1] and DELPHI [2] collaborations. The limits on the anomalous quartic gauge couplings in the framework of the Effective Field Theories have been studied in many different multi-boson final states at CMS and ATLAS. The study of the $W\gamma\gamma$ final state has been performed both by the CMS [3] and ATLAS [4] collaborations at a centre of mass energy of $8\,TeV$ with a dataset corresponding to an integrated luminosity of $19\,fb^{-1}$. The limiting factor on the measurement of the process cross section in these two analyses was the low statistics. For this reason, this measurement is performed using data collected by the CMS experiment during the whole Run 2 of LHC at a centre of mass energy of $13\,TeV$, corresponding to approximately $137\,fb^{-1}$. In this way, the uncertainty connected to the statistics can be substantially reduced and competitive limits on the values of the anomalous couplings are expected to be achievable.

For this analysis, only the electron and muon final states of the W decay, which are the ones with the lowest background contributions, have been taken into account. The event selection has been studied and all the required corrections to the Monte Carlo simulations needed to improve their agreement with data have been applied. The main background contributions to the $W\gamma\gamma$ final state, which are the misidentification of jets as photons and the misidentification of electrons as photons, have been studied in detail by developing a data-driven approach, validated in single-photon control regions. The data-driven approach gives a better prediction for these processes when compared to the Monte Carlo simulations. The contribution of the other, subleading, background sources has been evaluated with Monte Carlo simulations.

The presence of anomalous gauge couplings can be tested by generating Monte Carlo events where contributions linked to beyond the Standard Model effects are taken into account. The comparison of these new simulations with the distributions obtained from data can probe the presence of anomalous couplings or set new limits on their strength. For this reason, dedicated simulation samples have been studied and events have been produced.

The final measurement for the cross section and the extraction of the limits on the anomalous couplings have been performed taking into account many possible systematic uncertainties on the object reconstruction, on the correction factors and on the methods implemented and used. The results that have been obtained show a great improvement if compared to previous $W\gamma\gamma$ analyses and are competitive with more recent results in other multi-boson channels.

In the first chapter, a brief historical review of the theoretical and experimental discoveries that led to the formulation of the Standard Model is presented. Particular focus is placed on the electroweak sector and the multi-boson production, alongside the theoretical aspects of proton collisions. The unresolved issues of the Standard Model and prospects of searches for new physics in the effective field theory parametrisation are discussed.

The second chapter focuses on the experimental aspects of high energy collisions. After a brief introduction on phenomenological concepts, a detailed description of the Large Hadron Collider and of the Compact Muon Solenoid detector is given. In the third chapter, the object reconstruction performed by both online and offline algorithms is discussed. Close attention is given to the reconstruction of the particles that are fundamental for this analysis. The fourth chapter depicts the main aspects of the Monte Carlo simulations, with particular regard to the matrix element generation with MadGraph5_aMC@NLO and the hadronisation with Pythia 8.

The analysis workflow is described in chapter five. After the description of the datasets and simulations, the event selection is discussed. The study of the background sources to the $W\gamma\gamma$ measurement follows, with great focus on the description of the data-driven methods implemented in order to estimate the jets and electrons misreconstructed as photons. The review of the systematic uncertainties and their treatment precedes the measurement of the cross section and the discussion of the results. Finally, the last chapter describes the computation of the limits on anomalous quartic gauge couplings in the effective field theory framework.

CONTENTS

The Standard Model is so complex it would be hard to put it on a T-shirt - though not impossible; you'd just have to write kind of small.
— Steven Weinberg

1 THE STANDARD MODEL OF PARTICLE PHYSICS

The Standard Model of particle physics is the most general theory to date able to describe the fundamental constituents of matter and their interactions. It has been developed starting from the 1930s and has been the standard theory of particle physics for the last fifty years. The theoretical predictions based on this theory have been marvelously validated by experimental results obtained with a wide range of designs, from the decay of heavy nuclei to the very high energy collisions studied at the CERN Large Hadron Collider.

In this chapter, a brief introduction to the historical motivation and theoretical and experimental paths that have been followed to get to the present day theory is presented. The electroweak and strong interactions are then discussed, with particular focus on the aspects that have the greatest impact on the collider measurements. A small discussion of the Higgs mechanism and the discovery of the Higgs boson is also given. Eventually, a selection of the so far unexplained features of the model and the possibilities to study the presence of new physics is presented, with particular focus on the aspects to which this analysis is more sensitive to. The general phenomenology of multi-boson interactions is detailed.

1.1 THEORETICAL BACKGROUND

At the beginning of the 20th century, only the electron, the proton and the photon had been discovered and the structure of the atom was not the one accepted at the present time. While looking at the interactions among particles, the description of the electromagnetic interaction had been proved multiple times, while the only known process linked to the weak interaction was the nuclear β decay, which was believed to be a two-body decay of the form $N_1 \rightarrow N_2 + e^-$. However, the observation performed both by Otto Hahn and Lise Meitner [5–7] and by James Chadwick [8] showed a continuous spectrum for the decay electron, while a mono-energetic one was expected from a two body decay. Further observations showed that also the angular momentum was not conserved and the Pauli exclusion principle was violated.

Instead of accepting the failure of the conservation laws that had been verified so well up to that moment (as initially suggested by Niels Bohr [9]), Wolfgang Pauli proposed an original solution [10]: the β decay did not involve two particles in the final state, but three. From conservations laws, the third mysterious particle should have been neutral, weekly interacting and with spin 1/2. The particle was named *neutrino* by Enrico Fermi, after the discovery of the heavier (but still neutral) neutron by Chadwick in 1932 [11]. It has to be remarked the cutting edge character of the prediction of the existence of the neutrino: more than 20 years were needed before the experimental discovery from Frederick Reines and Clyde Cowan in 1956 [12].

In 1934 Fermi proposed a theoretical framework to describe the β decay of the neutron [13]. By postulating the existence of the neutrino, Fermi formulated the quantum field theory for fermion fields and introduced the creation and annihilation operators. He overcame also the idea that the final state particles of a nuclear decay had to be present in the nucleus itself from the beginning. By using an analogy with the electromagnetic interactions, Fermi exploited the idea of the electromagnetic potential A^μ produced by the current of charged particles j^μ as a starting point to introduce the quantised fermion fields of the electron $\psi_{(e)}(x)$ and the neutrino $\psi_{(\nu)}(x)$. By considering a static density for the nucleons and by using the isospin operators $\tau_\pm$, Fermi derived the Hamiltonian for the β decay as

$$H_I = \frac{G_F}{\sqrt{2}}\left[\tau_- \psi^\dagger_{(e)}(x)\psi_{(\nu)}(x) + \tau_+ \psi^\dagger_{(e)}(\nu)\psi_{(e)}(x)\right], \tag{1}$$

where G_F is the Fermi coupling constant and the symbol † indicates the hermitian adjoint.

The form of Equation 1 was subsequently rewritten in the most common V-A form which maximally violates the invariance under space inversion. This fact led to the hypothesis that maybe also the the parity or the charge inversion could be violated. The experimental proof of parity violation came soon after, with the experiment of Chien-Shiung Wu [14]. The known properties of universality and conservation allowed for a rewriting of the interaction Hamiltonian as

$$H_I = \frac{G_F}{\sqrt{2}} J^\mu(x) J^\dagger_\mu(x) \quad \text{with} \quad J^\mu(x) = l^\mu(x) + h^\mu(x), \tag{2}$$

where the leptonic part was written as a function of the fields of the known leptons as

$$l^\mu(x) = \bar{\nu}_{(e)}\gamma^\mu(1+\gamma_5)e(x) + \bar{\nu}_{(\mu)}\gamma^\mu(1-\gamma_5)\mu(x). \tag{3}$$

The determination of the hadronic part of the Hamiltonian has been done by exploiting some hypothesis on the conserved vector currents and on the universality of the interactions. In particular, Nicola Cabibbo wrote an explicit form for $h_\mu(x)$ by introducing the strangeness changing currents [15] as

$$h_\mu(x) = \cos\theta h^{(\Delta S=0)}_\mu(x) + \sin\theta h^{(\Delta S=1)}_\mu(x), \tag{4}$$

where the hadronic part is split among conserving and non-conserving parts and θ is known as the *Cabibbo angle*. In this way, the weak current was interpreted as part of a SU(3) octet, which matched the experimental evidences. While the conservation of the vector current led to the determination of the vector part of the weak hadronic current, no approximate axial symmetry seemed to be present in the spectrum of hadrons. The correct interpretation of the phenomenon came from the work of Yoichiro Nambu [16] that showed how the dynamics of strong interactions is approximately invariant under chiral transformations, but the symmetry is spontaneously broken. From this hypothesis, the pions are the corresponding Nambu-Goldstone bosons of the broken symmetry.

These arguments led Murray Gell-Mann [17] to the postulation of a group structure as

$$U(3) \times U(3)\ U(1) \times U(1) \times SU(3) \times SU(3)\,, \tag{5}$$

which corresponds to 18 (approximately) conserved currents and as many charges. Without going to the theoretical details of the description, some results can be directly derived from the group structure. The vector $U(1)$ group results in an exact symmetry from which the baryon number conservation law is derived. The chiral $SU(3) \times SU(3)$ part is spontaneously broken as $SU(3)_R \times SU(3)_L \to SU(3)_V$, where $SU(3)_V$ is known as the *flavour group*.

Despite being an elegant prescription and despite the fact that it led to many different predictions, it became clear from the beginning that the Fermi theory could not have been the complete description of the weak interactions. In 1936 Markus Fierz [18] proved that the cross section for the neutrino scattering on protons at high energies increases with the neutrino energies and the higher order terms in the expansion would exceed the lower order ones. This feature makes the theory *non-renormalisable* and hence it cannot be interpreted as a fundamental theory, but as an *effective field theory*.

The fundamental step to the formulation of the Standard Model was the introduction of gauge theories. Gauge invariance had been known since the development of electrodynamics, where a redundant set of variables had to be mathematically reduced in order to retain the physical interpretation. However, the transformation of the vector potential in classical electrodynamics is an example of an internal symmetry, which is a transformation that does not depend on space-time. The local version of the gauge symmetries is a prerogative of quantum mechanics and comes from the phase of the wave function or that of the quantum fields. The power of local gauge invariance has been proved for the first time by Vladimir Aleksandrovich Fock [19] who demonstrated that the electromagnetic interactions can be derived from Schrödinger equation as a consequence of gauge invariance.

The work that brought gauge theories to particle physics has been done by Chen Ning Yang and Robert Mills in 1954 [20]. This fundamental paper introduced the $SU(2)$ non-abelian gauge theory as a tentative description of the strong interactions. However, the first application was found in the description of the weak and electromagnetic interactions. Starting from the Hamiltonian introduced by Fermi, Julian Schwinger [21] proposed the idea of an intermediate vector boson in the interaction, hence rewriting the Hamiltonian as

$$H_I = g J^{\mu}(x) W^{-}{}_{\mu}(x) + h.c.\,, \tag{6}$$

where g is a dimensionless coupling constant. The driving principle to write such an equation was to express the weak interaction in a form as close as possible to the electromagnetic one. The main differences between Schwinger's hypothesis and electromagnetism are that the vector boson in this formulation is massive (the weak interaction has a limited range) and is charged. As for the Fermi hypothesis, however, this formulation of the weak interaction was again not renormalisable. By exploiting Yang-Mills theories, Schwinger assumed a $SU(2)$ structure for the weak interaction with two massive charged bosons and the photon coupled to V-A currents.

The key turning point in the development of the theory came from Sheldon Lee Glashow [22]. His work proposed for the first time a unified description of the weak and electromagnetic interactions based on a $SU(2) \times U(1)$ structure with two charged and two neutral gauge bosons. In this picture, the photon is a linear combination of the fields associated with $U(1)$ and the third generator of the $SU(2)$ group as a function of an angle θ.

Gauge invariance, however, requires that the vector bosons of the Yang-Mills theory have zero mass, which is the opposite of what seems to happen for the weak interaction. The solution came from the spontaneous symmetry breaking mechanism (Section 1.2.1) and allowed Steven Weinberg [23] and Abdus Salam [24] to give the final formulation of the Standard Model for the leptons.

The leading order divergences in the expansion of the Fermi theory were solved by introducing a mass term for the quarks. The next-to-leading order divergences were harder to remove. The experimental evidences that excluded transitions with $\Delta S = 2$ and the absence of observation of the flavour changing neutral current posed limits on the possible theory workarounds. The idea proposed by Glashow, Iliopoulos and Maini [25] proved to be the correct one: the model should have not three quarks, but four. The introduction of the charm quark allowed to rewrite the charged weak current as

$$J_\mu(x) = \bar{U}(x)\gamma_\mu(1-\gamma_5)\mathbf{C}D(x) \tag{7}$$

with

$$U = \begin{pmatrix} u \\ c \end{pmatrix} \quad D = \begin{pmatrix} d \\ s \end{pmatrix} \quad \mathbf{C} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \tag{8}$$

where θ is the Cabibbo angle. The introduction of a fourth quark meant that more diagrams should be added to the computation of the scattering amplitudes. If the flavour symmetry was exact, the contribution of the u-type diagrams would cancel that of the c-type ones, but, since the flavour symmetry is broken, this translates into a difference in the quark masses and sets an upper bound of few GeV for the value of the c quark mass. The final contribution in terms of the renormalisability of the electromagnetic and weak interactions came from the work of Gerard 't Hooft [26, 27].

All the theoretical developments just described led to the formulation of the Standard Model, that will be described in detail in the next section.

1.2 THE STANDARD MODEL

The present day formulation of the Standard Model is based on a $SU(2) \times U(1)$ group structure with four gauge bosons: the $W^{\pm}$, Z and the photon. By defining the three generators of $SU(2)$ as T_i with $i = 1, 2, 3$ and the generator of $U(1)$ as Y, it is possible to write the operator for the electric charge as $Q = T_3 + \frac{1}{2}Y$. In this way, the gauge group is defined.

The next step in order to construct the gauge theory is the definition of the matter fields. The basic unit for the structure of the theory is a family of spin-1/2 chiral fermions. The representation of the lepton fields is

$$\Psi_L^i(x) = \frac{1}{2}(1+\gamma_5)\begin{pmatrix}\nu_i(x)\\ \ell_i^-(x)\end{pmatrix} \quad i = 1,2,3\,, \tag{9}$$

where the index i refers to the electron, the muon or the tau. The right-hand components can be written as singlets in $SU(2)$ as

$$\ell_{iR}^-(x) = \frac{1}{2}(1-\gamma_5)\ell_i^-(x)\,. \tag{10}$$

It is important to note the absence of a right-handed field for the neutrino. This is due to the fact that, as of today, there are no experimental evidence of the presence of such a field. The consequences of the absence of the right-handed neutrino fields are the fact that the neutrino are massless and the leptonic numbers are separately conserved in the interactions. While the latter is an experimentally proven fact (as of now), the former is not. The oscillation of neutrinos, observed experimentally [28–30], is a clear sign that the mass of the neutrino is not zero and exposes the formulation of the Standard Model to a tension between theoretical framework and experimental results.

The hadronic part of the model is written as a function of the elementary quark fields. As it is known both from prediction and experiments, all the quarks are massive (differently than in the theory of leptons, where neutrinos are not) so both left and right-handed fields must be introduced. The fact that the quarks are massive means also that there can be a mixing among them. Another difference from the lepton fields is the fact that each quark has an associated *colour* charge, and hence every quark has three different species and as many fields. The quark fields can be written as

$$Q_L^i(x) = \frac{1}{2}(1-\gamma_5)\begin{pmatrix}U_i(x)\\ D_i^-(x)\end{pmatrix}\,, \tag{11}$$

where the index i runs among the three families $U^i = u, c, t$ and $D^i = d, c, b$ and the summation on the three colours is understood.

The final ingredient is the introduction of a scalar field to allow for the spontaneous symmetry breaking mechanism. As explained in Section 1.2.1, the simplest choice is to have a doublet Φ containing a ϕ^+ and a ϕ^0 scalar fields.

By using these ingredients, it is then possible to write the most general Lagrangian invariant under $SU(2) \times U(1)$ symmetries that is also renormalisable. The renormalisability requirement imposes the presence of fields with dimension equal or smaller than four (this feature is important in the discussion of the Effective Field Theories formalism reported in Section 1.6.1).

Many physical consequences can be derived from the theoretical structure that has just been built. The breaking of the symmetry produces the generator of the electromagnetic group Q already described and sets the masses of the gauge bosons. Three of them are massive ($W^\pm$ and Z) while the photon is massless. By calling $\overrightarrow{W}$ the

gauge fields associated to $SU(2)$ and B the one associate with $U(1)$ and by naming g and g' the corresponding coupling constants, the two neutral bosons can be written as

$$Z_\mu = \sin\theta_W B_\mu - \cos\theta_W W^3_\mu \quad A_\mu = \cos\theta_W B_\mu + \sin\theta_W W^3_\mu \,, \tag{12}$$

with $\tan\theta_W = g'/g$, where θ_W is known as the *Weinberg angle*. The masses eigenvalues can be expressed as

$$m_W = \frac{vg}{2} \,, \qquad m_Z = \frac{v\left(g^2+g'^2\right)^{1/2}}{2} = \frac{m_W}{\cos\theta_W} \,, \qquad m_A = 0 \,, \tag{13}$$

where v is the vacuum expectation value of the scalar field of the Higgs mechanism.

The masses of the fermions come from the Yukawa coupling with the scalar field given by

$$\mathcal{L} = \sum_{i=1}^{3} \left[-G_i \left(\bar{\Psi}^i_L \ell^-_{iR} \Phi + \text{h.c.} \right) + G^i_u \left(\bar{Q}^i_L U^i_R \tilde{\Phi} + \text{h.c.} \right) \right] + \sum_{i,j=1}^{3} \left(\bar{Q}^i_L G^{ij}_d D^j_R \Phi + \text{h.c.} \right) \,, \tag{14}$$

where the summation runs over families and G are the Yukawa coupling constants which are proportional to the masses of the relative fermions. It must be noted that, with this mechanism, all fermions except for neutrinos acquire mass. The previous expression has been derived by diagonalising the up-type quarks. By diagonalising also the down-quark masses it is possible to obtain the Cabibbo-Kobayashi-Maskawa [31] matrix, which expresses the strength of the mixing among the quark families in weak interactions. By defining $c_{ij} = \cos\theta_{ij}$ and $s_{ij} = \sin\theta_{ij}$, the matrix can be parametrised as

$$\begin{aligned} \text{CKM} &= \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{-i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{-i\delta} & s_{32}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{-i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{-i\delta} & c_{23}c_{13} \end{pmatrix} \\ &= \begin{pmatrix} 0.97419 \pm 0.00022 & 0.2257 \pm 0.0010 & 0.00359 \pm 0.00016 \\ 0.2256 \pm 0.0010 & 0.97334 \pm 0.00023 & 0.0415 \pm 0.0011 \\ 0.00874 \pm 0.00037 & 0.0407 \pm 0.0010 & 0.999133 \pm 0.000044 \end{pmatrix} \end{aligned} \tag{15}$$

where the measured experimental values are taken from Reference [32]. From this expression, it is clear that the mixing between the first and the second families or the second and the third is stronger than the one between the first and the third. An experimental consequence of this structure is the decay of the top quark that produces nearly all the times a b quark and a W boson.

Finally, another important outcome of the structure of the theory is the presence of vector boson self-couplings. This behaviour is linked to the non-abelian nature of the Yang-Mills structure of the theory and is of key importance for the present work. A more detailed discussion is reported in Section 1.6.

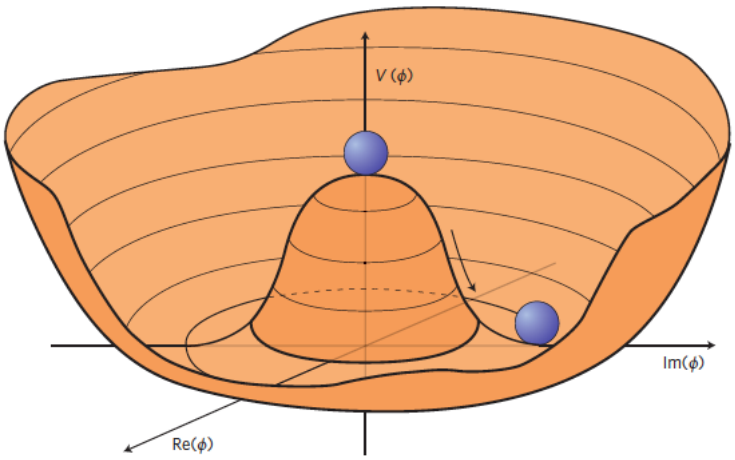


Figure 1: depiction of an effective *Mexican hat* potential that leads to spontaneous symmetry breaking [37].

1.2.1 *The Higgs mechanism*

The mechanism through which particles acquire mass has a key role in the structure of the Standard Model. The work initiated in the 1960s introduced a novel way to exploit spontaneous symmetry breaking in order to explain the fact that the particles of the Standard Model have mass and was proved by the discovery of a Higgs boson reported by the CMS [33] and ATLAS [34] collaborations in 2012.

As discussed in the previous section, the idea of spontaneous symmetry breaking was introduced in the context of particle physics by Nambu in 1960 and connected to the breaking of the global chiral symmetry that could explain the low-energy interactions of pions. In 1961, Goldstone [35] applied the concept of spontaneous global symmetry breaking to a $U(1)$ theory with a single complex field ϕ that results in a potential with a *Mexican hat* shape (Figure 1). In this situation, the minimum of the potential is not at the origin and hence the system is unstable and must collapse in one specific direction, breaking its original symmetry. The fact that the bottom of the brim has the same potential value means that there is no preferential direction to which the system would converge and changing this value does not cost energy. As it was proven by Goldstone, Salam and Weinberg in 1961 [36], the spontaneous symmetry breaking of a global symmetry must be accompanied by the presence of one (or more) bosons, named *Nambu-Goldstone bosons*.

While this is true for the spontaneous breaking of a global symmetry, this result does not necessarily apply to the breaking of a gauge symmetry. The spontaneous breaking of gauge symmetries was introduced in 1964 by Englert and Brout [38], then by Higgs [39] and also by Guralnik, Hagen and Kibble [40]. They demonstrated how a spinless Nambu-Goldstone boson and a gauge boson of an exact local symmetry could be combined into a single massive vector boson in a fully relativistic theory, hence finding a mechanism that was able to provide the mass terms for the Standard Model vector bosons. The reason for the particle associated to the mechanism being known as *Higgs boson*, despite the work of many different physicists, is due to the

fact that a second paper by Higgs [39] was the only one mentioning explicitly the presence of a massive scalar particle associated to the potential.

The concept of spontaneous symmetry breaking is best introduced in the context of a single complex scalar field, that can be written as

$$\phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2)\,, \tag{16}$$

where ϕ_1 and ϕ_2 are real fields. The corresponding Lagrangian can be written as

$$\mathcal{L} = (\partial_\mu \phi)^*(\partial^\mu \phi) - V(\phi) \quad \text{with} \quad V(\phi) = \mu^2(\phi^*\phi) + \lambda(\phi^*\phi)^2\,, \tag{17}$$

which is invariant under the global $U(1)$ symmetry $\phi \to \phi' = e^{i\alpha}\phi$. The potential has the Mexican hat shape for $\mu^2 < 0$ and an infinite set of minima can be defined as

$$\phi_1^2 + \phi_2^2 = \frac{-\mu^2}{\lambda} = v^2\,. \tag{18}$$

By choosing the minimum to be in the real direction (i.e. $(\phi_1, \phi_2) = (v, 0)$) and by expanding the complex scalar field around the minimum, it is possible to rewrite the Lagrangian as

$$\begin{aligned} &\mathcal{L} = \frac{1}{2}(\partial_\mu \eta)(\partial^\mu \eta) - \frac{1}{2}m_\eta^2\eta^2 + \frac{1}{2}(\partial_\mu \xi)(\partial^\mu \xi) - V_{int}(\eta, \xi) \\ &\text{with} \quad V_{int}(\eta, \xi) = \lambda v \eta^3 + \frac{1}{4}\lambda\eta^4 + \frac{1}{4}\lambda\xi^4 + \lambda v \eta \xi^2 + \frac{1}{2}\lambda\eta^2\xi^2\,, \end{aligned} \tag{19}$$

which represents a scalar field η with mass $m_\eta = \sqrt{2\lambda}v$ and a massless scalar field ξ, which is the Nambu-Goldstone boson of the theory. The excitations of the massive scalar field are in the direction where the potential is quadratic, while the ones for the massless field correspond to excitations on the directions in which the potential does not change.

The same procedure can be followed in the mechanism of spontaneously breaking a local gauge symmetry by replacing the partial derivatives with the covariant derivatives that depend on a massless gauge field B_μ, as

$$\partial\mu \to D_\mu = \partial_\mu + igB_\mu\,, \tag{20}$$

with the Lagrangian being

$$\mathcal{L} = (D_\mu \phi)^*(D_\mu \phi) - V(\phi^2) \quad \text{with} \quad V(\phi) = \mu^2\phi^2 + \lambda\phi^4\,. \tag{21}$$

By using the same procedure as before, a massive scalar field η and a massless ξ are obtained. Moreover, the B_μ field acquires a mass term hence providing the needed solution to the original problem. The fact that the massless field acquires a mass term introduces a fictional degree of freedom that suggests that the fields appearing in the previous formulation are not the physical ones. It is possible to match the correct number of degrees of freedom by eliminating the $\xi(x)$ field by applying the *unitarity gauge*. In this situation, the complex scalar field $\phi(x)$ can be written as being entirely real

$$\phi(x) = \frac{1}{\sqrt{2}}(v + \eta(x)) \equiv \frac{1}{\sqrt{2}}(v + h(x)) \tag{22}$$

where $h(x)$ is used to indicate the so-called *Higgs field*. The Lagrangian can then be rewritten as

$$\begin{aligned}\mathcal{L} = &\underbrace{\frac{1}{2}(\partial_\mu h)(\partial^\mu h) - \lambda v^2 h^2}_{\text{massive h scalar}} - \underbrace{\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}g^2 v^2 B_\mu B^\mu}_{\text{massive gauge boson}} \\ &+ \underbrace{g^2 v B_\mu B^\mu h + \frac{1}{2}g^2 B_\mu B^\mu h^2}_{\text{h, B interactions}} - \underbrace{\lambda v h^3 - \frac{1}{4}\lambda h^4}_{\text{h self interactions}},\end{aligned} \tag{23}$$

where the gauge boson has acquired mass $m_B = gv$ and the mass of the Higgs boson is $m_H = \sqrt{2\lambda}v$.

The approach for the Standard Model is the same, but since the symmetry to be broken is the $SU(2) \times U(1)$ and since three Goldstone bosons are needed in order to get the longitudinal degrees of freedom of the $W^\pm$ and Z bosons, the simplest solution is to use two complex scalar fields placed in an isospin doublet

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}. \tag{24}$$

Without entering into the details of the derivation, the result has the same properties as the one of the previous example. The important outcomes are the expressions for the masses of the Standard Model particles. The mass of the W boson can be written as $m_W = \frac{1}{2}g_W v$ and it is then proportional to the constant of the $SU(2)$ gauge interaction g_W and the vacuum expectation value of the Higgs field v. The photon is massless, while the mass of the Z boson is given by $m_Z = \frac{1}{2}v\sqrt{g_W^2 + g'^2}$, where g' is the coupling constant of the $U(1)$ group. An important relation that can be obtained from these expressions is that the W and Z boson masses are linked to the Weinberg angle by

$$\frac{m_W}{m_Z} = \cos\theta_W. \tag{25}$$

The coupling of the Higgs boson to both the W and the Z bosons is proportional to the boson masses: $g_{HWW} = g_W m_W$ and $g_{HZZ} = g_Z m_Z$. Eventually, also the coupling of the Higgs boson to the fermions depends on the masses of the fermions

$$g_f = \sqrt{2}\frac{m_f}{v}. \tag{26}$$

1.3 EXPERIMENTAL EVIDENCES

The goodness of the Standard Model predictions has been verified countless times in the past sixty years by many different experiments with different techniques. What emphasises even more the breakthrough nature of the Standard Model is that most of the experimental discoveries came after the theoretical predictions and were found to be in perfect agreement with the theory. The particles that constitute the Standard Model are summarised in Figure 2.

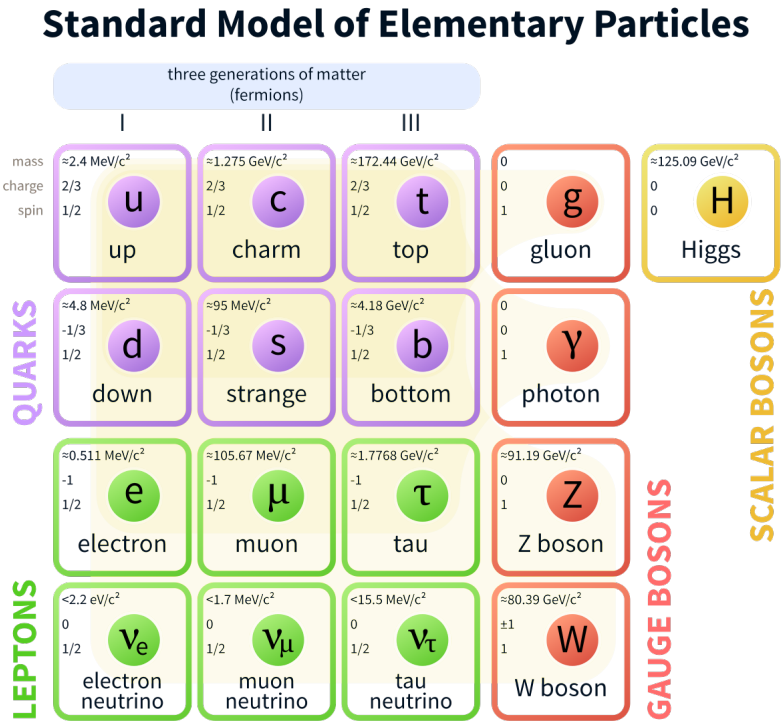


Figure 2: particle content of the Standard Model. The quarks (purple) and leptons (green) are divided into the three families. The vector bosons are presented in red and the Higgs boson in yellow. Mass, charge and spin of the particles are also reported.

The electroweak sector of the Standard Model had to be proven by proving the existence and studying the properties of the vector bosons. The first evidences for the presence of neutral currents came from neutrino beams experiments in the 1970s that studied neutrinos scattering with the aid of bubble chambers [41]. The direct discovery of the Z and the W bosons came from the experiments at the CERN Super Proton Synchrotron (SPS) collider in 1983 [42]. The SPS collider was at the time the most powerful particle accelerator in the World, colliding a proton and an antiproton beam at a centre of momentum energy of more than 400 GeV. Carlo Rubbia, who coordinated the UA1 experiment and promoted the construction of the Sp$\bar{p}$S collider, and Simon Van der Meer, who developed the stochastic cooling technique [43] fundamental to have colliders that could reach such high energies, received the Nobel Prize in physics in 1984.

The detailed study of the properties of the W and Z bosons was performed at the Large Electron-Positron (LEP) collider at CERN in the 1990s. Among the many results obtained in ten years of data collection and analysis, a precise measurement of the mass of the two vector bosons, that drives the present time World averages, was performed [44, 45]. From the study of the Z boson lineshape, an indirect evidence that the number of light neutrinos families had to be three was produced [46–48]. The study of the Z boson asymmetries allowed for a thorough test of the Standard Model [45]. The combination of all the measurements allowed to perform the *global*

electroweak fit, thanks to which the values of the theoretically unknown parameters of the Standard Model were determined with great precision and predictions on the mass of the Higgs boson were made.

As discussed in Section 1.1, the presence of a fourth quark was hypothesised in the theory in order to cancel the infinities in the scattering amplitude computations. The discovery of the fourth quark, named *charm*, came from two independent experiments lead by Burton Richter [49] and Samuel Ting [50] in 1974. The two groups discovered a narrow resonance in the $e^+e^- \to e^+e^-$ and the $pp \to e^+e^-$ scattering processes respectively that was identified as a $c\bar{c}$ bound state and named it J/ψ. Richter and Ting were awarded the Nobel prize in physics in 1976.

The third generation of quarks was a theoretical consequence of the Kobayashi-Maskawa extension of the Cabibbo matrix that had been introduced to explain possible CP violating processes. The direct discovery of the down-type quark of the third family, named *bottom*, came from the experiment lead by Leon Lederman at Fermilab in 1977 [51]. As for the charm quark, also the bottom quark was discovered from the identification of a $b\bar{b}$ resonance with a mass of 9.46 GeV, named Υ. The discovery was recognised with the Nobel prize to Lederman in 1988.

The evidence for the last missing quark, the *top*, came in 1994 from the results of the CDF [52] and D0 [53] experiments at the Tevatron accelerator at Fermilab. The reason why so many years passed between the discovery of the bottom and the top quarks is linked to the mass of the latter. The top quark, in fact, with its mass of 172 GeV, is the heavier particle of the Standard Model. In order to be able to produce it, a huge centre of momentum energy had to be reached in the collisions and the technology to do so was available only in the 1990s, when the Tevatron proton-antiproton collider reached the record-breaking centre of momentum energy of 1.8 TeV. Differently from the other quarks, the top quark does not form bound states and decays before the hadronisation process. The discovery has been done by studying the $t\bar{t}$ pair production and the decay of the top quark into a W boson and a bottom quark.

The last undiscovered piece to complete the Standard Model puzzle was the Higgs boson. Before the actual discovery, many indirect information were obtained from precision measurements of the Standard Model parameters at LEP and at Tevatron. Before the start of the Large Hadron Collider program in 2008, the most likely region where the Higgs boson could be found was restricted to the 115-130 GeV mass interval. The first direct observation was reported in 2012 by the CMS [33] and ATLAS [34] collaborations, that found strong evidence of the presence of a new particle with a mass of 125 GeV. After the discovery, many precision measurements have been performed in order to check if the properties of the new particle matched the ones of the Higgs boson of the theory. Probably, the most striking one is the value of the Yukawa couplings of the Higgs boson with the fermions, summarised in Figure 3 [54]. As it can be seen, the measured value of the couplings is in perfect agreement with the theoretical prediction.

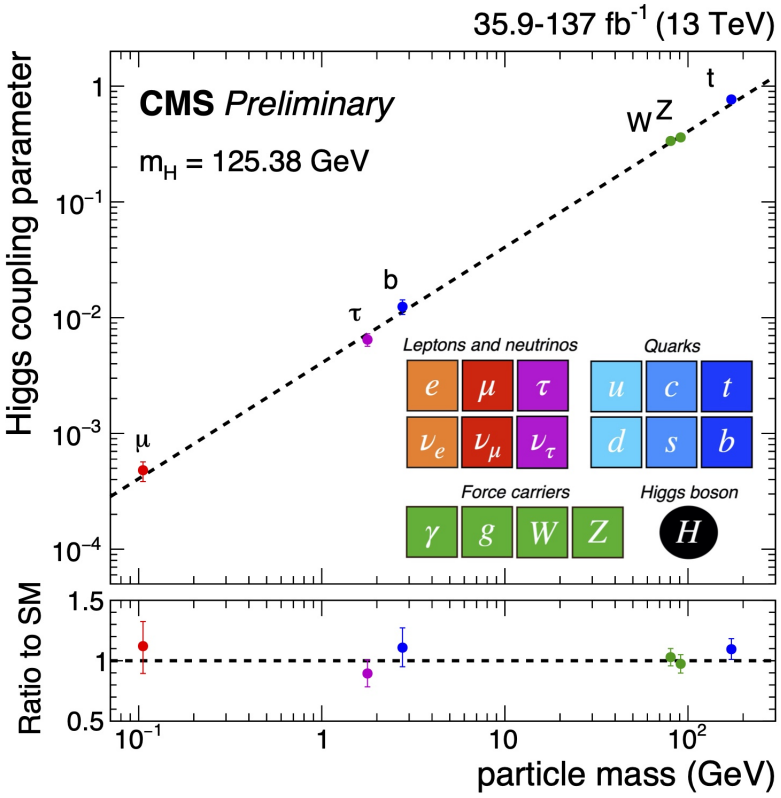


Figure 3: summary of the measurements of the Higgs boson couplings to fermions performed by the CMS experiment. The theoretical prediction is represented with the dotted line, the measured quantities are the coloured points [54].

1.4 BEYOND THE STANDARD MODEL

The Standard Model has proven to be a very general theory able to describe a lot of phenomena. The agreement between the theoretical predictions and the experimental measurements is overwhelming in a range of energies that goes from few MeV to the TeV scale. There are, however, some observations that are not predicted by the theory.

The most notable one is the fact that the ingredients of the Standard Model constitute approximately 5% of the total mass-energy of the Universe. The hints of the presence of another kind of matter, called *dark matter* since it does not interact electromagnetically, come from many cosmological observations. The first evidence of the presence of dark matter derives from the observation of the galaxy rotational curves [55], where the revolution speed does not follow the prediction of gravitational theories if the mass of the galaxy was only the one of luminous objects. Another evidence is linked to the effect of *gravitational lensing*, where the light coming from distant cosmological objects can be seen after being bent by the presence of other massive objects (such as clusters of galaxies) in its trajectory towards Earth. As for the rotation speed of the galaxies, also the mass obtained from the lensing effect exceeds that of the object that is producing the lensing effect [56]. One of the most striking observations comes from the shape of the *baryonic acoustic oscillations* of the angular spectrum

of the cosmic microwave background. The average matter density of the Universe and then the amount of dark matter can be obtained form these measurements [57].

By studying neutrinos coming form different sources, such as the Sun [29, 58, 59], nuclear reactors [60, 61] or those produced by the interaction of cosmic rays with the atmosphere [62], it has been shown that their flavour can change while they are travelling from the source to the detector. This process is known as *neutrino oscillations* and involves all the three families. The physical explanation of the oscillation of the neutrinos and hence the mixing among different families, was first formulated by Bruno Pontecorvo in 1957 [63]. The description of the phenomenon is based on the superposition of mass states, which is in plain contrast with the Standard Model prediction of massless neutrinos.

As discussed in Section 1.3, the mass of the Higgs boson has been measured to be of the order of 125 GeV. From a theoretical point of view, the value of the mass of the Higgs boson is affected by the presence of very large quantum corrections (mostly due to virtual top quark loops) which are much larger than the actual value of the Higgs boson mass. To have such a small mass for the Higgs boson compared to the magnitude of the corrections means that there must be a fine tuning of the Standard Model parameters in order to almost completely cancel the contribution of these higher order corrections. This issue is generally referred to as *hierarchy problem* or *naturalness problem*.

If the Big Bang created the same quantity of matter and anti-matter, the question of where did all the antimatter go is still an unanswered one. From conservation laws and by knowing the annihilation processes, there is in principle no reason why only one kind of matter should have survived in the Universe. The problem is generally referred to as *matter anti-matter asymmetry* and its explanation is not found within the Standard Model framework. One of the most promising solutions is the one connected to the CP violation, which can explain how a tiny unbalance between the quantity of matter and antimatter could have resulted in the structure of the Universe that we observe.

Many possible models have been developed in the last fifty years in order to try to generalise the Standard Model and include these unexplained features. The most famous batch of models are those of the *supersymmetric theories* (SUSY). The supersymmetric models predict the existence of new fundamental particles that are the counterpart of those observed in the Standard Model framework. Without going in the detail of the theory, supersymmetry claims to have a solution for the issues not explained by the Standard Model. The supersymmetric particles could be those that are currently categorised as dark matter and the quantum corrections connected to their presence would naturally cancel the contributions of the Standard Model particles, hence solving the hierarchy problem. As for the Standard Model, the masses of the particles are parameters that cannot be derived from the theory but must be measured in experiments.

Since the LHC started operations, a lot of energy and personpower has been invested in the search for these new undiscovered particles. A priori, there are two ways that can be followed: search for a direct signature of the new particle or measure known quantities with a high precision. The direct search has the great advantage

that, if the particle exists, its signature is clearly visible (as for the Higgs boson discovery). However, supersymmetric particles are "heavy", which means that a high centre of momentum energy must be achieved for them to be produced. What is worst, it is not possible to determine a priori how "high" this "high energy" is: there is a possibility that the new particles have such a great mass that they are far beyond the reach of the present time accelerators.

The other possibility is to perform high precision measurements of the Standard Model parameters and see if the observed value is in disagreement with the theoretical prediction. If this is the case, the disagreement can be connected to the presence of interactions between the undiscovered particles and the Standard Model ones. While not providing a direct evidence, this method has the great advantage that signatures of undiscovered phenomena can be measured even at energies much smaller than the scale of new physics.

As of the time these words are being written, no evidences of beyond the Standard Model physics have been found.

1.5 PROTON STRUCTURE AND QUANTUM CHROMODYNAMICS EFFECTS

Particle accelerators are built with the goal of producing high energy collision events that can give rise to new particles that can then be studied. The events that are marked as "interesting" are those where there is a great momentum transfer between the two colliding particles and the final state is characterised by the presence of particles that are different from the colliding ones. These kind of events are referred to as *deep inelastic scattering* (DIS) events. In this section, the basic phenomenology of the DIS events is briefly discussed, with particular attention to the aspects that led to the understanding of the proton structure. The resulting quark model is introduced before the discussion of the quantum chromodynamics (QCD) effects. The measurement of the parton distribution functions (PDFs) and their impact on high energy physics analyses is detailed.

1.5.1 *Deep inelastic scattering and the parton model*

The study of the structure of the proton can be performed while studying DIS events in electron-proton collisions. This has both the experimental and theoretical advantage that the electron is a fundamental particle and is used as a probe, while the proton has a complex structure that must be understood. The cross section for an electron scattering with a proton can be written as [64]

$$\frac{d^2\sigma}{dQ^2 d\nu} = \frac{\pi\alpha^2}{4k^2 \sin^4(\theta/2)} \frac{1}{kk'} \left[W_2 \cos^2(\theta/2) + 2W_1 \sin^2(\theta/2) \right] , \tag{27}$$

where Q^2 is the momentum transfer between the scattering particles, ν is the energy difference between the incoming and the scattered electron in the proton rest frame, θ is the scattering angle, k is the energy of the incoming electron and k' is the energy of the scattered electron ($\nu = k - k'$). An event is categorised as DIS one if both Q^2 and

ν are large. $W_1(Q^2,\nu)$ and $W_2(Q^2,\nu)$ are called *structure functions* and are related to the electric and magnetic form factors of the non pointlike proton.

In order to understand the behaviour of the DIS events it necessary to understand the dependences on the structure functions. In principle, the two functions depend both on Q^2 and ν. However, James Bjorken in 1969 [65] proposed a simplified description in which, for $Q^2 \to \infty$ and $\nu \to \infty$, the dependence of the structure functions relied on a single scaling variable defined as

$$x = Q^2/2M\nu\,, \tag{28}$$

where M is the mass of the proton. With this hypothesis, the structure functions are finite and can be rewritten as $MW_1(Q^2,\nu) \to F_1(x)$ and $\nu W_2(Q^2,\nu) \to F_2(x)$ as a function of x, named *Bjorken scaling* variable.

The physical interpretation of this theoretical model was developed by Richard Feynman [66]. At high momentum transfer, the electron is probing lengths that are shorter than the dimension of the proton itself. The object on which the electron scatters is not the proton, but a free point-like constituent called *parton*. In principle it seems unreasonable that a constituent of the proton, subject to the strong force, should be free but, as will be discussed in the next section, this is exactly what happens for quarks. The parton carries only a fraction of the proton momentum $p^\mu_{\text{parton}} \simeq f p^\mu_{\text{proton}}$ and Feynman demonstrated that the fraction of momentum f is equivalent to the Bjorken scaling variable x.

The first experimental evidences for an internal structure of the proton came from experiments at SLAC in 1967 [67]. The first observation was that in the DIS region the electron rebounded with less energy than in the elastic scattering case, indicating that the particle on which it scattered was not the proton itself but one of its constituents. The results proved that there was in general good agreement with the prediction of the structure functions being dependent only on x, however there were hints to a Q^2 dependence especially for low values of x, suggesting a deeper feature of the phenomenon. From electron scattering on nuclei of deuterium (made of two bounded particles that resemble the structure of the partons bounded in the proton) a value of $x = 1/2$ was found, as expected (the proton and the neutron have almost the same mass which is half the one of the deuterium nucleus). Applying the same procedure, the partons in the proton where found to have $x = 1/3$ and a model with three *valence* partons was developed. However, since the first experiments, the observed behaviour suggested a more complex structure for the proton.

By making the assumption, strongly supported by the experimental results, that the partons of the model just described are indeed the quarks of the Standard Model, it is possible to write the structure functions of the electron proton scattering as

$$F_2^{ep} = x\left(\frac{4}{9}[u(x)+\overline{u}(x)] + \frac{1}{9}[d(x)+\overline{d}(x)+s(x)+\overline{s}(x)]\right)\,, \tag{29}$$

where $u(x)$ is the parton distribution function for the up quark, $\overline{u}$ for the up anti-quark and so on. For simplicity of the notation, only the three lightest quarks are

written explicitly. Since the proton has charge +1 and has no strangeness, it is possible to derive the following relations

$$\int_0^1 dx\,[u(x)-\overline{u}(x)]=2 \quad \int_0^1 dx\,[d(x)-\overline{d}(x)]=1 \quad \int_0^1 dx\,[s(x)-\overline{s}(x)]=0\,, \tag{30}$$

from which the idea of the excess of quarks over antiquarks is introduced. The physical interpretation of these relations is that, besides the three valence quarks, there exist also *sea quarks* that are present into the proton in $q\overline{q}$ pairs that can also enter the scattering process. The model with valence and sea quarks has a much better agreement with the experimental results, but is not complete. By measuring the sum of the contribution of all the parton distribution functions, i.e.

$$\int_0^1 dx\,x[u(x)+\overline{u}(x)+d(x)+\overline{d}(x)+s(x)+\overline{s}(x)]=1-\epsilon\,, \tag{31}$$

it can be found that it does not match unity. The remaining contribution, responsible for the presence of ϵ, is linked to the presence of gluons inside the proton.

Experimental evidences for the sea-quark model came from data collected at SLAC in the period 1970-73 [67] where both electron-proton and electron-neutron scattering processes were studied. In particular, from the study of the ratio of the two cross sections, the hypothesis of no internal constituents was immediately discarded ($\sigma_n/\sigma_p < 1$). The fact that the ratio was close to unity for small values of x suggested that in this regime proton and neutrons had a similar behaviour, meaning that the electron was scattering the same particles which are the quarks from the sea. On the other hand, for larger values of x, the behaviour was different hence proving the evidence that valence quarks carry the majority of the momentum of the proton or the neutron. In 1981 the CHARM experiment [68] measured $1-\epsilon = 0.44 \pm 0.02$, meaning that roughly half of the proton total momentum is carried by gluons.

1.5.2 *QCD corrections and PDF evolution*

Asymptotic freedom

The basic hypothesis of the parton model is the assumption that the partons themselves can be considered as free while inside the proton. As discussed earlier, this idea seems counterintuitive, since the quarks are subject to the action of the strong force. However, the property of *asymptotic freedom* for the quarks in the proton guarantees the basic assumption of the parton model.

The cross section for the $e^+e^- \to q\overline{q}$ can be written as [69]

$$\sigma = \sigma_{point}(1+\alpha_s/\pi)\,, \tag{32}$$

where σ_{point} is the point-like $e^+e^- \to$ hadrons cross section and α_S is the strong coupling constant. By expanding this expression at higher order, the cross section can be written as

$$\sigma \propto \sigma_{point}\left[1+\frac{\alpha_s/\pi}{1+\alpha_s b\,\ln(s/\mu_R^2)}\right] \quad \text{with} \quad b=\frac{33-2N_f}{12\pi}\,, \tag{33}$$

where s is the square of the total e^+e^- energy, μ_R is the renormalisation mass scale and N_f is the number of flavours "active" at a certain energy scale.

The introduction of the renormalisation scale μ_R follows the same path used for the renormalisation of quantum electrodynamics (QED). By looking at the situation in QED, the introduction of μ_R makes the electric charge dependent on the renormalisation scale itself, which of course is meaningless since the electric charge and the cross section are observable that can be measured in experiments. By defining $S \equiv \sigma/\sigma_{\text{point}}$, the dependence on μ_R is removed by imposing

$$\frac{dS}{d\mu_R} = \left(\frac{\partial}{\partial \mu_R}\bigg|_{e_{\mu_R}} + \frac{de_{\mu_R}}{d\mu_R}\bigg|_{e_0} \frac{\partial}{\partial e_{\mu_R}} \right) S(|Q^2|/\mu_R^2, e_{\mu_R}) = 0\,, \tag{34}$$

where e_{μ_R} is the value of the electric charge at a mass scale μ_R. The equation can be rewritten in terms of μ_R^2 and $\alpha = e^2/4\pi$ in the form

$$\left(\mu_R^2 \frac{\partial}{\partial \mu_R^2}\bigg|_{\alpha_{\mu_R}} + \beta(\alpha_{\mu_R}) \frac{\partial}{\partial_{\alpha_{\mu_R}}} \right) S(|Q^2|/\mu_R^2, \alpha_{\mu_R}) = 0 \quad \text{with} \quad \beta(\alpha_{\mu_R}) = \mu_R^2 \frac{\partial \alpha_{\mu_R}}{\partial \mu_R^2}\bigg|_{e_0}\,. \tag{35}$$

This equation is generally referred to as *renormalisation group equation* and its solution is

$$\alpha(|Q^2|) = \frac{\alpha_{\mu_R}}{1 - \frac{\alpha_{\mu_R}}{3\pi} \ln(|Q^2|/\mu_R^2)}\,. \tag{36}$$

The same procedure can be followed in the QCD case, where the β-function is defined as

$$\beta_s = \mu_R^2 \frac{\partial \alpha_s}{\partial \mu_R^2}\bigg|_{\text{fixed bare } \alpha_s}\,, \tag{37}$$

and the solution to the renormalisation group equation is given by

$$\alpha_s(|Q^2|) = \frac{\alpha_s(\mu_R^2)}{1 + \alpha_s(\mu_R^2) b \, \ln(|Q^2|/\mu_R^2)}\,, \tag{38}$$

where b is defined in Equation 33. The important thing to notice while comparing the QCD result to the one for the QED is the sign at the denominator. For the QCD solution, the value of the coupling constant decreases as Q^2 becomes larger. This means that at small distances (which is equivalent to great momentum transfers) the coupling is weaker and so the quarks inside the proton can be effectively treated as free. The property of asymptotic freedom is at the core of all the perturbative treatment of the QCD.

Scaling violations

Despite being a good low-energy approximation, the dependence of the structure functions and hence of the PDFs only from the Bjorken scaling variable does not provide the complete description of the phenomenon. The violation of the scaling dependence is due to higher order corrections in QCD.

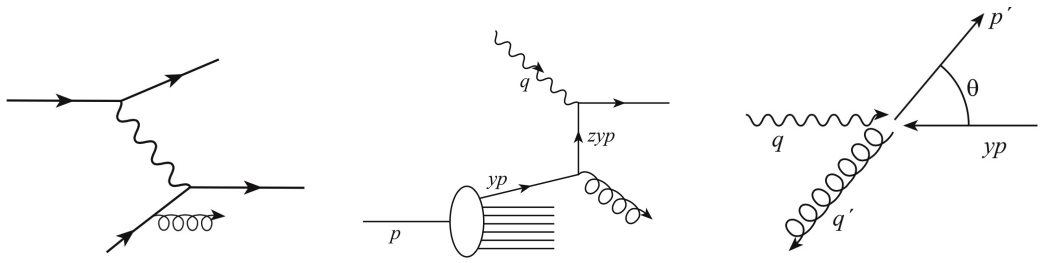


Figure 4: electron-quark scattering with gluon emission. On the left the complete process, while on the centre a zoom on the second vertex. The diagram on the right is the same vertex but referred to the centre of momentum system.

The first order QCD corrections to the tree-level electron-quark scattering calculations are linked to the emission of a single gluon (Figure 4). The structure function of the parton can be written as [69]

$$2\hat{F}_1 = \frac{\hat{s}}{4\pi^2\alpha}\int_{-1}^{1} d\cos\theta \frac{4}{3}\frac{\pi e_i^2 \alpha\alpha_s}{\hat{s}}\left(-\frac{\hat{t}}{\hat{s}} - \frac{\hat{s}}{\hat{t}} + \frac{2\hat{u}Q^2}{\hat{s}\hat{t}}\right), \tag{39}$$

where s, t and u are the Mandelstam variables (see the following) and the hat symbol $\hat{}$ indicates quantities that refer to the parton. By looking at Figure 4 centre, the relation between the structure function of the parton and that of the nucleon can be derived. The quark coming from the nucleon carries a momentum fraction yp and after the gluon emission it has a momentum of zyp, with $y, z < 1$. By exploiting the fact that $Q^2 = 2M\nu$, $Q^2 = -q^2$ and assuming that the quark is on-shell after the emission (i.e. $(zyp)^2 = M^2$) and neglecting all masses, it can be seen that

$$q^2 + 2zyp \cdot q = 0 \rightarrow zy = \frac{Q^2}{2p\cdot q} = \frac{Q^2}{2M\nu} = x, \tag{40}$$

and by using the Callan-Gross relations [70] this can be rewritten as

$$\frac{F_2}{x} = 2F_1 = \sum_i \int_0^1 dy\, f_i(y) \int_0^1 dz\, 2\hat{F}_1^i \delta(x - yz), \tag{41}$$

where $f(i)$ are the PDFs of the quarks.

To get the corrections to the parton model, the Mandelstam variables of Equation 39 must be written explicitly. By moving to the centre of momentum system, letting k be the module of the momentum of the incoming particles and k' be the one of the outgoing ones, it is possible to write

$$\begin{aligned}
\hat{s} &= 4k' = (yp + q)^2 = Q^2(1-z)/z \\
\hat{t} &= (q - p')^2 = -2kk'(1-\cos\theta) = -Q^2(1-\cos\theta)/2z \\
\hat{u} &= (q - q')^2 = -2kk'(1+\cos\theta) = -Q^2(1+\cos\theta)/2z
\end{aligned} \tag{42}$$

with $z = Q^2/(\hat{s} + Q^2)$. By looking at the expression of Equation 39, it is clear that there is a divergent term of the form

$$\sim \int^1 \frac{d\cos\theta}{1-\cos\theta}\,, \tag{43}$$

which means that the divergence arises for $\cos\theta = 1 \rightarrow \theta = 0$. This is a *collinear divergence* because it happens when q and p' are parallel. It is easily solved by reintroducing the mass terms in the equation. If m is the mass of the quark, it is possible to rewrite the previous equation as

$$\sim \int^1 \frac{d\cos\theta}{(1+2m^2z/Q^2)-\cos\theta} \quad \overset{m^2\to 0}{\longrightarrow} \quad \ln\left(\frac{Q^2}{m^2}\right), \tag{44}$$

from which it can be seen that the quark mass regulates the divergence. The uncancelled mass singularity is the essential feature of the scaling violation.

It is now possible to compute the integral in Equation 39 retaining only the terms proportional to $\hat{t}^{-1}$ (the divergent ones). For only one species of quark, the QCD correction contributes with a term

$$\frac{e_i^2\alpha_s}{2\pi}\int_x^1 \frac{dy}{y} q(y)\left[P_{qq}(x/y)\ln(Q^2/m^2) + C(x/y)\right] \quad \text{with} \quad P_{qq}(z) = \frac{4}{3}\left(\frac{1+z^2}{1-z}\right), \tag{45}$$

where $P_{qq}(z)$ is called *splitting function* or *splitting kernel*. P_{qq} is the probability that a quark, having radiated a gluon, has momentum z after its emission. $C(x/y)$ is a term that involves everything that does not have a mass singularity. Eventually, it is possible to write a correction term to the quark PDF $q(x)$ that does not depend anymore only on the Bjorken variable x as

$$q(x) + \frac{\alpha_s}{2\pi}\int_x^1 \frac{dy}{y} q(y)\left[P_{qq}(x/y)\ln(Q^2/m^2) + C(x/y)\right], \tag{46}$$

which directly expresses the violation of the scaling behaviour. This term, however, has an uncancelled divergence that is not physical, since the PDFs can be measured experimentally. The procedure that brings physical sense to this expression is discussed in the following.

PDF evolution

The divergence that has just been derived is, once more, connected to the collinearity of the emission. The important point that must be noted is that when the emission is collinear, the two particles have a very small relative momentum and so, as derived earlier, the interaction between them is very strong. This in turn means that the perturbative calculations that have been employed until now have no meaning in this regime. The idea is to divide the term in a calculable and an uncalculable part

$$\ln(Q^2/m^2) = \ln(Q^2/\mu_F^2) + \ln(\mu_F^2/m^2)\,, \tag{47}$$

where the term depending on m is not calculable. μ_F is referred to as *factorisation scale* and can be interpreted as the mass scale that separates the perturbative short-distance physics to the non-perturbative long-distance one. Partons emitted at small transverse momentum ($< \mu_F$) enter the non calculable part of the structure function. By defining the uncalculable term as

$$q(x, \mu_F^2) = q(x) + \frac{\alpha_s}{2\pi} \int_x^1 \frac{dy}{y} q(y) P_{qq}\left(\frac{x}{y}\right) \left[\ln\left(\frac{\mu_F^2}{m^2}\right) + C\left(\frac{x}{y}\right) \right], \tag{48}$$

the structure function can be written as

$$2F_1(x, Q^2) = e_i^2 \int_x^1 \frac{dy}{y} q\left(y, \mu_F^2\right) \left\{ \delta\left(1 - \frac{x}{y}\right) + \frac{\alpha_S}{2\pi} P_{qq}\left(\frac{x}{y}\right) \ln\left(\frac{Q^2}{\mu_F^2}\right) \right\} \equiv e_i^2 q(x, Q^2). \tag{49}$$

For the same reasons as previously discussed, in order to have a physical meaning, the structure function $F_1(x, Q^2)$ must not depend on the factorisation scale μ_F. By differentiating Equation 49 with respect to μ_F^2 and by setting it equal to zero, one obtains

$$\mu_F^2 \frac{\partial q(x, \mu_F^2)}{\partial \mu_F^2} = \frac{\alpha_S(\mu_F^2)}{2\pi} \int_x^1 \frac{dy}{y} P_{qq}(x/y) q(y, \mu_F^2), \tag{50}$$

which is known as the DGLAP equation after the names of its authors Dokshitzer [71], Gribov, Lipatov [72], Altarelli and Parisi [73]. It is a fundamental equation in QCD that allows to compute the PDFs at a certain value of the factorisation scale once they are known at a fixed value μ_0^2. The theoretical description of the evolution of the PDF has been the key to precision measurements at hadron colliders.

1.5.3 *Experimental features*

The theoretical model just outlined results in a satisfying structure for the proton and provides a good description of the DIS scattering interactions. However, the key ingredient for the computations, the shape of the PDFs, cannot be derived theoretically but must be measured from collision experiments. With the knowledge coming from one or more experiments, it is possible to perform theoretical predictions for other experiments at different energies by exploiting the DGLAP equation.

The most complete early determination of the PDFs came from the experiments at the HERA accelerator at DESY [74]. As discussed, the best experimental setup to study the behaviour of the partons is to have an electron-proton collider where the electron is used as a probe for the proton structure. The HERA accelerator collided a 27.5 GeV electron beam with a 920 GeV proton one.

The most recent determination of the shapes for the PDFs comes from an inclusive analysis of data collected at many different experiments. For the NNDPF3.1 PDF set [75], which is among the most recent ones and is the one used largely in the simulated samples implemented in this work, data from DIS collisions (such as those collected at HERA) are analysed together with data from the Tevatron and the LHC. The resulting PDF sets are shown in Figure 5. As it can be seen, there are two clearly

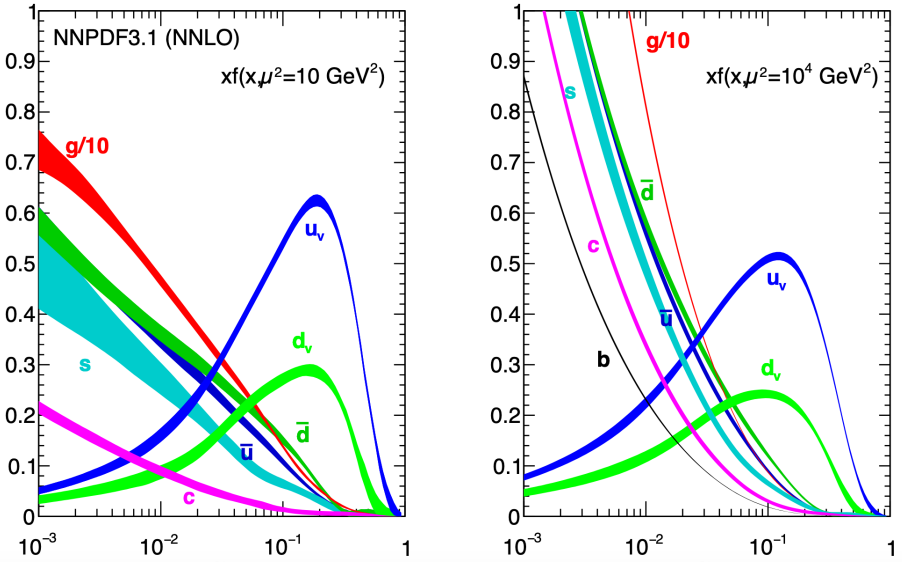


Figure 5: NNPDF3.1 PDF sets evaluated at μ_F =10 GeV (left) and μ_F =10 TeV (right).

separate behaviours for the PDFs. The PDFs for the valence up and down quark peak for $x \simeq 0.3$, while those for the sea quarks and the gluons are more important at low values of the Bjorken variable, in agreement with the theoretical prediction. The other important point to be made is that when increasing the collision energy, the sea quarks and especially the gluons get more and more predominant if compared to the valence quarks. This is the reason why the majority of the collisions at the LHC (which has at the present time a centre of momentum energy of 13 TeV) happen between gluons.

1.6 MULTI-BOSON PRODUCTION

The electroweak structure of the Standard Model allows for the interaction between the gauge bosons themselves. In particular, since the gauge bosons carry weak charge, vertices that contain three or four gauge bosons are allowed by the theory. Triple boson vertexes can be studied in vector boson fusion (VBF) processes, where the $q\overline{q}$ pair produces a single boson that is then attached to the triple vertex; the final state is characterised by the presence of two vector bosons. Quartic boson vertexes can be studied either in processes with three bosons in the final state (tri-boson processes) or in vector boson scattering (VBS) events, where two gauge bosons are emitted by the incoming quarks and scatter with each other. In VBS events, the final state is characterised by the presence of two vector bosons and two remnant jets from the colliding quarks.

The study of multi-boson processes has been carried out in the past both at LEP and at the Tevatron. However, given the fact that these are rare processes, the very high statistics provided by the LHC sets the perfect environment for precision studies. While at previous experiments the limited amount of collected data restricted the

field of study to di-boson processes, enough tri-boson events have been recorded at the LHC hence allowing the study of these processes for the first time.

The motivations to study multi-boson processes with high precision are manifold. Since the effect of multi-boson interactions is a direct consequence of the non-abelian structure of the theory, measuring their properties allows for a direct test of non-abelian gauge theories. Even though this kind of study has been carried out in the past, the multi-boson processes are able to test a region with higher typical energy and with unprecedented detail.

The W, Z and γ bosons are the particles that have the strongest coupling to the Higgs (let alone the top quark): a precision study of their relative interactions is a definitive test of the electroweak symmetry breaking mechanism that could shed light also on the hierarchy problem. The link between the multi-boson production and the electroweak symmetry breaking mechanism is the magnitude of the vector boson scattering amplitude that would grow as a function of the energy (and then diverge) if no Higgs loops were taken into account. A measurement of the process cross section and its properties can improve the understanding of the Higgs couplings.

Multi-boson processes have final states that can coincide with those of searches for new particles or with other precision measurements. For example, the $q\overline{q} \to ZZ$ process has the same final state as the Higgs decay in two Z bosons. The detailed study of the former and a precise determination of its cross section are fundamental to get a better description of the characteristics of the latter.

Multi-boson processes are also an indirect probe of the presence of new physics, as will be discussed in detail in the next section.

A very precise experimental description of the multi-boson processes must be matched by a precise theoretical calculation of their cross sections. At the present time, most of the cross sections are computed at next-to-next-to-leading order in α_S and at next-to-leading order in α_{EW}.

1.6.1 *Effective field theories and anomalous couplings*

As explained in Section 1.4, the presence of subtle effects of new physics can be probed by measuring known quantities with high precision: if a discrepancy is found between the theoretical Standard Model prediction and the measurement, this can hint to the presence of unknown particles or interactions. The observation of these new effects can be connected to a yet unknown particle that could be explained by a yet unknown model. The basic idea of the Effective Field Theories (EFTs) is to provide model-independent extensions of the Standard Model that allow for the interpretation of the results in a general way.

One of the most notable implementation of the EFT approach is the Fermi theory of the electroweak interactions [13], already discussed in Section 1.1. The description provided by Fermi assumed a contact interaction with a certain coupling strength for the description of the β decay. This approach is predictive and useful to make computations in the low energy regime. While going at higher energy, the full description of the electroweak interactions mediated by the W boson is needed to perform the calculations. This example introduces another important feature of EFTs: they are

valid only for an energy range which is smaller than the scale of the new, not yet described physics phenomena (the Fermi theory works well for energy scales much smaller than M_W).

The EFTs [76] provide a general prescription that does not make a commitment to a certain model and that allows to quantify the accuracy for the exclusion of new physics processes in the cases for which no direct evidence is found. The basic idea of the method is to extend the Standard Model Lagrangian by including higher order corrections. It is reasonable to expand the Lagrangian in a way that preserves the most important features of the Standard Model. The S-matrix should keep its properties of unitarity, analyticity and so on, and the main symmetries of the Standard Model (the Lorentz invariance and the $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge symmetry) should be maintained. Moreover, it must be possible to recover the Standard Model as a limit of the new theory, since we know that its predictions are correct as far as its particle contents and interactions are concerned. In order to take into account all these features, an effective quantum field theory is needed.

The Standard Model is the most general theory that can be written by including the known particles that interact following a $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge symmetry. The operators of its Lagrangian have dimension equal or smaller than four. The EFT expansion of the Lagrangian adds operators with dimension of powers of mass. By making the required assumption that the scale of the new physics is large if compared to the experimentally accessible energies, these higher dimension operators are suppressed if compared to the Standard Model ones. From this consideration, the operators that count the most are those with dimension five, followed by those of dimension six and so on. For this reason, the searches must be focused on the lower dimensionality expansion of the Standard Model.

The extended Lagrangian can be written as

$$\mathcal{L}_{BSM} = \mathcal{L}_{SM} + \sum_{i,d} \frac{c_i}{\Lambda^{d-4}} O_i \,, \tag{51}$$

where O_i are the $d-$dimensional operators, Λ is the energy scale at which the new physics appears and c_i are the Wilson coefficients, dimensionless parameters that indicate the strength of the coupling between the new physics particles and the Standard Model ones. As it can be seen from the functional expression, the Standard Model is recovered in the limit $\Lambda \to \infty$, in the situation where no new physics exists.

The operators are built from the low energy fields. In order to be able to build all the possible operators, the further assumption that no other particles exist in the Standard Model at low energies must be made. If it were not like this, some of the possible operators would not be written in the expansion. There exist only one operator of dimension five [77] and is responsible for the generation of Majorana masses to the neutrinos. The interesting operators in multi-boson analyses are those of dimension six and eight [76]. Dimension six operators are typical of di-boson production process, while dimension eight can be probed in vector boson scattering (VBS) processes [78] or in process with three gauge bosons in the final state.

A parallel formalism to EFTs is that of anomalous gauge couplings (aGCs). The idea is that the strength of the multi-boson interaction, allowed by the non-abelian gauge

structure of the electroweak theory, can be modified by the presence of new particles (that introduce loop corrections) and the variation in the interaction strength is parametrised by an anomalous component to the gauge coupling itself. The anomalous gauge coupling framework aims to expand the Lagrangian by writing the possible trilinear and quadrilinear interactions in terms of form factors as a function of the interacting particles four-momenta. It must be noted that it is possible to reframe these form factors as the operator described in the previous paragraph for the EFT. A measurement of the coupling strength of the EFT operators can therefore be interpreted in terms of searching for anomalous gauge couplings. Remembering the features of a generic EFT, if no discrepancy is found between the theory and the prediction (i.e. the coupling of the higher-dimension operators are zero, which in turn means that no aGC contributions are observed), it is still possible to compute the limits on the couplings strength hence providing a further, more stringent test of the consistency of the electroweak sector of the Standard Model.

The $W\gamma\gamma$ process that is the object of this study is sensitive to the presence of quartic gauge couplings. For this reason, the coupling of the 14 different dimension eight operators in the EFT framework to which this process is sensitive to has been studied. While these are derived from theoretical assumptions, one should expect that the experimental measurement will not have a good enough sensitivity to compute stringent limits for all of them. Denoting by Φ the Higgs doublet, by $W^{i}_{\mu\nu}$ the $SU(2)_L$ field strength from which

$$\hat{W}_{\mu\nu} \equiv \sum_{j} W^{j}_{\mu\nu} \frac{\sigma^{j}}{2} \tag{52}$$

can be defined (σ^{j} are the Pauli matrices), by $B_{\mu\nu}$ the $U(1)_Y$ field strength and by defining the covariant derivative as

$$D_{\mu}\Phi = \left(\partial_{\mu} - igW^{j}_{\mu} \frac{\sigma^{j}}{2} - ig'B_{\mu} \frac{1}{2} \right) \Phi \,, \tag{53}$$

the functional form of the operators can be obtained. The full list of operators that one can expect to be constrained by the $W\gamma\gamma$ process and their functional form is presented in Table 1. For a more detailed description, one can refer to Reference [79].

1.6.2 *The $W\gamma\gamma$ production in proton-proton collisions*

The study of the $W\gamma\gamma$ process is interesting in many different ways. The measurement of its cross section with high precision constitutes a precision test of the electroweak sector of the Standard Model. The process is also a possible background source to other precision measurements (such as the associate production of a W and a Higgs boson) and for searches of new physics.

The Feynman diagrams for the $W\gamma\gamma$ process are represented in Figure 6. As it can be seen, the process proceeds through a four-particle vertex, allowed by the non-abelian structure of the Standard Model. As described in the previous section, the strength of this coupling can be modified by the presence of new physics processes

$\mathcal{L}_{M,0} = \mathrm{Tr}\left[\hat{W}_{\mu\nu}\hat{W}^{\mu\nu}\right] \times \left[(D_\beta\Phi)^\dagger D^\beta\Phi\right]$	$\mathcal{L}_{T,0} = \mathrm{Tr}\left[\hat{W}_{\mu\nu}\hat{W}^{\mu\nu}\right] \times \mathrm{Tr}\left[\hat{W}_{\alpha\beta}\hat{W}^{\alpha\beta}\right]$
$\mathcal{L}_{M,1} = \mathrm{Tr}\left[\hat{W}_{\mu\nu}\hat{W}^{\nu\beta}\right] \times \left[(D_\beta\Phi)^\dagger D^\mu\Phi\right]$	$\mathcal{L}_{T,1} = \mathrm{Tr}\left[\hat{W}_{\alpha\nu}\hat{W}^{\mu\beta}\right] \times \mathrm{Tr}\left[\hat{W}_{\mu\beta}\hat{W}^{\alpha\nu}\right]$
$\mathcal{L}_{M,2} = \mathrm{Tr}\left[B_{\mu\nu}B^{\mu\nu}\right] \times \left[(D_\beta\Phi)^\dagger D^\beta\Phi\right]$	$\mathcal{L}_{T,2} = \mathrm{Tr}\left[\hat{W}_{\alpha\mu}\hat{W}^{\mu\beta}\right] \times \mathrm{Tr}\left[\hat{W}_{\beta\nu}\hat{W}^{\nu\alpha}\right]$
$\mathcal{L}_{M,3} = \mathrm{Tr}\left[B_{\mu\nu}B^{\nu\beta}\right] \times \left[(D_\beta\Phi)^\dagger D^\mu\Phi\right]$	$\mathcal{L}_{T,5} = \mathrm{Tr}\left[\hat{W}_{\mu\nu}\hat{W}^{\mu\nu}\right] \times B_{\alpha\beta}B^{\alpha\beta}$
$\mathcal{L}_{M,4} = \mathrm{Tr}\left[(D_\mu\Phi)^\dagger \hat{W}_{\beta\nu}D^\mu\Phi\right] \times B^{\beta\nu}$	$\mathcal{L}_{T,6} = \mathrm{Tr}\left[\hat{W}_{\alpha\nu}\hat{W}^{\mu\beta}\right] \times B_{\mu\beta}B^{\alpha\nu}$
$\mathcal{L}_{M,5} = \mathrm{Tr}\left[(D_\mu\Phi)^\dagger \hat{W}_{\beta\nu}D^\nu\Phi\right] \times B^{\beta\mu}$	$\mathcal{L}_{T,7} = \mathrm{Tr}\left[\hat{W}_{\alpha\mu}\hat{W}^{\mu\beta}\right] \times B_{\beta\nu}B^{\nu\alpha}$
$\mathcal{L}_{M,6} = \mathrm{Tr}\left[(D_\mu\Phi)^\dagger \hat{W}_{\beta\nu}\hat{W}^{\beta\nu}D^\mu\Phi\right]$	
$\mathcal{L}_{M,6} = \mathrm{Tr}\left[(D_\mu\Phi)^\dagger \hat{W}_{\beta\nu}\hat{W}^{\beta\mu}D^\nu\Phi\right]$	

Table 1: list of the dimension-8 operators that affect the $W\gamma\gamma$ quartic coupling.

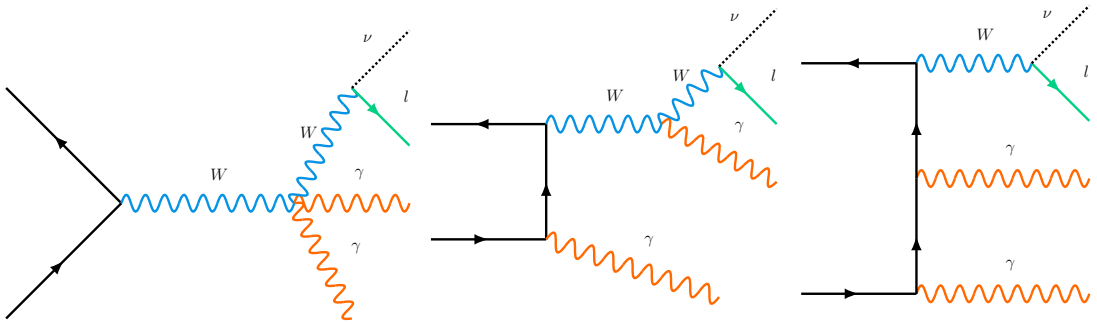


Figure 6: example of the three topologies for the tree level $W\gamma\gamma$ production.

and the result can be interpreted in the EFT framework. Besides being a precision measurement of the Standard Model, the analysis of the $W\gamma\gamma$ process can be interpreted also in terms of searches for new physics.

The cross sections for the multi-boson processes are very small, which in turn means that to be able to measure this final state, a large amount of data is needed. The $W\gamma\gamma$ production, which is the process with the largest cross section among the triboson productions, has been studied in the past both by the CMS [3] and ATLAS [4] collaborations. The measurement has been performed by using data collected up to 2013 at a centre of momentum energy of $8\,TeV$ corresponding to an integrated luminosity of almost $20\,fb^{-1}$. Both these measurements were able to measure the cross section and set limits on some of the anomalous couplings listed before. However, given the limited statistics, the significance needed to claim the observation of the process has not been reached.

By performing the analysis with the data collected by the CMS experiment from 2016 to 2018 at a centre of momentum energy of $13\,TeV$ corresponding to almost $140\,fb^{-1}$, the statistical error can be reduced and the precision obtained by the measurement can be greatly improved. Moreover, as it will be discussed later in Chapter 6, the sensitivity to the presence of anomalous gauge couplings is enhanced at high energies or transverse momentum. By having data collected at a higher centre of mass

energy, the high momentum tail of the distribution is more populated if compared to previous studies and a better sensitivity on the anomalous couplings is expected.

There is no promised road leading to definite results. What's important is how to keep open as many options as possible.
— Makoto Kobayashi

2 LHC AND THE CMS DETECTOR

The data analysed in this work were collected by the Compact Muon Solenoid (CMS) experiment at CERN Large Hadron Collider (LHC) between 2016 and 2018. After a brief introduction to the fundamentals of accelerator physics, a description of the accelerator complex at CERN is given with particular focus on the LHC. A brief section is dedicated to the future developments of the LHC itself. The CMS experiment with all its subdetectors and trigger system is then described in detail.

2.1 BASICS OF ACCELERATOR PHYSICS

As it is known from special relativity, in order to produce heavy particles a large amount of energy is needed. The best way that has been found so far to produce heavy particles is to accelerate charged particles and then smash them into a target or with each other. A particle accelerator is a machine that exploits electromagnetic fields to accelerate and increase the energy of particles.

Particle accelerators can be build in a linear or in a circular configuration. A linear accelerator is made of a straight pipe, where vacuum is produced, and a set of drift tubes. The drift tubes provide the electric field that accelerates the particles. If the length of the tubes and the radio-frequency voltage applied to them are correctly chosen, the particles receive a continuous acceleration while passing toward the beam pipe. The particles accelerated in this way can collide on a target or head-on with another beam of particles accelerated in the same way but coming from the opposite direction. A linear accelerator, however, has two main drawbacks. The first one is that to accelerate more and more the particles, the length of the machine must be increased. The other one is that, in the configuration just described, only one interaction point for each accelerator can be exploited to build physics experiments.

To overcome these problems, circular accelerators have been developed. The clear advantage of a circular accelerator is that particles keep passing through the same path and hence can be accelerated many times by using a single radio-frequency electric field. The main drawback of a circular accelerator is the fact that the particles must be maintained on a circular orbit by exploiting magnetic fields. Modern circular accelerators are referred to as *synchrotrons*, since both the magnetic field and the cavity frequency must be increased and synchronised as the speed of the particles increases. While in linear accelerators the maximum reachable energy is limited by the length of the machine, in circular accelerators the limit is its radius: when the energy and speed of the particles increase, a stronger magnetic field is needed to maintain them in a stable orbit. To go to higher energies, more intense magnetic fields or a larger ring are needed.

In principle, all charged particles and nuclei can be accelerated. However, unstable particles could decay during the acceleration process, hence spoiling the collision

event. For this reason, the only particles that have been used up to the present time in particle accelerators are electrons and protons or, although for different purposes, stable nuclei. Machines that exploit electrons and protons have different points of strength and weaknesses and are hence exploited for different purposes.

In the Standard Model, electrons are fundamental particles. This means that by measuring the energy of the two colliding beams (one of electrons and of positrons) in a circular accelerator, the total initial energy of the collision is known. This is a great advantage since conservation laws can be applied to derive quantities in the final state. For example, if studying the decay of a W boson into a lepton and a neutrino, the energy and direction of the neutrino can be reconstructed from the knowledge of all the other particles in the event as a missing energy and the invariant mass of the boson can be derived. Another advantage of knowing with high precision the energy of the collision is that it can be tuned in order to study in high detail the characteristics of a resonance, as it has been done, for example, with the Z boson at the LEP accelerator at CERN in the 1990s. The main disadvantages of an electron machine is that electrons lose a lot of energy via bremsstrahlung. A charged particle in a circular trajectory, in fact, loses energy by emitting photons in the direction tangent to the trajectory. The energy loss is inversely proportional to the fourth power of the mass of the particle: light particles, such as the electron, lose much more energy that heavy ones.

Protons are almost two thousands times as massive as electrons, and hence the loss of energy due to synchrotron radiation is 10^{13} times smaller. A higher mass means not only a drastic reduction of the energy lost via synchrotron radiation, but also a higher energy reachable by the particle beam in a machine with the same radius as one with electrons. The drawback is the fact that protons are not elementary particles, but are made of quarks: the energy of the collision does not correspond to the energy of the proton, but is the energy of the two colliding quarks. Each quark carries a fraction of the energy of the proton that can be estimated, statistically, by using the parton distribution functions (as explained in Section 1.5). For this reason, it is not possible to exploit conservation laws of the overall quantities. The energy on the plane perpendicular to the beam line, however, is close to zero before the collision and it must be the same after it hence driving to the usage of conservation laws of transverse quantities. On the other side, the fact that the centre of momentum energy is different between each collision is the most important feature of a hadron collider because in this way, by recording many collisions, a wide range of energies is naturally scanned and the presence of possible unknown resonances can be tested. For this reason, hadron machines are often referred to as *discovery machines* while electron accelerators are identified as *precision machines*.

Besides the energy at the centre of momentum, another fundamental quantity in the physics of accelerators is the luminosity. It is defined as

$$\mathcal{L} = \frac{f_{rev} \cdot n_{bunch} \cdot N_p^2}{4\pi\sigma\sigma_y} \tag{54}$$

where f_{rev} is the revolution frequency of the particles, n_{bunch} is the number of particles bunches in the accelerator, N_p is the number of particles per bunch, and σ_x and σ_y are the transverse dimensions of the beam.

2.2 THE LARGE HADRON COLLIDER

The Large Hadron Collider [80] is the largest and most powerful colliding machine in the world. It is built in a 27 km long underground tunnel in the region around the Swiss-French border close to the city of Geneva. It is hosted in the tunnel built in the 1980s for the Large Electron-positron Collider (LEP). It is a proton-proton collider: each of the two beams has been accelerated up to 6.5 TeV, giving a centre of momentum energy of 13 TeV. The beams are forced to collide in four points along the accelerator ring where the main LHC experiments are located. CMS (Compact Muon Solenoid) and ATLAS (A Toroidal LHC ApparatuS) are big, multipurpose, high energy experiments: given their complexity and the variety of subsystems they can collect physics data that can explore many different fields of particle physics, from precision studies to new particle searches.

Before reaching the LHC, the energy of the protons is increased by exploiting a chain of accelerators (Figure 7). Protons are obtained from the ionisation of hydrogen atoms and then injected into the Linac, which is a linear accelerator that increases the energy of the particles up to 50 MeV. After this first stage of linear acceleration, the protons are injected into a series of three circular accelerators. After the Booster, that brings the energies of the protons to 1.4 GeV, the protons are injected in the 628 m long Proton Synchrotron that accelerates them to 25 GeV. The protons are then injected in the 7 km long Super Proton Synchrotron (SPS), the machine that in the 1980s allowed for the discovery of the W and Z bosons, that brings their energy up to 450 GeV. Up to this stage, a single beam of protons is accelerated. After the SPS, by means of bending magnets, the proton bunches are inserted in the LHC ring in two beams rotating in opposite directions.

In the LHC, the particles travel in two beam pipes where an ultra-high vacuum is performed in order to limit the presence of impurities that can spoil the proton beams. Since the beams inside the accelerator are needed to be sufficiently clean for many hours, the pressure of the vacuum should be of the order of 10^{-11} mbar. To reach these values of pressure, the beam pipe is cooled down to 1.9 K so that residual gases present in the pipe condense and adhere to the walls of the pipe itself. The acceleration of the protons is performed by two systems of superconducting radio frequency cavities working at 400 MHz. When the machine started operations, in 2009, the energy of each proton beam was of 3.5 TeV, while during the Run 2, between 2016 and 2018, the energy of each beam was raised to 6.5 TeV.

To maintain the protons in the circular orbit, 1232 dipole magnets as the one schematised in Figure 8 are placed along the tunnel. Differently from all the previous accelerators, the LHC magnets are superconducting magnets. This challenging requirement was needed because, in order to maintain in circular orbit 7 TeV protons, each dipole magnet has to operate a field of around 8 T which can be reached only by exploiting superconducting materials. Each of the 15 m long dipole magnets is made

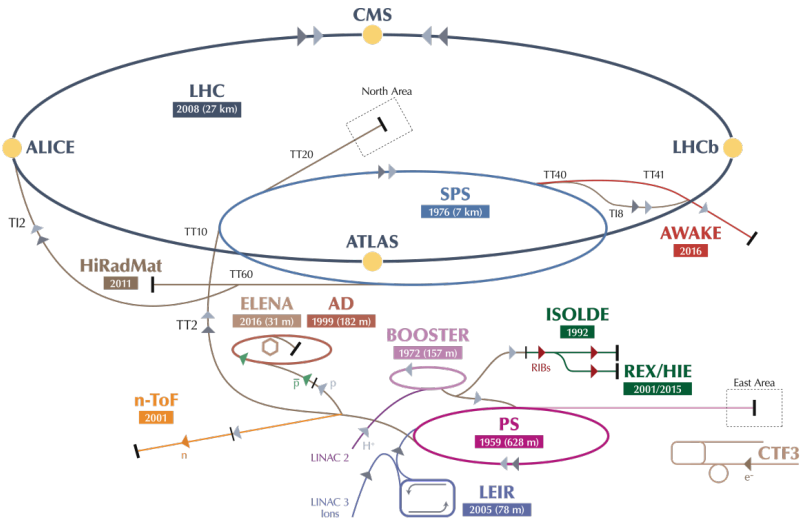


Figure 7: scheme of the CERN accelerator complex. The protons are first injected in the Linac before passing to circular accelerators (Booster, PS, SPS) and eventually being injected in the LHC.

of niobium-titanium (NbTi) and is operated at 2 K. To cool the magnets at such a low temperature, superfluid helium is employed. Besides the dipoles that bend the protons in the trajectory, 392 quadrupole magnets are implemented to squeeze the beam especially before the interaction points. A smaller beam results in a higher proton density and hence a higher number of collision events and consequently a higher luminosity.

During the LHC Run 2, there were 2556 proton bunches in the ring during the physics fills, each one with approximately 10^{11} protons. This corresponds to an instantaneous luminosity peaking at $2 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. In three years of operation at a centre of momentum energy of 13 TeV the machine delivered approximately $140\,\mathrm{fb}^{-1}$ of integrated luminosity.

2.2.1 *Towards the future: high luminosity and beyond*

Among the most interesting searches foreseen for the near future in high energy physics, the study of multi Higgs production and the searches for possible new particles are at the top of the list. These investigations, however, are strongly limited by the present time statistics, which amounts to $140\,\mathrm{fb}^{-1}$ for the Run 2 and will go up to approximately $300\,\mathrm{fb}^{-1}$ after the Run 3 (expected to end in 2024). The small cross section for the production of two Higgs bosons and the small signal-to-noise ratio make the statistics collected by the LHC insufficient to perform precision studies. Possible new resonances are expected to be found at the TeV scale where the spectrum is less populated and the events are rarer; physics searches would then benefit a lot from a higher statistics.

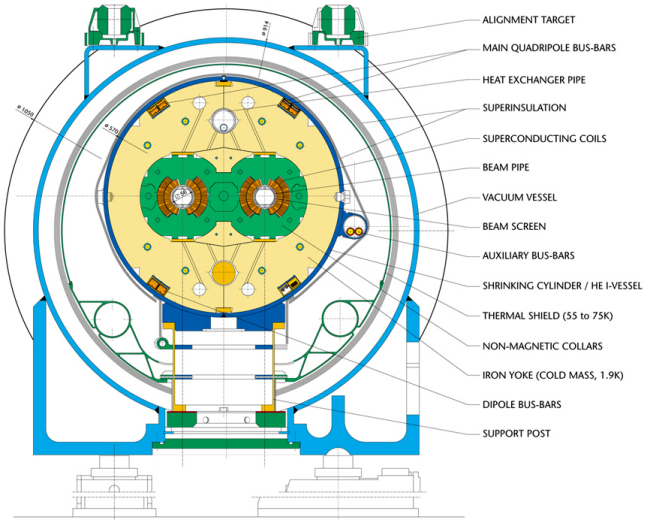


Figure 8: section of a LHC dipole magnet. The structure of the two beam pipes, the superconducting magnets and the cooling system is highlighted.

For these reasons, the future development of the LHC, the High-Luminosity LHC, will focus on an increase of the luminosity and hence of the amount of collected data. The goal is to record an integrated luminosity of approximately $3000\,\mathrm{fb}^{-1}$ during the ten years of the project. To do so, new magnets will be needed to squeeze the beam more than it is done at the present time and *crab cavities* will be installed in order to have more head-on collisions of the proton bunches [81]. In an environment with higher luminosity, the requirements of readout capabilities and radiation hardness of the detector components will become more and more stringent. For this reason, the LHC experiments are already planning and testing new electronics and new detectors that will be installed for this groundbreaking project.

There are still ongoing discussions on which machine will be the successor of the LHC. The discovery of the Higgs boson pushes toward the construction of an electron collider where the characteristics of this fundamental particle can be studied with high precision (as it has been done for the W and Z bosons at LEP). The future collider could hence be a $100\,\mathrm{km}$ electron circular collider that could then be upgraded to a hadron collider in the same way as LEP has been upgraded to LHC. This plan has the clear advantage of building a single tunnel facility to first study in detail the Higgs boson characteristics and then move to higher energies (around $100\,\mathrm{TeV}$) to test the presence of new yet undiscovered particles.

2.3 COORDINATE SYSTEM

The quantities that are measured in a collision at one of the LHC experiments are referred to a particular coordinate system. The origin of the system is placed at the centre of the detector, where the two beams collide. The $\hat{z}$ axis lies on the plane of the accelerator and is tangent to the direction of the beam. By looking at the accelerator ring from above, the axis is directed anticlockwise. The $\hat{x}$ axis lies on the intersection

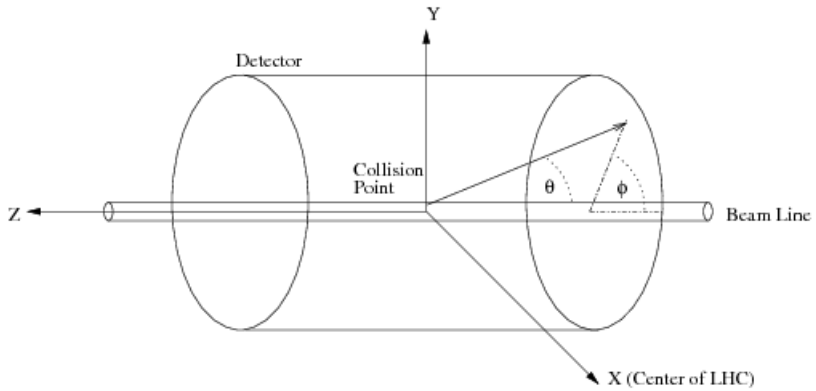


Figure 9: schematic illustration of the coordinate system for one of the LHC experiments. Besides the cartesian plane, also some important angular variables are defined.

of the plane where the accelerator lies and the plane perpendicular to the beampipe; it points toward the centre of the ring. The $\hat{y}$ axes is perpendicular to the two axes just defined and has the direction needed to build a right-hand reference system (i.e. it points upwards). A schematic representation of the coordinate system is presented in Figure 9.

From this definition, it is possible to derive some angular quantities that are useful in the description of the event:

- $r = \sqrt{x^2 + y^2}$ which describes the position relatively to the interaction point on the plane perpendicular to the beamline;
- $\phi = \tan^{-1}(x/y)$ is the azimuthal angle defined to be zero in the positive direction on the $\hat{x}$ axis and growing towards the positive direction on the $\hat{y}$ axis;
- $\theta = \arctan(y/z)$ is the polar angle relative to the beamline.

Instead of using the polar angle, it is useful to define the pseudorapidity as:

$$\eta = -\ln\left(\tan\left(\theta/2\right)\right) . \tag{55}$$

This is done in order to have a quantity that is additive and Lorentz-invariant. The value of η is equal to zero in the plane transverse to the beamline and grows to infinity towards the direction of the beams.

As already described, the energy at the centre of momentum of the collision is not known due to the fact that the particles that actually collide are quarks and not protons. Many useful quantities are defined in the plane perpendicular to the beamline and for this reason are referred to as "transverse" quantities, such as the transverse momentum $p_T = p \sin\theta$ and the transverse energy $E_T = \sqrt{p_T^2 c^2 + m^2 c^4}$.

2.4 THE COMPACT MUON SOLENOID EXPERIMENT

The Compact Muon Solenoid (CMS) experiment is one of the four main experiments of the LHC. Together with ATLAS, it is a general purpose experiment: this means

that a large variety of measurements in particle physics can be performed with this apparatus. In particular, the CMS experiment has made some pioneering studies in searches for new particles (such as the Higgs discovery) and supersymmetry, while performing high precision measurements of Standard Model processes and also digging in the very promising field of the physics of the bottom quark.

To perform such high precision measurements, a high performance detector and reconstruction procedure are needed. The CMS detector, schematically depicted in Figure 10, is built as a cylinder closed on the sides around the beamline in order to cover the vast majority of the volume where the particles coming from the collisions can travel. The central feature of the detector is a superconducting magnet that curves the particles. The track of the charged particle is reconstructed by the silicon tracker, which is the innermost detector built just outside the beampipe. From the trajectory information, the particle momentum can be inferred. Just outside the silicon tracker, the electromagnetic and hadronic calorimeters allow for the measurement of the leptons, photons and hadron energies. The outermost detectors are the muon chambers. Given the fact that the superconducting magnet is in the inner part of the detector and that the magnetic field lines are closed with steel yokes on the outside, the direction of the magnetic field changes from inside and outside the solenoid, making the muons travel in a characteristic "s" shaped trajectory (as depicted in the well known CMS logo).

In this section, the CMS detector and its subsystems are described. For a more detailed and technical description, one can refer to Reference [82].

2.4.1 *The superconducting magnet*

The superconducting magnet (Figure 11) is the central feature of the CMS detector. It is 12.5 metres long and has a diameter of 6 metres. The design performance allows to have a magnetic field of 4 T with a stored energy of 2.6 GJ at full current.

The superconducting material exploited in the magnet is niobium-titanium (NbTi), which has been arranged in Rutherford-type cables co-extruded with pure aluminium. The mechanical structure outside the superconducting material is made of aluminum. The choice of the materials was dictated by the need to limit the shear stress level inside the winding and prevent mechanical damages on the insulating system. The final design is the one of a self-supporting conductor. The winding is composed of four layers instead of the one or two layers structure that has been implemented in previous high energy experiments; this configuration was necessary to generate a field as strong as 4 T. The magnet is operated at 4.6 K in order to reach the superconducting regime of the NbTi.

The structure of the detector is supported by an iron structure built on the outside of the subdetectors. The yoke is composed of 11 elements weighting from 400 to nearly 2000 tonnes and can be moved around in the cavern where the detector is located in order to easily allow maintenance and upgrade procedures. Besides the structural function, the iron yoke constitutes the return element of the magnetic field.

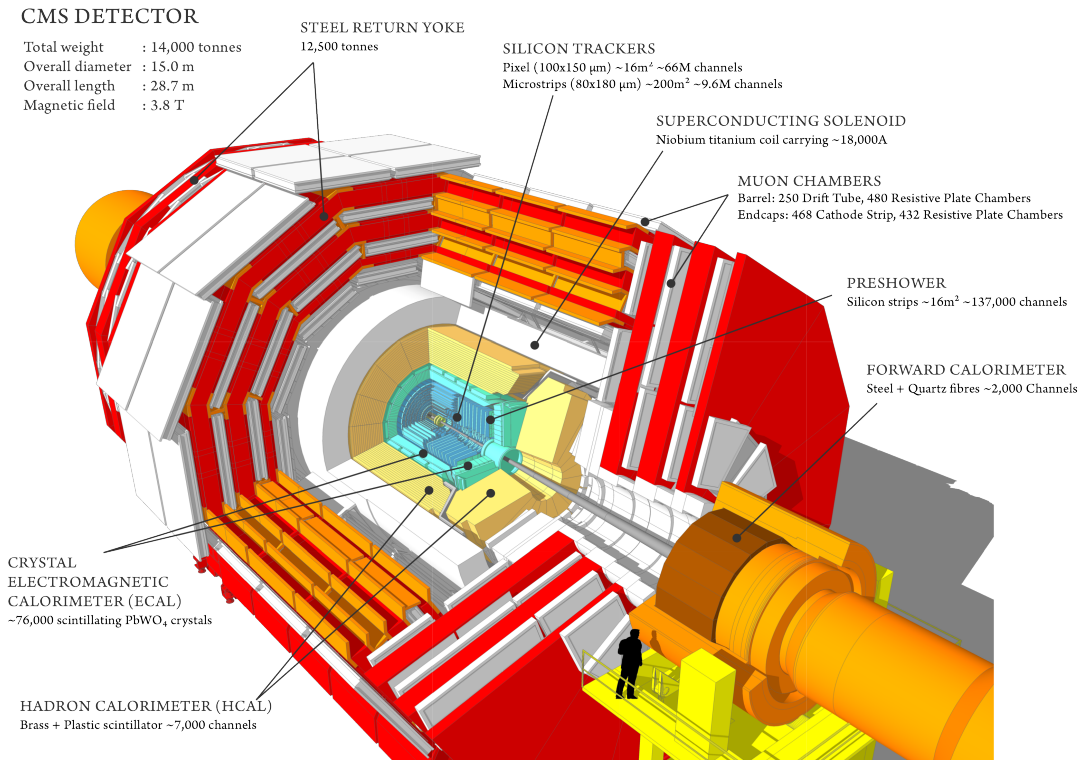


Figure 10: schematic representation of the CMS detector. The main subsystems are briefly described in the labels.

2.4.2 *The tracking system*

One of the key points in the reconstruction of the kinematics of a particle is the determination of its momentum that can be inferred from the curvature of the track of the particle. The goal of the CMS silicon tracker is the reconstruction of the tracks of the charged particles with a precision as high as possible by taking into account the pileup and the particle occupancy.

The requirements that have to be matched in the construction of the tracker are manyfold and some of them are in contrast with each other. To cope with the high number of interactions in the same bunch crossing (up to 60 collision vertices) and with the very high rate of the collisions (one every 25 ns), a detector with a high granularity and a fast response is needed. This is in conflict with the principle of having as little material as possible to avoid multiple scattering, bremsstrahlung and photon conversion. Being the tracking system the closer one to the beamline, it is also the one that is more subject to radiation damage.

Given all the aforementioned requirements, a silicon detector with a pixel structure in the inner part and a strip structure in the outer part was designed and built. The pixels guarantee a higher granularity in the region where the occupancy is higher, while the strips allow for a reduction of the number of readout channels (being bigger and hence covering more volume) and of the overall costs.

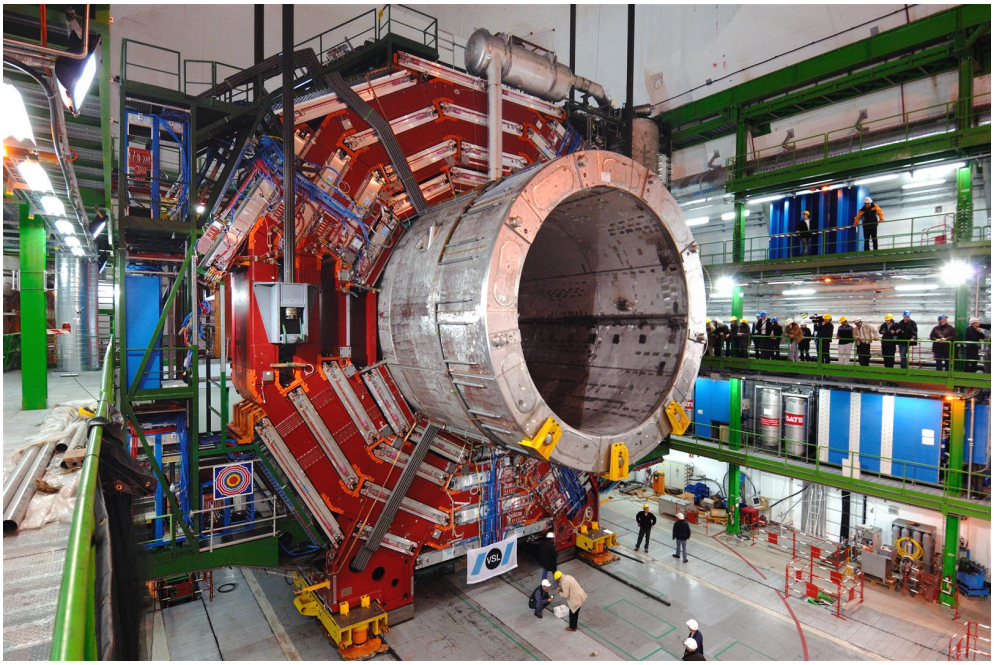


Figure 11: picture of the central part of the CMS detector just lowered in the underground cavern. There can clearly be seen both the superconducting magnet (the big silver cylinder) and the iron yoke structure in red [83].

Regarding the radiation damage, there are three coexisting factors that have been taken into account for the construction of the detector. A ionising particle can create positively charged holes that affect mostly the front-end chips. Non-ionising energy deposits can damage the silicon crystal lattice and this in turn increases the leakage current and changes the doping and hence the depletion voltage. To avoid positive feedbacks on the self heating of the silicon sensors and to reduce the effect of the leakage currents, the tracker is operated at low temperatures (around $-10\,^{\circ}\mathrm{C}$). The design of the tracker is optimised to have a signal-to-noise ratio of at least 10:1 and has been built to be fully operational for 10 years of data acquisition.

The pixel detector [85] is responsible for the determination of the seed of the tracks for charged particles and for the reconstruction of the secondary vertexes. Since its upgrade in 2017, it is made of four barrel layers (Figure 12), extending from 3 to 16 cm from the beamline, and three endcap discs placed at $z = \pm 29.1$ cm, $z = \pm 39.6$ cm and $z = \pm 51.6$ cm. It covers the pseudorapidity range up to $|\eta| < 2.5$. Each layer consists of pixels of the size of $100 \times 150\ \mu\mathrm{m}^2$. Since the silicon sensors work with charge collection of drifting electrons, the influence of the 4 T solenoid magnet must be taken into account. The choice was to place the pixels in a configuration so that the direction of the electron drift is perpendicular to the direction of the magnetic field. In this way, the charge collection happens in more than one adjacent pixel and this allows for a better spatial resolution that can get to a precision better than $15 - 20\ \mu\mathrm{m}$, depending on the process under study. A similar approach is implemented in the forward detectors that are tilted in order to exploit the charge spreading and hence increasing the spatial precision. The readout of the pixels is performed by ASICs

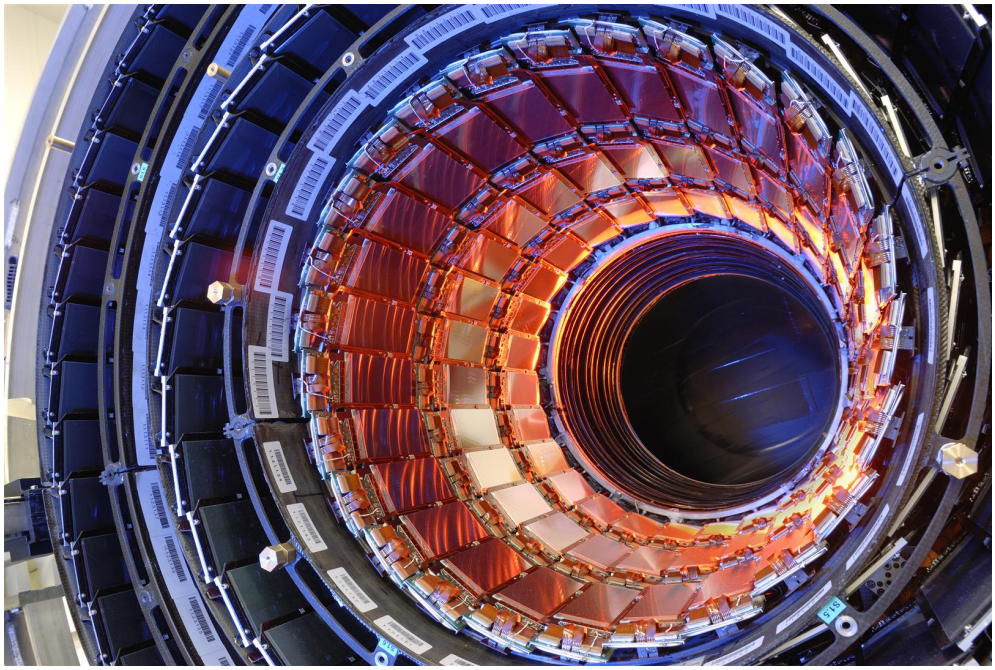


Figure 12: picture of the inner tracker of the CMS experiment. The different layers of silicon strip detectors can be seen [84].

bump bonded to the sensors that provide both amplification and buffering of the charged signal. A zero suppression is applied for pixel unit, meaning that only signals that are above a give threshold are read.

The outermost part of the tracker is made of silicon strips. It is composed of 15148 detector modules with 15 different geometries for a total of 24244 strip sensors, which amount to a total active area of $198\,m^2$. Each module is made of either one thin $320\,\mu m$ sensor or two thick $500\,\mu m$ sensors and is supported by either a carbon or a graphite frame, depending on the position of the module in the tracker. The inner barrel is placed from 25 to $50\,cm$ from the beamline, divided into four layers. The strip pitch goes from 80 to $120\,\mu m$. Two inner endcap discs are placed on both side of the inner barrel and guarantee hermetical coverage up to $\eta = 2.5$. The outer barrel tracker is made of a wheel that occupies the radius from 55 to $115\,cm$ from the beamline. It is made of six layers that are slightly displaced in ϕ one from the other in order to achieve a full coverage of the volume. Nine endcap disks, going up to $280\,cm$ in both direction from the $z = 0$ position close the outer cylinder. The silicon modules are here mounted in a "petal" structure to fully cover the phase space.

The alignment of the structure is controlled with an infrared laser that allows for a precision of $100\,\mu m$ in the positioning of the modules. The main tool for alignment monitoring, however, is the one based on the track reconstruction. Both collision and cosmic muons tracks are exploited in order to obtain a unique set of alignment constants. By means of the global Millepede [86] and a Kalman filter based [87] algorithms, the linear system of equations with almost a hundred thousand unknowns can be solved and the alignment of the detector can be determined.

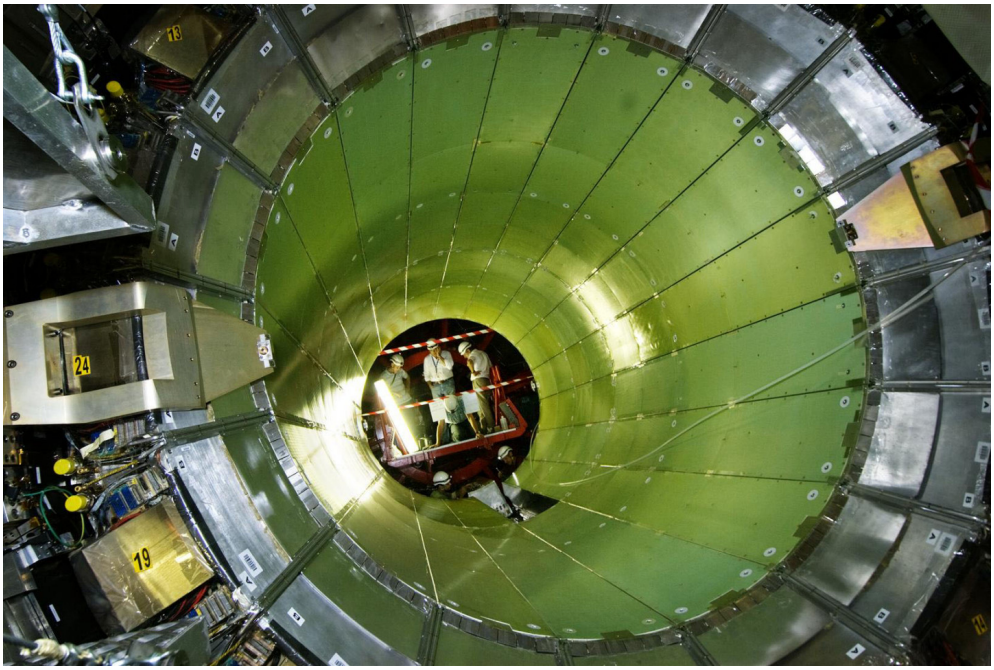


Figure 13: picture of the barrel section of the electromagnetic calorimeter at a late stage of construction. As it can be seen, the last crystal supermodule is being inserted in the structure [88].

2.4.3 *The electromagnetic calorimeter*

The electromagnetic calorimeter (ECAL) (Figure 13) is the next-in-line subdetector that the particles encounter after the tracker. It is an hermetic homogeneous calorimeter made of 75848 lead tungstate ($PbWO_4$) crystals. As for the tracker, it has a central cylindrical part called barrel and two endcaps perpendicular to the beamline on the sides. Its main function is to measure the energy and the position of electrons and photons.

The choice of the lead tungstate as the material of the crystals has many advantages. The high density of the material and the short radiation length ($0.89\,cm$) allowed to build a smaller detector while guaranteeing the needed accuracy in the measurement of high energy electrons and photons. Moreover, the Molière radius of the material is $2.2\,cm$ hence resulting in a high granularity of the calorimeter itself giving an excellent spatial determination of the particle position. Another parameter that has to be taken into account in the construction of the calorimeter is the time response of the crystals: lead tungstate ones have a scintillating time that allows for the collection of about 80% of the light in $25\,ns$ which is the time between two consecutive LHC bunch crossings. The light output is of the order of 4.5 photoelectrons per MeV; to collect as much light as possible, the internal reflection of the crystals has been optimised by polishing the sides of the crystals. However, the barrel crystals have a truncated pyramidal shape (to cope with the cylindrical geometry) that makes the light collection different at different depth of the crystal itself. To solve the problem, one lateral face has been depolished hence changing the reflective properties. Like

all the other detectors, also the ECAL crystals have to withstand a high radiation environment. Radiation damage from ionising radiation increases the light absorption in a wavelength-dependent way. A monitoring system is in place which exploits the light of two lasers to monitor the loss in transparency of the crystals at different wavelengths and allows to compute corrections in the measured energy.

The barrel section of the detector covers the pseudorapidity region up to $|\eta| = 1.479$ and is made of 61200 crystals mounted on a quasi-projective geometry towards the collision point, so that the probability for a track to be collinear with a crack between the crystals is reduced. Each of the barrel crystals has a $26\times26\,\mathrm{mm}^2$ front face area and a length of 23 cm, which corresponds to nearly 26 radiation lengths. The face of the crystal is 1.29 m away from the beamline. The crystals are contained in an aluminum structure that resembles the shape of an hive and are organised in submodules and supermodules.

The two endcaps are placed perpendicularly to the beamline at 3.15 m from the collision point and cover the pseudorapidity region $1.479 < |\eta| < 3.0$. The endcap crystals have an area of $30\times30\,\mathrm{mm}^2$ and a length of 22 cm corresponding to nearly 25 radiation lengths.

Operating the detector at a constant known temperature is fundamental since the number of scintillation photons and the amplification system of the photodetectors are both temperature dependent. A water-flow based cooling system is capable of maintaining the detector temperature to 18 °C with a variation smaller than ±0.05 °C.

The readout of the light produced by the interaction of a particle with the crystal is performed using avalanche photodiodes for the barrel part of the detector and vacuum phototriodes for the endcaps. In addition to the requirements on readout speed and radiation hardness, they should amplify the small light yield coming from the crystals and be insensitive to particles passing through them. The readout is performed by on-detector electronics by reading supermodules of 5×5 crystals together. The information on the energy deposit in the tower and on the lateral profile of the shower are sent to the Level 1 trigger that accepts or rejects the signal in 3 µs. The selective readout algorithm classifies the ECAL signal interest depending on the intensity of the energy deposit; this is done in order to have the possibility to suppress the amount of signal to be processed by exploiting a different lower threshold. As for the tracker, also the ECAL readout exploits zero suppression to avoid filling the trigger bandwidth with uninteresting signals.

The calibration of the calorimeter has a twofold nature: a global one, where the absolute energy scale is determined, and a relative one, where the response of different crystals is compared. The important features that the detector must have are uniformity within the detector itself and stability in time, so that the data recorded at different times can be compared. The main difference in energy response is given by the different light yields in scintillation between different crystals in different supermodules. The intercalibration procedure exploits cosmic rays and particles produced in the collision (mainly isolated electrons coming from the W decay) that allow for the determination of calibration coefficients for the different supercrystals. Another possible calibration technique is the reconstruction of the invariant mass of known particles, such as $\pi^0 \to \gamma\gamma$.

As already mentioned, the transparency of the crystals changes during the period of data taking due to radiation damage. This change in transparency is η-dependent, being minimum in the barrel region (order of some percent) and maximum in the endcaps regions closer to the beamline, where it can go up to tens of percent. Besides this damaging process, the effect of annealing must be taken into account. The damage provoked by the irradiation is "cured" at the ECAL operating temperature: this effect results in an oscillation of the crystal transparency between the runs and the periods when the LHC bunches are refilled. To measure the evolution of the crystal transparency, light coming from laser pulses is injected in the system and the response of the photodiodes is measured. Two blue lasers, with wavelength very close to the scintillation emission peak, and an infrared one are used. This procedure allows to determine the transparency of the crystals with an accuracy of the per mille level.

The energy resolution of ECAL can be parametrised as

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \,. \tag{56}$$

S is the stochastic term and it has contributions from event-to-event fluctuations in the lateral shower containment, photostatistics contributions and fluctuation in the energy deposited in the preshower absorber (where it is present). C is the constant term and it has contributions coming from the non-uniformity of the longitudinal light collection, from possible intercalibration errors and from the leakage of the shower energy from the back of the crystals. In particular, leakage from the rear side of the crystals has been proven to be negligible both in terms of energy resolution and in terms of giving direct signals in the photodetectors. N is the noise term that takes into account contributions from electronics, digitisation and pileup noises. In particular, the signal shaped by the preamplifier will extend in time over several consecutive LHC bunch crossings: a multi-amplitude algorithm is exploited to account for this effect. The typical performance of the calorimeter allows for a resolution better than 0.45% for electrons up to 120 GeV.

2.4.4 *The hadron calorimeter*

The hadron calorimeter (HCAL) is a fundamental detector in the measurement of hadronic jets energy that is in turn a key element in the determination of the missing transverse energy of an event and hence in the determination of the presence of neutrinos. The calorimeter is placed outside ECAL, at a radius of 1.77 m from the beamline and fills the volume up to the superconducting solenoid. The limited amount of space available in this region drove the choice to add an outer part of the hadron calorimeter outside the solenoid in order to measure also very energetic showers that would not be contained in the inner part. As for the other detectors, the hadron calorimeter has a structure divided in barrel and endcaps.

The barrel calorimeter, covering the region up to $|\eta| < 1.3$, is a sample calorimeter made of absorbing plates alternated with scintillating material. The choice for the absorber has been driven by the fact that the detector operates in a high-intensity

magnetic field and hence a non-magnetic material had to be adopted. Given this requirement and the need for a high-density material, the choice was to use 14 brass plates plus two steel plates in the front and the rear of the calorimeter in order to guarantee structural strength. The last layer of absorber is thicker in order to avoid as much as possible shower leakages outside the calorimeter. This structure amounts to nearly 6 interaction lengths (λ_I) plus another $1.1\,\lambda_I$ coming from the ECAL material.

The readout is performed with 70000 plastic scintillators, chosen for their long-term stability and moderate radiation hardness. The scintillators planes are interposed between the absorbers and in particular the first layer is located in front of the inner steel plate so that the showers can be sampled before their interaction with the calorimeter in order to estimate the energy lost in the dead material between ECAL and HCAL. Each tray of scintillators has small stainless steel tubes that are used to carry a radioactive source for calibration purposes. The scintillators are coupled to wavelength shifting fibres that are then spliced into clear fibres that are subsequently arranged into readout towers and connected to hybrid photodiodes.

The hadron endcap calorimeter covers the region $1.3 < |\eta| < 3$, which is approximately 13% of the solid angle and contains about 34% of the particles produced in the final state. As for the barrel calorimeter, brass plates are exploited as absorbing material, adding up to 10 interaction lengths. The geometry of the plates is optimised in order to reduce the cracks between the barrel and the endcap parts of HCAL. The structure is longitudinally segmented, which means that there are two or three division in depths of the calorimeter that are read out separately. This is important since this detector, and especially the towers closest to the beamline, are exposed to high level of radiation; having a segmented structure allows to estimate the degradation of the scintillators and hence a possible differential correction of their read out.

As already pointed out, some high energy showers could not be contained in the barrel calorimeters that are built inside the solenoid. For this reason, the hadron calorimeter is extended outside the magnet guaranteeing a total depth of at least $11.8\,\lambda_I$.

The last section of the hadron calorimeter is the forward calorimeter, which is placed around the beamline at a distance of $\pm 11.2\,\mathrm{m}$ from the interaction point. The forward calorimeter covers the region up to $|\eta| = 5$ and is subjected to the highest values of radiation of the whole CMS detector. This hostile environment leads to the necessity of building a highly radiation resistant apparatus. This was the reason behind the choice of building a Cherenkov detector with quartz fibres as active medium and steel absorbers. As for the endcap part of HCAL, also the forward detector is built in two longitudinal segments: this makes it possible to distinguish the signature of an electromagnetic or hadronic particle, since the former deposits large fraction of its energy in the first part of the detector, while the latter releases it almost equally in the two sectors. Besides the measurement of particles and jets, the forward calorimeter is also exploited for the determination of the collected luminosity.

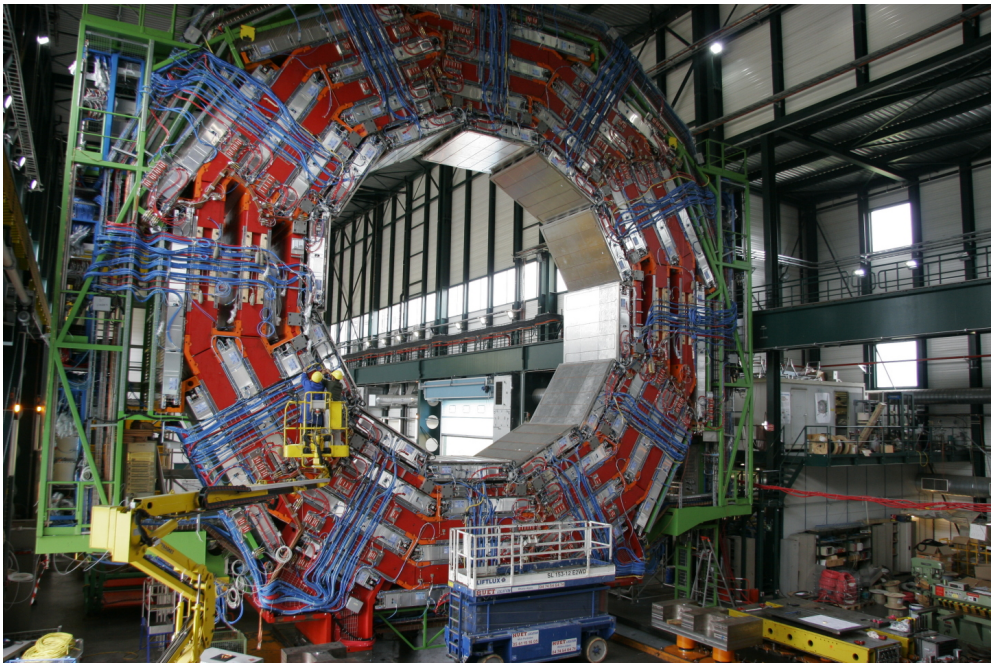


Figure 14: muon chambers installed in one iron barrel wheel [89].

2.4.5 *The muon system*

The last detector that a particle generated in a collision might encounter is the muon system (Figure 14). While almost all the particles lose all their energy in the calorimeters, muons are very penetrating and survive the interaction with the material of the inner detectors and can then be measured in the outermost part of the apparatus. The goal of the muon system is to identify muons, measure their momentum and use the information from the interactions for trigger purposes. Given the need to reconstruct muons in the entire kinematic range and the dimensions of the outer region of CMS, inexpensive, reliable and robust detectors had to be implemented, hence leading to the choice of gaseous particle detectors.

As for the other subsystems, also the muon chambers are divided in a barrel and in two endcaps. The barrel region is the optimal one in terms of uniformity of the magnetic field, low rate of produced muons and low neutron-induced background. For this reason, drift chambers with standard rectangular drift cells are used. The segmentation of the iron yoke constrains the dimension of the chambers to have 2.4 m long wires. The distance between the wires has been chosen so that the maximum drift path is 21 mm, which corresponds to a drift time of 380 ns in the argon and carbon dioxide mixture. The drift tube chambers that cover the pseudorapidity region up to $|\eta| = 1.2$ are organised in four stations, where the first three chambers measure the muon coordinate in the bending plane $(r - \phi)$, while the last one measures the z direction, i.e. the direction along the beamline. It is important to have an identification of the bunch crossing from which the muons come, and this has been done by shifting the drift cells by a half-cell one from the other in order to be able to easily measure the time of the muons crossings. Test beams have estimated that the spatial resolution is

of the order of 170 µm, the uncertainty on the reconstructed time is smaller than 5 ns and the overall efficiency of the cell is generally above 99.5%. The trigger efficiency, intended as the ability of correctly assign a muon to its bunch crossing, has been measured to be close to 100%.

In the endcap regions, the muon system uses cathode strip chambers that are better suited to be used in this environment, where fast response, fine segmentation and radiation resistance are needed. One of the big advantages in using this setup is that precise gas temperature and pressure control are not required, making this the right choice in a region where radiation fluxes are higher and the magnetic field is not as uniform as in the barrel. The pseudorapidity coverage goes up to $|\eta| = 2.4$, matching the same coverage of the tracker and ECAL. The cathode strip chambers are multi-wire proportional chambers made of 6 anode planes interleaved with 7 cathode planes and cover an area of about 5000 m^2. Test beams measured the time resolution to be around 11 ns for a single plane and smaller than 5 ns when using all the six planes in a chamber. The spatial resolution has been measured to be of the order of 80 µm when exploiting the full cell, much better than the 150 µm of the design goal.

Even though muons are very penetrating particles, the great amount of material they traverse before interacting with the muon chambers results in multiple-scattering interactions that can spoil the muon reconstructed momentum of about 9% in the central region of the detector for a 200 GeV muon. To be able to recover this effect and get a determination of the particle momentum at the percent level, a global momentum fit is performed exploiting also the tracks reconstructed by the tracker while the muons were passing through it.

Both the drift tubes and the cathode strip chambers are suitable for triggering on the muon transverse momentum, providing good efficiency and a high background rejection even if used independently from the other subsystems of CMS. However, the very high background rates and the high pileup make the measurement of the beam crossing time crucial in order to assign the reconstructed muons to the correct interaction. For this reason, a dedicated trigger facility, consisting of resistive plate chambers, has been designed for both barrel and endcaps. The resistive plate chambers provide a fast response that is independent from the other gaseous detectors and hence helps the resolution of possible ambiguities when multiple hits are recorded in a single chamber. The resistive plate chambers are built in a redundant configuration for the innermost muon stations in order to allow for the trigger algorithm to work also in the case of low momentum muons that will likely stop before reaching the outer part of the detector.

2.4.6 *The trigger system*

Proton-proton collisions in the LHC happen every 25 ns corresponding to a bunch crossing rate of 40 MHz. With the instantaneous luminosity level of the accelerator during Run 2, the pileup level goes up to 60 collisions per bunch crossing. The resulting information coming from the tracks of all the produced particles is too large to be processed and stored for analysis purposes. Moreover, the vast majority of the

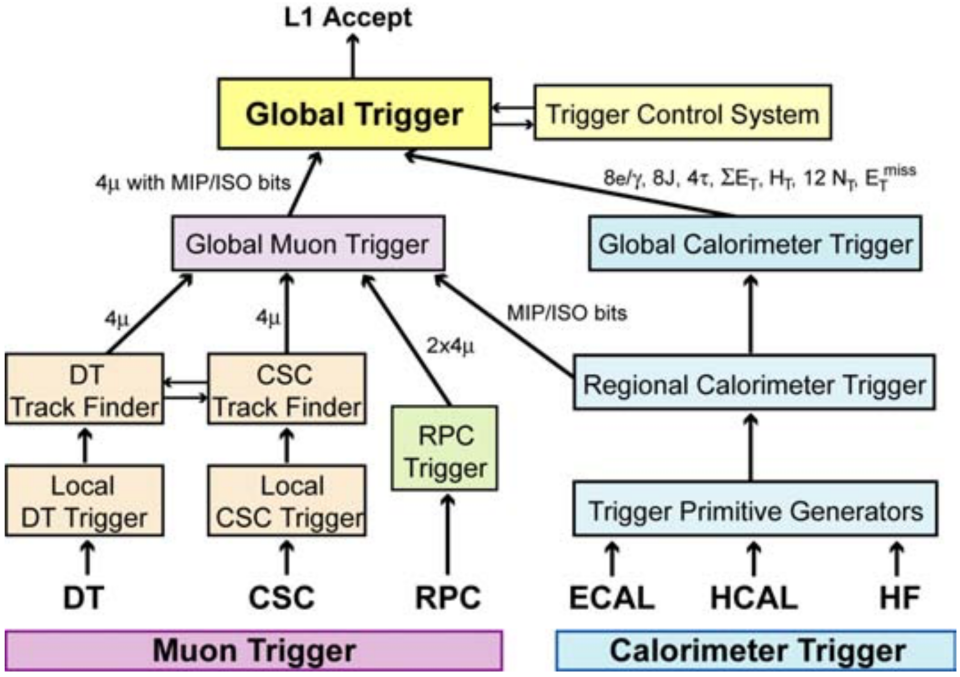


Figure 15: scheme of the architecture of the Level-1 Trigger.

collision events do not contain interesting information to be used in physics analysis. For these reasons, a trigger system has been implemented in order to reduce the amount of data to be saved and processed and to retain only the information of those collisions that appear to be interesting for physics analyses.

The CMS trigger is divided into two steps: the Level-1 (L1) trigger and the High-Level Trigger (HLT). The first consists of programmable electronics that must be very quick in the choice of which events to pass to the second that is based on software systems. The L1 trigger aims to reduce the output data rate to 100 kHz by exploiting segmented data from the calorimeters while holding the full resolution data in the front-end electronics memory. The HLT uses the full information coming from the read-out electronics and performs calculations and reconstructions that are close to those that are performed offline while doing physics analyses.

The L1 Trigger exploits information coming from local, regional and global components, as schematised in Figure 15. The local trigger uses information coming from energy deposits in the calorimeters or hit patterns from the muon chambers. The Regional Trigger is able to combine these information and reconstruct partially the signature of electrons or muons based on energy and momentum. The Global Calorimeter and the Global Muon trigger make use of the information coming from the entire detector and pass it to the Global Trigger that eventually accepts or rejects the event. To take this decision, the characteristics of particles and jets are analysed alongside total and missing transverse energies. If the event is accepted, it is then passed to the HLT for further evaluation.

When action grows unprofitable, gather information;
when information grows unprofitable, sleep.
— Ursula K. Le Guin

3 PARTICLE RECONSTRUCTION

The proton proton collisions at the LHC produce a shower of particles that interact with the different subsystems of the CMS detector. The first step towards a data analysis is to be able to identify and reconstruct ideally all the particles in the final state of the collision. The response given by the electronics after the interaction of a particle with the detector must be translated into objects that can be used in physics analysis. At CMS, the particle flow algorithm [90] has been developed to reconstruct isolated particles, jets and secondary kinetic quantities, such as the missing transverse energy. The information coming from the different detectors is combined in order to identify and reconstruct the properties of the detected particles.

Although having been developed in the 1990s for experiments at LEP [91], the particle flow algorithm had not been implemented in hadron colliders experiments before the LHC era. The fine spatial granularity of all the CMS detectors provides good separation between individual particles allowing for a global description of the event with high resolution and efficiency.

In this chapter, the reconstruction of the main objects needed for the $W\gamma\gamma$ analysis, namely electrons, muons, photons and jets, are discussed. The improvement made from the previous version of the algorithm and their effects on the overall events reconstruction are reviewed.

3.1 TRACK RECONSTRUCTION AND PARTICLE TRACKING

Particles are reconstructed from the information on their tracks in the tracker and in the muon chambers and from the energy deposits in the calorimeters. The first implementation of the track reconstruction algorithm was mainly oriented towards muon and particles with secondary vertexes, such as τ leptons and b quarks [91]. The main goal of the algorithm was to identify high energy particles and this was done in three stages exploiting a Kalman Filter [92] approach. A seed for the track was determined from few silicon tracker hits compatible with a track and then extrapolated to a trajectory by exploiting the information obtained from the tracking system in the direction of the track initiated by the seed. Finally, the full information was fitted in order to compute the characteristics of the charged particle, mainly momentum and collision vertex.

At CMS, the goodness of such a reconstruction algorithm can be tested by exploiting two variables: efficiency (defined as the fraction of reconstructed tracks that match at least 50% of the simulated hits in the tracker while matching less than 50% of unassociated simulated hits) and misreconstructed rate (defined as the fraction of reconstructed tracks that cannot be associated with simulated ones, hence tracks obtained from hits that do not come from a single real particle). By requiring two hits in successive pixel layers, at least eight hits in total and with at most one missing

hit in the tracker, a misreconstruction rate at the level of the percent is measured, at the expense of a low efficiency (70-80% for charged pions to be compared to 99% for muons). The main physical motivation behind this low level of efficiency is the fact that hadrons, especially low-energy ones, can undergo a nuclear interaction with the tracking material, spoiling the subsequent track reconstruction pattern. On the other side, high energy hadrons are usually part of a jet and it is impossible for the algorithm to uniquely assign a given hit to a single particle in the jet, hence further reducing the efficiency in reconstructing these particles. A hadron that is not reconstructed from the tracker but that leaves energy in the calorimeter is reconstructed as a neutral hadron by the calorimeters. The reconstruction performed by the calorimeters is in this way biassed: the direction of the hadron will be wrongly identified (if it were a true neutral hadron it would not have bent in the magnetic field) and the efficiency and energy resolutions will be degraded. For a typical event, the jet energy and the angular resolution would be degraded by about 50% from the non-reconstructed track in the tracker.

To increase the efficiency to reconstruct tracks in the tracker, the requirements on the number of hits and on the minimum energy of the particles can be reduced. However, relaxing these requirements increases exponentially the number of possible reconstructed tracks which in turn gives a probability to have a misrecontructed track that is an order of magnitude higher than before. Having a set of misreconstructed tracks that are randomly placed in the detector and that correspond to random energy particles spoils the reconstruction of the energy of the event.

The solution to these problems (get higher efficiencies while having a low misreconstruction rate) has been found in what is called *iterative tracking*. The basic idea is to start from the reconstruction of well recognised tracks and then, after having removed the hits associated to those tracks, reconstruct objects with selection criteria that are more and more relaxed at each step. The great advantage of this method is that, after each iteration, the number of remaining hits decreases, limiting the effect of the combinatorial misreconstruction. In order to reduce the rate of misreconstructed tracks, a set of quality criteria based on track seeds, fit χ^2, tracks pointing to the vertex and number of hits is implemented. The first three iteration of the algorithm search for tracks which have three pixel hits as a seed with additional requirement on the position of the reconstructed vertex; this allows to relax the requirement on the number of hits and on the energy. The next two iterations try to reconstruct tracks with one or two missing hits in the pixel detector. Then, very displaced tracks are looked for before trying to reconstruct the core of jets. The last steps are oriented to increase the muon-tracking reconstruction efficiency. As it can be seen from Figure 16, the iterative procedure identifies almost half of the tracks with transverse momentum greater than 1 GeV that were not reconstructed with the standard algorithm.

The original reconstruction of electrons was based on a seeding algorithm that used their energy deposits in ECAL (*clusters*). Starting from the position of the energy deposit in the calorimeter, the track of the electron was inferred in the tracker. An electron passing through the silicon of the tracker, however, is likely to interact with it and emit bremsstrahlung photons. To have a better reconstruction of the particle, what are called *superclusters* were exploited: regions of the electromagnetic

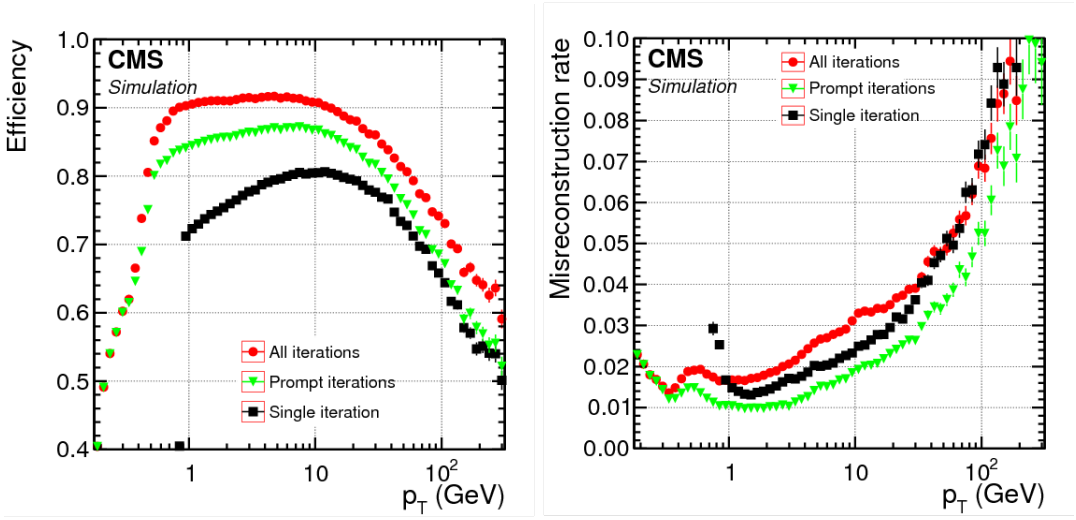


Figure 16: efficiency (left) and misreconstruction rate (right) obtained by using the single iteration algorithm (black) or the iterative one while using only tracks with at least one hit in the pixel detector (green) or while using all the iterations, including those without a seed (red) [90].

calorimeter narrow in η and wide in ϕ. This is a direct consequence of the fact that the solenoidal magnetic field bends the charged particles in the ϕ direction and so, if a bremsstrahlung photon is emitted, its energy is recorded in a calorimeter tower displaced in ϕ from the one of the electron. Given the high occupancy of the tracker for a high pileup collision and the fact that an electron could be part of a jet, the extrapolation of the track from the calorimeter deposit to the tracker leads to a substantial misreconstruction rate.

To overcome this issues, a tracker-based approach has been developed in the particle flow algorithm instead of the ECAL-based one. The fact that an electron radiates a photon can be exploited to distinguish it from a charged hadron. The algorithm proceeds in different ways as a function of the amount of energy emitted by the electron. If the electron radiates a small amount of energy, its trajectory is slightly modified and the χ^2 fitting procedure will give a good track reconstruction that can be directly matched with the calorimeter cluster. Since the radiated energy is small, the momentum obtained from the track reconstruction and the energy reconstruction in the calorimeter are required to give a compatible result. If there is a soft emission, there is still a possibility that the track is correctly reconstructed even though the fitting procedure will be spoiled. If the radiated energy is remarkable, the track reconstruction will likely fail and just a few hits will be associated with the particle trajectory. This issue can be overcome by exploiting a Gaussian-sum filter algorithm [93] that allows for a good description of sudden kinks in the particle trajectory, as those produced by a hard emission. Eventually, a boosted-decision-tree (BDT) classifier assigns a score by exploiting information from number of hits, displacement between the track extrapolation in ECAL and the closest energy cluster, difference between the standard Kalman filter fit and the GSF one, χ^2 of the GSF track and energy loss. By using this

improved tracked-based algorithm in combination with the ECAL-based one, the efficiency for electron reconstruction is almost doubled.

Differently from the other charged particles, muons do not entirely rely on the tracker to reconstruct the trajectory of the particle, but can exploit the hits in the muon detectors. The interplay between the muon chambers and the tracker leads to three different ways to reconstruct muons:

- Standalone muons: they are reconstructed exploiting only the information of muon detectors. Seeds of the track are identified by building segments in the drift tubes and in the cathode strip chambers and then exploited for pattern recognition in the muon spectrometer. After seeding, all the hits in the DT, CSC and RPC are exploited to build the standalone muon track.

- Global muons: the standalone track obtained from the muon chambers is matched to a reconstructed track in the tracker. The hits from both the tracker and the muon detectors are then combined and fitted altogether to get a global muon track.

- Tracker muons: they are reconstructed following an inverse procedure as the one used for global muons. A track in the tracker is extrapolated to the muon chambers to check if there is at least one matching segment.

For most of the cases, global muons give the better efficiency in track reconstruction. If the particle, however, has a momentum of $10\,GeV$ or less, tracker muons have a higher efficiency since the contribution of muon chambers is reduced due to possible multi-scattering events that occur to the muon in the steel return yoke.

3.2 CALORIMETER CLUSTERS

Since neutral particles do not interact with the tracker, the only way to measure their energy is to exploit the information coming from the calorimeters. In principle, if an energy deposit does not match a track, it means that it comes from a neutral particle. The neutral particle reconstruction is not so straightforward if the neutral deposit is close to, or overlaps with, one or more charged-particle deposits. In this situation, the only way to obtain the energy of the neutral particle is to rely on the information of the track to infer the energy of the charged particle and get the one of the neutral particle by subtraction from the calorimeter response. To be able to exploit this procedure, a precise calibration of the calorimeters is fundamental in order to get a good estimate of the energy of the neutral particle.

As previously described, the cluster algorithm of the electromagnetic calorimeter tests the energy detected by the towers with a series of different thresholds (one for the highest energy tower, another for the closest neighbours and so on). The energy measured by the calorimeter for an incoming photon is smaller than the actual energy of the particle, since some of the interactions can give an energy value in a tower that is below the threshold for that specific tower, effectively losing that energy contribution. To get a calibration term that takes into account this effect, a simulation of single-photons events can be exploited. The generated particles are processed

through the GEANT4 simulation [94] of the CMS detector. As expected, the correction is close to unity for high energy photons (where the threshold effect is negligible), while accounts for almost 20% for low energy photons. To check the goodness of the results obtained from simulations, events with a neutral pion decaying into a pair of photons have been studied, giving an agreement on the reconstructed pion mass with respect to the real value at the percent level.

A neutral hadron deposits its energy partly in the electromagnetic calorimeter and the remaining part in the hadronic calorimeter. Although the electromagnetic calorimeter has already undergone a procedure of calibration, a further recalibration must be implemented since the response to hadrons of the ECAL is different than the one for photons and, furthermore, the calorimeter response depends on the fraction of energy released by the hadron in it. An iterative intercalibration procedure based on simulated events that takes into account the threshold effects in ECAL and the relative loss of energy in one or the other calorimeter is implemented to get a more precise calibration coefficient. This procedure allows to restore the linearity of the calorimeter response even for the low energy regions where, if not applied, the difference between true and reconstructed energy could be of the order of 40%.

3.3 PARTICLE RECONSTRUCTION

The passage of a particle through the experiment is generally recorded by different subdetectors. The first step towards a reconstruction that exploits all the available information is then to try and find which of these different outputs are linked to the same particle. The difficulty of this procedure comes from the high occupancy in terms of number of particles per solid angle and from the limited granularity of the different subdetectors. Moreover, effects such as trajectory kinks and the production of secondary particles further increase the probability of assigning a detector output to the wrong particle.

A "link algorithm" has been developed and implemented in order to couple pair of elements coming from different detectors. Elements such as tracks or calorimeter clusters are paired by looking at possible counterparts in a region with small angular distance in the (η, ϕ) plane in order to reduce the computation time. Two compatible elements are then classified by defining a distance between them.

The first link that the algorithm seeks is the one between a track from the tracker and a cluster in a calorimeter. To do so, the track is extrapolated into the calorimeters. For ECAL, its position is evaluated at a depth corresponding to the maximum of a typical electromagnetic shower, while for HCAL it is extrapolated at a depth corresponding to one interaction length. To take into account the presence of gaps and cracks between different instrumented modules, the track position is widened to the size of one cell in each direction. If the extrapolated position is found to be inside a cluster, the track is matched to it. A distance is computed between the extrapolated track position and the centre of the cluster in the (η, ϕ) plane. If more than one track is found that points towards a cluster, the one with the smallest distance is associated to it.

Given the high probability for an electron to emit a bremsstrahlung photon, a tangent to the electron trajectory is extrapolated to the electromagnetic calorimeter for each layer of the tracker. If the position extrapolated from the tangent coincides with a cluster in the calorimeter, the cluster is associated to the original track as a potential bremsstrahlung photon. Besides bremsstrahlung emission, also pair production can happen in the tracker detector material. For this reason, a dedicated algorithm that matches two tracks originating from the same spot in the tracker is implemented.

Besides the track matching to the calorimeter clusters, also the energy deposits in the two calorimeters are connected one to the other. By starting from the ECAL, which has a higher granularity and is then more precise in the determination of the position, an energy cluster that matches its position is searched for in the HCAL. As for the track matching, if more than one HCAL cluster is found, the one with the smallest distance is selected.

Further links are built for secondary vertexes, where at least three tracks are associated to a displaced vertex and at most one of them can be an incoming track (if the decayed particle was a charged one). As already described, tracks from the tracker are associated to tracks in the muon chambers to build global and tracker muon objects.

The reconstruction algorithm follows a step-by-step reconstruction. First, muons are reconstructed and the information on tracks and clusters associated to them is removed from the particle flow algorithm. Then electrons and isolated photons are reconstructed and, as for muons, their associated hits and clusters are discarded. Hadrons follow in the reconstruction with further attention to possible nuclear interactions. Finally, when all the other elements are reconstructed, global quantities, such as the missing transverse energy, are computed. This last step allows to have a double check on the previous reconstruction and is fundamental to spot possible misreconstructed particles and correct them.

3.3.1 *Muons*

The reconstruction of isolated muons is the most straightforward. Once a global muon candidate is found, the energy recorded by the calorimeters and the transverse momentum of the tracks inside a cone of $\Delta R = 0.3$ are checked. If the transverse energy and the sum of the transverse momentum of the tracks is smaller than 10% of those of the muons, the particle is accepted as an isolated muon. This simple procedure allows the rejection of the majority of muons inside jets.

The most probable misreconstruction effect is that of a muon identified as a hadron and viceversa. A charged hadron whose shower is not fully contained in the hadronic calorimeter will produce some hits in the muon system (punch through effect) and can be reconstructed as muon. Since a small part of the shower leaks in the muon system, the algorithm reconstructs the muons and assigns the rest of the energy in HCAL to spurious neutral particles. On the other hand, if there is no matching between the calorimeter clusters and the track in the muon system, a muon is reconstructed as a charged hadron. To reduce these two effects, the reconstruction of a muon that is not isolated is more stringent than that for the isolated ones. A tight muon selection [95] is applied and in addition three segments from the muon chambers or the

tracker track and the calorimeter deposits must satisfy the muon hypothesis. If the muon fails the tight selection, the possible particle is recovered if the standalone track matches a large number of hits in the muon system (at least 23 hits in the drift tubes or 15 hits in the cathode strip chambers).

Given the fact that the track is measured by two separated subdetectors, some different choices can be made to decide which information has to be used to determine the particle transverse momentum. If the p_T is smaller than 200 GeV, the track from the tracker is exploited to determine the particle transverse momentum, since for these values of energies the bending performed by the magnetic field allows a good determination of the curvature radius. For higher energies, the particle momentum is determined from the track fit (using only the tracker or both the tracker and the muon chambers) with the lowest χ^2. The elements exploited for this particle identification are then removed from the successive steps of the reconstruction. A further refinement of the muon selection is performed while building global quantities, see Section 3.3.4.

3.3.2 *Electrons and photons*

The interaction of an electron or a photon with the tracker and calorimeter material is very similar: electrons emit bremsstrahlung photons that convert into an electron-positron pair that can further emit photons and so on. For this reason, the reconstruction of electrons and photons is done at the same step of the algorithm. An electron is identified if an ECAL cluster matches a GSF track but at most two tracks (if many tracks point to the same cluster it is likely to be a jet and not an isolated electron). A photon is reconstructed if a deposit in the electromagnetic calorimeter with E_T greater than 10 GeV is found to not match a GSF track from the tracker. For both electrons and photons, the energy deposit in the hadronic calorimeter in a radius of $\Delta R = 0.15$ around the position of the electromagnetic cluster must be less than 10% of the ECAL energy measurement. This requirement reduces the possibility of misidentifying a jet for an electron or a photon.

The energy of the photon is identified with the energy of the ECAL cluster after the application of corrections due to possible missed energy, as previously described. The direction of the photon is taken to be the one of the calorimeter cluster. The energy of the electron is obtained by combining the information from the GSF track and the calorimeter deposit, while the direction is chosen to be the one of the tracker track.

Additional requirements are applied to both the selections. The collection of the isolated electrons is further refined with a boosted decision tree algorithm that exploits additional information, such as the ratio between the ECAL and HCAL energy, goodness of fit of the track and number of hits in the tracker. Photons must be isolated from other tracks and must have a small ratio of energy released in the hadronic calorimeter over the one released in the electromagnetic calorimeter.

As for muons, the elements used for the electron and photon reconstruction are masked for the remaining steps of the algorithm. Since the information exploited by the particle flow algorithm to reconstruct electrons and photons could be not ideal for

some specific analyses, the full information on both the particle candidates is saved for further study.

3.3.3 *Hadrons*

Once muons, electrons and photons are reconstructed, what is left are likely neutral and charged hadrons both from jet fragmentation and quark hadronisation and non-isolated photons. As previously discussed, photons leave the majority of their energy in the electromagnetic calorimeter, while hadron leave a small amount of their energy in ECAL and the majority in HCAL. This evidence would lead to the idea that, when a matching with a track is not found, the ECAL deposits should be associated with a photon while the HCAL ones to a neutral hadron. However, the total amount of energy released by both neutral and charged hadrons in ECAL is around 25% of the jet energy, which means that, if the track association fails and charged hadrons are identified as neutral, the energy deposit that actually comes from a hadronic jet would wrongly be associated to a non-isolated photon. The algorithm has been refined so that ECAL clusters (without an associated track) that have no matching HCAL cluster are reconstructed as photons, otherwise the algorithm identifies the particle as a neutral hadron.

HCAL clusters with one or more associated tracker tracks are reconstructed as charged hadrons. The energy estimate to perform the calibration is taken to be the larger between the sum of the track momenta and the sum of the raw calorimeter deposits. If the energy obtained from the calorimeters is larger than the one obtained from the tracks, this is a hint of the presence of one or more neutral particles in the jet. In the end, each track is reconstructed as a hadron with an energy corresponding to the transverse momentum obtained from the track while assuming its mass to be the one of the charged pion. If no neutral particles are identified in the cluster (i.e. the energy reconstructed from the tracks and the one obtained from the calorimeters are compatible), both track and calorimeter information are exploited in order to have a smooth description for the whole energy range (at low energy the track information is more important, while at high energy the calorimeter clusters dominate).

If the energy of the calorimeters is much smaller that the one obtained from the track information, this is a hint of the presence of a muon that has not been reconstructed in the previous dedicated steps. If this is the situation, a new search for muons is performed with relaxed requirements in order to find some more muons in the event while reducing the possibility of misidentification.

3.3.4 *Missing transverse energy*

Given their incredibly small interaction cross section, neutrinos produced in the collisions cannot be detected by the CMS experiment. Many processes of interest, among which the one studied in this analysis, have however a neutrino in their final state. By exploiting the laws of conservation, it is possible to obtain some information on these invisible particles. As described in Section 2.1, the total energy of the colli-

sion is not available in proton-proton accelerators (differently from what happens in electron-positron machines), but the total transverse energy and momentum before the collision are compatible to zero and so must be after. The presence of a neutrino can then be identified by the presence of a missing transverse energy (MET) in the final state that would be necessary to restore the zero-value balance. Having reconstructed all the particles in the event by means of the particle flow algorithm, it is possible to define the missing transverse energy as

$$\vec{E}_T^{\text{miss}} \equiv -\sum \vec{p}_T \,, \tag{57}$$

where the sum is performed over all the reconstructed particles.

Even though the particle flow algorithm reconstructs the particles in the event by exploiting many different combined information and is optimised for the identification of a lot of different final states configurations, it can happen that one or more particles are misreconstructed and this could give an artificially inflated value of the MET. This issue is important in precision measurements, but is critical when performing a search for BSM particles that typically do not interact with the detector material. In order to find and fix these issues, a post processing algorithm is implemented.

The most likely case is to have reconstructed a high energy particle that never existed that in turn gives a high value for the missing transverse momentum. The post processing algorithm looks for these artificially high momentum particles, checks if the direction is correlated to the one of the MET (if this is the particle that gives the majority of the missing transverse momentum, then its direction is likely to be opposite to the MET one) and retrospectively reconstructs the particle.

The most probable event that causes the reconstruction of an unnaturally high MET is the presence of a cosmic muon that traverses the detector in time with a LHC collision. Even though the CMS experiment is located 100 m underground, the flux of cosmic muons is not zero and in-time interactions can happen. To avoid the cosmic muon to be reconstructed as part of the event, the hits and energy deposits associated to it are removed if the reconstructed interaction point is more than 1 cm away from the actual interaction point of the hard collision. This requirement reduces the contribution of cosmic muons, but potentially removes from the event muons coming from decay of b quarks. In this configuration, however, the removal of the muon contribution would increase the value of the MET (while what the algorithm is trying to do is to reduce it by removing fake particles) and it is then possible to retain this true contribution in the event.

Another possible effect that enlarges the value of the MET is the presence of a charged hadron that releases part of its energy in the muon chambers. The algorithm in this situation would first reconstruct a muon (from tracker, calorimeter and muon system information) and then a charged hadron (since the track information has been masked by the algorithm). The result is then that two particles are reconstructed while only one existed. The post processing algorithm is able to identify such situations and perform again the reconstruction of the hadron while removing the muon, hence identifying only the correct particle.

The opposite process can also happen. A muon is not reconstructed as a muon because it fails some of the requirements and is not recovered in the hadron recon-

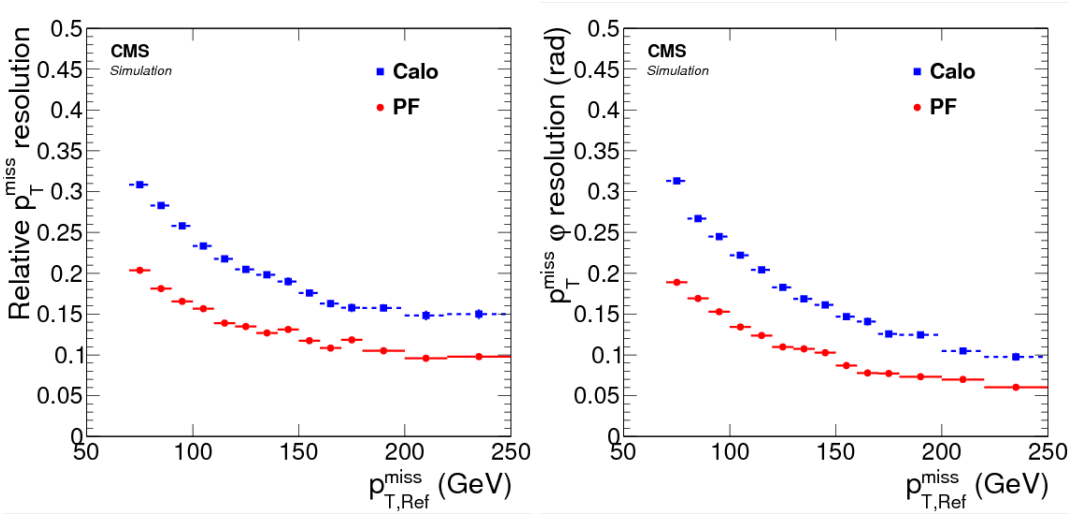


Figure 17: relative resolution on the missing transverse momentum (left) and on the determination of its direction (right) when using the calorimeter quantities (blue) or the particle flow fully reconstructed objects (red).

struction process because it overlaps with a neutral hadron. What is left after the muon reconstruction step is then a track and a deposit in the hadronic calorimeter that the algorithm reconstructs as a charged hadron instead of a neutral hadron and a muon. The post processing algorithm identifies this particle from the information on the missing transverse energy and tries to disentangle the contribution of the two particles from the reconstructed one in order to reduce the high value of the MET.

Before the implementation of the particle flow algorithm, calorimetric information was exploited in the reconstruction of the jets and in turn to compute the MET. The sum over the momenta of all the particles was performed in terms of deposits in the calorimeter cells and the momenta of jets was corrected by a coefficient that took into account effects of the calorimeter response. Figure 17 shows the advantages of using the full particle flow algorithm instead of calorimeter quantities: the relative resolution both in the missing transverse momentum and in the determination of its direction are greatly improved with the new approach.

The generation of random numbers is too important to be left to chance.
—Robert R. Coveyou

4 EVENT GENERATORS

The simulation of collision events with Monte Carlo methods is a fundamental tool in high energy physics. By implementing theoretical computations and by matching them to the detector response, it is possible to build precise simulated samples of events that can then be compared with the experimental data. In particular, Monte Carlo simulations allow to identify the contribution of known processes while searching for new physics or allow to test which theoretical prediction is the one that matches the experimental result. The discovery of the Higgs boson shows well the implementation of both these concepts. Known background processes, such as ZZ in four leptons, have been simulated and taken into account in the distributions. Many different signal samples, with variable values for the mass of the Higgs boson, have been generated and compared to the data to infer which was the actual mass of the boson.

Besides physics analysis, Monte Carlo simulations are fundamental tools for detector studies. As seen in the previous chapter, many efficiency studies can be performed for the detector performances in order to compute correction factors that allow a better event description. Moreover, simulation are fundamental tools in the development of future detectors and accelerators. It is, in fact, possible to simulate the design of a future machine and see what would be its output to physics collision in order to optimise its layout.

As far as analyses are concerned, the simulation of an event proceeds in subsequent steps. The first step of the simulation is the generation of a hard scattering process, which is an event having a large momentum transfer between the incoming partons. This is, for example, associated to the production of a heavy particle. Once the effect of the parton distribution functions is taken into account, the generation of the diagrams of the process is pretty straightforward. The process can be simulated at different orders of precision: leading order (LO) if only tree-level diagrams are considered, next-to-leading order (NLO) if also one loop processes and real emission diagrams are calculated, next-to-next-to-leading order (NNLO) if two loops processes are taken into account and so on. At the present time, the NLO computation is available for the vast majority of the simulated processes, while NNLO corrections have been computed for many processes of interest. Some processes, for which the cross section must be known with a very high precision, have a next-to-next-to-next-to-leading order (N^3LO) calculation.

The following step of the simulation is the parton shower. As it is known from the QCD theory, coloured particles can emit gluons. This is exactly what happens in a typical LHC collision both with the incoming quarks (initial state radiation) and with the produced particles (final state radiation). Gluons, however, carry themselves colour charge and can in turn emit other gluons giving rise to a particle shower.

The parton shower starts from the hard scattering particles and proceeds with the emission of particles with increasingly lower momentum.

The emission of very low energy partons, however, is not described by perturbative QCD. For quarks and gluons in these low momentum configuration, particle confinement takes the advantage over fragmentation and hadrons are produced. In order to simulate this process, hadronisation models are implemented. The simulation of the decay of the particles is also implemented in this step.

It must be noted that two of the partons from the two incoming protons interact to give the hard scattering process. The other quarks and gluons from the same protons can interact with themselves, giving rise to the production of soft hadrons. These processes are simulated alongside the main scattering interaction and are known as underlying events.

These steps are implemented, even if with some differences, by many of the present time general purpose event simulators. In the following, a detailed description of the MadGraph5_aMC@NLO event generator and the Pythia particle shower is given. After being produced in the collision and (possibly) decay, the particles reach the detector and interact with it. The simulation of these interactions is performed with Geant4 and will be briefly described in the following.

4.1 THE MADGRAPH5_AMC@NLO GENERATOR

The computation of tree level and especially higher order amplitudes and cross sections has been, and to some extent still is, a complex problem. The complexity of the computations and the theoretical challenge of taking into account and cancelling possible singularities required a great effort from the theoretical community. Up to some years ago, the different cases had generally ad-hoc solutions and were implemented one by one. This method, however, became a limiting factor as soon as the LHC started its operations: the collected and analysed data had such a high precision that NLO calculations became fundamental for all the signal and background processes. Moreover, searches for new physics with indirect techniques require a high precision theoretical modelling.

A general solution allowing to compute the full expansion in α_S is not available. What has been achieved is a general recipe to compute NLO corrections for all possible processes, which means that the computation of the amplitudes for process up to the NLO precision can be performed algorithmically by using a software. This fundamental step in the improvement of Monte Carlo simulations has been achieved by exploiting a universal formalism for the cancellation of infrared singularities [96–100] and an algorithmic approach to the computation of the renormalised one-loop amplitudes [101–111]. Moreover, NLO computations can be matched to different parton shower event generators, thus providing all the needed information for the event simulation.

MadGraph5_aMC@NLO [112, 113] is a general purpose matrix-element based event generator. It is able to compute any number of observables for any LO or NLO processes that can also be matched to parton showers.

The basic idea is that the structure of the cross section is independent of the theory and the perturbative orders that are taken into account. It can thus be written in an algorithmic way (i.e. with some code instructions) for any desired process. By giving as input the masses and the number of the particles, the phase space can be defined. The infrared subtractions, that were the main problem for the computation of NLO amplitudes, can be reduced to a small number of cases that must be computed and their result can be used for all the different situations. Of course, the matrix elements depend on the theory (i.e. the Lagrangian of the model) and from the process that is being generated. However, the full information can be passed to the code by providing the Feynman rules for the process that can be directly derived from the model Lagrangian. With these information, MadGraph5_aMC@NLO is able to write the matrix elements by computing the Feynman diagrams and from those compute the amplitudes.

While computing the matrix elements for the process, only tree-level techniques are implemented. This is possible because loop diagrams can be "cut" with the L-cutting procedure [114], which means that one-loop diagrams are reduced to tree level ones by cutting the propagator line. Feynman diagrams are generated from Feynman rules that are passed to the generator in the Universal FeynRules Output (UFO) [115] format. FeynRules is a package that is able to obtain the Feynman rules having as input the fields, parameters and Lagrangian of the physical model.

When computing Feynman diagrams for processes with many particles, the problem of the rapid increase of the number of possible diagrams arises. The solution implemented in MadGraph5_aMC@NLO is to compute helicity amplitudes with fixed polarisation for the external particles and then implement colour decomposition. By doing so, the number of needed computations grows linearly with the the number of diagrams instead of quadratically. Moreover, diagrams are categorised based on their substructures: if a substructure is common to many of them, it is computed only once, significantly decreasing the computation time.

Once all the diagrams are generated, the process amplitudes can be computed by the ALOHA package [116]. After the combination of the external wave functions and the determination of the currents, the amplitudes of different diagrams are added and squared by applying the previously computed colour matrix. The final result for the process amplitude is eventually obtained.

Even by implementing all the described optimisations, the process of events generation is nevertheless very time consuming. Samples of million of events, as those generally used for analyses at LHC, can take days or even weeks to be produced. The situation is worsened if one thinks that, usually, the effect of theoretical or modelling uncertainty has to be studied and hence simulated. The most efficient way to store the information on scale and modelling variations is to generate a single set of events in the Les Houches format (LHE) [117, 118] and then reweight them on an event-by-event basis. In this way, the simulation time is reduced while all the information is generated and saved.

A certain process, for example the production of a Z boson, can be characterised at matrix level by the presence of some partons that result in jets associated to the vector boson in the final state. For the way in which the simulation is implemented,

however, some partons (and hence some jets) can be included also during the showering process. The so-called merging procedure has been developed in order to avoid double counting of partons included in these two separate steps. With this method, it is possible to merge the generated events and obtain a truly inclusive sample. For LO processes, the scheme provided by Michelangelo Mangano (MLM) [119] exploits a k_T-jet clustering algorithm in order to reject showered events with specific conditions. At NLO, the situation is much more complicated and has been solved by applying the FXFX approach [120] developed by Rikkert Frederix and Stefano Frixione.

MadGraph5_aMC@NLO implements many other features, such as the computation of particles spin correlation treated with MadSpin, that are not discussed here since they would go beyond the scope of this brief introduction. For a detailed description of all the possibilities and submodules of MadGraph5_aMC@NLO, one can refer to Reference [113].

4.1.1 *Events in presence of anomalous couplings*

In order to produce a sample with MadGraph5_aMC@NLO, a series of basic ingredients is needed. In particular, they are:

- UFO model: it translates the Lagrangian into Feynman rules to be passed to the generator. There are many different models that can be implemented, both SM ones (with different settings, such as the number of massive quarks) and BSM ones, that implement different possible extensions of the SM.
- Process card: defines the process that needs to be generated, which quarks are used to make jets, the number of jets in the final state and the order of the loop computation for each subprocess. The typical configuration is the one to have NLO computations for low jet multiplicity processes, while using LO for many jet subprocesses.
- Run card: defines the kinematic cuts on the particles, the matching and merging schemes, the PDF set to be used for the event generation and other sets used to reweight the events, the parameters of the reconstruction algorithm (such as the dimension of the jet clustering cone).

As explained in Section 5.1.2, CMS centrally produced simulations have been used in this analysis for the signal and background events. However, to test the presence of anomalous couplings, appropriate samples must be produced and compared to the SM ones. Since the production and the settings for these samples are strongly analysis dependent, the anomalous couplings samples have been produced as part of this work.

The main characteristic of the anomalous coupling samples is that the underlying model is not anymore the SM, but its EFT extension, as described in Section 1.6.1. Starting from the configuration for the centrally produced MadGraph5_aMC@NLO $W\gamma\gamma$ sample, the UFO model has been set to the AQGC_ALL_UFO_new model [79, 121–124]. The information on the process and run card have been left as it is in order to cover the same phase space as the standard signal process.

Parameter	Range of scan TeV^{-12}
$f_{M,0}/\Lambda^4$	-150, 150
$f_{M,1}/\Lambda^4$	-40, 40
$f_{M,2}/\Lambda^4$	-100, 100
$f_{M,3}/\Lambda^4$	-200, 200
$f_{M,4}/\Lambda^4$	-200, 200
$f_{M,5}/\Lambda^4$	-200, 200
$f_{M,6}/\Lambda^4$	-250, 250
$f_{M,7}/\Lambda^4$	-200, 200
$f_{T,0}/\Lambda^4$	-20, 20
$f_{T,1}/\Lambda^4$	-20, 20
$f_{T,2}/\Lambda^4$	-40, 40
$f_{T,5}/\Lambda^4$	-40, 40
$f_{T,6}/\Lambda^4$	-40, 40
$f_{T,7}/\Lambda^4$	-40, 40

Table 2: summary of the normalised anomalous couplings coefficients and the range of variation for their scan.

As described in Section 1.6.1, the $W\gamma\gamma$ process is sensitive to a certain number of dimension-8 operators. A series of event-by-event weights has been implemented in the MadGraph5_aMC@NLO configuration in order to generate events that have the strength of these operators different from zero. In this way, by exploiting the reweighting procedure, the physics effect of the presence of any of the operators can be tested. The value of the normalised Wilson coefficients has been varied in the range reported in Table 2. The values of the range for the scan of these parameters have been chosen based on results from previous measurements.

4.2 THE PYTHIA GENERATOR

The matrix element generation produces a limited number of particles directly resulting from the hard scattering process. These particles, however, do not describe a physical final state for the collision, which is generally characterised by a complex multi-particle nature. In order to obtain a final state that matches the one produced in collisions, the matrix element calculations must be enriched with initial state radiation (ISR), final state radiation (FSR), multiparton interactions, fragmentation and hadronisation. A rigorous theoretical description is not available for all these processes and hence some phenomenological models tuned on experimental data must be implemented. All these features are implemented in the Pythia framework [125].

PYTHIA 8 is the most recent version of this framework and is the one used for all the samples exploited in this analysis. The matrix element processes generated with MADGRAPH5_aMC@NLO produce events in the LHE format; these events are then directly passed to PYTHIA for the hadronisation. Despite being used mainly to hadronise matrix level events obtained with other Monte Carlo generators, PYTHIA can be used standalone also to generate matrix level events and then shower them. As described in Section 5.1.2, some of the samples used in this analysis are generated in this way.

One of the main features of PYTHIA is the possibility to add initial and final state radiation to the hard process. ISR and FSR are contributions of the emission of elementary particles from the incoming and outgoing partons. As explained in Section 1.5.2, the evolution of a shower can be described in terms of the DGLAP splitting kernels $P(z)$ both for the emission of quarks and gluons and for the emission of photons or fermion-antifermion pairs. The emission of radiation is described at LO in PYTHIA by solving the differential equations based on the splitting kernels as a function of the transverse momentum of the emitted particle. The transverse momentum dependency is described by using a shower evolution variable that is the analogue of the factorisation scale for a parton shower. The treatment of the two steps, although similar, has some differences. For FSR the decreasing p_T evolution follows the physical time: a parton is branched into two other particles at each step. For ISR, on the contrary, the evolution goes backwards in time: the hard scattering parton is seen as the daughter of a mother particle that previously split. This backward evolution means that the information on the PDF of the particle in exam must be translated to the one of its mother. For both ISR and FSR, the branching strength is set by the value of the strong coupling constant $\alpha_S(M_Z)$ that in principle can be different for the two situations (of course, there is only one value of α_S in nature, but the difference for the two splitting kernel can be parametrised with two different values of α_S).

The addition of the ISR and FSR particles introduces non-physical coloured states. The mechanism implemented to transform these particles into colourless particles is called hadronisation and in PYTHIA is based on the Lund string fragmentation framework [126, 127]. The basis of the model is the idea that a pair of coloured particles are subject to a linear confinement which means that the field strength between the two particles increase linearly with distance. This behaviour is translated into a string model. The exemplification is that of a flux tube connecting a $q\overline{q}$ pair with a string tension (the amount of energy per unit length) of the order of $k \approx 1\,\mathrm{GeV/fm}$. As the distance between the two quarks increases, so does the potential energy of the string up to the point that it is more energetically convenient to break the string into a $q'\overline{q}'$ pair. In this way, two colour singlets $q\overline{q}'$ and $q'\overline{q}$ are formed. If the invariant mass of one or both of these colour singlets is large enough, the splitting and hadronisation process further occurs until the only left states are on-shell particles. Since the various splittings are in principle independent, the process can be described in both directions (from colour charged partons to hadrons or vice versa). Starting from the hadrons and going backwards has the clear advantage that the value of the masses of the particles can be fixed to their physical values.

Hard scattering events involve two partons coming from two protons (in the case of proton-proton collisions), however the other partons from the same protons can also collide with each other. In most of the cases, these are soft collisions and are referred to as underlying events. Even though suppressed in rate, it can happen that more than one hard scattering collision happens for the same proton-proton interaction. PYTHIA is able to simulate both these contributions following the procedure described in Reference [128]. The presence of other scattering processes for the same proton-proton collision has a direct effect on the hadronisation procedure. In fact, coloured particles produced in soft collisions are correlated with the hard scattering one and colour reconnection can move around the colour charge. A colour bookkeeping system is implemented in PYTHIA in order to take into account this effect.

In order to implement the different theoretical and phenomenological models that are needed to describe the process, some physical parameters must be given as input to the PYTHIA simulation (an example already discussed is the value of the strong coupling constant). The set of input parameters, obtained from comparisons with data, is called a tune. The goal of the tune is to attempt to deliver a global description of the data. CMS has developed a set of tunes based on its physics data, the most recent one and the one implemented in the majority of the samples used in this analysis is the CP5 tune [129]. The CP5 tune uses the NNPDF3.1 PDF set [75], which is a NNLO set, with the value of the strong coupling constant set to $\alpha_S(M_Z) = 0.118$ and orders the ISR radiations in terms of rapidity (instead of transverse momentum). Besides the aforementioned quantities, the tune constrains also the threshold for multi-parton interactions, the overlap distributions of the two incoming protons (described with a double gaussian function) and the range of the colour reconnection probability, which is connected to the multiplicity of final state particles.

The procedure described in this chapter aims to give a general picture of the simulation procedure, but of course is not a detailed description of all the event generators possibilities. While the focus of the text has been placed on MADGRAPH5_aMC@NLO and PYTHIA, many other possibilities to generate high energy physics collision events are available. For this work, besides the already cited generators, also samples obtained with Sherpa [130, 131] and POWHEG [132–134] have been used. Detailed information can be found in the referenced papers.

4.3 THE GEANT4 SIMULATION

The matrix level event simulated with MADGRAPH5_aMC@NLO and then showered with PYTHIA contains the full description of the physics collision and of the final states. It is a comprehensive and detailed simulation of the hard scattering collision and the underlying events and is suitable to be used for analyses. Of course, the output of this procedure is a description of the process in vacuum. What is done, therefore, is to simulate the interaction of the final state particles with the detector. This is done with a GEANT4 simulation of the CMS experiment [135], that provides a detailed description of the electromagnetic and hadronic interactions with the de-

tector, tools for building the detector geometry and to simulate the response of the different sensitive elements while keeping track of the history of the particles.

The output of the event generators are the four vectors of the different particles that are traced by GEANT4. The geometry of the different subdetectors and a detailed description of the material content (both active and dead) is provided by an XML-based detector description. This choice has the advantage that the description of the detector response can be configured at different run times and with different conditions.

Being interfaced directly with the matrix-element and hadronisation Monte Carlo simulation, GEANT4 is able to keep the history of the particles both at generator level and after the decays and interactions with the detector. The Monte Carlo generation is used also to simulate the contribution of pileup events which is chosen based on the luminosity profile and by considering the intrinsically statistical nature of the process. Besides the pileup linked to the hard interaction under examination (in-time pileup), also the pileup contribution from previous and successive bunch crossings (out-of-time pileup) is taken into account.

Each of the subdetectors has different requirements in terms of simulation. The tracker needs a high degree of precision in the description of the active and passive material and a very precise treatment of electron bremsstrahlung and δ-ray emission. The electromagnetic calorimeter simulation is based on the accurate representation of the geometry of the system (not only of the single crystals, but also of supermodules, support structure and cooling system) and on the description of the electromagnetic processes. The key features for the hadron calorimeter are connected to a faithful description of the noise and the energy response as a function of the energy of the jets themselves. Eventually, the goodness of the simulation of the muon detectors is driven by the description of the muon bremsstrahlung, nuclear effects and multiple scatterings, besides the requirements on the modules geometry and alignment.

The goodness of the description is tested by comparing the prediction to test beam and collision events. For the description of the muon system, also cosmic ray interactions are exploited. In general, taking into account the precise description of the detector and the possibility to change the details of the simulation based on the different data taking conditions, the effect of the interaction of the particles with the detector is well reproduced.

Software and cathedrals are much the same - first we build them, then we pray.
— Samuel Redwine

5 $W\gamma\gamma$ CROSS SECTION MEASUREMENT

In this chapter, the analysis of the W$\gamma\gamma$ process is described in detail. The importance and the possible breakthroughs of the study of this interaction have already been introduced in Section 1.6.2. After the description of the datasets collected by the CMS experiment at 13 TeV and of the Monte Carlo simulation exploited for the analysis, a detailed account of the event selection is performed. The implementation of residual corrections between data and simulation is then reported. The study of the background, with particular focus to the ones estimated with ad-hoc data driven techniques follows. Before the extraction of the cross section value, the treatment of the possible sources of systematic uncertainty is described in detail.

5.1 DATASETS AND SIMULATIONS

5.1.1 *Data*

The data analysed in this work were collected by the CMS experiment in the period from 2016 to 2018, known as *Run 2*. In the following, the reference to "Run 2 data" is used to indicate the full dataset. The analysis has been performed in the electronic and muonic decay channel of the W boson and hence the *SingleElectron* and *SingleMuon* datasets have been used. These datasets contain mostly events with a single high energetic electron or muon selected online by the trigger system. The luminosity for the collected data is equal to 35.9 (35.9) fb^{-1} in 2016, 41.5 (41.5) fb^{-1} in 2017 and 59.4 (59.7) fb^{-1} in 2018 for the single electron (muon) datasets. The small difference in the luminosity between the two sets is connected to some periods of data taking (called *runs*) where some of the subdetectors were switched off and the particle identification and reconstruction could not be efficiently performed. The luminosity for the Run 2 dataset amounts to 136.8 fb^{-1} for the electron dataset and to 137.2 fb^{-1} for the muon one. The detailed summary of the datasets analysed is presented in Table 3. The samples of events collected during each year are divided in periods, called *eras*, that correspond to possibly slightly different data-taking conditions.

As described in Section 2.4.6, the events to be collected are selected based on a trigger system that is able to reject uninteresting background events from the hard scattering signal ones. Since in CMS many different analyses are performed exploiting the same collected data, a set of different trigger paths is implemented in order to record a large variety of events that can be easily categorised in the analysis workflow. The presence of a W boson is characterised, in the leptonic decay channel, by a high energetic electron or muon. For this reason, High Level Trigger single lepton paths have been used in this work to retain only those events where the trigger is fired explicitly by a highly energetic electron or muon. As already pointed out, the main objective of the trigger system is to reduce the number of processed collisions

2016		2017		2018	
Dataset	fb^{-1}	Dataset	fb^{-1}	Dataset	fb^{-1}
SingleElectron-RunB	5.8	SingleElectron-RunB	4.8	EGamma-RunA	13.7
SingleElectron-RunC	2.6	SingleElectron-RunC	9.6	EGamma-RunB	7.1
SingleElectron-RunD	4.2	SingleElectron-RunD	4.2	EGamma-RunC	6.9
SingleElectron-RunE	4.0	SingleElectron-RunE	9.3	EGamma-RunD	31.7
SingleElectron-RunF	3.1	SingleElectron-RunF	13.5		
SingleElectron-RunG	7.6				
SingleElectron-RunH	8.7				
Total 2016	35.9	Total 2017	41.5	Total 2018	59.4

2016		2017		2018	
Dataset	fb^{-1}	Dataset	fb^{-1}	Dataset	fb^{-1}
SingleMuon-RunB	5.8	SingleMuon-RunB	4.8	SingleMuon-RunA	14.0
SingleMuon-RunC	2.6	SingleMuon-RunC	9.6	SingleMuon-RunB	7.1
SingleMuon-RunD	4.2	SingleMuon-RunD	4.2	SingleMuon-RunC	6.9
SingleMuon-RunE	4.0	SingleMuon-RunE	9.3	SingleMuon-RunD	31.7
SingleMuon-RunF	3.1	SingleMuon-RunF	13.5		
SingleMuon-RunG	7.6				
SingleMuon-RunH	8.7				
Total 2016	35.9	Total 2017	41.5	Total 2018	59.7

Table 3: datasets exploited for the analysis in the W electronic (top) and muonic (bottom) channel. The datasets are divided into years and eras. The luminosities for each sample and for the full year are presented.

to a rate that is acceptable for the computing facilities. For this reason, there is a minimum threshold in transverse momentum of the triggering lepton that is requested by the trigger system. A selection based on the trigger paths presented in Table 4 have been implemented in the analysis. They correspond to the lowest values in transverse momentum for single electron and muon triggers. As it can be seen, for the electron channel the threshold has been increased from 27 GeV for 2016 to 32 GeV in 2017 and 2018 in order to cope with the higher pileup rate due to the higher instantaneous luminosity.

5.1.2 *Monte Carlo simulations*

As it will be described later in more detail, the event selection aims to retain as much signal events as possible while limiting the number of background ones. However, no matter how good the selection, some background events will still be present in

	Electron HLT trigger path	Muon HLT trigger path
2016	HLT_Ele27_WPTight_Gsf	HLT_IsoMu24 OR HLT_IsoTkMu24
2017	HLT_Ele32_WPTight_Gsf_L1DoubleEG	HLT_IsoMu24 OR HLT_IsoMu27
2018	HLT_Ele32_WPTight_Gsf	HLT_IsoMu24

Table 4: trigger paths for the single electron and single muon channel analyses. The numbers are the threshold for the triggering particle in GeV.

the signal region and so, in order to be able to compare the observed data with the expectation, both signal and background events must be simulated with Monte Carlo techniques. The main background contributions in the $W\gamma\gamma$ signal region can be divided into processes that have two real photons (such as the $t\gamma\gamma$ or the $t\bar{t}\gamma\gamma$ production) and processes where photons are selected as a consequence of a misreconstruction procedure (such as the $Z\gamma$ process where one of the two electrons coming from the Z decay is reconstructed as a photon). Besides the $W\gamma\gamma$ signal region, also the single-photon $W\gamma$ and inclusive W control regions are exploited in this analysis and their backgrounds must also be simulated. As for the signal region, the background contributions in the $W\gamma$ control region come from events with a true generated photon (such as $Z\gamma$, γ+Jets, $t\gamma$ and tri-boson $WW\gamma$ and $WZ\gamma$ production) or from events where a photon is obtained as the result of a misreconstruction. The main background sources in the inclusive W control region come from Z+Jets events (also known as Drell-Yan events), single and double top quark production, multi-boson processes (WW, WZ, ZZ, WWW, WWZ and WZZ) and multi-jet events, that are described in detail in Section 5.4.1.

The $W\gamma\gamma$ signal sample is generated at NLO with MadGraph5_aMC@NLO and showered with Pythia8. The single and double-photon background samples, namely the $W\gamma$ and $Z\gamma$, the $t\gamma$, $t\bar{t}\gamma$ and $t\bar{t}\gamma\gamma$ samples are generated in the same way. The inclusive W+Jets and Z+Jets samples, the tri-boson samples (WWW, WWZ, WZZ, ZZZ, $WW\gamma$ and $WZ\gamma$) are also generated at NLO with MadGraph5_aMC@NLO and showered with Pythia8. The W+Jets ad Z+Jets samples are generated separately as a function of the jet multiplicity in the final state. This choice allows to have a large statistics also in the region with one or two jets, while, if the process were to be generated in a single sample, the contribution of these events would be statistically suppressed by the smaller cross section values. The γ+Jets samples are generated at leading order with MadGraphMLM and showered with Pythia8. As for the inclusive samples and for the same reason, also for this case the generation is split in different samples following a categorisation based on the scalar sum of the jets transverse momenta H_T. The single top contribution both in the s- and t-channel is generated with Powheg [136] and showered with Pythia8. Finally, the di-boson samples (WW, WZ and ZZ) are generated with a Pythia8 only simulation. A summary of the samples is reported in Table 5.

As described in Section 4.2, Pythia8 is initialised with a set of parameters referred to as a tune. For the aforementioned samples, the CP5 tune [129] has been implemented for the 2017 and 2018 samples, while the CUETP8M1 tune [137] has been used for

Dataset name	Generator order	Cross section pb	Cross section order
$W\gamma\gamma$	NLO	2.001	NLO
$W\gamma$	NLO	199.5	NNLO
$Z\gamma$	NLO	131.6	NNLO
$t\gamma$	NLO	3.055	NLO
$t\bar{t}\gamma$	NLO	4.216	NLO
$t\bar{t}\gamma\gamma$	NLO	0.01687	NLO
γ+Jets - $40 < H_T < 100$	LO	18570	LO
γ+Jets - $100 < H_T < 200$	LO	8601	LO
γ+Jets - $200 < H_T < 400$	LO	2194	LO
γ+Jets - $400 < H_T < 600$	LO	257.8	LO
γ+Jets - $600 < H_T < Inf$	LO	84.92	LO
Drell-Yan + 0 jet	NLO	4833.39	NNLO
Drell-Yan + 1 jet	NLO	889.66	NNLO
Drell-Yan + 2 jets	NLO	354.17	NNLO
W + 0 jet	NLO	50084.9	NNLO
W + 1 jets	NLO	8262.7	NNLO
W + 2 jets	NLO	3179.1	NNLO
WW	LO	118.7	LO
WZ	LO	47.13	LO
ZZ	LO	16.523	LO
WWW	NLO	0.2154	NLO
WWZ	NLO	0.1676	NLO
WZZ	NLO	0.05701	NLO
ZZZ	NLO	0.01473	NLO
WWγ	NLO	0.2316	NLO
WZγ	NLO	0.04345	NLO
$t\bar{t}$	LO	491.3	NNLO
t t-channel	LO	136.02	NLO
$\bar{t}$ t-channel	LO	80.95	NLO
t tW-channel	LO	35.85	NNLO
$\bar{t}$ tW-channel	LO	35.85	NNLO

Table 5: list of the Monte Carlo signal and background samples. Besides the process, the order at which the sample is generated is reported. The value of the cross section and the order at which it is computed are also presented.

the 2016 samples. The PDF set used in the event simulation also differs between the 2016 and the 2017/2018 samples. In 2016, the NNPDF3.0 [138] set, which is a NLO precision set, is used. For the other samples, the NNLO precision NNPDF3.1 [75] set is implemented.

The value of the cross section for the samples has been obtained in different ways. The cross sections for the inclusive W+Jets and Z+Jets samples have been computed using the "Fully Exclusive W and Z Production" FEWZ3.1 [139] framework, which is a computation that includes, in a single step, both the NNLO QCD correction and the NLO electroweak corrections to the W and Z processes. The cross sections for the $W\gamma$ and the $Z\gamma$ samples have been computed explicitly for this analysis using the MATRIX framework [140]. The generator level phase space used in the Monte Carlo generation was reproduced in the MATRIX configuration card, in order to obtain a result that matches the simulated sample. The value of the NNLO cross section is compatible with the NLO one, but the associated uncertainty is more than halved, giving a better precision in the result of the computation. The precision for the cross section for the other samples is the same as the generation order of the process.

5.1.3 *Photon overlap removal*

The generation of the events is performed in two steps: matrix element computation and parton shower. The process of interest is defined and computed at matrix level and then the contribution from initial and final state radiation is added before proceeding to the hadronisation. This procedure might introduce an ambiguity when many simulations are used at the same time. To give an example, the W+Jets sample does not include photon contributions at matrix level (the process is like the one of Figure 18 on the top left), but a photon can be added in the parton shower step as a contribution from ISR or FSR (Figure 18 top right). If the added photon is sufficiently energetic and isolated from the other particles, the obtained final state is the same as the one generated with a $W\gamma$ process at matrix level (Figure 18 bottom). This means that the number of simulated single-photons events if both the W+Jets and the $W\gamma$ sample are used at the same time is counted twice. The same happens, of course, for the $W\gamma\gamma$ signal events that can come from the $W\gamma\gamma$ simulation, but also from $W\gamma$ events with one ISR/FSR photon or from W events with two such photons. Eventually, even if the contributions are smaller and could be negligible, the same effect is present for the $t\bar{t}\gamma\gamma$ events that get contributions from the $t\bar{t}\gamma$ and $t\bar{t}$ simulations.

To avoid double counting, a photon overlap removal procedure has been applied at generator level for all the samples. The idea is to perform a selection of the events at generator level (i.e. using particles before the detector simulation) that is able to identify the presence of possible photons and count their number. After having established the number of photons at generator level, the event is retained or discarded accordingly to the number of photons that are nominally produced for that particular sample (for example, the W+Jets sample should have no photons, while the $W\gamma$ sample should have one of them).

In order to count the number of photons, a generator-level selection is applied. The photons are required to have a transverse momentum greater than 20 GeV and to be

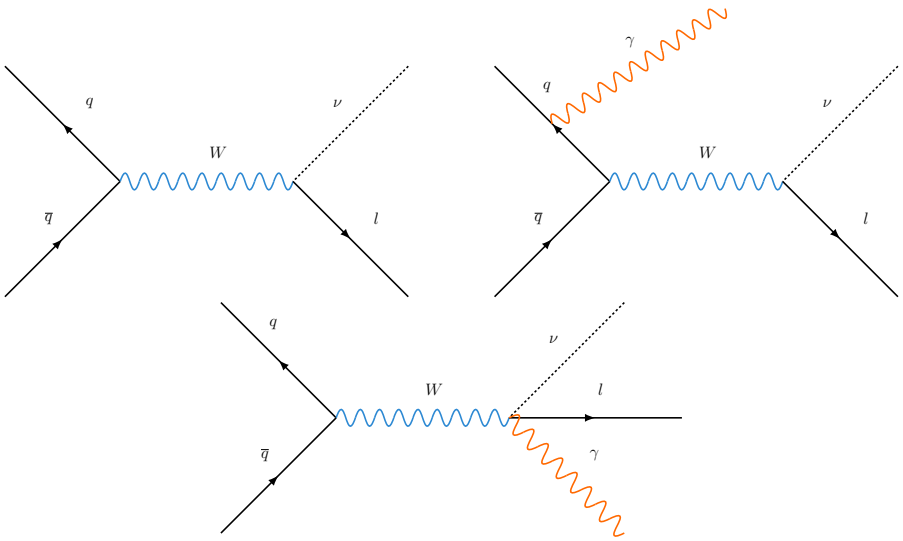


Figure 18: possible W diagrams without (top left) and with (top right) addition of an initial state radiation photon. As a comparison, the matrix level production for Wγ is also sketched (bottom).

in the detector acceptance region with $|\eta| < 2.5$. The photon candidate is required to be a prompt particle, which means to be produced in the hard scattering interaction, and to be stable, which means that it did not undergo decays and interaction chains. The goal of this particular selection is to avoid counting photons that resemble photons produced at matrix level, which means that the contribution from photons that are inside jets or are a product of successive emission in the parton shower should not be considered in this procedure. In order to go as much as possible in this direction, photons that are close to leptons or other stable particles are not taken into account. If the photon is inside a $\Delta R = \sqrt{(\phi_\gamma - \phi_p)^2 + (\eta_\gamma - \eta_p)^2}$ cone of radius 0.4 around a lepton with $p_T > 15\,GeV$ or inside a cone of radius 0.1 around all the other stable particles it is not selected. If more than one photon is found in the event, it must be at a distance of at least $\Delta R = 0.4$ from the previously selected one, otherwise the candidate with lower transverse momentum is discarded.

After the generator-level selection just described, the number of photons likely resembling prompt ones is established. If the sample is an inclusive one, such as the W+Jets or the Z+Jets, the event is discarded if one or more photons are identified in this way. Following the same logic, for the single-photon samples, such as Wγ and Zγ, the event is discarded if the number of generator level selected photons is different from one. For the W$\gamma\gamma$ signal sample, the event is retained only if two or more photons are selected.

A validation of the procedure can be seen in Figure 19. The plots show the distribution of the transverse mass of the W boson in events with one reconstructed photon. The plot on the left is built by using only the inclusive W+Jets sample. The contribution labelled as Wγ is taken from the inclusive sample itself (it is made of the events for which one photon passes the generator-level selection described above)

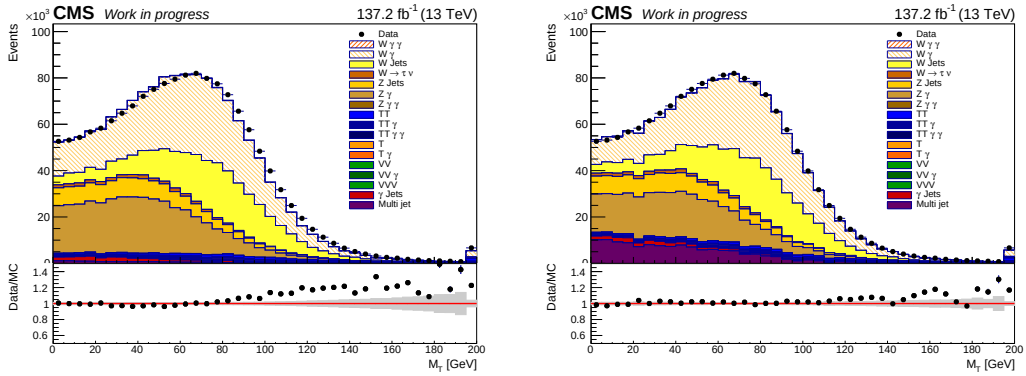


Figure 19: W boson transverse mass in events with one reconstructed photon. The plot on the left exploits the inclusive W+Jets sample for both the estimation of the W and the Wγ contributions, while the plot on the right uses the inclusive W+Jets sample for the W events and the exclusive Wγ sample for the events with one reconstructed photon. The grey bands in the bottom plots are the statistical uncertainties of the Monte Carlo simulations.

and is just shown in a different colour for clarity. The plot on the right, however, uses the W+Jets sample for the events with no selected photons at generator level and the Wγ exclusive sample for the events with one photon selected at generator level. As it can be seen, the description of the process is similar in the two plots, showing that the splitting of the inclusive sample in the components with zero or one photons gives a similar result as using the W+Jets sample for the events with zero photons and the Wγ sample for the events with one photon. The plot on the right has a better data-simulation agreement and a smaller statistical uncertainty; this is linked, of course, to the implementation of the exclusive sample that has a much better predictive power for the single-photon events.

Treatment of τ leptons

As it will be explained in the next section, only the electron and muon final states of the W decay will be used for this analysis. The events where the W decays into a τ lepton and a neutrino are then treated as background. As for the photon overlap removal procedure, these events are subsets of the inclusive W+Jets sample and of the exclusive Wγ and Wγγ samples. The contribution of τ leptons is removed from the signal contribution and treated separately by using a generator level selection, similarly to what has been done for the photons. Among all the leptons with transverse momentum greater than 15 GeV, τ leptons are selected. The Monte Carlo simulation retains the information on all the history of the generated process and it is possible to check what is the particle from which the τ decayed: if it matches a W boson, the event is categorised as τ background, removed from the inclusive sample and added into a dedicated background contribution.

5.2 EVENT SELECTION

The $W\gamma\gamma$ final state is characterised by the presence of the W decay products and two photons. For this analysis, only the muon and the electron W decays are treated as signal. The $W \to qq'$ decay has a very low signal-to-noise ratio since it is easily faked by the di-jet production. The τ decay channel has the same branching ratio as the electron and muon ones, but it is hard to reconstruct since there are two neutrinos in the final state: one coming from the W boson decay and one from the τ decay. In the following, if not specified otherwise, the term "lepton" will be used to indicate only electrons and muons, and not taus. The W electronic and muonic decays have a branching ratio of nearly 11% and are characterised by the presence in the final state of a lepton with high transverse momentum and a non negligible amount of missing transverse energy (defined in Equation 57), connected to the presence of the undetected neutrino. The event selection proceeds first with the identification of final states object and then, by using this information, with the reconstruction of the full final state.

Electrons

Reconstructed electron candidates are selected if they have a transverse momentum greater than 15 GeV and fall inside the acceptance region of the electromagnetic calorimeter, which means that they have $|\eta| < 2.5$. The gap region between barrel and endcaps, corresponding to $1.442 < |\eta| < 1.566$ is excluded due to the small reconstruction efficiency.

In general, the same quantities are used to select the electrons in the barrel and in the endcaps. However, being the reconstruction in the endcaps region less accurate than the one in the barrel region, the requirements are usually looser for the former. A requirement on the distance of the reconstructed vertex from the position of the primary vertex of the interaction is performed by imposing the distance between the two to be small both in the transverse plane and along the beam direction.

The electron reconstruction in the calorimeter is characterised by the lateral dimension of the shower, expressed in terms of the η variable: a narrow shower is likely to come from an isolated electron, while a wide one can indicate the presence of other particles in its proximities. For this reason, a selection on the lateral dimension of the shower ($\sigma_{i\eta i\eta}$) is applied.

The reconstructed track and the energy deposit in the calorimeter must be close to each other, which hints to the fact that the two reconstructed objects have a common origin. For this reason, the difference in angular distances between the two, $\Delta\eta$ and $\Delta\phi$, are required to be small.

An electron releases the greater part of its energy in the electromagnetic calorimeter, while only a small amount is normally released in the hadronic calorimeter. Hadronic jets have the opposite behaviour. The ratio of the electron energy released in HCAL over the the release in ECAL (H/E) is then required to be small. The H/E threshold is computed as a function of the detected energy in the ECAL supercluster since it is more probable that a high energy electron releases more energy in the HCAL

if compared to a lower energy one. The H/E selection takes into account also the possible contribution coming from pileup that directly affects the energy measured in a certain region of the calorimeters. For this purpose, a dependency on ρ, defined as the event-specific average pileup energy density per unit area in the (η, ϕ) plane, is introduced.

Another fundamental quantity, that has a strong discriminating power in identifying electrons over jets, is the relative isolation of the electron with respect to the other reconstructed particles. The energy of an electron is estimated both by the energy deposit in ECAL and by the reconstruction of the track in the tracker. To pass the selection, the two energy values must give a compatible result. Eventually, the electrons are required to have no more than one missing hit in the tracker and not to come from conversion processes. A summary of all the selection cuts and the applied threshold is presented in Table 6. The average efficiency for this electron selection is approximately 70%.

Muons

Reconstructed muons candidates are selected if they have a transverse momentum greater than 15 GeV and fall inside the tracker and muon chambers acceptance regions, i.e. their pseudorapidity must be smaller than 2.4.

Detector misalignment, software reconstruction bias and uncertainties in the measurement of the magnetic field can spoil the determination of the muon momentum. To take into account for these eventualities, the muon momentum is corrected following the "Rochester" correction approach [141].

As for the electrons, the reconstructed vertex must be close to the primary vertex of the collision both in the transverse plane and in the beamline direction. To be selected for the analysis, a muon must be reconstructed both as a global and a particle-flow muon. The track fitting procedure must give an accurate description in terms of χ^2, which means that the successive hits describe a particle curvature compatible with the presence of a muon. The candidates are required to have at least one associated hit in the muon chambers and at least two muon station segments (see Sections 3.1 and 3.3.1 for a detailed description of the muon reconstruction). Moreover, to exploit both the information from the muon chambers and the one on the tracker, at least one hit in the pixel detector and 5 hits overall in the tracker are required. A summary of the selection cuts for the muons is presented in Table 7. The efficiency for the muon selection ranges from 75% for the low muon transverse momentum and low pseudorapidity regions, to about 90% for muons in the $30 < p_T < 40$ GeV region and is above 95% for muons with transverse momentum greater than 40 GeV.

Photons

Reconstructed photon candidates are selected if they have a transverse momentum greater than 20 GeV and they fall inside the acceptance region of the electromagnetic

Electrons	
$p_T > 15\,GeV$	
$\lvert\eta\rvert < 1.442$ OR $1.566 < \lvert\eta\rvert < 2.5$	
Barrel	Endcaps
$d_{xy} < 0.05$	$d_{xy} < 0.10$
$d_z < 0.10$	$d_z < 0.20$
$\sigma_{i\eta i\eta} < 0.0104$	$\sigma_{i\eta i\eta} < 0.0353$
$\lvert\Delta\eta\rvert < 0.00255$	$\lvert\Delta\eta\rvert < 0.00501$
$\lvert\Delta\phi\rvert < 0.022$	$\lvert\Delta\phi\rvert < 0.0236$
$H/E < 0.026 + 1.15/E_{SC} + 0.0324\rho/E_{SC}$	$H/E < 0.0188 + 2.06/E_{SC} + 0.183\rho/E_{SC}$
$Iso_{rel} < 0.0287 + 0.506/p_T$	$Iso_{rel} < 0.0445 + 0.963/p_T$
$1/E - 1/p < 0.159$	$1/E - 1/p < 0.0197$
Missing hits $\leqslant 1$	Missing hits $\leqslant 1$
Pass conversion veto	Pass conversion veto

Table 6: selection cuts applied for the electron reconstruction.

Muons
$p_T > 15\,GeV$
$\lvert\eta\rvert < 2.4$
Rochester corrections
$d_{xy} < 0.2$
$d_z < 0.5$
Track fit χ^2/ndof < 10
Muon chamber hits $\geqslant 1$
Muon station segments $\geqslant 2$
Global muon
Particle-flow muon
Pixel hits > 1
Tracker layers hits > 5

Table 7: selection cuts applied for the muon reconstruction. See text for the description of the Rochester corrections.

Photons	
$p_T > 20\,GeV$	
$\lvert\eta\rvert < 1.442$ OR $1.566 < \lvert\eta\rvert < 2.5$	
$\lvert m_{e,\gamma} - 91.2\rvert > 5$	
$\lvert m_{\gamma,\gamma} - 91.2\rvert > 5$	
Barrel	Endcaps
Pixel seeds = 0	Pixel seeds = 0
$\sigma_{i\eta i\eta} < 0.01015$	$\sigma_{i\eta i\eta} < 0.0272$
$H/E < 0.02197$	$H/E < 0.0326$
$Iso_{chg} < 1.141$	$Iso_{chg} < 1.051$
$Iso_{neu} < 1.189 + 0.01512 p_T + 2.259 \cdot 10^{-5} (p_T)^2$	$Iso_{neu} < 2.718 + 0.0117 p_T + 2.3 \cdot 10^{-5} (p_T)^2$
$Iso_\gamma < 2.08 + 0.004017 p_T$	$Iso_\gamma < 3.867 + 0.0037 p_T$

Table 8: selection cuts applied for the photon reconstruction.

calorimeter. As for electron candidates, the gap region between the ECAL barrel and endcaps is discarded.

The photon is required to pass a pixel seed veto requirement, which means that no track seeds in the silicon tracker can be associated to the energy deposit in the electromagnetic calorimeter. As for the electrons, the H/E and the lateral shower shape dimension $\sigma_{i\eta i\eta}$ variables are used in order to reduce the possible contamination from jets. The average efficiency for this photon identification is approximately 80%.

The particle-flow reconstructed photon is required to be isolated from neutral and charged hadrons and from other photons. If more than one photon is selected in the same event, the one with lower transverse momentum is discarded if it falls inside a $\Delta R = 0.4$ cone around the highest transverse momentum one. This cut is specifically implemented to reduce the contamination of photon pair production for example from the decay of a neutral pion. Photons are required to be isolated from all the electrons and muons that pass the nominal selection previously described. If the photon candidate falls within a $\Delta R = 0.4$ cone around a selected lepton, it is discarded in order to reduce the bremsstrahlung contribution.

As it will be discussed later in more detail, if the track of an electron is badly or not reconstructed, the electron can be misidentified as a photon. The greater contribution to these events for the W analysis comes from the decay of a Z boson into two electrons where one of the two electrons is reconstructed as a photon. In order to reduce the contribution from this background as much as possible, the invariant mass of a candidate photon and the candidate electron is required to be at least 5 GeV away from the nominal Z mass value (M_Z =91.2 GeV). Even if more unlikely, both the electrons coming from the Z decay can be misreconstructed as photons. In the same fashion as just described, if two photons are found to have an invariant mass compatible with the Z mass, they are discarded. The selection cuts for the photons are summarised in Table 8.

Overall event selection

An event passes the W boson selection if an electron with transverse momentum greater than 35 GeV or a muon with transverse momentum greater than 30 GeV is identified. As described earlier, single electron and single muon trigger requirements are implemented in order to match the online selection. Besides the already mentioned requests, the selected lepton must be the one that triggered the event. To reduce the contamination from Z+Jets and $t\bar{t}$ backgrounds, no other electron or muon besides the one that fired the trigger can be present in the event. An event is characterised as a $W\gamma$ event if, besides the just mentioned requests, one photon is reconstructed following the aforementioned criteria. An event is characterised as a $W\gamma\gamma$ event if two or more such photons are identified.

5.3 DATA-SIMULATION CORRECTIONS

Monte Carlo simulations are an incredibly powerful instrument to check what the outcome of a hard process will be. Besides the generation of the matrix level and hadronised events, the simulation of the particle interactions with the detector is described as well. The goodness of the description of the particle interactions with the detector, however, is limited by the accuracy of the modelling of the detector itself. Effects such as miscalibrations, inefficiency and degrading performances are difficult to simulate both in a space and time dependent way. It is, in fact, not unlikely that one portion of one subdetector has a malfunction or that the response of a certain crystal changes during the data acquisition periods. To take into account these effects, it is then necessary to compute some corrections based on the collected data and then apply them to the simulated events in order to get an overall better description of the underlying physics. The corrections that have been applied to the simulations used for this analysis are linked to the pileup simulation, to the particle reconstruction efficiency and to the energy resolution.

5.3.1 *Pileup corrections*

As already explained, an average of 30 collisions in 2016 and of 50-60 collisions in 2017 and 2018 occurred at every bunch crossing during the LHC Run 2. One of these collisions could be the hard scattering interaction that drives the studied process, while the others are soft collisions that are not interesting in terms of physics, but that influence the response of the detectors. For this reason, the simulation must include also these events.

What is currently done at CMS is referred to as *PreMixing*. Minimum bias events, i.e. events produced by a soft (low p_T) inelastic proton-proton interaction, are generated. They include only the contribution of the pileup and are successively mixed with a hard scattering process using a reference value for the minimum bias cross section of $69.2 \pm 3.2\,mb$ as measured in dedicated CMS runs. After the addition of these processes, the simulation of the event is completed.

The estimation of the pileup is linked to the measurement of the luminosity and to the measurement of the interaction vertices. This procedure gives rise to an estimate of the average pileup interactions for each run of data collection. The distribution of the number of multiple interactions for individual events follows a Poissonian distribution around this average. To correct the simulation for this contribution, each event is weighted with a factor that takes into account the statistical information just described.

5.3.2 *Monte Carlo scale factors*

The events selection described in the previous section relies on a series of criteria that are applied both to the data and the simulation. The efficiency of these requirements, i.e. the probability that an event is above or below the threshold, can be different in data and simulation and hence the final distributions can differ. To take into account this possibility, the events in the simulations that pass the selection are reweighted by considering possible differences in the efficiency. Each selected event in the simulation is corrected by efficiency correction factors for the reconstructed objects (electrons, photons and muons) computed with the procedure described in the following paragraph. Since generally the simulation gives a good description of the real events, the value of the scale factors is usually close to unity.

To compute the value of the efficiencies, the so called *Tag&Probe* method has been implemented. The idea is to exploit a known process and to check what is the efficiency for a certain particle to pass or fail the selection criteria. What has been done for the electrons, for example, is to exploit the decay of the Z boson into two electrons, which is a very clean process. A set of selection cuts is applied to the first electron of the Z decay, that tags the event. A second electron, called probe, is selected with a set of selection cuts that resembles the one of the analysis, with the exception of the cut that is under study. The events are then categorised as *pass* if the probe satisfies the selection cut under investigation, otherwise they are categorised as *fail*. It is then possible to compute the efficiency as

$$\epsilon = \frac{n_{\text{pass}}}{n_{\text{pass}} + n_{\text{fail}}} \,. \tag{58}$$

This is done both in data and simulation and the scale factor can then be computed as previously explained. In order to take into account possible dependencies on the region of the detector where the particle was reconstructed and on the energy of the interacting particle itself, the scale factors are computed as a function of the particle transverse momentum and pseudorapidity.

For this analysis, electrons are corrected for their identification, reconstruction and trigger efficiencies. Muons are corrected for their identification, isolation and trigger efficiencies. Photons, finally, are corrected for their identification efficiency and for the efficiency of the removal of tracker seeds associated to the energy deposit in the calorimeter.

5.4 BACKGROUND CONTRIBUTIONS

Three main regions are analysed and exploited in this work: the inclusive W and the single-photon Wγ control regions and the W$\gamma\gamma$ signal one. The inclusive control region is fundamental in order to evaluate the correct application of all the data-simulation corrections. Since no photons are explicitly selected, it is easier to evaluate the consistency between data and simulation for the electron and muon reconstruction. The main backgrounds for this region are Z+Jets events, where one of the two decay leptons of the Z is not reconstructed, which are obtained from simulation, and multi-jet events, that are estimated with a data-driven technique. Both the single and the double-photon regions are affected by backgrounds where a photon is reconstructed while it was not present at generator level. These misidentified photons come from electrons whose tracks were not reconstructed and jets initiated by a neutral particle with a large electromagnetic component. Both these contributions are estimated with a data-driven technique. Finally, in all the three regions there are minor background contributions coming from single and double top, di-boson and tri-boson production, possibly associated with one or more photons. The associated production of a jet and one or more photons accounts for some of the background events in the Wγ and W$\gamma\gamma$ regions. All these contributions are estimated from Monte Carlo simulations.

5.4.1 *Multi-jet background sources*

The most probable proton-proton interaction process that happens at the LHC is the scattering of two quarks into two quarks. The final state quarks hadronise and produce two energetic jets. If one of the two jets has a great electromagnetic component, it can be misreconstructed as an electron. If the other jet is outside the acceptance of the detector, it is not reconstructed. This kind of events resemble exactly the presence of a W boson, with one energetic electron-like deposit and a substantial amount of missing transverse energy. The generation throughout Monte Carlo simulations of such processes is not convenient, since a lot of multi-jet events need to be generated and only a small number of them passes the selection criteria to be identified as a leptonic W boson decay. For this reason, a data driven approach has been used to estimate this background contribution.

The idea of the method is to obtain the distribution for the multi-jet background from data in a control region and then rescale and use it in the signal region. The multi-jet events are characterised by the presence of a jet that is misidentified as an electron. To build a region enriched in such events, the same electron selection as the nominal one is applied, except for the requirement on the electron isolation that is now inverted. A non isolated electron, in fact, is more likely coming from a jet. Since the control region is enriched in multi-jet events for which there is no simulation available, a considerable discrepancy between data and simulations is observed. This behaviour can be seen, for example, from the distribution of the W boson transverse mass in the multi-jet control region, reported in Figure 20. The missing contribution needed to match data and simulation comes exactly from the multi-jet background.

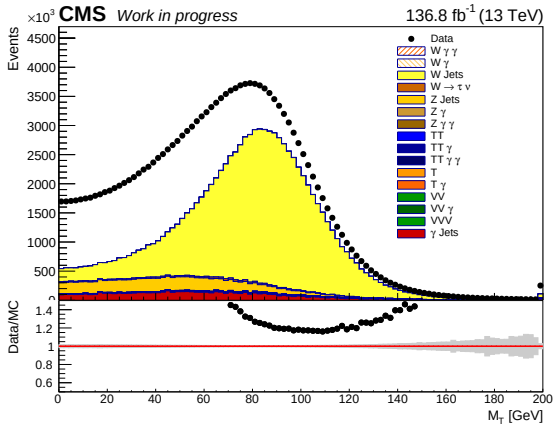


Figure 20: W boson transverse mass in the multi-jet control region. The data are reported with the black dots while the Monte Carlo contributions are stacked with the filled areas. The discrepancy between data and simulation is linked to the presence of the multi-jet background.

From this distribution (and from its analogue for all the other quantities) the template for the multi-jet background can be estimated as the difference between data and the sum of the distributions from simulation.

Having obtained the shape of the distribution, it is necessary to rescale it in order to get the correct prediction in terms of number of events in the signal region. To do so, the template is added to the sum of the Monte Carlo contributions in the signal region (Figure 21, left) and is fitted to the data. Since the contribution from the multi-jet background is larger for low values of the transverse mass, the fit is performed in the region with M_T <40 GeV. The fitting procedure fixes the amount of the other backgrounds and fits the normalisation of the multi-jet distribution to the data. To have a better description of the possibly different behaviours in the barrel and endcap regions, two separate fits are performed. The final result, where the multi-jet background is stacked with all the other simulations and compared to data, is presented in Figure 21 on the right. As it can be seen, the agreement between data and simulation in all the transverse mass spectrum is very good.

The same procedure is applied also for the muon decay channel. The selection for the control region follows the same idea as the one for the electron decay channel, which means that the isolation of the muon is inverted while all the other selection criteria are the same as in the signal region. As can be seen from Figure 22, the contribution from jets reconstructed as muons is much smaller if compared to the electron case.

5.4.2 *Electron-photon misidentification background contribution*

An electron is reconstructed when a track in the tracker is matched to an energy deposit in the electromagnetic calorimeter. As described in Section 3.3.2, the track must satisfy some matching requirements in order to be associated to the deposit and to result in a reconstructed electron. If any of these requirement fails, the unmatched

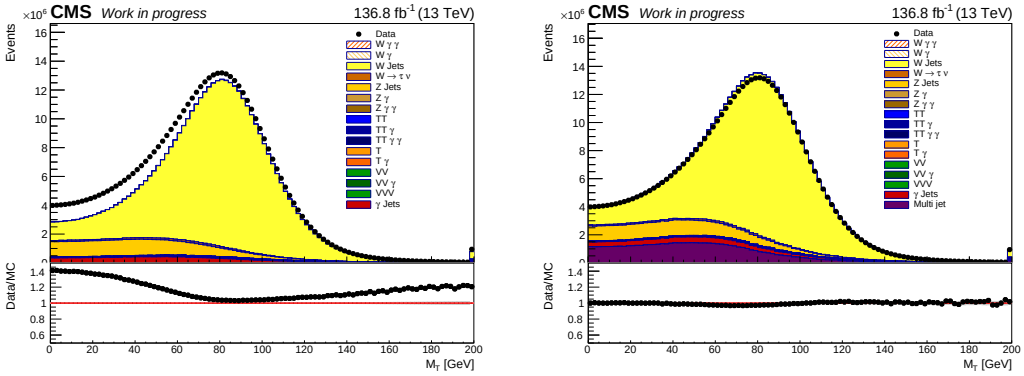


Figure 21: W boson transverse mass distribution for the electron decay channel without (left) and with (right) the contribution of the multi-jet background as derived from the control region. The plot shows great agreement between data and simulation in the final distribution.

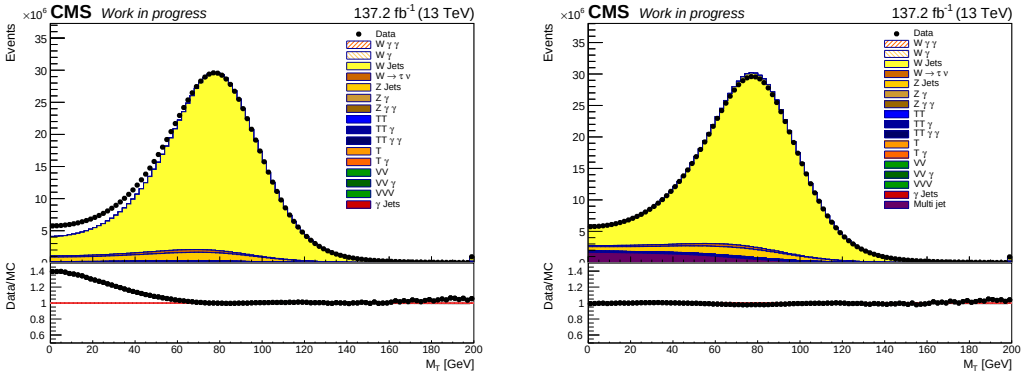


Figure 22: W boson transverse mass distribution for the muon decay channel without (left) and with (right) the contribution of the multi-jet background as derived from the control region. The plot shows great agreement between data and simulation in the final distribution.

energy deposit in the electromagnetic calorimeter is likely reconstructed as a photon. As for the multi-jet background, the simulation of these events is not very effective or precise and their presence is accounted for with a data driven method.

The most likely way for an electron to be reconstructed as a photon while applying a W boson selection is that the hard process was indeed a Z decay. A Z boson can decay into two electrons and if the track of one of the two electrons is not reconstructed, the event resembles exactly a $W\gamma$ one. The same reasoning applies for the $Z\gamma$ production that can fake a $W\gamma\gamma$ event. The Monte Carlo simulation both for the Z+Jets and the $Z\gamma$ events does not give a satisfactory description of this process and must then be corrected. The events obtained from the Z or $Z\gamma$ simulations are kept, but their relative weight is corrected on an event-by-event basis in order to obtain a distribution that is able to properly describe this effect.

To check the different impact of the electron-photon misidentification in data and Monte Carlo, a distribution that is sensitive to this effect is needed. During the selection process, if the invariant mass of a candidate photon and the trigger electron is in the proximity of the nominal value of the Z boson mass, the event is discarded. By releasing this requirement, it is possible to build a "fake" Z mass peak with one electron and a photon that is likely to come from an electron, which is exactly the process under investigation. This fake-Z is reconstructed in different intervals of the photon transverse momentum and pseudorapidity, in order to take into account possible variation in the photon spectra.

The distributions obtained in this way are fitted both in data and in simulation with a signal plus background model. The peak is fitted with a template obtained from Monte Carlo events with one reconstructed electron and one reconstructed photon that matches a generator level photon, which is the signature of the effect under study. The background is fitted with an analytical function,

$$f(x) = \mathrm{erfc}(\beta(\alpha - x)) \cdot e^{-\Gamma(x-x_0)}, \tag{59}$$

that describes a background shape that first increases due to threshold effects and then decreases exponentially. The erfc function is defined as the complementary of the error function. An example of the fitting procedure is presented in Figure 23, while the whole set of fits is added for completeness in Appendix A.

After the fitting procedure, it is possible to obtain the number of events for the fake-Z peak both in data and in simulation. In order to remove possible dependencies from the cross section value (to which the Monte Carlo simulations are normalised to) or from other effects due to efficiency differences, the number of events in the fake-Z peak is normalised to the number of events for the real Z peak both in data and in simulation. After this procedure, it is then possible to compute a correction factor as

$$f(p_T, \eta) = \frac{N^{e\gamma}_{DATA}(p_T, \eta)/N^{ee}_{DATA}(p_T, \eta)}{N^{e\gamma}_{MC}(p_T, \eta)/N^{ee}_{MC}(p_T, \eta)} \tag{60}$$

where $N^{e\gamma}$ is the number of events obtained from the fit of the peak for the fake-Z distribution, while N^{ee} is the number of events obtained for the real Z peak. The correction factors are derived separately for the three years of data taking in order

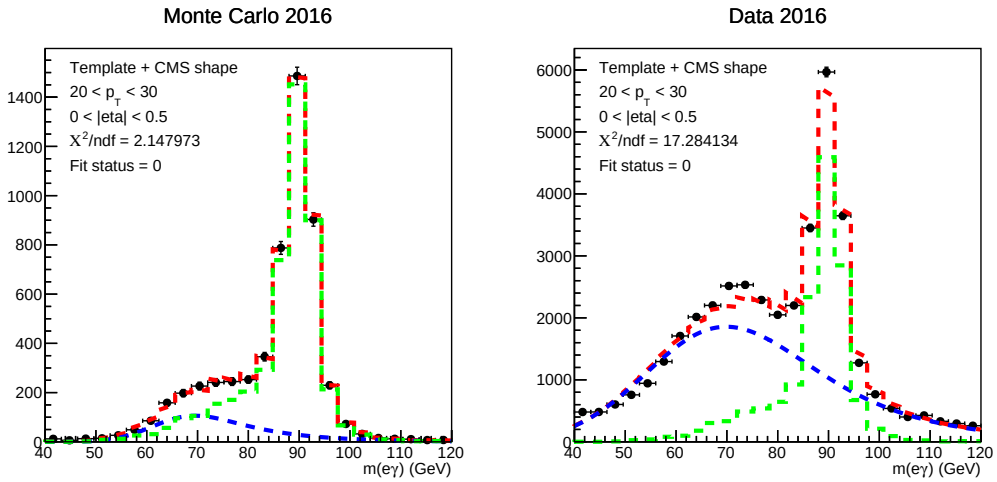


Figure 23: example of fitted fake-Z mass distributions in Monte Carlo (left) and data (right) for the $0 < \eta < 0.5$ and $20 < p_T < 30$ region for the 2016 data. The template in green is obtained from Monte Carlo events where the electron and the photon selected at reconstructed level are required to match respectively an electron and a photon at generator level. The blue curve is an analytical function that describes the background and is explicitly written in Equation 59. The red curve is the sum of the two. The black dots represent the distribution of events with an electron and a photon selected at reconstructed level. The full set of fitted distribution is presented in Appendix A.

to take into account possible variations between the different periods. The values of the correction factors are summarised in Figure 24 for the 2018 dataset, the full set of scale factors is reported in Appendix A.

While running the analysis on the Monte Carlo simulations, all the events with a reconstructed photon that originates from a generator level electron are weighted with the correction factors. While, as already stated, the vast majority of the electron-photon misidentification events come from Z or $Z\gamma$ events, it is not possible to exclude that other processes produce such an effect. For this reason, all the simulations are reweighted with the procedure just described.

A significative closure test can be obtained by looking at the distribution of the electron plus photon invariant mass in the $W\gamma$ selection before and after the correction, as shown in Figure 25. The distribution on the left is the invariant mass before the reweighting procedure. One can argue that the ratio plot of the two is not too different from unity, but what must be noted is the contribution that comes from the multi-jet background. As it can clearly be seen, in the left plot the multi-jet background peaks under the Z peak, which is an artefact. The multi-jet background, in fact, being made of events with more than one jet (one of them reconstructed as a photon), should not show a resonant structure. It does so because some of the contribution it is estimating does come from the electron-photon misidentification process. After the correction, this is accounted for by the Z+Jets Monte Carlo contribution and the multi-jet background has the expected flat distribution.

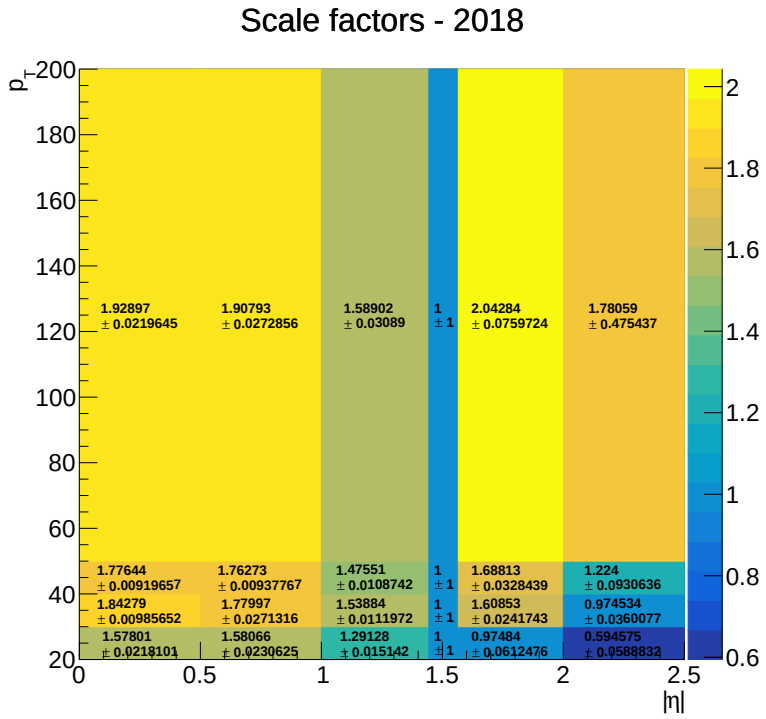


Figure 24: electron-photon misidentification background scale factors as a function of the photon transverse momentum and pseudorapidity for the 2018 data. The central values are obtained with the template fitting procedure, while the uncertainties are computed following the method described in Section 5.5.4.

5.4.3 *Jet-photon misidentification background contribution*

Neutral jets are characterised by the presence of an energy deposit in the electromagnetic and in the hadronic calorimeter and by the absence of a matching track from the tracker. One such jet can be initiated, for example, by an energetic neutral pion. The neutral pion can decay into two photons before reaching the ECAL. If the two photons are not resolved, what is reconstructed is an energy deposit in the electromagnetic calorimeter that is not associated with a track, which is exactly the same signature of a photon.

The jet-photon misidentification background is the one that contributes the most in the W$\gamma\gamma$ signal region. In principle, this process is simulated in the Monte Carlo generation, but, in the same way as for the electron-photon misidentification, the description is not guaranteed to match the process that happens in real collisions. Moreover, the number of simulated events that pass the W$\gamma\gamma$ selection for this background process is very small, giving a limited statistical prediction power. For these reasons, this background contribution has been estimated in a data-driven way.

The idea of the method is to exploit the information on a variable that is powerful in discriminating real photons from jets, such as the photon isolation. As stated before, a "true" photon is likely to be isolated, while a photon coming from a jet is more likely to be accompanied by other particles. The isolation variable is then able

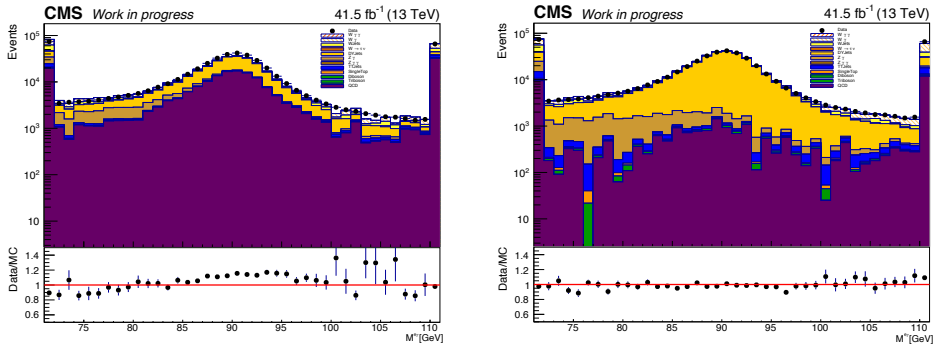


Figure 25: electron plus leading photon invariant mass for the Wγ selection before (left) and after (right) the correction for the electron-photon misidentification background. In particular, the Z+Jets distribution is presented in yellow, while the multi-jet background is presented in purple.

to differentiate between a region that is more likely dominated by the signal and one dominated by the jet-photon misidentification background. By exploiting this information both in data and in simulation, it is possible to build a matrix that can link the number of events in the isolated or not isolated regions with the composition of the signal region in terms of true or misidentified photons.

From the isolation distributions (schematically presented in Figure 26), it is possible to compute the number of isolated (tight) or not isolated (loose) reconstructed photons. The distribution for the Monte Carlo simulation is built with reconstructed photons that match a generator level photon: in this way, only the behaviour of true photons is taken into account and the probability for a true photon to be isolated can be estimated. Reconstructed photons in data samples contain both events with true photons and events where a jet is misidentified as a photon. In order to retain as much as possible the latter, the distributions obtained from Monte Carlo simulations with a reconstructed photon that matches a generator level one are subtracted from the distributions obtained from data. In this way, it is possible to build the probability for a jet to be isolated. With these information, the probability for a photon to be isolated in simulation (ϵ) and in data (f) can be computed as

$$\epsilon = \frac{N^T_{MC}}{N^T_{MC} + N^L_{MC}} \qquad f = \frac{N^T_{DATA}}{N^T_{DATA} + N^L_{DATA}}, \tag{61}$$

where N^T_{MC} is the number of photons that pass the isolation selection and N^L_{MC} is the number of photons that fail the isolation selection. MC refers to counts in the simulations, while DATA refers to count in data after the subtraction of real photon contributions from simulations. These probabilities are computed in six intervals of the photon transverse momentum and separately for the barrel and endcaps regions. Of course, $1-\epsilon$ $(1-f)$ indicate the probability for a photon to be not isolated in simulation (data). From these coefficients, it is possible to build a matrix that can be

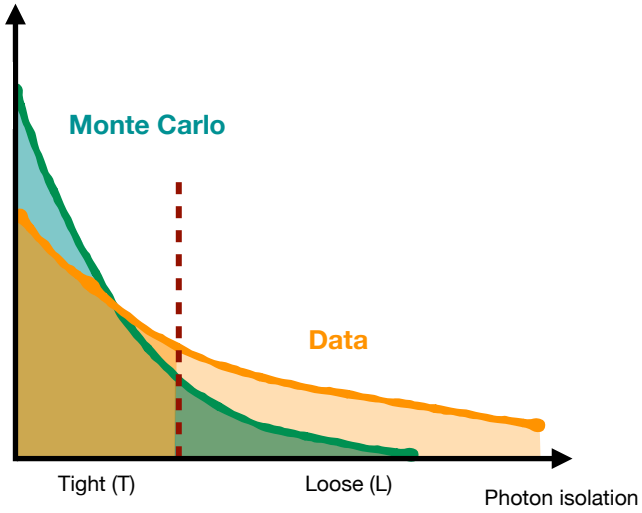


Figure 26: exemplification of the distributions of the photon isolation in data (orange) and in simulation (green). The red dotted line indicates the value of the nominal cut for the photon isolation in the event selection. The shaded areas with different colours indicate the isolated and not isolated events for both data and simulation.

used to estimate the contribution of the jet-photon misidentification background in the single-photon control region as

$$\begin{pmatrix} N_T \\ N_L \end{pmatrix} = \begin{pmatrix} \epsilon & f \\ 1-\epsilon & 1-f \end{pmatrix} \cdot \begin{pmatrix} N_\gamma \\ N_j \end{pmatrix} . \tag{62}$$

The left-hand-side vector is built from the number of isolated or non-isolated photons in data for the $W\gamma$ selection. The matrix exploits the coefficients just derived, and contains the information on the isolation probability. The vector on the right-hand-side contains the information of the composition of the events: N_γ are events with a true photon, while N_j are events where the reconstructed photon comes from a jet. By inverting the matrix it is then possible to estimate the composition of the signal sample in terms of real or fake photons. Since the matrix has been built from the single-photon control region and is applied to it, this procedure gives back the identity and is then just a closure check of the mathematical correctness of the method.

Having computed the probabilities for a single photon to be isolated (or not), they can be combined to estimate the contribution of the jet-photon misidentification background in the $W\gamma\gamma$ signal region from

$$\begin{pmatrix} N_{TT} \\ N_{TL} \\ N_{LT} \\ N_{LL} \end{pmatrix} = \begin{pmatrix} \epsilon_1\epsilon_2 & \epsilon_1 f_2 & f_1\epsilon_2 & f_1 f_2 \\ \epsilon_1(1-\epsilon_2) & \epsilon_1(1-f_2) & f_1(1-\epsilon_2) & f_1(1-f_2) \\ (1-\epsilon_1)\epsilon_2 & (1-\epsilon_1)f_2 & (1-f_1)\epsilon_2 & (1-f_1)f_2 \\ (1-\epsilon_1)(1-\epsilon_2) & (1-\epsilon_1)(1-f_2) & (1-f_1)(1-\epsilon_2) & (1-f_1)(1-f_2) \end{pmatrix} \cdot \begin{pmatrix} N_{\gamma\gamma} \\ N_{\gamma j} \\ N_{j\gamma} \\ N_{jj} \end{pmatrix} \tag{63}$$

where the indices for the ϵ and f coefficient indicates if they are computed for the leading or subleading photon. The vector on the left-hand side is built from events

with a W boson and two photons where the photons are both isolated (TT), either the leading or the subleading one are isolated while the other is not (TL and LT) or both of them are not isolated (LL). These numbers of events can be obtained straightforward from the data sample by dividing them in the four possible categories. The events coming from the electron-photon misidentification background are subtracted in this phase since the estimation of the background has been done separately. The number of events in the right-hand-side vector shows the composition of the data in terms of true photons and misidentified ones: $N_{\gamma\gamma}$ are the events where both the candidates are true photons, $N_{\gamma j}$ and $N_{j\gamma}$ are those events where one of the photons comes from a jet and N_{jj} are those events where both the photons come from a jet. By solving the system, it is possible to obtain the number of background events in the signal region as

$$N_{TT}^{j\to\gamma} = \epsilon_1 f_2 \cdot N_{\gamma j} + f_1 \epsilon_2 \cdot N_{j\gamma} + f_1 f_2 \cdot N_{jj} \ . \tag{64}$$

The method does estimate both the shape and the normalisation of the jet-photon misidentification background. For a certain distribution, in fact, the events in a single bin are categorised as a function of their photons momenta and pseudorapidities. The matrix is built by using the corresponding ϵ and f in terms of (p_T, η) and then inverted to estimate the contribution of the background. This procedure is repeated for all the events that populate one bin of the distribution of interest and the resulting background contribution for that interval is the sum of all of them. By repeating the procedure for all the bins of the distribution, the distribution of the jet-photon misidentification background is found.

In principle, the ϵ and f coefficients can be estimated both in the electron or muon W decay channel. However, since the probability for a photon to be misidentified does not depend on the decay channel of the boson and since the muon channel is cleaner (the signal to noise ratio is higher and there is almost no contribution from the electron-photon misidentification), the estimation of this background has been always performed by using the matrix computed from the muon channel.

5.5 SYSTEMATIC UNCERTAINTIES

The measurement of the cross section is affected by three different kind of uncertainties: a statistical uncertainty coming from the limited amount of data and Monte Carlo simulated events, a theoretical uncertainty that is linked to the modelling of the process and a series of systematic uncertainties. The systematic uncertainties are connected to the method that has been implemented to estimate a particular contribution. For the scale factors, for example, the fitting procedure uses a certain analytical function to fit the distributions, but another reasonable one could have been implemented. The fact that one choice or the other is made changes the final result. For this reason, it is necessary that for each applied correction an uncertainty is associated to the chosen method. In the following, the systematics evaluated in this analysis are described in detail.

5.5.1 *Global quantities*

Luminosity

At CMS, the luminosity is measured in few different ways, both online and offline. The offline measurements exploit the occupancy of the pixel detector or make use of a van der Meer scan technique [142]. The measurement can also be performed online by using the forward hadronic calorimeters. The value of the luminosity that is used to normalise the Monte Carlo simulations (in order to compare them to the collected data) comes with an uncertainty that is linked to these measurement techniques. For the full Run 2 measurement, the overall luminosity is 136.8 (137.2) fb^{-1} for the electron (muon) channel with an uncertainty of 1.8%. The uncertainty on the luminosity affects only the normalisation of the samples and hence is applied after the whole event selection and data-simulation corrections.

It must be stressed that the uncertainty on the luminosity has an effect not only on the final overall distributions, but also on the estimate of the jet-photon misidentification background. The f coefficients are obtained from distributions in data after the subtraction of the true photon contributions from Monte Carlo samples. For this reason, a different value of the luminosity corresponds to a different normalisation of the Monte Carlo distributions and to a different computed value for the f coefficients.

Pileup

As described before, the pileup distribution is obtained by using a statistical approach that exploits the average number of primary vertices and the measurement of luminosity. The pileup events are added to the nominal hard scattering process with the Pre-Mixing technique and the minimum bias events are generated according to the value of the measured minimum bias cross section. To take into account the uncertainty on its value, the full analysis has been run again while considering the minimum bias cross section augmented or diminished by its estimated uncertainty (4.6%). Since the method used to measure the minimum bias cross section and its uncertainty is the same for all the three years of data taking and is independent of the process under examination (electron or muon W decay channel), the uncertainty is considered as correlated between years and between decay channels.

5.5.2 *Scale factor uncertainties*

As previously described, electrons, muons and photons are corrected for a set of scale factors. Since the procedure uses both data and simulations, the usage of a different Monte Carlo simulation gives different results for the value of the scale factors. Moreover, the procedure can assume a certain analytical shape to describe the distributions, that can be chosen in a different way. All these effects introduce an uncertainty for each value of the scale factors as a function of the transverse momentum and pseudorapidity of the particle. Since the value of the scale factor changes the weight of the single event, the analysis is performed again for each of the uncorrelated scale factors that have been applied.

While some scale factors are naturally uncorrelated between different years or different channels, some other are not. The reconstruction and identification efficiencies for electrons and photons are treated as fully correlated among different years, since the method to estimate their contribution is the same. The pixel seed veto efficiency, on the other hand, is treated as uncorrelated between the three years of data taking, since the pixel detector response and the running conditions differ from one year to the other. The electron trigger scale factor uncertainties are negligible if compared to the ones just described. The muon identification, isolation and trigger efficiencies are treated as fully correlated among the three years since the method used to compute them and the detector performances are comparable.

5.5.3 *Monte Carlo statistics*

Even though it is not properly a systematic uncertainty, the statistics of the Monte Carlo samples has a direct impact on the uncertainty of the final measurement. If the number of generated events that pass the selection is small, the uncertainty connected to the estimation of the background will be large. The statistical uncertainty on the simulations is taken into account directly from the main flow of the analysis. As it will be discussed later, this uncertainty has a particularly large contribution for the electronic W$\gamma\gamma$ channel. This is linked to the fact that the Zγ simulation is exploited to evaluate the electron-photon misidentification background, which is one of the main backgrounds for the electron channel, but a small number of events passes the W$\gamma\gamma$ selection. Even though not negligible, this uncertainty does not limit the accuracy of the final measurements.

5.5.4 *Background uncertainties*

The data-driven procedures for the multi-jet, the electron-photon misidentification and the jet-photon misidentification backgrounds follow a precise flow to get an estimate of these background contributions, as described before. Different choices can be made for each of these methods, leading to possibly quite different results. The choice that has been made to identify the best way to estimate the systematic uncertainties is focused on the differences among the whole set of possible variations of the reference methods.

Electron-photon misidentification

The determination of the electron-photon misidentification background relies heavily on the fitting procedure for the fake-Z peak described in Section 5.4.2. The usage of the template extracted from the simulation is a reasonable choice to fit the peak of the distribution, but it is of course not the only one. To estimate the systematic

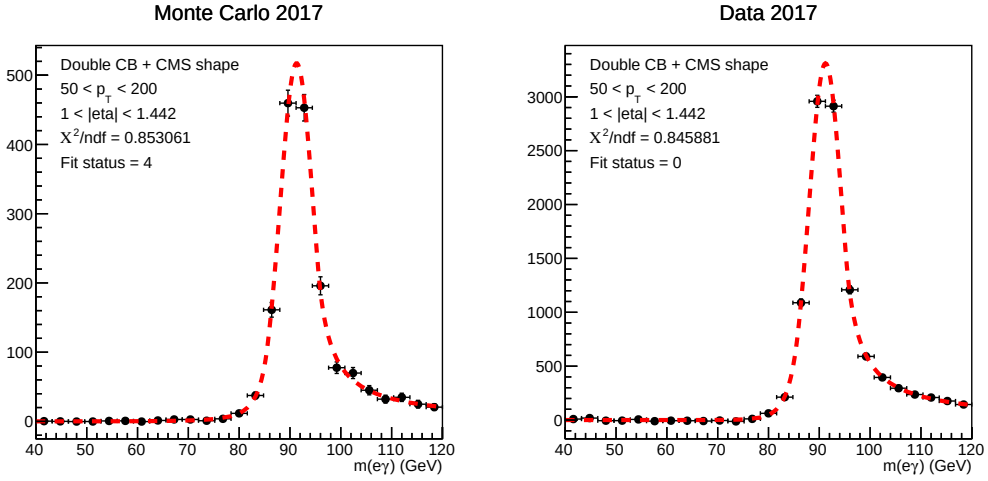


Figure 27: example of fitted fake-Z mass distributions in Monte Carlo (left) and data (right) for the $1 < \eta < 1.442$ and $50 < p_T < 200$ region for the 2017 data. The peak is fitted here with the double-sided Crystal-Ball function, while the background contribution has already been estimated with the standard procedure and subtracted from the data distributions.

uncertainty, the peak has been fitted with a double sided Crystal-Ball function, whose analytical expression is

$$f(x) = \begin{cases} \left(\frac{n_1}{|\alpha_1|}\right)^{n_1} \cdot e^{-\alpha_1^2/2} \cdot \left(\left(\frac{n_1}{|\alpha_1|} - |\alpha_1|\right) - \frac{x-\bar{x}}{\Gamma}\right)^{-n_1} & \text{for} \quad \frac{x-\bar{x}}{\Gamma} < -\alpha_1 \\ e^{-(x-\bar{x})^2/2\Gamma^2} & \text{for} \quad \frac{x-\bar{x}}{\Gamma} > -\alpha_1 \wedge \frac{x-\bar{x}}{\Gamma} < \alpha_2 \\ \left(\frac{n_2}{|\alpha_2|}\right)^{n_2} \cdot e^{-\alpha_2^2/2} \cdot \left(\left(\frac{n_2}{|\alpha_2|} - |\alpha_2|\right) + \frac{x-\bar{x}}{\Gamma}\right)^{-n_2} & \text{for} \quad \frac{x-\bar{x}}{\Gamma} > \alpha_2 \end{cases} \tag{65}$$

Its expression is particularly suitable to describe this distributions since it allows for an asymmetric distribution both in the shape of the peak region and in the shape of the tails. The background contribution has been estimated with the same function of Equation 59 with the parameters fixed to those obtained for the nominal fit. An example of the fitting procedure with the double Crystal-Ball function is presented in Figure 27. The standard procedure to evaluate the correction factors is performed in the same way as for the reference method. In this way, a new set of scale factors is obtained. The central value of the scale factor is taken as the one obtained from the standard procedure, while its error is taken as half the difference between the standard procedure and the one just described. As for the other scale factors, the analysis workflow is re-run for the upward and downward variations of the correction factors. The electron-photon misidentification scale factors are computed not only for the reference method, but also for the pileup and photon scale factors variations.

Jet-photon misidentification - Choice of the control region

The jet-photon misidentification background estimation relies on the determination of the ϵ and f coefficients in the single-photon control region. The description of the jet misidentification is in principle independent of the hard scattering event that is being taken into account. For this reason, the choice of the region where to study the isolation properties of the photon should not influence the values of the ϵ and f coefficients. To estimate the systematic uncertainty for this background, the same procedure as the standard one is applied for events in the multi-jet control region. As previously described, these are events where the leading lepton is not isolated, while the selection for the photon is the same as the nominal one. The systematic uncertainty is computed as half the difference between the results obtained with the standard procedure and with the one that uses the multi-jet control region.

Jet-photon misidentification - Shape effects

Other issues that must be addressed while considering the jet-photon background estimation come from the contribution of the Monte Carlo backgrounds that are subtracted from the distribution of the data in the method. Since the method relies on the count of the number of isolated and non-isolated photons, a different shape of this distribution can change the values of the ϵ and f coefficients. Moreover, the ϵ and f coefficients are estimated as a function of the photon transverse momentum and pseudorapidity: a difference in the spectra of these quantities can change the final result. To account for these possible effects, a different Monte Carlo simulation has been used for the $W\gamma$ and $Z\gamma$ backgrounds, that are the ones that contribute the most in the evaluation of the matrix coefficients. Instead of the nominal $Z\gamma$ MadGraph5_aMC@NLO simulation, a sample generated with Sherpa has been used. In order to avoid effects due to the relative normalisation of the two samples (already taken care by the uncertainty on the sample cross section described in the following), the Sherpa sample has been normalised to the MadGraph5_aMC@NLO one. In this way, only possible differences in the spectra of the particles are considered. The same procedure has been applied for the $W\gamma$ sample but, since a Sherpa simulation was not available, the $Z\gamma$ one has been used in its place, given the fact that the photon distributions are not expected to be strongly dependent on the associated boson and that normalisation effects are not considered at this stage. The systematic uncertainty due to the shape of the Monte Carlo distributions is computed as half the difference between the results obtained with the standard procedure and with the one just described.

$W\gamma$ and $Z\gamma$ cross sections

Each Monte Carlo sample needs to be normalised to its cross section and to the luminosity in order to be compared with data. The uncertainty on the cross section values enters then directly in the final result, where the contribution of the $W\gamma\gamma$ signal is evaluated from the background-subtracted data. For this work, the NNLO MATRIX [140] computation of the cross section has been used (instead of the one

at NLO precision of the MadGraph5_aMC@NLO sample) for the two background processes that contribute the most both to the estimation of the jet-photon misidentification background and to the final cross section measurement, which are the $W\gamma$ and $Z\gamma$ simulations.

The computed value of the cross section section, however, can still show a discrepancy between data and simulation. This fact is exploited to evaluate the systematic uncertainty correlated to the cross section value: the $Z\gamma$ distribution is fitted to the data in a $Z\gamma$ signal region and the discrepancy between the fitted value and the one computed by MATRIX is taken as the systematic uncertainty. The $Z\gamma$ signal region is defined by using the same object selection as the one presented in Section 5.2. An event is tagged as a Z if two opposite side same flavours leptons are identified. The leading electron (muon) is required to have a transverse momentum of at least 35(30) GeV. The invariant mass of the Z boson is required to be $M_Z > 55$ GeV. The photon selection and hence the classification of an event as a $Z\gamma$ one remains the same. The $Z\gamma$ simulated sample is fitted to the data in the Z boson invariant mass distribution. The result of the fit shows that the cross section of the Monte Carlo sample should be reduced by 5% in order to match the description of the data.

The same procedure can be applied for the $W\gamma$ sample, where the transverse mass can be fitted to the data. The result of this procedure shows a 4% discrepancy between the predicted cross section and the fitted one. The application of this method to the $W\gamma$ simulation, however, can be biassed by the fact that the procedure is affected also by the fit result of the multi-jet background that can compensate the effect of the $W\gamma$ cross section. Given the fact that the result of the $Z\gamma$ fit is more reliable and that the result of the $W\gamma$ fit is not too different, the same 5% uncertainty obtained for the $Z\gamma$ simulation is applied for the $W\gamma$ one.

The other background contributions are also affected by the uncertainty on their cross section values, however, since their contributions are smaller, the effect is not studied in detail and the uncertainty is assumed to be three times larger than the one on the $Z\gamma$ sample, i.e. 15%.

The systematic uncertainties on the cross section values are computed as half the difference between the results obtained with the processing that uses the reference value of the cross section or the one varied as just described (which means $\pm 2.5\%$ for the $W\gamma$ and $Z\gamma$ samples and $\pm 7.5\%$ for the other samples). Since the computation of the cross sections is independent between the different processes, the uncertainties on the cross section values are treated as uncorrelated.

5.6 CROSS SECTIONS MEASUREMENT

The measurement of the production cross section for the $W\gamma\gamma$ process is one of the main results of this work. The measurement is performed in a fiducial phase space both for the electron and muon decay channels. The two results are then combined in order to get a comprehensive overall measurement. Besides the value of the cross section itself, also the signal strength and the significance for the observation of the process are measured.

5.6.1 *Blind analysis*

Before digging in the description of the final results, it is worth spending some words on the so-called "blinding" procedure that has been applied in the analysis. As underlined in the previous section, the choice of how a systematic uncertainty is computed has a certain degree of arbitrariness. For example, the choice of the selection requirements for a control region can be different or the fitting function can have a different analytical shape. The method among all the possible ones that is used to perform the study must be chosen by verifying that it is robust, unbiassed and that has a consistent physical interpretation. However, since in the end a choice needs to be made, it is easy to fall into the temptation of using the method that gives the better results instead of the most reasonable one.

In order to avoid this situation, all the analysis workflow has been built in a *blind* way, which means that the data in the signal region were not used while defining all the analysis framework. In this way, there is no bias towards having a "good" final result, but the process is driven only by the need to establish the best way to perform each study. Since a dataset is anyway needed in order to perform the tests and to check the robustness of the framework, what is usually done is to build an Asimov dataset starting from simulation. An Asimov dataset [143] is a fake-data sample that is built from the prediction of the Monte Carlo simulations. The analysis is then performed in this blind environment and "expected" quantities are evaluated. While the blind analysis has a great power in the evaluation of the impact of the systematics (if the real signal is not too different from the Monte Carlo simulation, the expected values of the systematic uncertainties are very close to the measured ones), it has of course no power in measuring the cross sections and signal strength, since they will always be equal to the ones obtained from the Monte Carlo simulation.

5.6.2 *Fiducial phase space*

The measured values that will be extracted for the cross sections refer to a defined set of selection criteria which define the so-called fiducial phase space. The fiducial phase space is defined at generator level and is built in order to be as close as possible to the one at reconstructed level: in this way, the behaviour of data and simulation at reconstructed level will not be too dissimilar from the one at generator level and the extrapolation from one to the other can be performed with minimal dependency on the generator being used.

Electrons and muons at generator level are selected if they have a transverse momentum greater than 15 GeV and are in the $|\eta| < 2.5$ region. The electrons and muons used for this selection are those generated at matrix level and then "dressed" with photons that fall inside a $\Delta R = 0.1$ cone around the particle itself. This procedure is performed in order to get a behaviour as close as possible to the one at reconstructed level, where collinear emissions are reconstructed together with the lepton since it is not possible to experimentally disentangle their signatures. Similarly as it has been done for the main selection and for the overlap removal procedure, photons are selected if their transverse momentum is greater than 15 GeV, if they are in the $|\eta| < 2.5$

region and if there are no other selected leptons and photons in a cone of $\Delta R = 0.4$ and and no other stable particles (neutrinos excluded) in a cone of $\Delta R = 0.1$. The event is classified as a $W\gamma\gamma$ one if exactly one lepton with $p_T > 30\,GeV$ and at least two photons that pass the cuts just described are selected.

The measured fiducial cross section can be compared to the one obtained from simulation in order to determine the signal strength which is a measurement of how much the prediction and the measurement differ. As already mentioned and as it will be discussed in the following section, there are both statistic and systematic uncertainties that affect the measured quantities. The theoretical fiducial cross section (i.e. the cross section obtained from the Monte Carlo simulation after the application of the fiducial phase space generator level cuts) is also affected by some uncertainties that are connected to the underlying model of the simulation.

The uncertainty on the theoretical fiducial cross section comes from the choice of the strong coupling constant, renormalisation and factorisation scales (see Section 1.5.2) and from the uncertainty on the PDF set (see Section 1.5). The impact of the variation of the renormalisation and factorisation scales is computed by doubling or halving their values in all the possible combinations of up, down and central esteems. The impact of the strong coupling constant is evaluated by looking at how much the central value for the PDFs changes if α_S is set to 0.116 or 0.120 instead of the reference value of 0.118. To evaluate the uncertainty of the PDF set, 100 variations are computed starting from the central value with the Hessian method [144]. By building the distributions of interest with all these variations, it is possible to get an envelope from which the uncertainty can be estimated.

To perform all these checks, a Rivet [145] routine has been developed [146]. The same selection as the generator level one implemented in the main analysis has been rewritten in the Rivet routine. The resulting distributions are in agreement with the one obtained with the main analysis at the percent level. By exploiting this routine, the aforementioned theoretical modelling uncertainties have been evaluated for the $W\gamma\gamma$ signal sample. The effect of the choice of the renormalisation and the factorisation scales give an uncertainty of 0.55% on the generator-level fiducial cross section, while the uncertainty due to the PDF variations is found to be 0.28%. The uncertainty linked to the variation of the strong coupling constant is smaller and amounts to 0.03%.

From this procedure, the value of the generator-level fiducial cross section is

$$\sigma_{l\nu}^{exp} = 18.88 \pm 0.12\,(\text{pdf} + \text{scale})\,\text{fb} \tag{66}$$

where the uncertainty is only due to the implementation of the theoretical model.

5.6.3 *Results*

The distributions of the transverse momentum of the di-photon system, of the transverse momentum of the leading and subleading photons are presented in Figure 28. The $W\gamma\gamma$ signal contribution is presented in orange, the electron-photon misidentification background in yellow, the jet-photon misidentification background in blue and the other irreducible backgrounds in green. The bottom panels of the plots show

Process	$W^{\pm}\gamma\gamma \to e\nu_e\gamma\gamma$	$W^{\pm}\gamma\gamma \to \mu\nu_\mu\gamma\gamma$
Expected signal	$248.4 \pm 5.6(\text{stat.}) \pm 17.4(\text{syst.})$	$499.6 \pm 8.2(\text{stat.}) \pm 33.4(\text{syst.})$
Jet-photon misid.	$918.1 \pm 23.2(\text{stat.}) \pm 181.8(\text{syst.})$	$1441.3 \pm 26.6(\text{stat.}) \pm 281.9(\text{syst.})$
Electron-photon misid.	$668.9 \pm 27.8(\text{stat.}) \pm 34.0(\text{syst.})$	$106.9 \pm 8.9(\text{stat.}) \pm 7.1(\text{syst.})$
Other irreducible	$216.8 \pm 10.7(\text{stat.}) \pm 20.3(\text{syst.})$	$285.8 \pm 10.8(\text{stat.}) \pm 24.5(\text{syst.})$
Predicted backgrounds	$1803.8 \pm 37.8(\text{stat.}) \pm 179.9(\text{syst.})$	$1834.0 \pm 30.0(\text{stat.}) \pm 279.4(\text{syst.})$
Total prediction	$2052.3 \pm 38.2(\text{stat.}) \pm 181.5(\text{syst.})$	$2333.6 \pm 31.2(\text{stat.}) \pm 279.4(\text{syst.})$
Data	$1987.0 \pm 44.7(\text{stat.})$	$2384.0 \pm 48.9(\text{stat.})$

Table 9: summary of the observed and predicted number of events with the full Run 2 statistics for the $W\gamma\gamma$ selection in the electron and muon channels. The quoted systematic uncertainty on the signal and backgrounds quantify the effect in terms of change of number of events for that particular sample obtained from the quadrature sum of the contribution of all the systematic variations.

the ratio between data and the sum of all the backgrounds and the signal. The error bars on the black dots are the statistical uncertainties of the data and the grey shaded area is the uncertainty on the sum of the predictions. As it can be seen, a general good agreement can be observed in the different distributions. As previously described, the main background contribution is the jet-photon misidentification background. The electron-photon misidentification background has a big impact in the electron channel, while it is negligible for the muon one. It is important to notice that the di-photon transverse momentum spectrum is populated up to high values of the transverse momenta, which will have a key role in the extraction of the limits on the anomalous couplings.

A breakdown of the number of events for the data and all the simulations is presented in Table 9. As it can be seen, the number of expected signal events in the electron channel is half the one for the muon channel with a similar contribution from the background. This observation hints to the fact that the electron channel will have a smaller prediction power if compared to the muon one and that the combination of the two will likely be dominated by the muon channel.

Three quantities that are linked to the measurement of the signal cross section are computed in this analysis: the value of the fiducial cross section itself, the signal strength and the signal significance. The meaning of the fiducial cross section has been described earlier: the results presented in this section should be compared with the theoretical value reported in Equation 66. The information on the agreement between the predicted and measured cross sections can be inferred from the signal strength μ defined as

$$\sigma^{\text{obs}} = \mu \cdot \sigma^{\text{exp}} . \tag{67}$$

The scaling factor μ links the measured cross section with the one predicted at matrix level by the simulation. It has a direct interpretation: if $\mu = 1$ the measured cross section coincides with the theoretical one (i.e. the outcome of the measurement from the data is in perfect agreement with the signal prediction), while if $\mu = 0$ the ob-

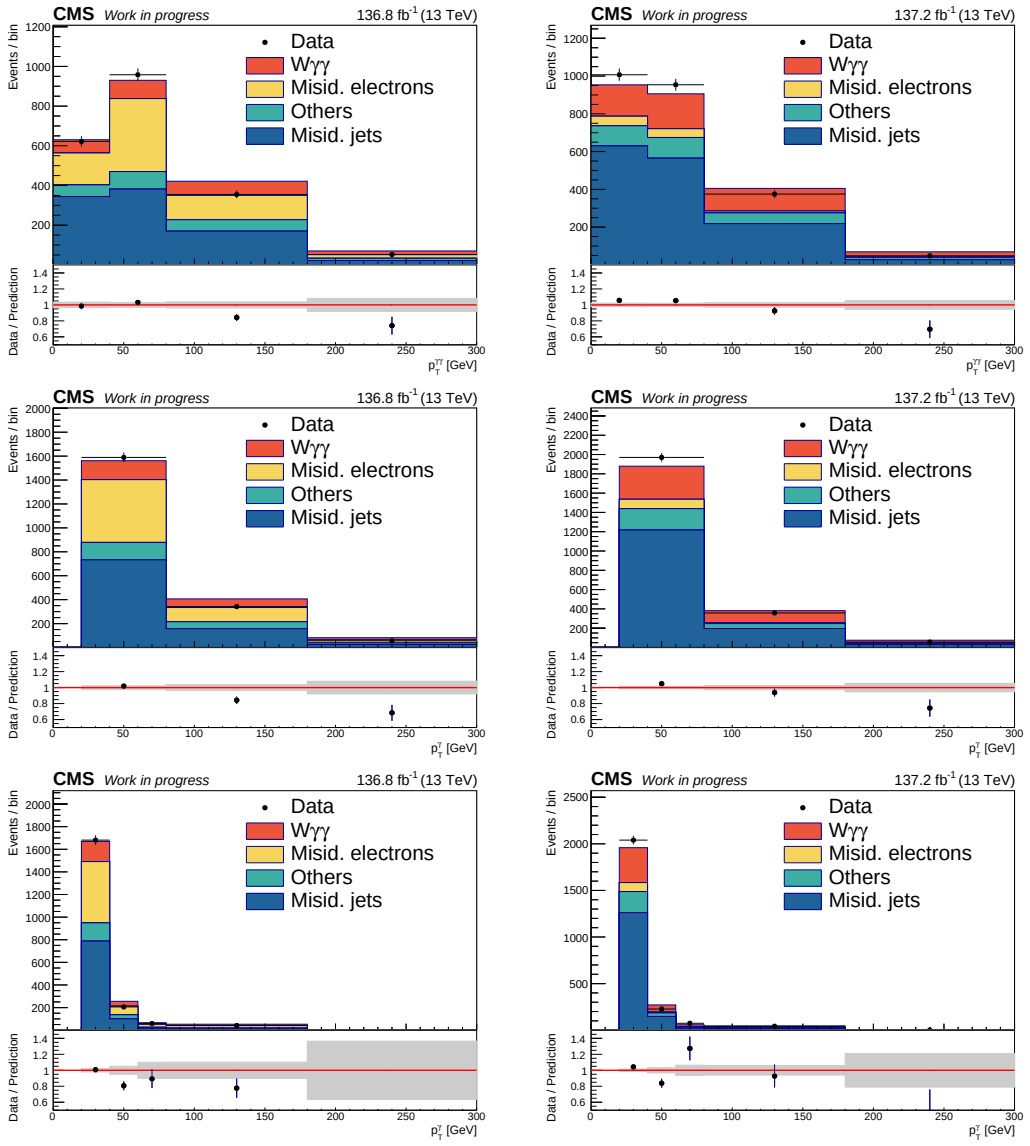


Figure 28: distribution of the transverse momentum of the di-photon system (top), of the leading photon (middle) and of the subleading photon (bottom) for the Wγγ selection in the electron channel (left) and in the muon one (right). Data are presented with black dots, signal and background contributions are stacked and depicted with coloured areas. The ratio between data and prediction is presented in the bottom panel of the plots, where the error bars on the black dots refer to the statistical uncertainty of the data while the shaded grey regions refer to the statistical uncertainty of the simulation.

served process is compatible with the background hypothesis only. If the data follow the Standard Model description provided by the Monte Carlo generator, the result is expected to be compatible with one after taking into account the statistical and systematic uncertainties.

The other quantity that is frequently used in high energy physics measurements is the significance σ. It is a conventional way of indicating the result of an hypothesis test. The null hypothesis is to have a background-only data sample and a p-value is computed to see with what degree of belief this hypothesis is discarded. The p-value, instead of being expressed in absolute terms, is expressed as the integral of a normal distribution in the region $[n\sigma, \infty]$. For example, if the measurement has a significance of 3σ it means that the background only hypothesis is discarded with a confidence level of 99.73%. The convention is to claim the "evidence" of a process if the significance is greater than 3σ and to claim the "observation" of a process if the significance is greater than 5σ.

To compute the values of the signal strengths, cross sections and significances, a maximum likelihood method that takes into account the contribution of all the systematics has been implemented. While performing a study of collisions events, the probability to observe n events while the expected number is ν can be expressed in terms of a Poissonian distribution

$$f(n, \nu) = e^{-\nu} \frac{\nu^n}{n!} . \tag{68}$$

To be more specific, the number of expected events can be rewritten in terms of expected background and expected signal events as $\nu = \nu_{bkg} + \nu_{sig}$. Since the goal of the procedure is to measure the signal strength μ and the cross section, the relationship can be rewritten by keeping in mind that $\nu_{sig} = \mu\sigma_{sig}L$ where σ_{sig} is the measured signal cross section and L is the luminosity of the sample.

This is, however, an oversimplified view of the problem that assumes a perfect knowledge of the expected number of events both in signal and background. As previously discussed, this is not the case since many systematic uncertainties can change the behaviour of the different contributions. The expression in Equation 68 should then be rewritten as

$$f(n, \nu, \vec{\theta}) = e^{-\nu(\vec{\theta})} \frac{\nu(\vec{\theta})^n}{n!} , \tag{69}$$

where $\vec{\theta}$ are the dependencies on the systematics of the model. The effect of the systematic uncertainties can be the same for the whole distribution under examination or it can have a different impact for each bin of the distribution; the latter situation is referred to as a shape uncertainty. All the systematics that have been previously discussed are of this kind (scale factors, for example, being transverse-momentum dependent, can change not only the normalisation of the distribution but also its shape). It could be argued that luminosity is a normalisation systematic but, since it affects directly the way the jet-photon misidentification matrix is built, it can lead to the estimation of a different background contribution and hence affect the overall shape of the backgrounds. The probability density function for the systematic uncertainties that have a shape effect is assumed to be a Gaussian.

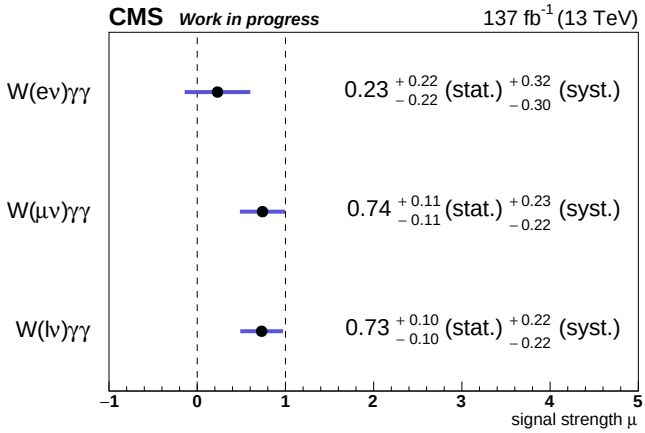


Figure 29: observed signal strengths with relative statistical and systematic uncertainty for the electron, muon and combined channels.

A likelihood function that takes into account the expected number of events in each bin of the distribution as a function of the systematic uncertainties and that is dependent on the probability density function of the systematic uncertainties themselves can be written as

$$\mathcal{L}(n, \nu, \vec{\theta}) = \prod_{i=0}^{N_{bins}} e^{-\nu_i(\vec{\theta}_i)} \frac{\nu_i(\vec{\theta}_i)_i^n}{n_i!} \times e^{-\vec{\theta}_i^2/2}, \tag{70}$$

where the index i refers to the bins of the distribution. The minimisation of the likelihood with respect to the value of μ (that enters the estimate of ν as previously mentioned) by considering the systematic uncertainties as nuisance parameters gives the best estimate of the signal strength.

By applying this statistical procedure, the values of the fiducial cross section and of the signal strength can be computed. The measured $W\gamma\gamma$ fiducial cross sections for the electron, muon and combined leptonic channels are

$$\begin{aligned} \sigma_{e\nu} &= 4.39^{+4.11}_{-4.11}(\text{stat.})^{+6.05}_{-5.53}(\text{syst.}) \pm 0.12(\text{pdf} + \text{scale})\ \text{fb} \\ \sigma_{\mu\nu} &= 13.99^{+2.15}_{-2.12}(\text{stat.})^{+4.35}_{-4.23}(\text{syst.}) \pm 0.11(\text{pdf} + \text{scale})\ \text{fb} \\ \sigma_{l\nu} &= 13.76^{+1.94}_{-1.91}(\text{stat.})^{+4.08}_{-4.06}(\text{syst.}) \pm 0.08(\text{pdf} + \text{scale})\ \text{fb} \end{aligned} \tag{71}$$

which correspond to measured signal strengths of

$$\begin{aligned} \mu_{e\nu} &= 0.23^{+0.22}_{-0.22}(\text{stat.})^{+0.32}_{-0.29}(\text{syst.}) \\ \mu_{\mu\nu} &= 0.74^{+0.11}_{-0.11}(\text{stat.})^{+0.23}_{-0.22}(\text{syst.}) \\ \mu_{l\nu} &= 0.73^{+0.10}_{-0.10}(\text{stat.})^{+0.22}_{-0.21}(\text{syst.}) \end{aligned} \tag{72}$$

The details of the impact of the different systematic uncertainties on these measurements are summarised in Table 10. The values of the computed signal strengths are presented also in Figure 29.

By looking at the values of the signal strength it can be seen that the measurement for the electron channel does not provide much information on the properties of the

Systematic source	$W(e\nu)\gamma\gamma$ (%)	$W(\mu\nu)\gamma\gamma$ (%)	$W(l\nu)\gamma\gamma$ (%)
Luminosity	< 1	2	2
Pile-up	2	< 1	< 1
Electron ID	4	< 1	< 1
Electron reconstruction	1	< 1	< 1
Electron trigger	1	< 1	< 1
Muon ID	1	< 1	< 1
Muon isolation	< 1	< 1	< 1
Muon trigger	< 1	< 1	< 1
Photon ID	3	2	2
Photon pixel seed veto	18	13	12
L1 ECAL trigger prefiring	< 1	< 1	< 1
Jet-photon misid.	21	19	17
Jet-photon misid. shape	14	12	12
Electron-photon misid.	4	< 1	< 1
$W\gamma$ theoretical cross section	3	3	3
$Z\gamma$ theoretical cross section	4	< 1	< 1
"Others" theoretical cross section	5	2	2
MC samples statistical uncertainty	18	7	8

Table 10: impact of each systematic uncertainty as estimated by the likelihood fit of the signal strength to the measured data. The results of the electron, muon and combined channel are presented.

process. The measured value is in fact compatible with zero within one sigma and with one within two sigmas, which means that it is not possible to exclude neither the only-background hypothesis, nor the Standard Model-like behaviour one. The signal strength smaller than one could have been inferred also from the plots in Figure 28 where the data are on average underestimated if compared to the simulation. The systematic uncertainty, as seen from Table 10, is dominated by the jet-photon background estimation and by the photon pixel seed veto scale factors. This comes as no surprise, since the jet-photon misidentification is the dominant background contribution (hence even a small variation causes a big impact on the signal extraction) and the photon scale factor uncertainty is enhanced due to the fact that two photons are selected in each event.

The situation is improved in the muon channel, where the measured signal strength is compatible with the Standard Model prediction within one sigma and is more than 3 sigmas away from the background-only hypothesis. The main systematic contributions are the same as for the electron channel, but, since the signal-to-noise ratio in the muon channel is higher, they have a smaller impact on the final measurement. The combined measurement of the two channels gives results that are similar to those obtained in the muon channel. This is reasonable and expected, since the better signal-to-noise ratio and overall sensitivity make the muon channel the one that definitely dominates in the combination.

From this procedure, the expected and measured significances can be computed. As previously mentioned, the expected significance is built while assuming that the data coincide with the Monte Carlo simulations. The measured (expected) significances in the electron channel is 0.6σ (2.7σ), while in the muon channel is 3.0σ (4.3σ). The combined measured (expected) significance is 3.1σ (4.5σ). For the same reason as discussed for the signal strength, the muon significance dominates the combination.

It doesn't make any difference how beautiful your guess is, it doesn't make any difference how smart you are, who made the guess, or what his name is. If it disagrees with experiment, it's wrong. That's all there is to it.
— Richard Feynman

6 LIMITS ON ANOMALOUS QUARTIC GAUGE COUPLINGS

The last chapter of this work is dedicated to the study of the anomalous gauge couplings. In Section 1.6.1, a theoretical discussion of the EFT has been presented alongside the direct implications for this study. In the following, the statistical procedure used to extract both expected and observed limits on the anomalous couplings is detailed. A comparison between the results obtained in this work, those obtained in the past $W\gamma\gamma$ analyses and those obtained by other CMS and ATLAS analyses performed with Run 2 data is presented.

6.1 EXTRACTION OF THE LIMITS

The AGC sample has been generated accordingly to the description of Section 4.1.1 and contain ten million events at generator level. The generation uses a reweighting procedure that assigns a different weight for each generated event for each value of the scan of the parameter under investigation. This procedure is useful so that with a single production of a sample of events, all the possible variations of the different coefficients are generated at once (considering that 16 parameters have been tested, each with a 16 value scan, 256 different samples should have been otherwise generated). For the first test, the couplings of all the operators have been set to zero and the Standard Model behaviour has been re-obtained.

For the estimation of the limits on the anomalous couplings, the distribution of the transverse momentum of the di-photon system has been considered. As shown in Figure 28, the spectrum is populated up to more than 100 GeV. As mentioned in the theoretical discussion (Section 1.6.1), the effect of the presence of possible new physics is enhanced while going at higher energies or higher transverse momenta. This fact is confirmed by the behaviour observed in Figure 30. The distribution of the di-photon transverse momentum as obtained from simulation with the Standard Model sample or with the AGC sample where only one operator is set to be different from zero is shown. As it can be seen, the two distributions have a compatible behaviour up to 150 GeV, while for higher values of the transverse momentum the AGC sample has a higher number of events if compared to the Standard Model one. This is exactly the expected behaviour that the presence of new physics at still not reachable energies has on the Standard Model measurements.

By keeping in mind the result of Figure 30, the procedure to set limits on the anomalous couplings checks if the behaviour of the observed data is compatible with the Standard Model or not. If it is not compatible, then effects of new physics are being observed. If it is compatible, however, it is possible to exclude the contribution of higher-dimensional operators and set limits on the possible value of their coupling constants.

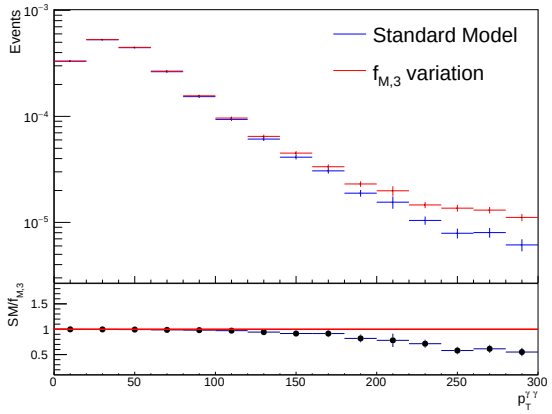


Figure 30: di-photon transverse momentum obtained from Monte Carlo simulation by assuming the Standard Model (blue) or by assuming the contribution of a particular higher-dimensional operator in the Lagrangian expansion (red). In particular, the red distribution has been obtained by imposing $f_{M,3}/\Lambda^4 = 100\,TeV^{-4}$.

As for the measurement of the cross section, a statistical method based on a likelihood scan has been implemented. Differently from the cross section measurement, in this case the parameter of interest is the value of the anomalous coupling. The first step towards the construction of the likelihood, is to define a signal model that depends on the value of the anomalous couplings. This can be done by fitting the ratio of the anomalous coupling yield to the Standard Model one as a function of the value of the anomalous coupling itself. The procedure can be better explained by visualising the result of Figure 31. The values on the abscissa correspond to the strength of the coupling (at zero the Standard Model is recovered), while on the ordinate the ratio between the number of events for a model with a modified coupling and the Standard Model is presented. As expected, the greater (in absolute terms) the value of the anomalous coupling, the greater the discrepancy from the Standard Model prediction. A parabola like the one presented in the figure is computed for each parameter scan and for each bin of the distribution of the di-photon transverse momentum. Since the effect is larger at higher values of the di-photon transverse momentum, the parabola will be "narrower" in the last bins of the distribution, while it will have a "wider" behaviour for the bins corresponding to lower values of the transverse momentum.

After computing the dependency of the signal model on the anomalous couplings, the same likelihood function of Equation 70 can be written. In this situation, however, the signal strength is no more an overall scaling factors, but has a dependency on the aGC parameter. It is possible to define a ratio of the likelihoods as

$$\lambda(\vec{\alpha}) = \frac{\mathcal{L}(\vec{\alpha}, \hat{\hat{\vec{\theta}}})}{\mathcal{L}(\hat{\vec{\alpha}}, \hat{\vec{\theta}})}, \tag{73}$$

where $\vec{\theta}$ are the nuisance parameters and $\vec{\alpha}$ are the anomalous coupling parameters. The effect of the nuisance parameters is described by a log-normal distribution.

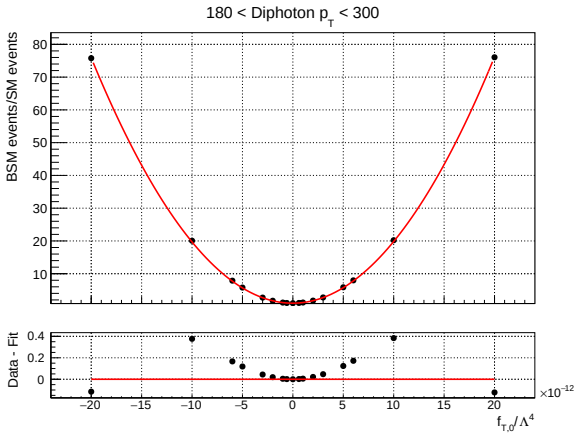


Figure 31: fitted parabola for the scan of the $F_{T,0}$ parameter in the last bin of the distribution of the di-photon transverse momentum. The bottom panel shows the ratio between the fitted distribution and the points.

The quantity $\lambda(\vec{\alpha})$ is the ratio of the likelihood maximised for a certain value of the coupling (with the optimal values of the nuisance parameters) over the maximum likelihood for the anomalous couplings and the nuisances. By defining a test statistic as

$$t_\alpha = -2 \ln \lambda(\vec{\alpha}) \,, \tag{74}$$

it is possible to determine the 68% and 95% confidence intervals for the values of the anomalous couplings. The distribution of t_α can be approximated with a χ^2 distribution by exploiting Wilks' theorem [147].

The W$\gamma\gamma$ process is subject to the possible contribution of up to 14 different dimension-8 operators, as described in Section 1.6.1. The experimental sensitivity, however, can limit the number of couplings that can be constrained by the measurement. As for the cross section measurement, also the computation of the limits on the couplings has been performed first in a blind configuration (only Monte Carlo simulations have been used) and then using the measured data. The results for the expected and observed limits on the anomalous couplings are presented in Table 11 for the electron and muon channel separately and in Table 12 for the combination of the two channels. As it can be seen, all the parameters are compatible with zero which means that no new physics effect is observed. As for the cross section measurement, the muon channel gives more precise predictions than the electron one. This is linked to the smaller impact of the systematics. For the same reason, the combination is dominated by the results obtained in the muon channel. The fact that the observed limits are slightly better than the expected ones can be linked to the behaviour of the distributions presented in Figure 28. As can be seen from the di-photon plots, data are lower than prediction for the last bin of the distribution. The presence of an excess of events in this region (connected to the presence of a non-zero anomalous coupling) is then more disfavoured in data than in prediction, hence providing a tighter constraint on the limits on the anomalous couplings.

Figure 32 and 33 show the comparison of the limits extracted in this work and in the previous W$\gamma\gamma$ analysis performed by CMS at 8 TeV [3]. The new results show an

	Electron decay channel		Muon decay channel	
	Expected (TeV^{-4})	Observed (TeV^{-4})	Expected (TeV^{-4})	Observed (TeV^{-4})
$f_{M,2}/\Lambda^4$	[−78.9, 78.5]	[−60.3, 59.1]	[−61.1, 61.1]	[−39.9, 39.7]
$f_{M,3}/\Lambda^4$	[−126, 125]	[−95.4, 96.2]	[−98.2, 99.4]	[−63.8, 65.4]
$f_{T,0}/\Lambda^4$	[−2.54, 2.54]	[−1.94, 1.94]	[−1.98, 1.98]	[−1.30, 1.30]
$f_{T,1}/\Lambda^4$	[−3.30, 3.26]	[−2.54, 2.50]	[−2.58, 2.54]	[−1.70, 1.66]
$f_{T,2}/\Lambda^4$	[−7.08, 7.08]	[−5.40, 5.40]	[−5.48, 5.56]	[−3.64, 3.64]
$f_{T,5}/\Lambda^4$	[−1.08, 1.16]	[−0.84, 0.92]	[−0.84, 0.92]	[−0.52, 0.60]
$f_{T,6}/\Lambda^4$	[−1.24, 1.32]	[−1.00, 1.00]	[−1.00, 1.08]	[−0.60, 0.68]
$f_{T,7}/\Lambda^4$	[−2.28, 2.36]	[−1.80, 1.80]	[−1.80, 1.80]	[−1.16, 1.16]

Table 11: expected and observed limits for the anomalous couplings in the electron and muon channels.

	Expected (TeV^{-4})	Observed (TeV^{-4})
$f_{M,2}/\Lambda^4$	[−57.3, 57.1]	[−39.9, 39.5]
$f_{M,3}/\Lambda^4$	[−91.8, 92.6]	[−63.8, 65.0]
$f_{T,0}/\Lambda^4$	[−1.86, 1.86]	[−1.30, 1.30]
$f_{T,1}/\Lambda^4$	[−2.38, 2.38]	[−1.70, 1.66]
$f_{T,2}/\Lambda^4$	[−5.16, 5.16]	[−3.64, 3.64]
$f_{T,5}/\Lambda^4$	[−0.76, 0.84]	[−0.52, 0.60]
$f_{T,6}/\Lambda^4$	[−0.92, 1.00]	[−0.60, 0.68]
$f_{T,7}/\Lambda^4$	[−1.64, 1.72]	[−1.16, 1.16]

Table 12: expected and observed limits for the anomalous couplings obtained by combining the electron and muon channels.

improvement of almost an order of magnitude if compared to the previous one. A comparison between this analysis and the results obtained by the $Z\gamma$ [148], WZ [149], WWW [150], $W\gamma$ [151], same sign WW [152], ZZ [153] and WV/ZV [154] performed by CMS with data collected at 13 TeV and the $Z\gamma$ [155] analysis by ATLAS with data collected at 8 TeV is shown in Figure 34. As it can be seen, for the T0, T1 and T2 operators the limits extracted in this work are comparable with some of the other analyses, such as the WWW, WZ and $Z\gamma$ ones, while they are generally less stringent than the other works. The results for the T5, T6 and T8 operators are competitive with the overall best results available.

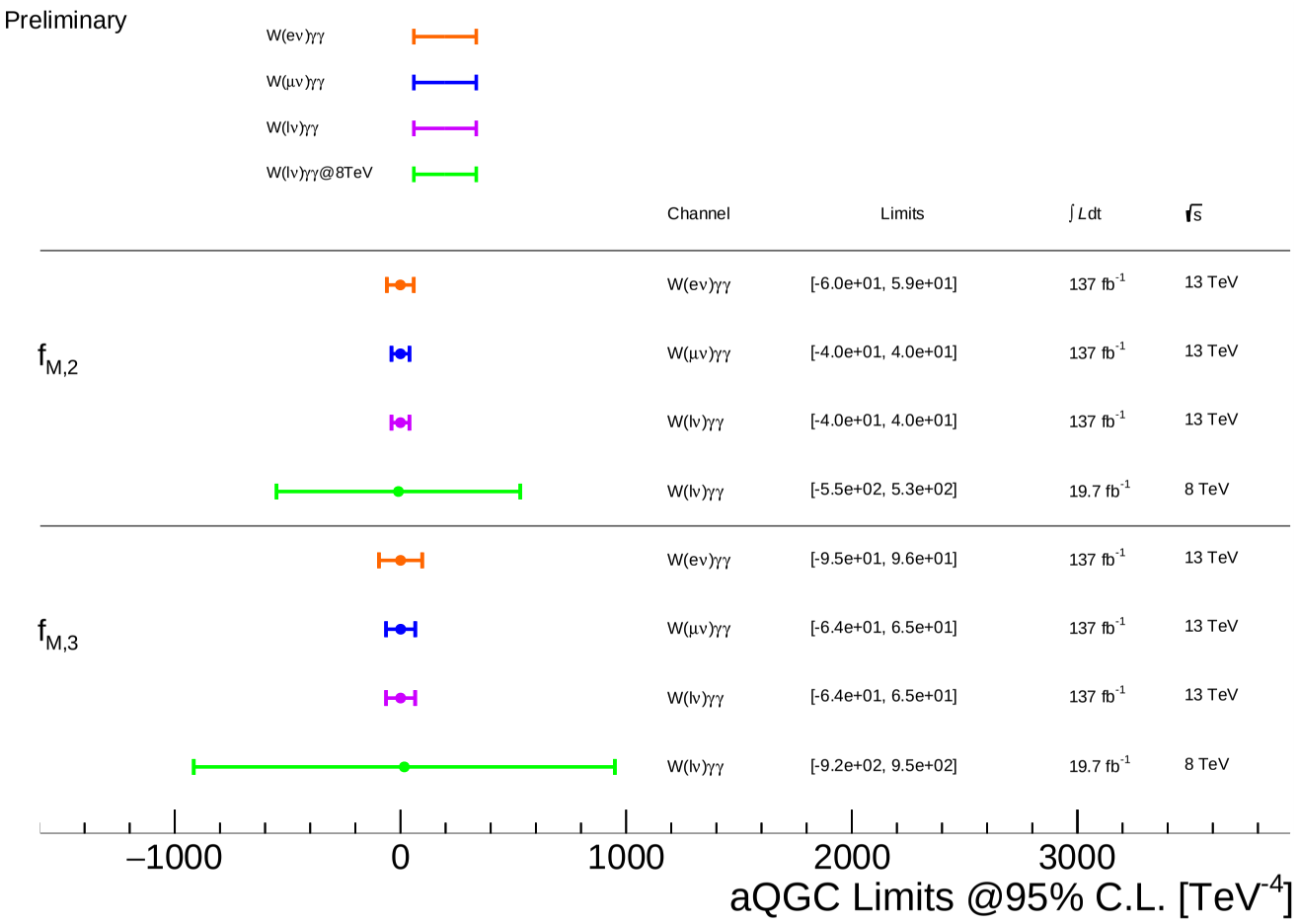


Figure 32: limits on the anomalous gauge couplings obtained by this analysis compared to the previous analysis performed by CMS at 8 TeV. The three channels where the limits have been extracted are presented separately: the electron channel in orange, the muon channel in blue and the combined channel in purple. The values obtained by the previous analysis are presented in green.

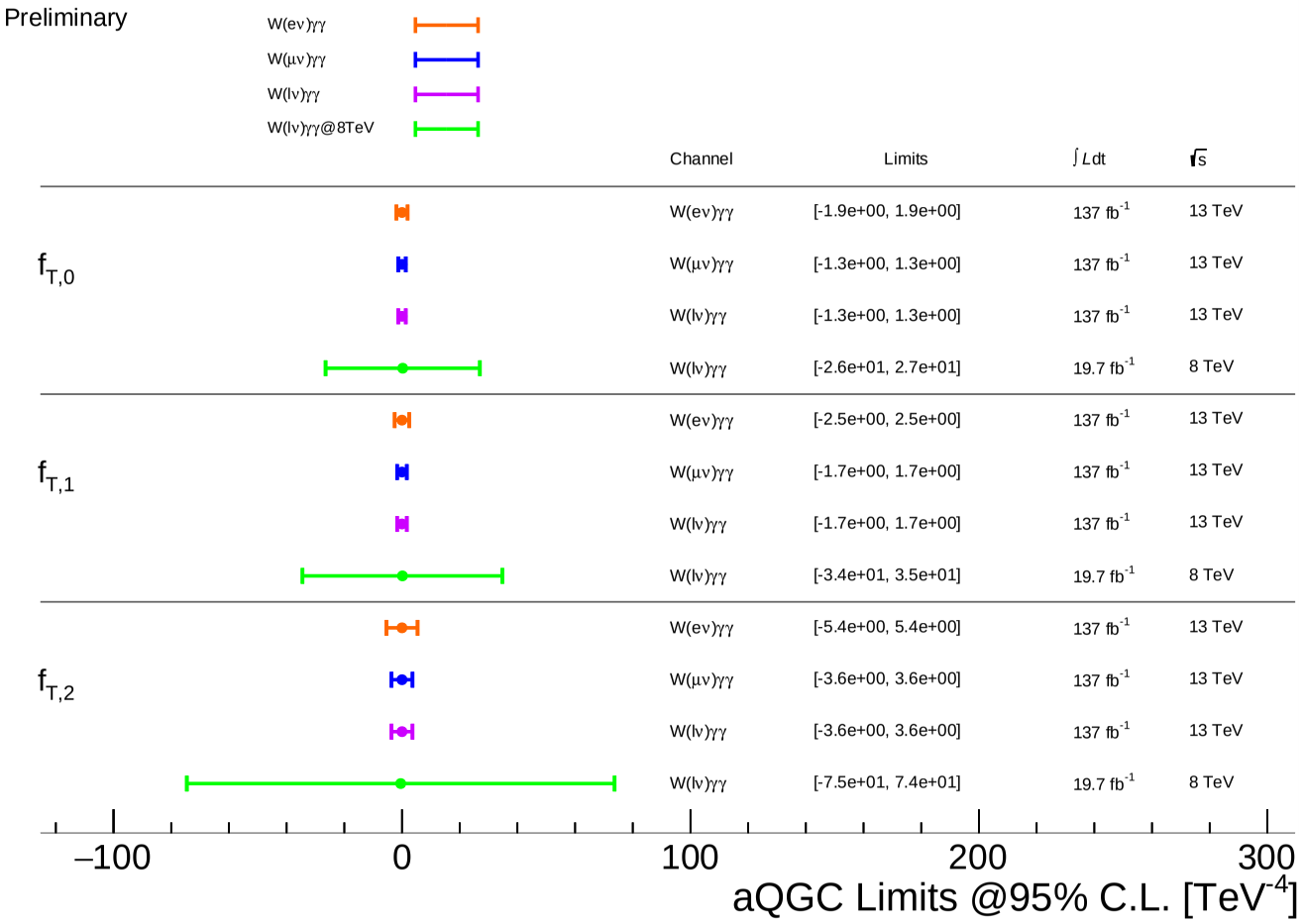


Figure 33: limits on the anomalous gauge couplings obtained by this analysis compared to the previous analysis performed by CMS at 8 TeV. The three channels where the limits have been extracted are presented separately: the electron channel in orange, the muon channel in blue and the combined channel in purple. The values obtained by the previous analysis are presented in green.

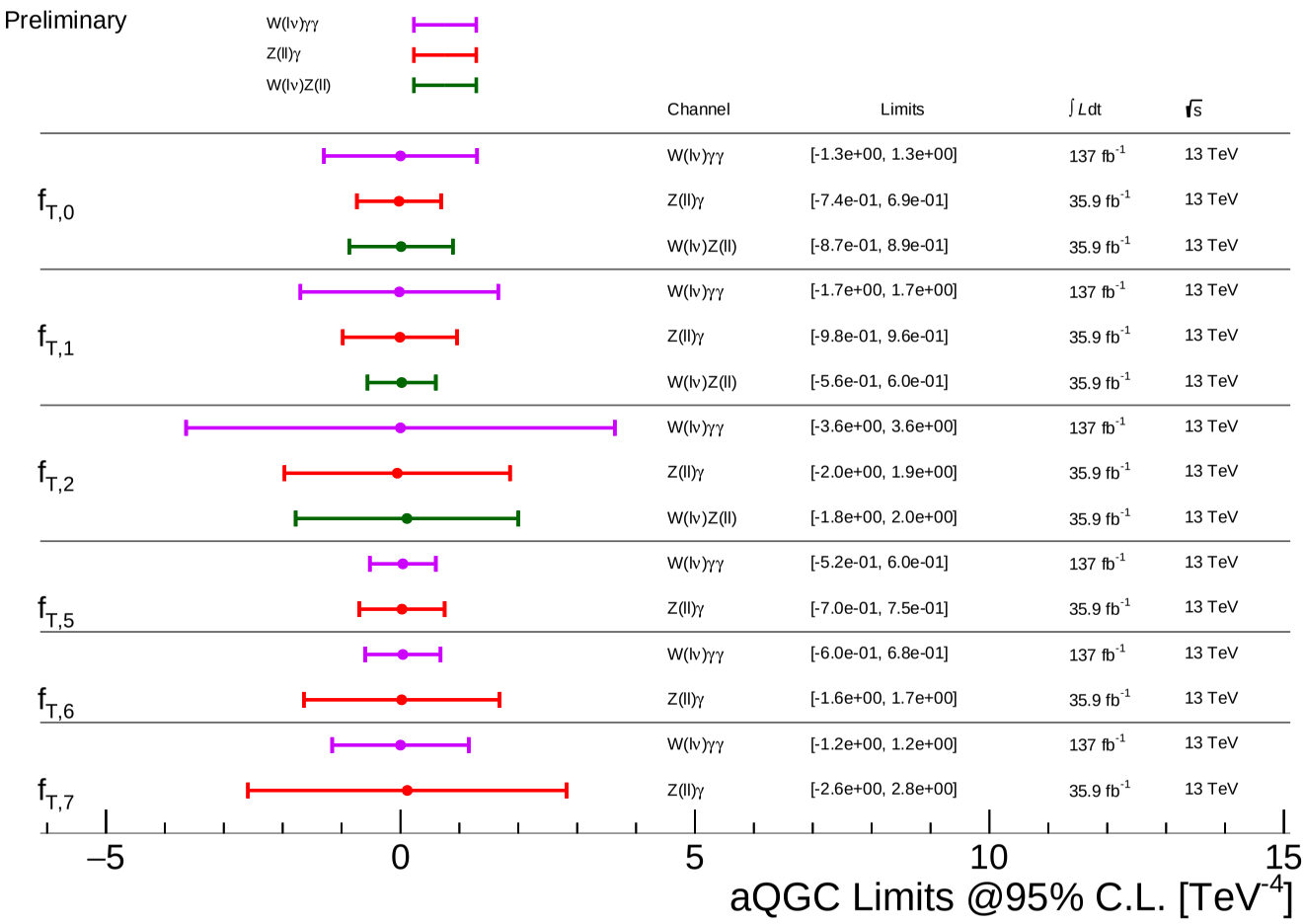


Figure 34: limits on the anomalous gauge couplings obtained by this analysis compared to those obtained by other CMS (green) and ATLAS (azure) analyses both with data collected at 8 and 13 TeV. Only the combined electron and muon channel is shown for this work in purple.

CONCLUSIONS

The $W\gamma\gamma$ production has been studied with data collected by the CMS experiment at a centre of momentum energy of 13 TeV corresponding to an integrated luminosity of $137\,fb^{-1}$. The event selection has been studied and optimised, while residual discrepancies between data and simulation have been corrected. A thorough study of the background contributions has been performed with particular focus on the contribution of electrons misidentified as photons and jets misidentified as photons. For these two contributions, a data driven approach has been developed and validated.

The systematic uncertainties that affect the final results have been studied and their overall impact on the measurement has been evaluated with a maximum likelihood technique. The cross sections and the signal strengths have been evaluated in the fiducial phase space in the electron and muon decay channels of the W boson. The combination of the two has also been computed. The measured signal strengths are equal to $0.23^{+0.22}_{-0.22}(\text{stat.})^{+0.32}_{-0.29}(\text{syst.})$ in the electron channel, $0.74^{+0.11}_{-0.11}(\text{stat.})^{+0.23}_{-0.22}(\text{syst.})$ in the muon channel and $0.73^{+0.10}_{-0.10}(\text{stat.})^{+0.22}_{-0.21}(\text{syst.})$ for the combination of the two channels. While the result of the electron channel is compatible both with the hypothesis of no signal and the hypothesis of a Standard Model-like signal, in the muon channel the hypothesis of no signal is strongly excluded. The combination is driven by the muon channel. These measured signal strengths correspond to observed (expected) significances respectively of 0.6σ (2.7σ) for the electron channel, 3.0σ (4.3σ) for the muon channel and 3.1σ (4.5σ) for the combination of the two. This is the first evidence for the $pp \to W\gamma\gamma$ process obtained by the CMS experiment.

Exclusion limits on the presence of anomalous quartic gauge couplings have been computed in the effective field theory framework. In particular, limits have been placed on the M2, M3, T0, T1, T2, T5, T6 and T7 operators. None of these measurements suggest a deviation from the Standard Model prediction. The observed limits are competitive with the ones obtained by other multi-boson analyses both at CMS and ATLAS, in particular for the T5, T6 and T7 operators.

Being the $W\gamma\gamma$ production a rare process (the cross section in the fiducial phase space is of the order of $20\,fb$), the collected amount of data is a limiting factor in the final measurement. The limited statistics does not only affect the number of signal events after the selection (corresponding to a few hundreds in each channel), but also the evaluation of the background contributions and the impact of the systematic uncertainties. The contribution of jets misidentified as photons, for example, is evaluated by exploiting control regions where the number of events is comparable to or even smaller than the signal region. The precision with which the contribution of this background is estimated is then statistically limited and so is the evaluation of its systematic uncertainties.

The data collected during the LHC Run 3 and most importantly during the High Luminosity LHC program are expected to be a factor ten more than those collected

during the Run 2. By exploiting the higher statistics, the future developments in the reconstruction algorithms and the novel machine-learning based optimisation for the analyses, a higher precision can be achieved both on the determination of the cross section and on the extraction of the limits on the anomalous couplings.

A

ELECTRON-PHOTON MISIDENTIFICATION SCALE FACTORS AND FITS

A.1 ELECTRON-PHOTON MISIDENTIFICATION FAKE FACTORS

The differential scale factors, computed as a function of p_T and η for the three years, used in this work are shown in Figure 35. For a detailed description of the scale factor determination see Section 5.4.2.

In order to decouple the effect of the pileup systematics from the uncertainty due to the fake factors computation, the electron photon misidentification fake factors are computed ad-hoc for the (up and down) variations of the pileup correction. The ratio of the fake factors obtained while varying the pileup with respect to the reference values is presented in Figure 36 for the up variation of the pileup and in Figure 37 for the down variation.

A.2 ELECTRON-PHOTON MISIDENTIFICATION FITS

After the plots for the value of the fake factors, the fits of the spurious Z mass reconstructed with an electron and a photon are presented. For a detailed description of the scale factor determination see Section 5.4.2.

The fits are performed with a template taken from Monte Carlo for the signal and a CMS shape for the background. The phase space is divided in bins of transverse momentum and pseudrapidity for both data and Monte Carlo and each of these distributions is fitted. The results for all the three years are presented. Some plots have a "NULL" χ^2 and this is due to the fact that some bins in the distribution have zero events with associated bin error of zero, which of course breaks the χ^2 calculation.

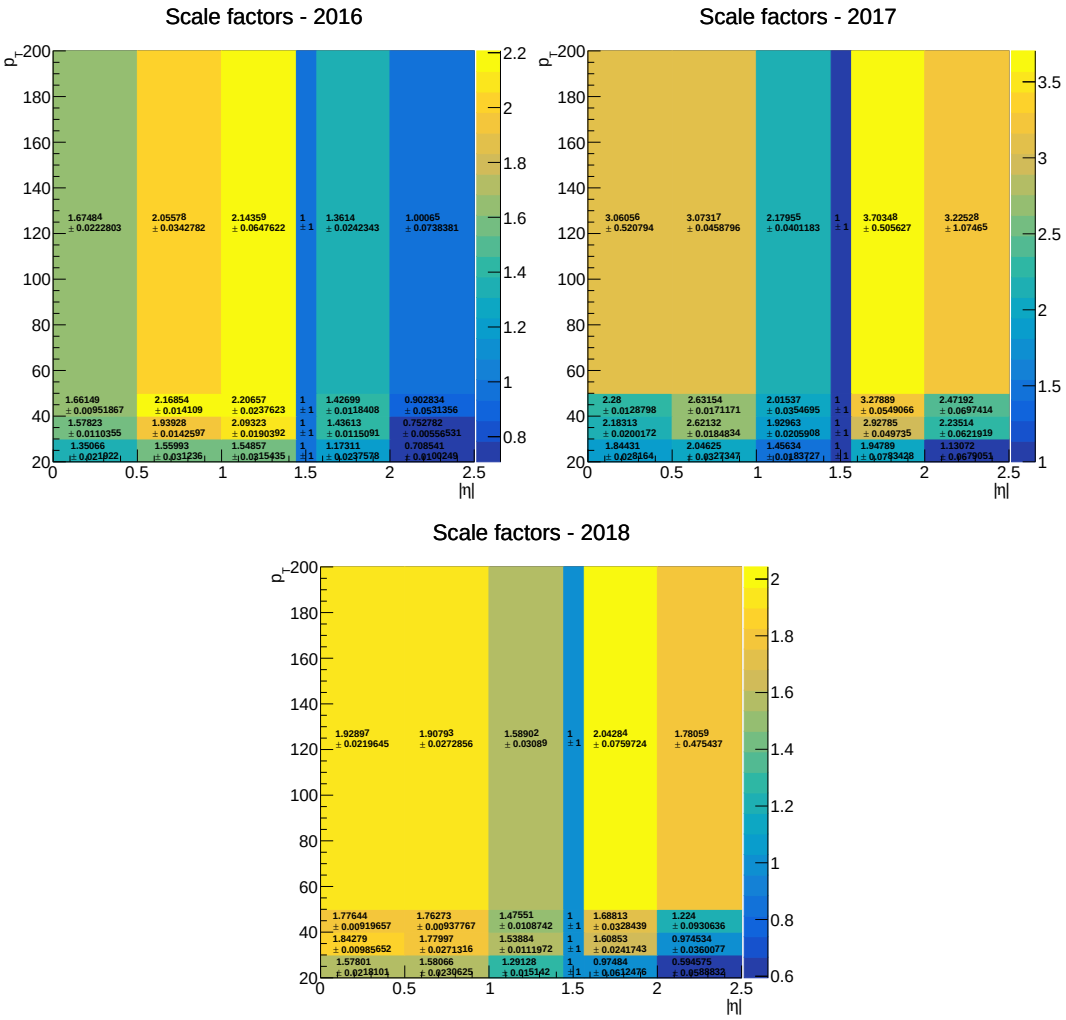


Figure 35: fake factors for the electron-photon misidentification. The central value is obtained with the template fit, while the uncertainty takes into account also the systematic effect computed by fitting the distribution with the double-sided Crystal-Ball.

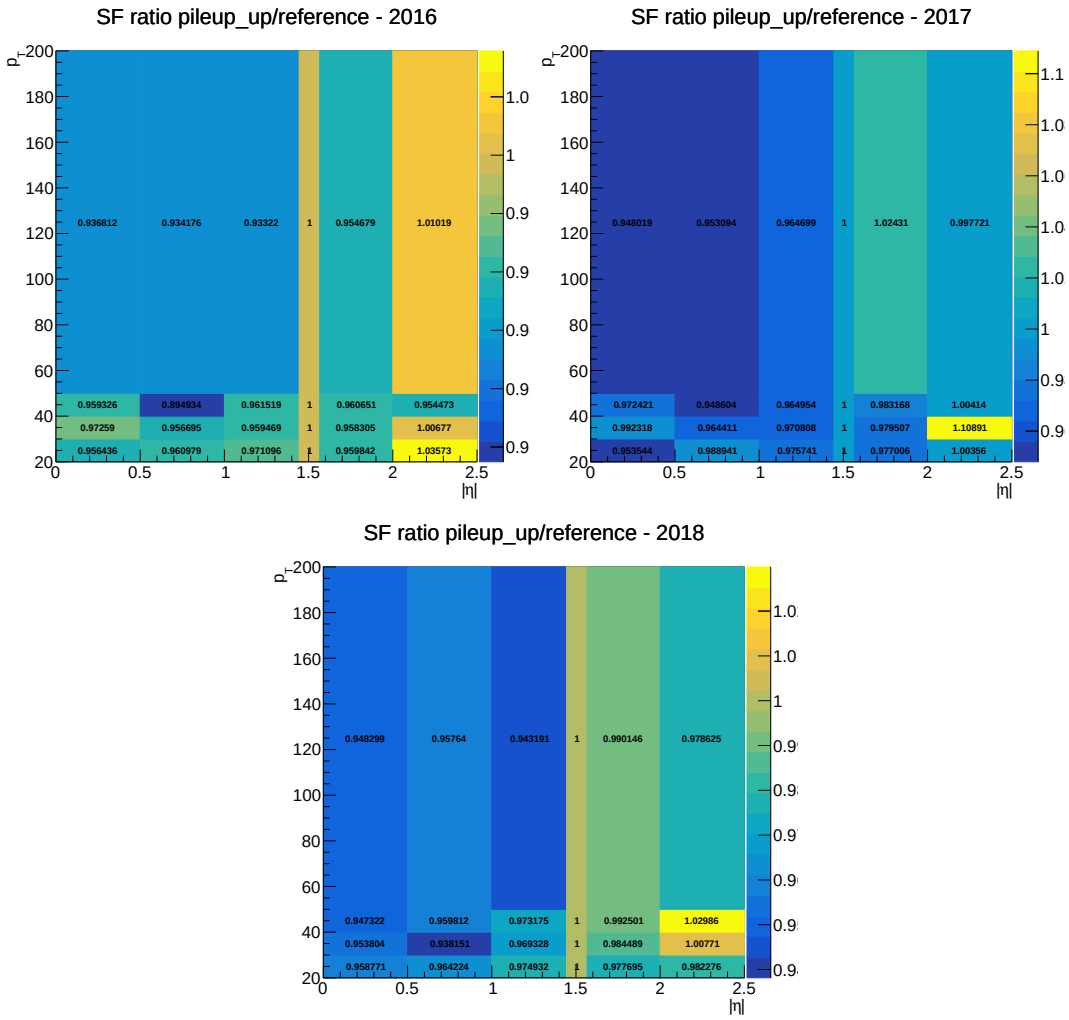


Figure 36: ratio of the fake factors for the electron-photon misidentification obtained while taking into account the up variation of the pileup systematic compared to the reference value.

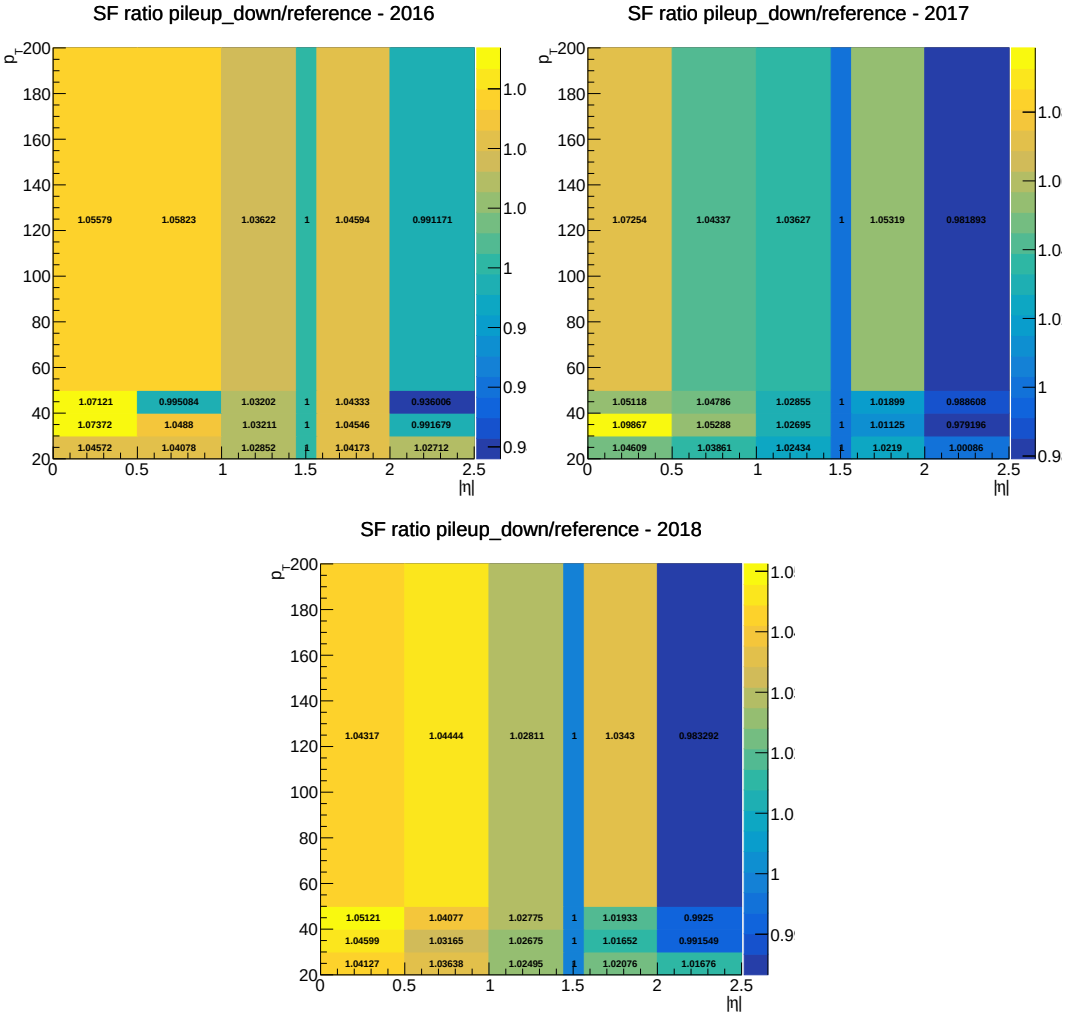
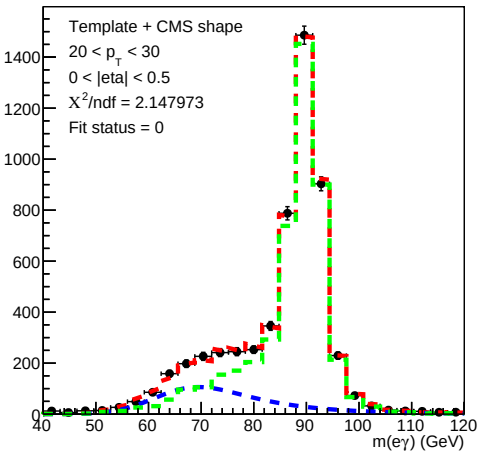
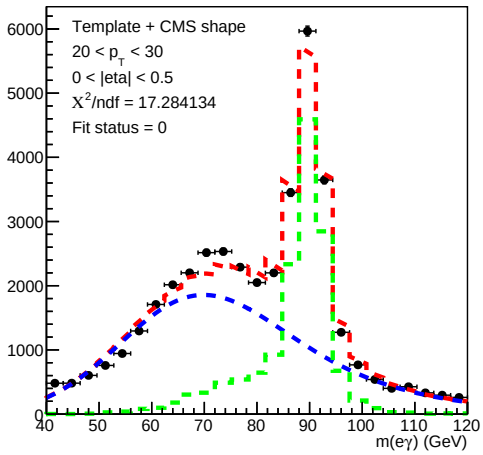


Figure 37: ratio of the fake factors for the electron-photon misidentification obtained while taking into account the down variation of the pileup systematic compared to the reference value.

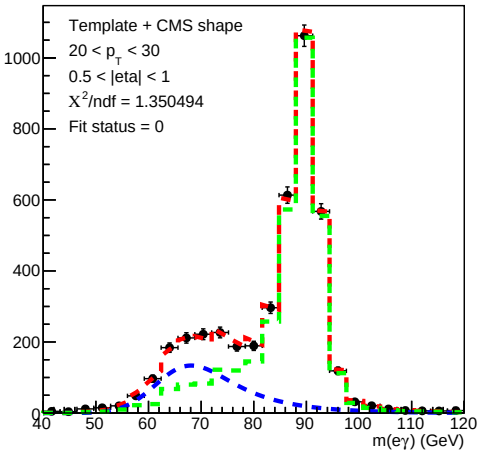
Monte Carlo 2016



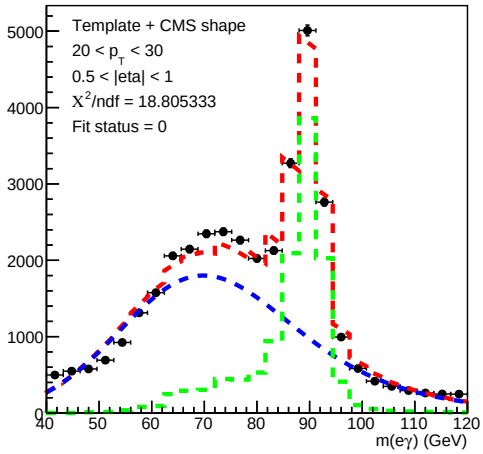
Data 2016



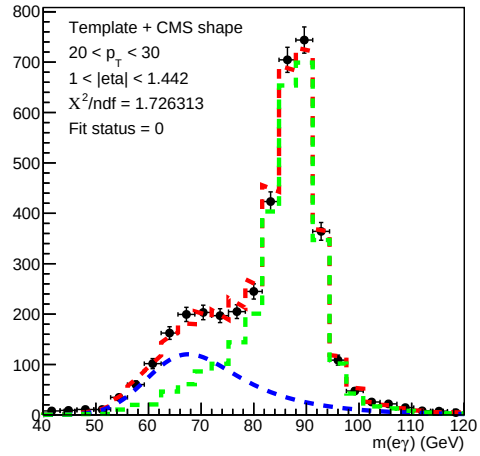
Monte Carlo 2016



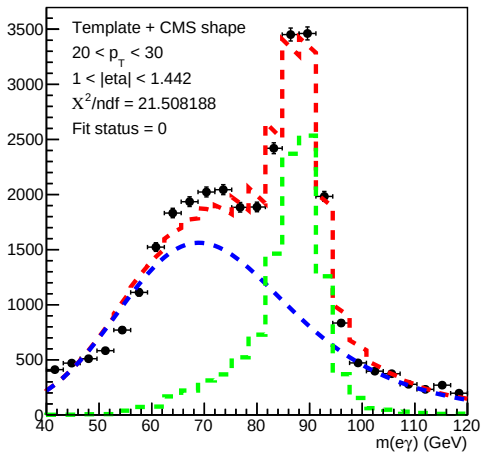
Data 2016



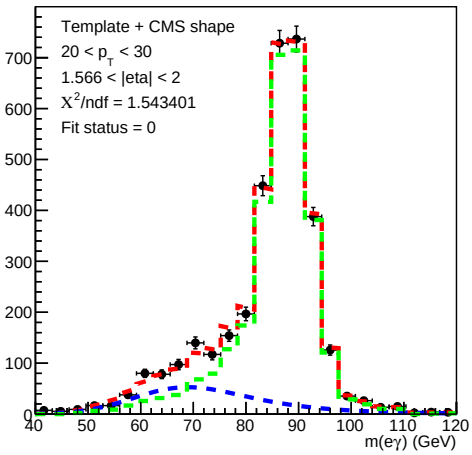
Monte Carlo 2016



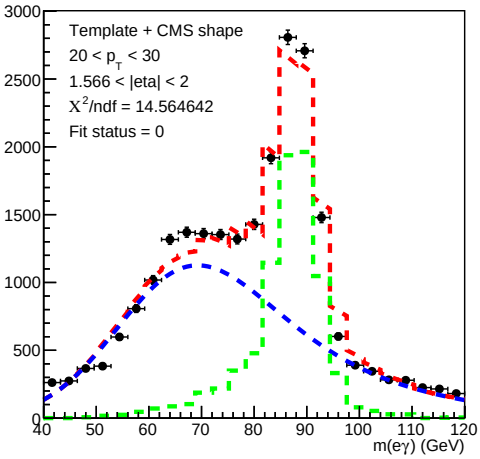
Data 2016



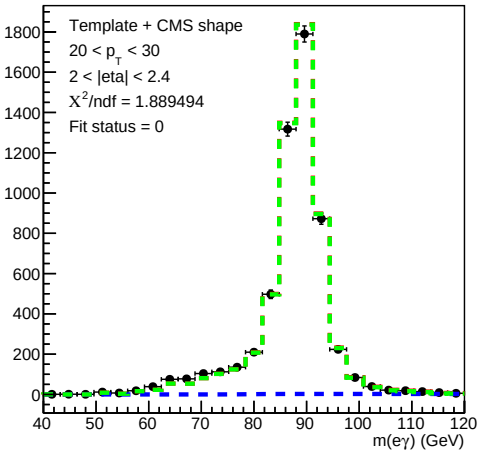
Monte Carlo 2016



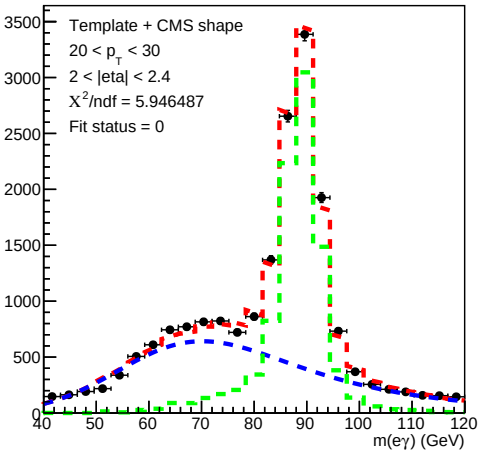
Data 2016



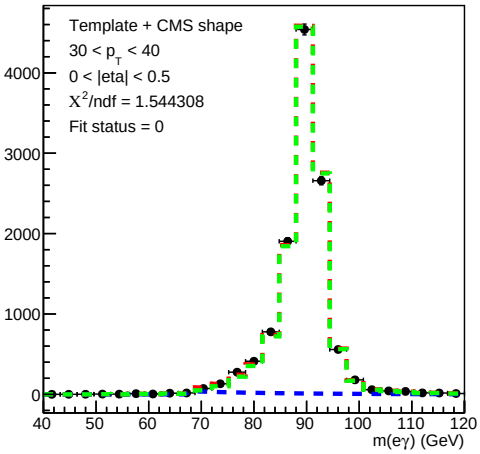
Monte Carlo 2016



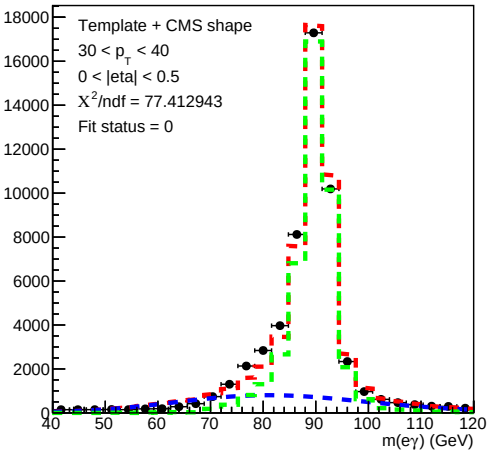
Data 2016



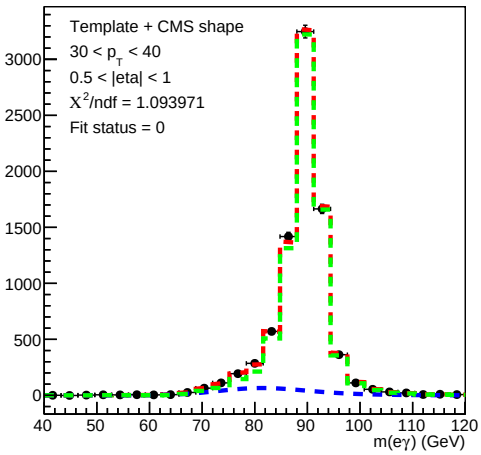
Monte Carlo 2016



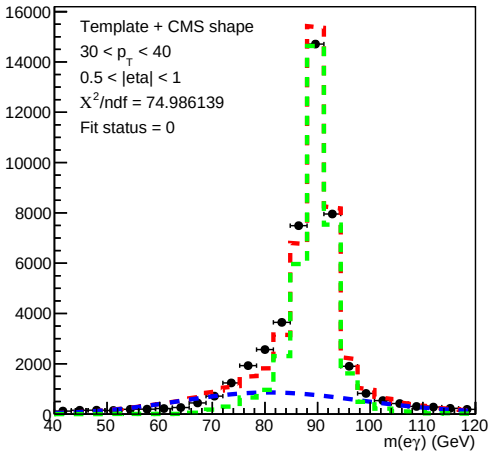
Data 2016



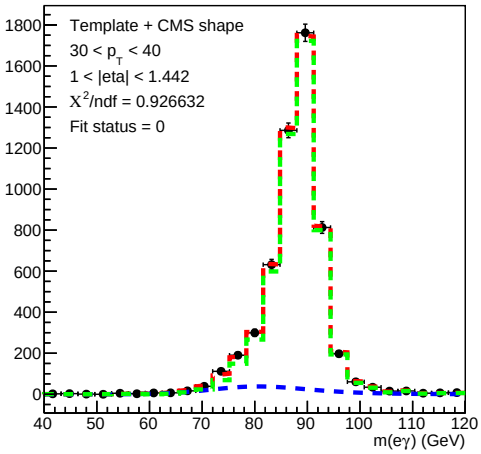
Monte Carlo 2016



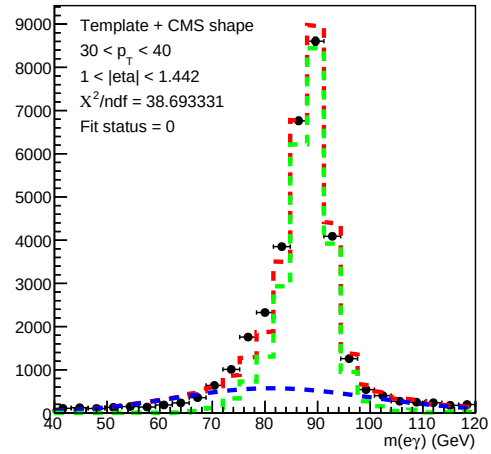
Data 2016



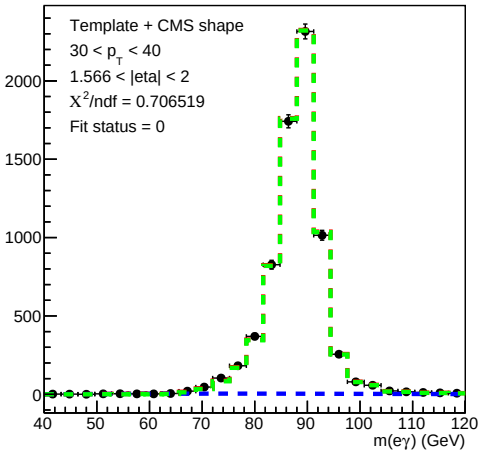
Monte Carlo 2016



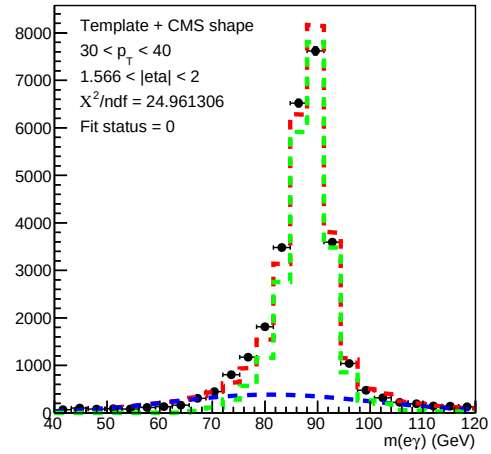
Data 2016



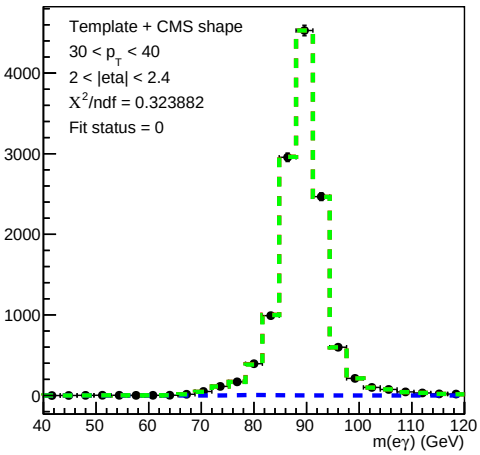
Monte Carlo 2016



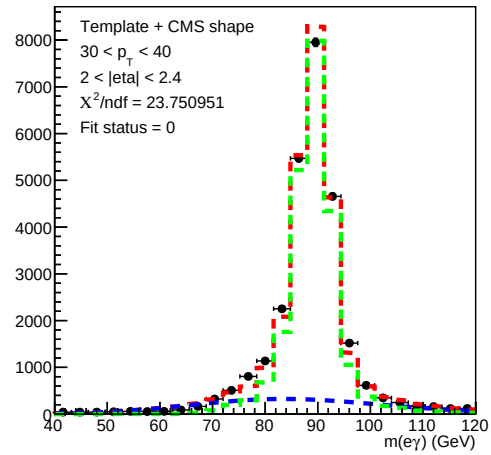
Data 2016



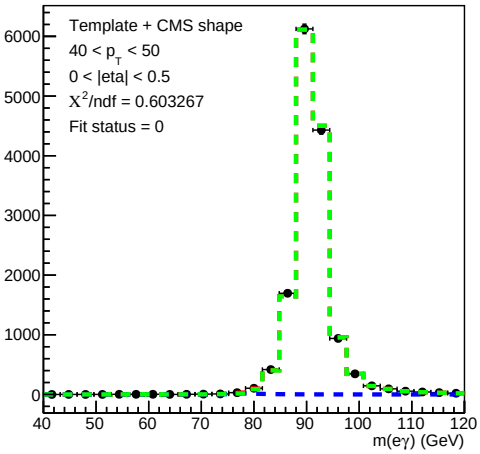
Monte Carlo 2016



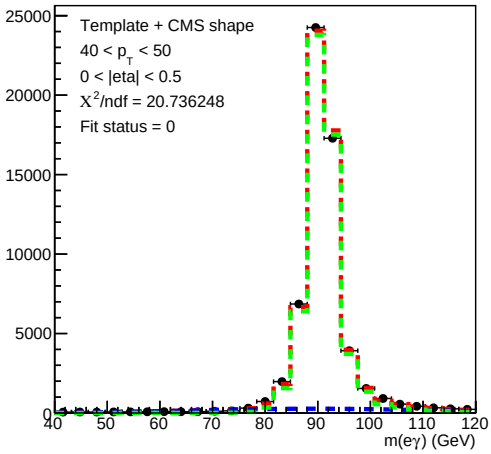
Data 2016



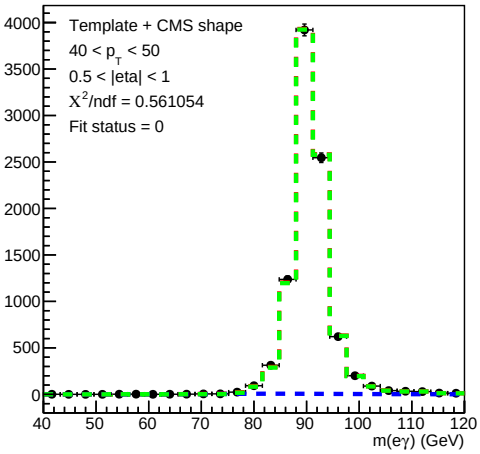
Monte Carlo 2016



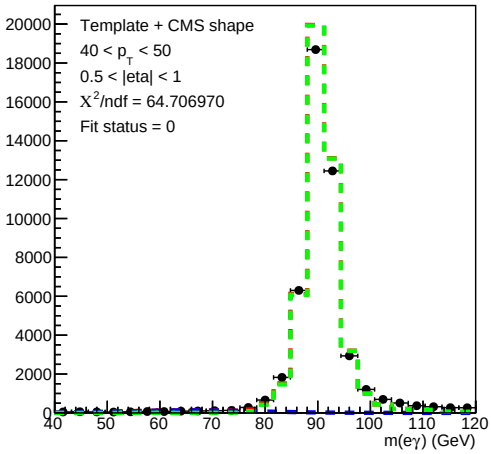
Data 2016



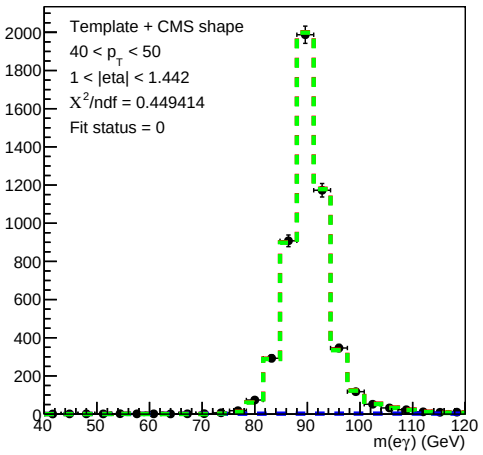
Monte Carlo 2016



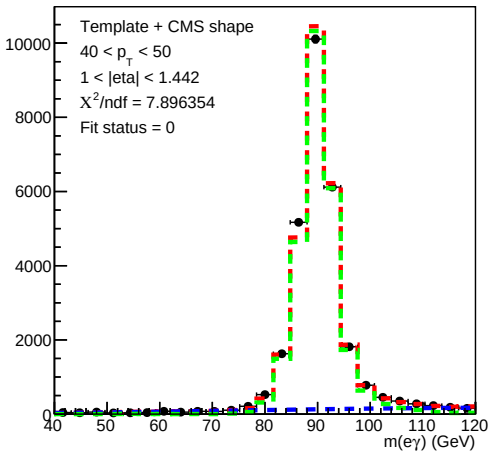
Data 2016



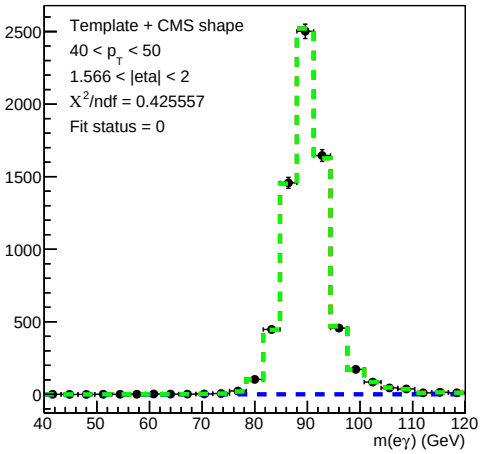
Monte Carlo 2016



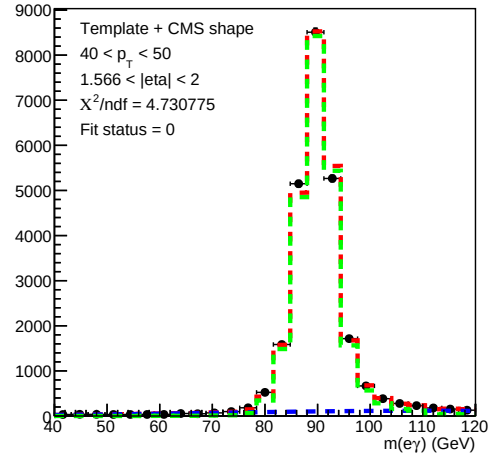
Data 2016



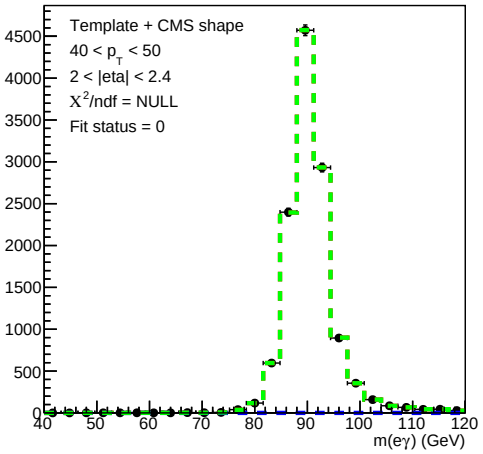
Monte Carlo 2016



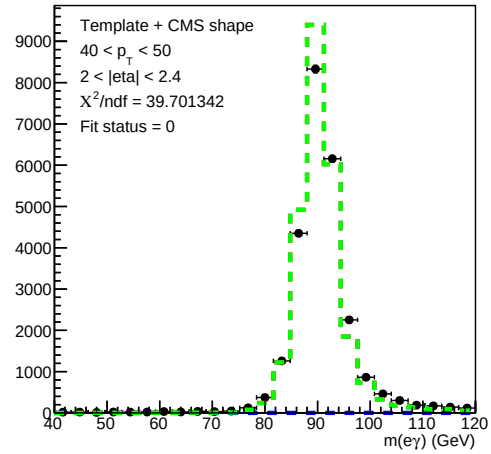
Data 2016



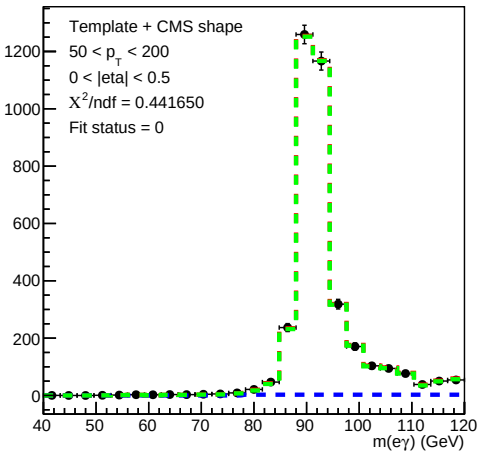
Monte Carlo 2016



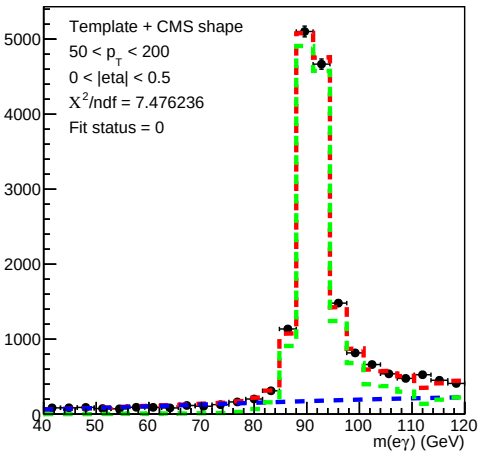
Data 2016



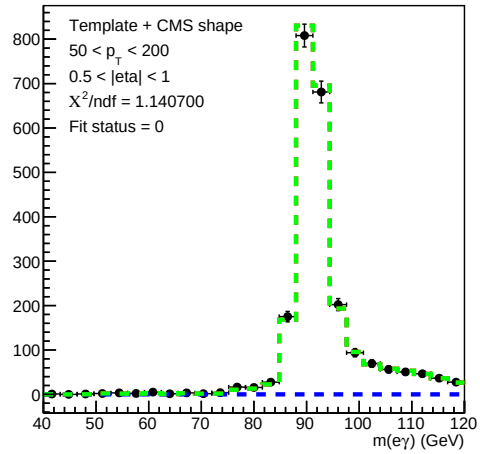
Monte Carlo 2016



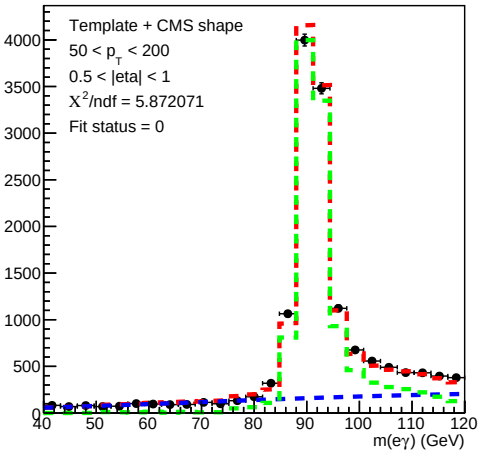
Data 2016



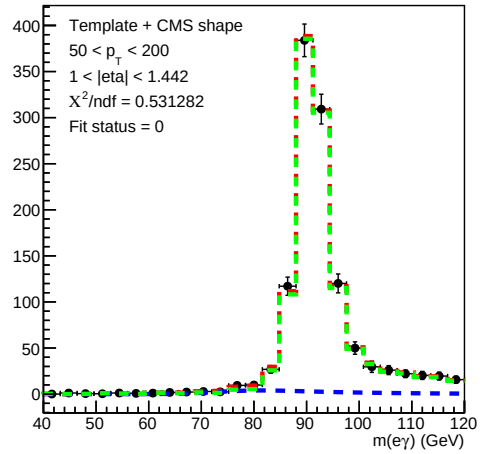
Monte Carlo 2016



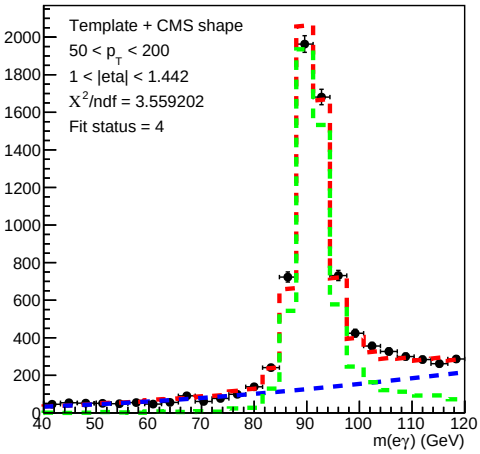
Data 2016



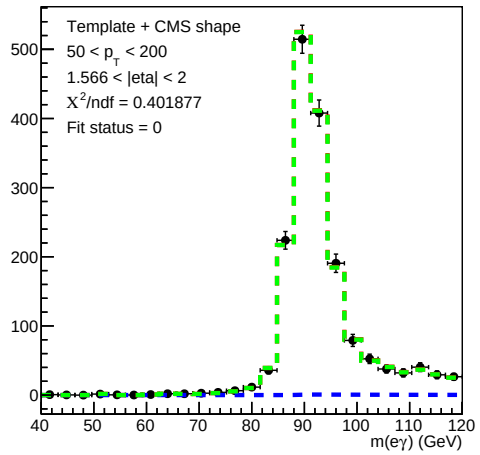
Monte Carlo 2016



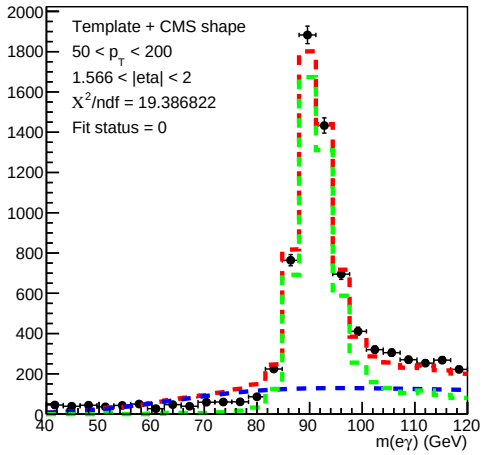
Data 2016



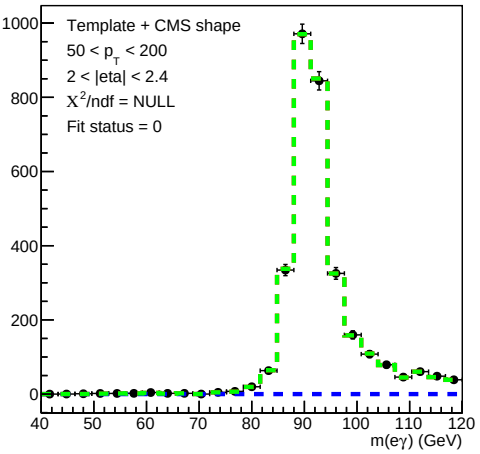
Monte Carlo 2016



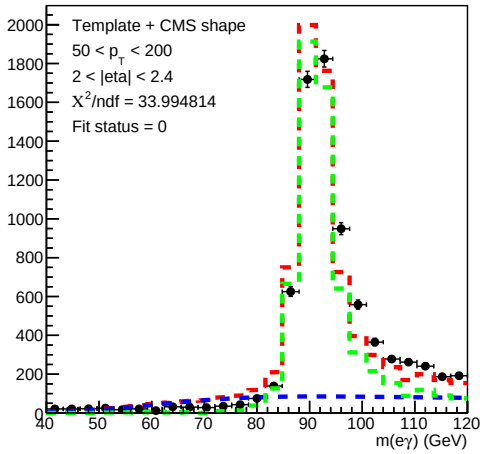
Data 2016



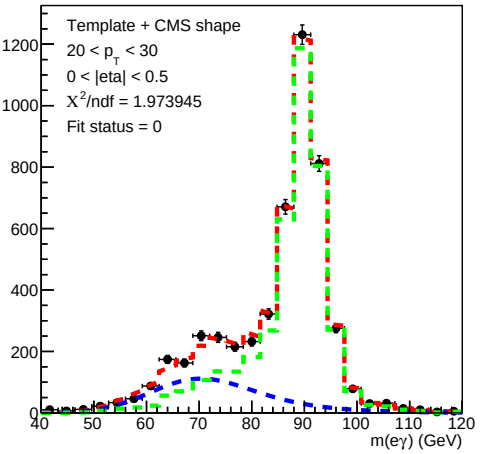
Monte Carlo 2016



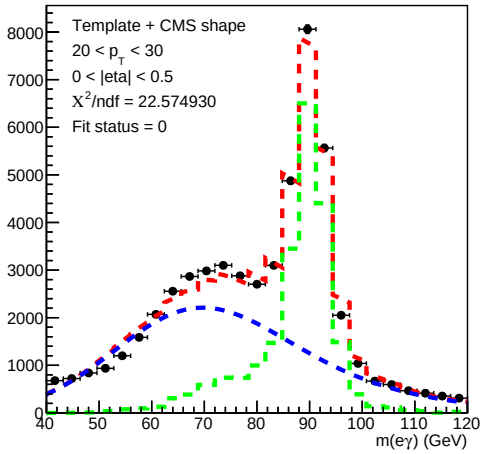
Data 2016



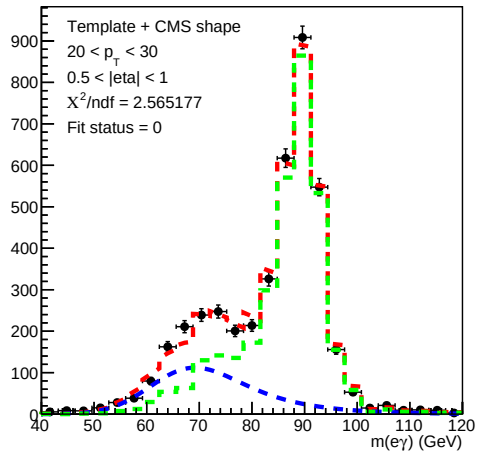
Monte Carlo 2017



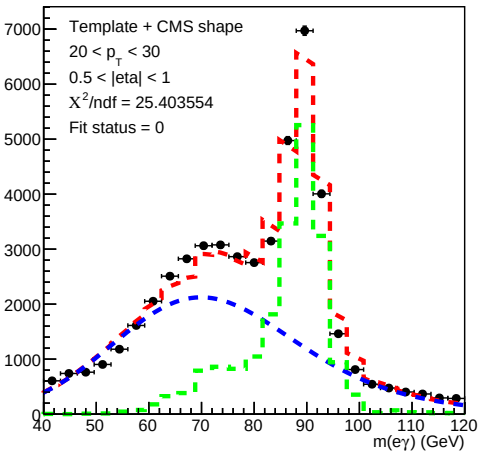
Data 2017



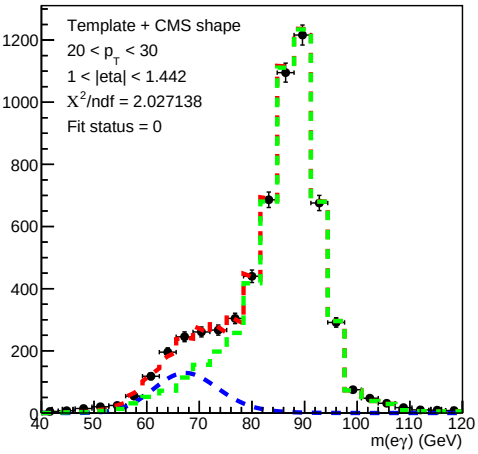
Monte Carlo 2017



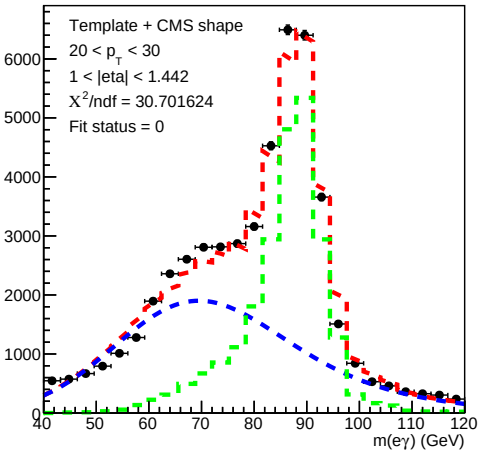
Data 2017



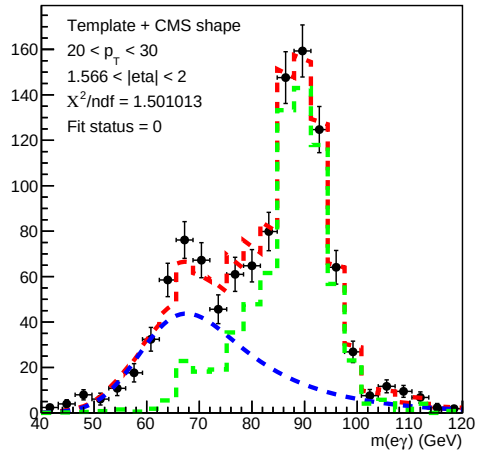
Monte Carlo 2017



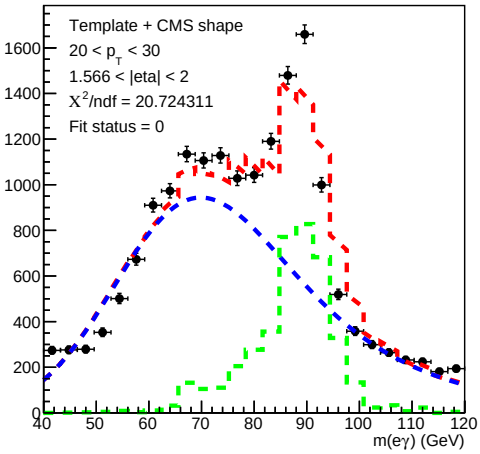
Data 2017



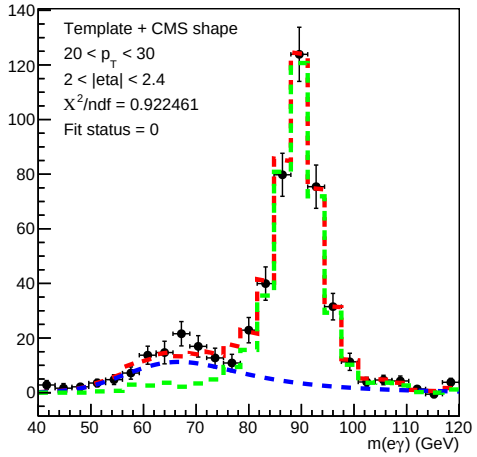
Monte Carlo 2017



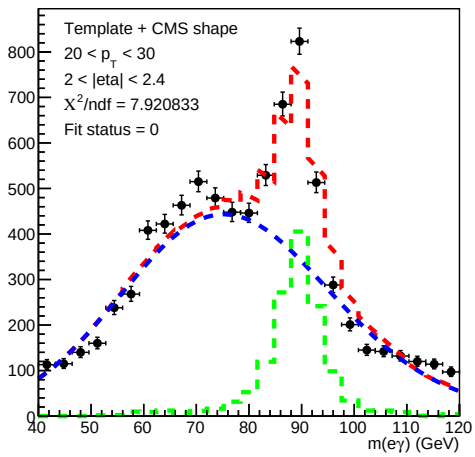
Data 2017



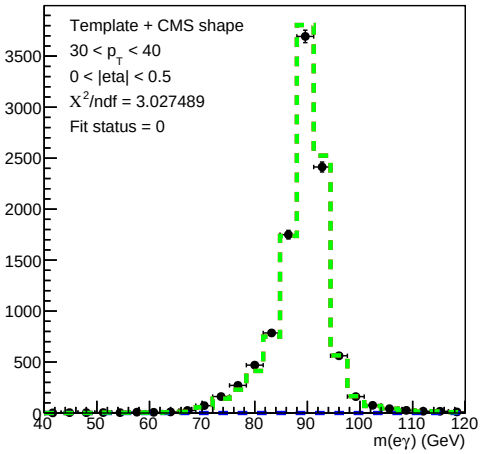
Monte Carlo 2017



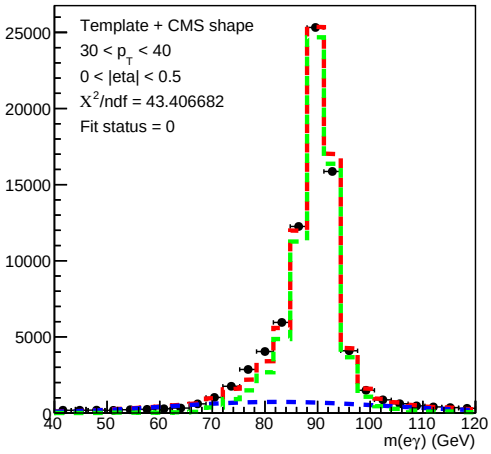
Data 2017



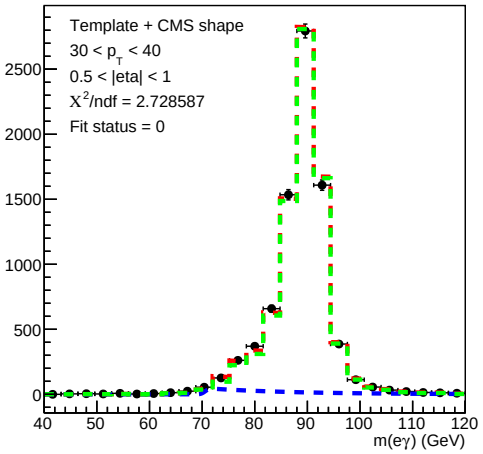
Monte Carlo 2017



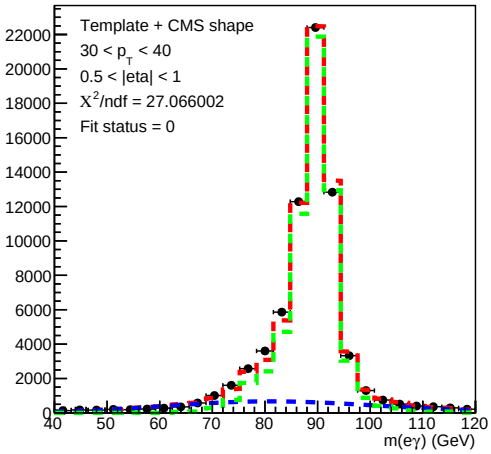
Data 2017



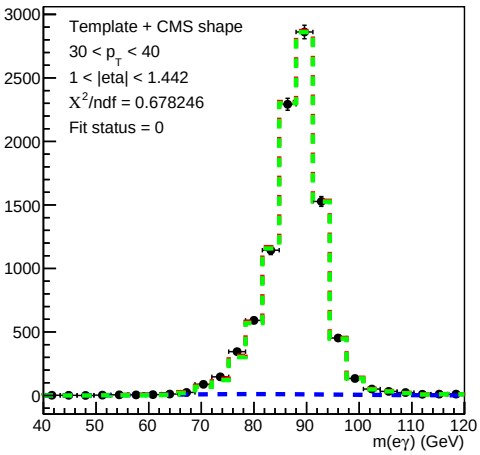
Monte Carlo 2017



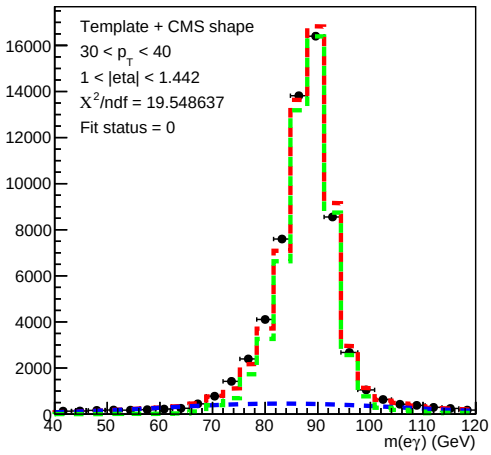
Data 2017



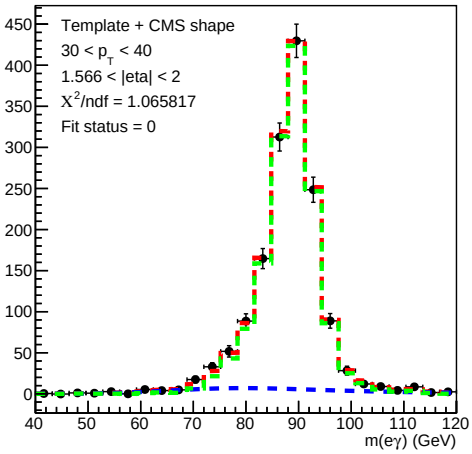
Monte Carlo 2017



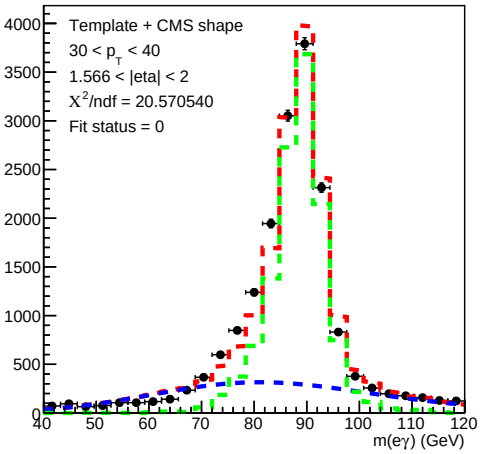
Data 2017



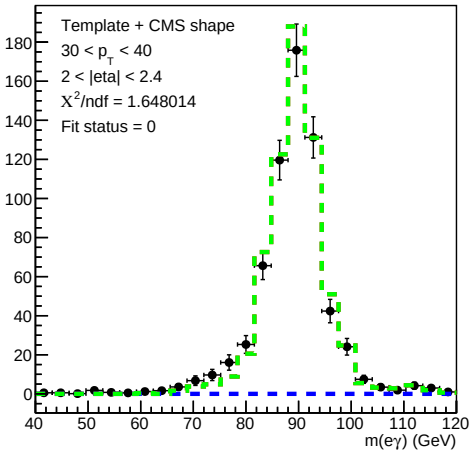
Monte Carlo 2017



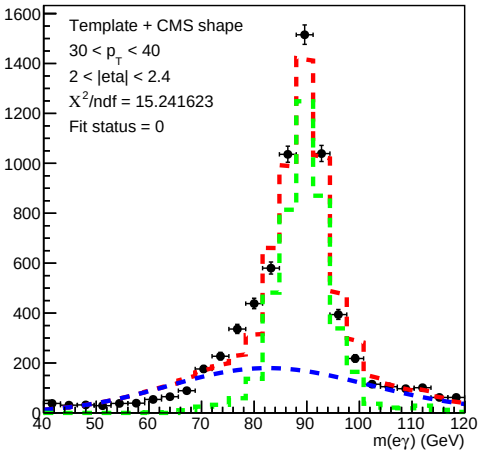
Data 2017



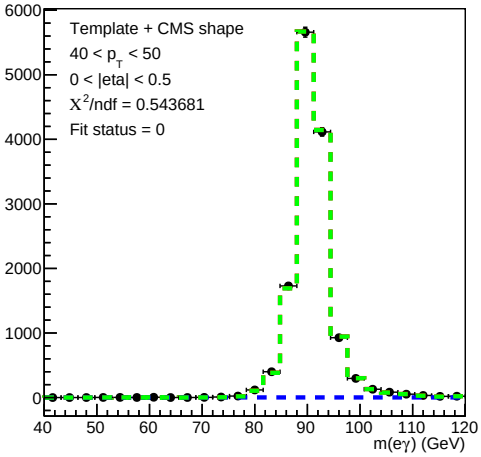
Monte Carlo 2017



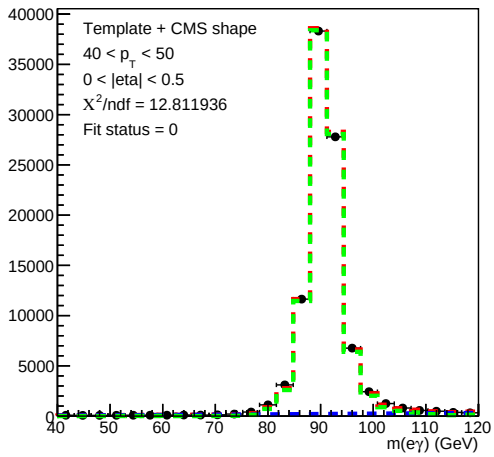
Data 2017



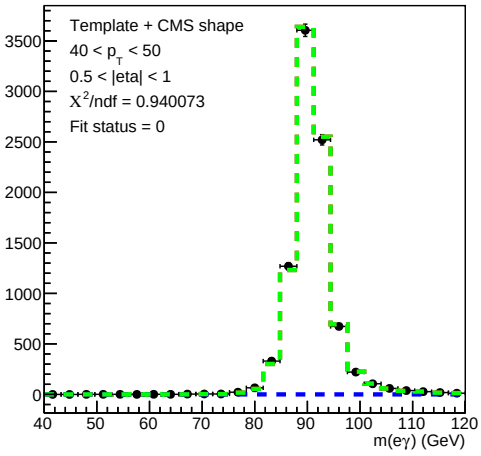
Monte Carlo 2017



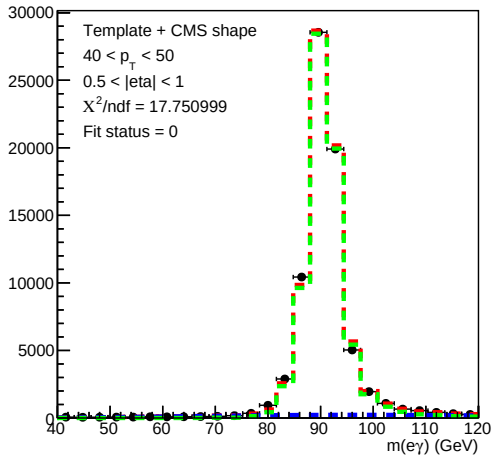
Data 2017



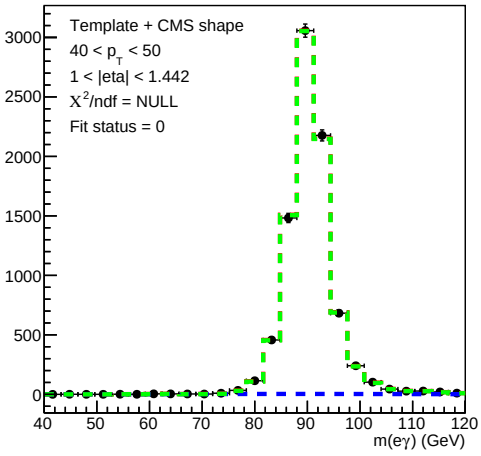
Monte Carlo 2017



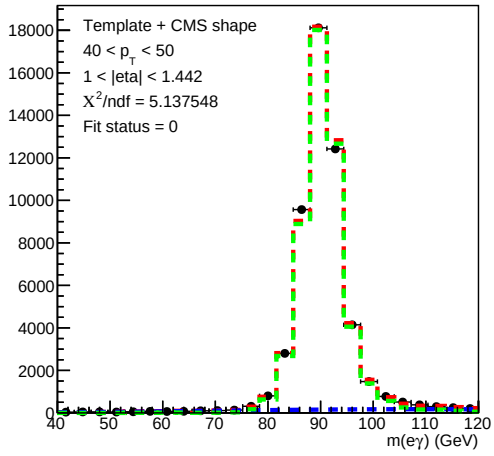
Data 2017



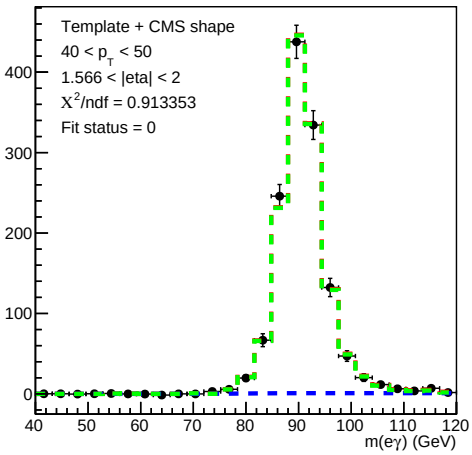
Monte Carlo 2017



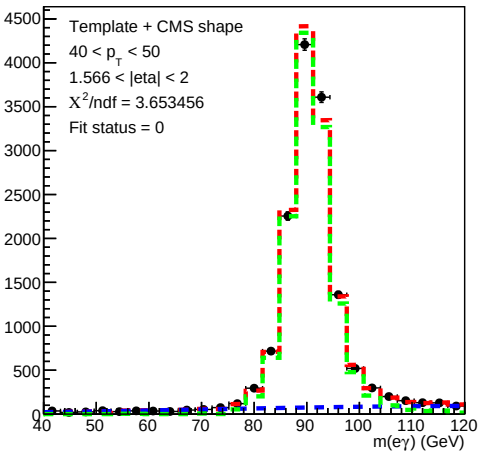
Data 2017



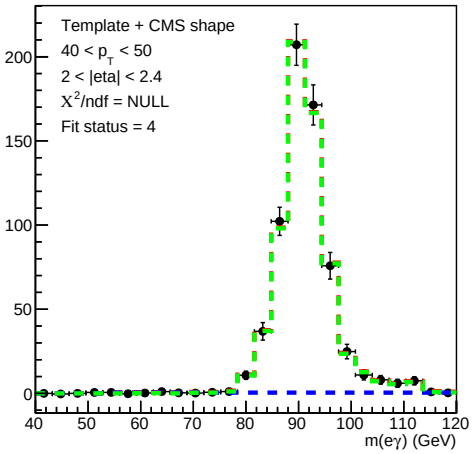
Monte Carlo 2017



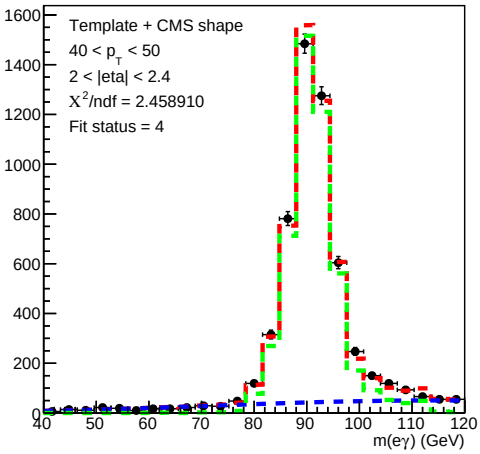
Data 2017



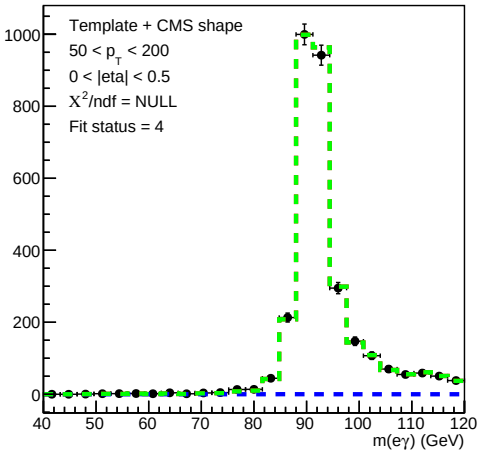
Monte Carlo 2017



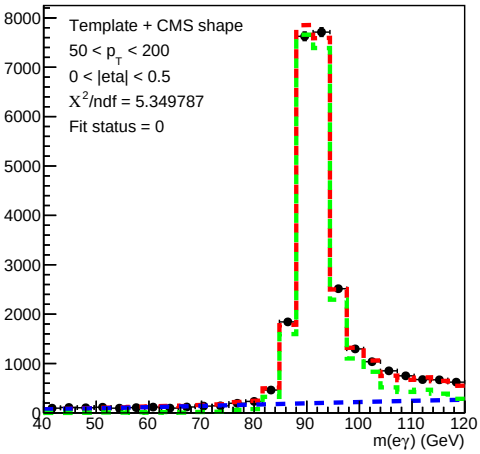
Data 2017



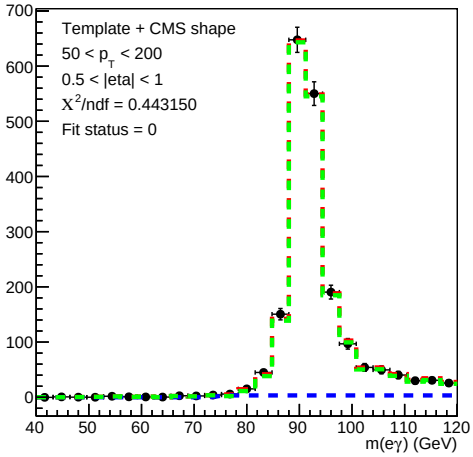
Monte Carlo 2017



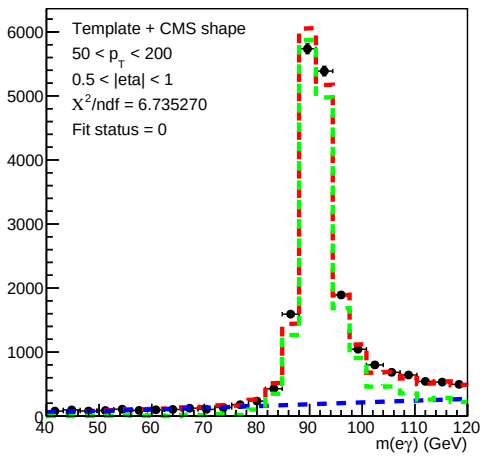
Data 2017



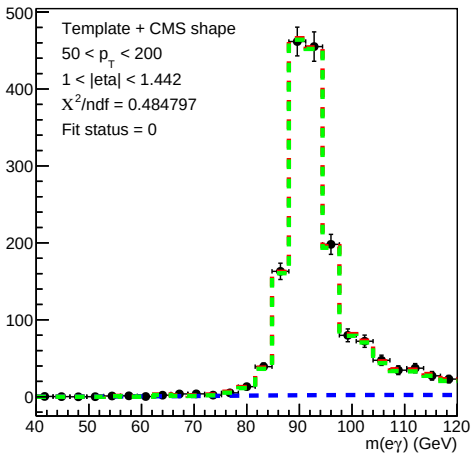
Monte Carlo 2017



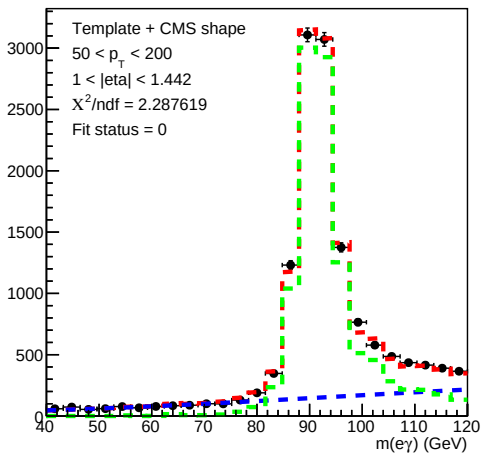
Data 2017



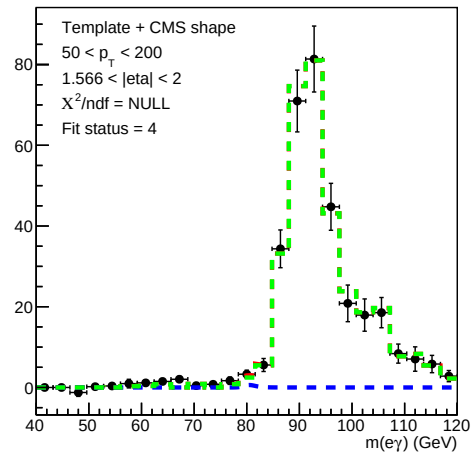
Monte Carlo 2017



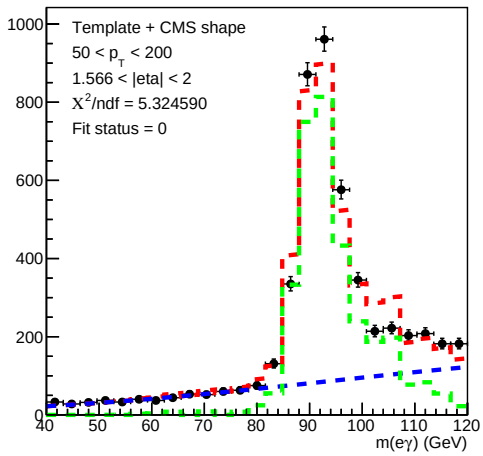
Data 2017



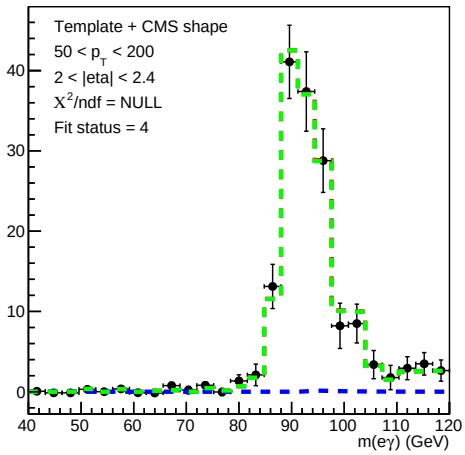
Monte Carlo 2017



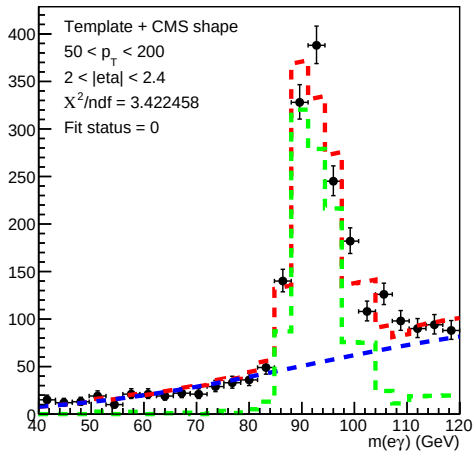
Data 2017



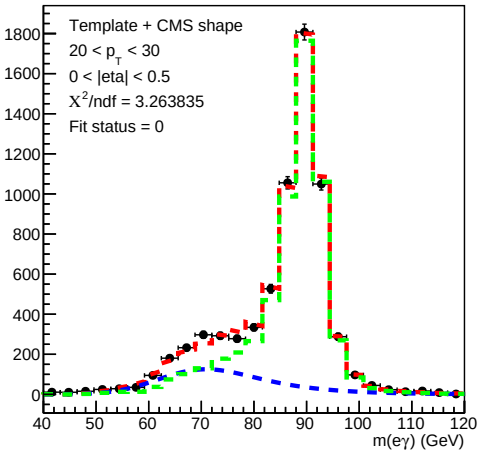
Monte Carlo 2017



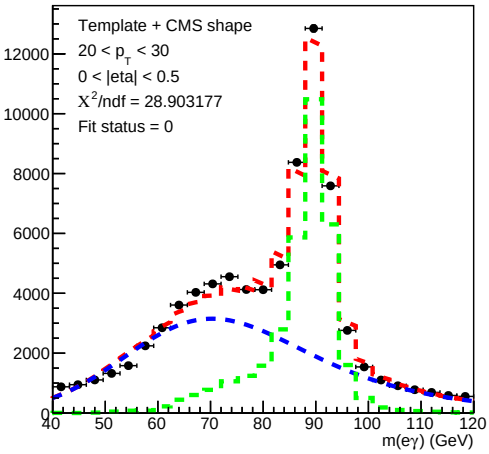
Data 2017



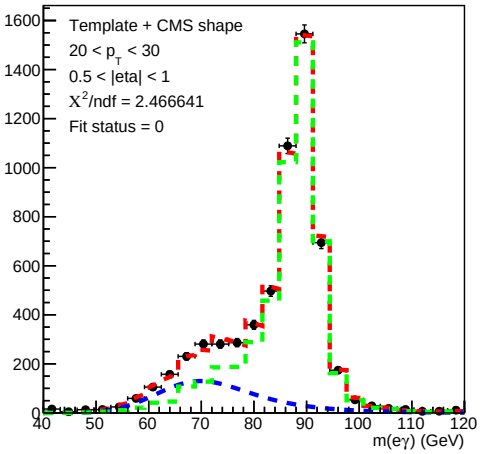
Monte Carlo 2018



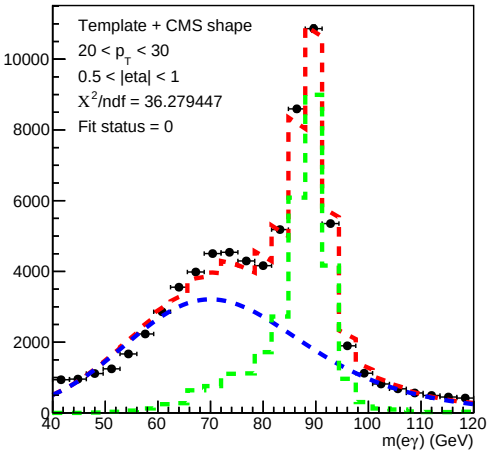
Data 2018



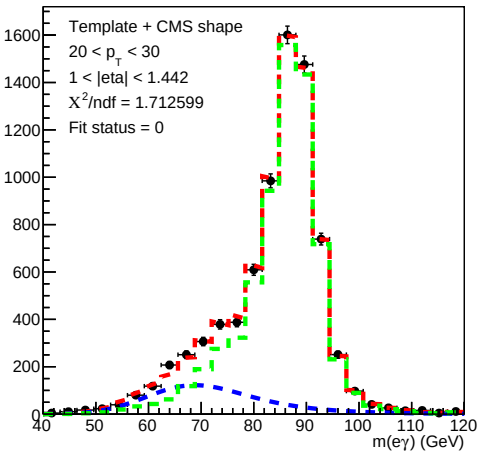
Monte Carlo 2018



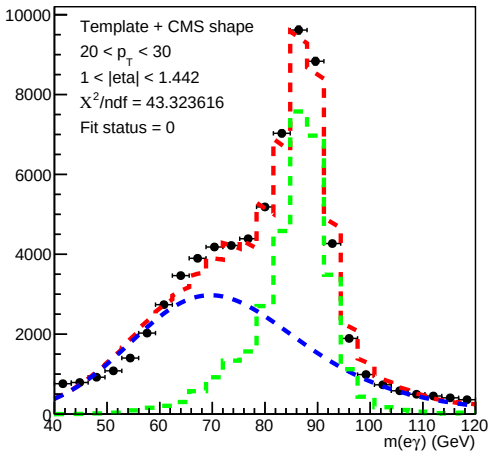
Data 2018



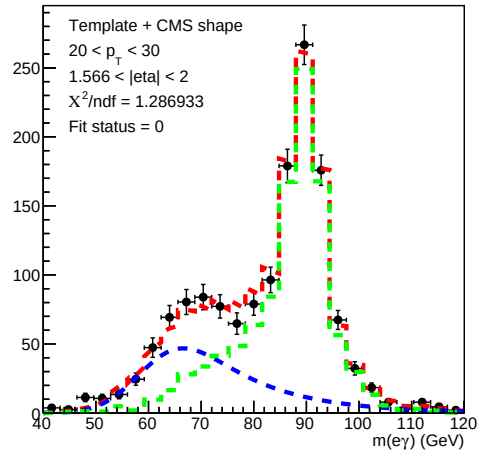
Monte Carlo 2018



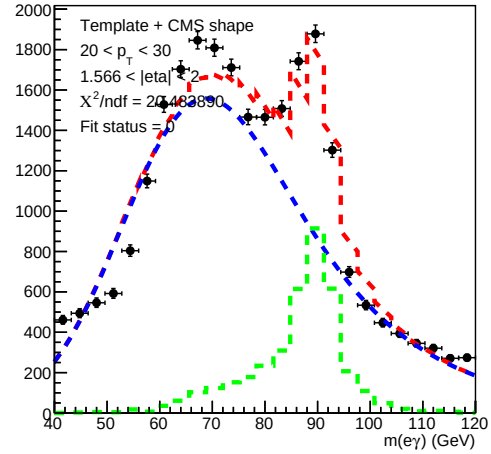
Data 2018



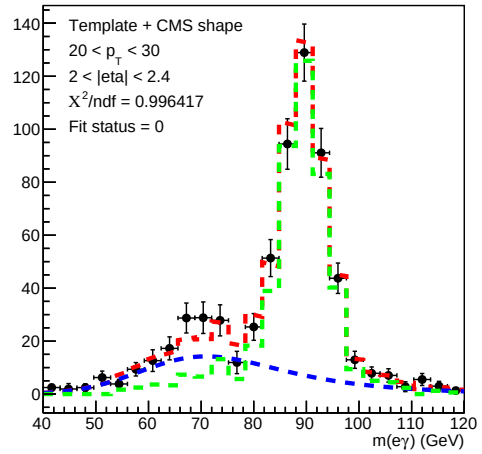
Monte Carlo 2018



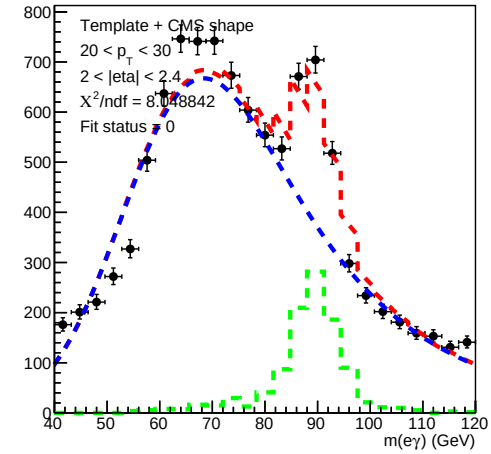
Data 2018



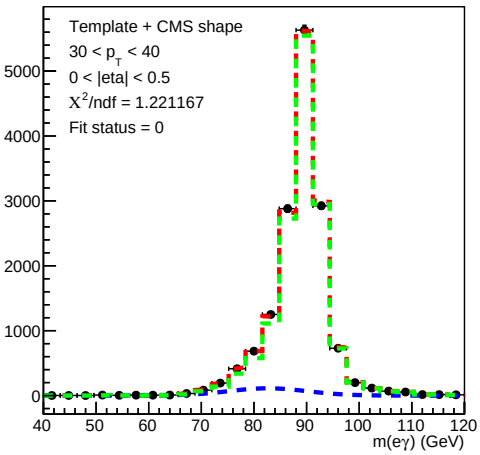
Monte Carlo 2018



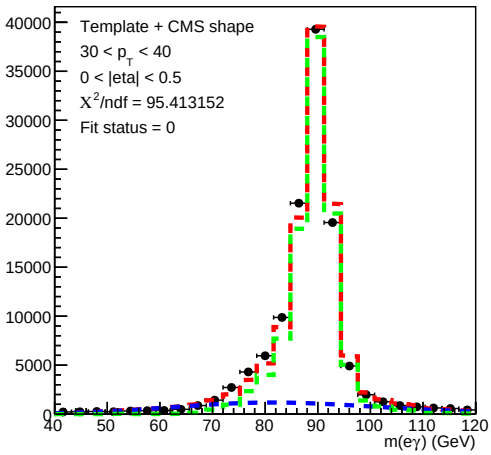
Data 2018



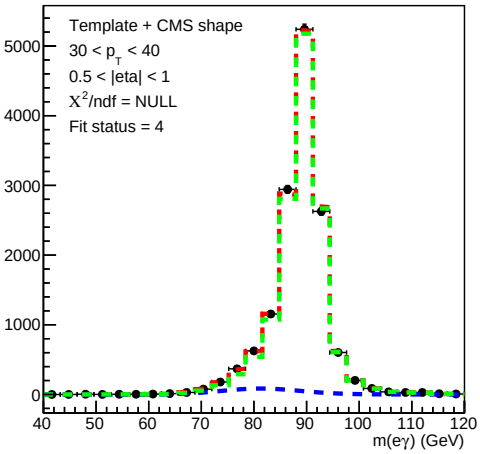
Monte Carlo 2018



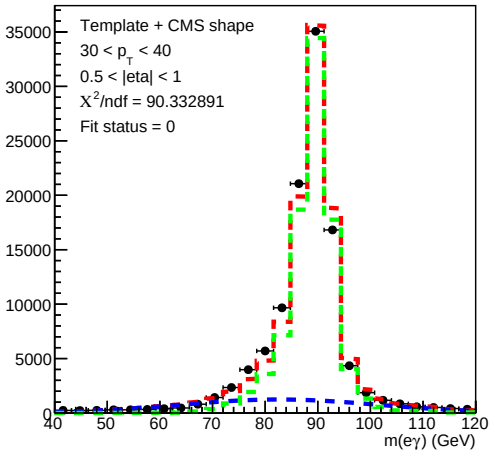
Data 2018



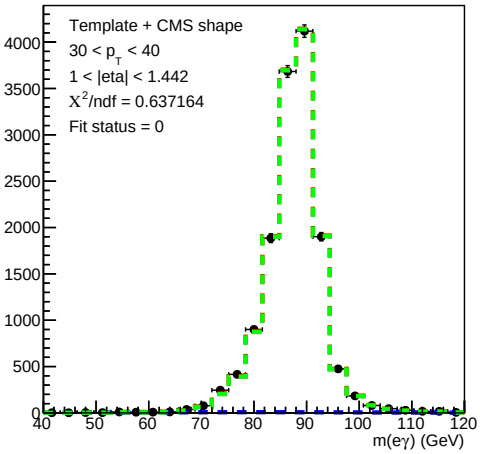
Monte Carlo 2018



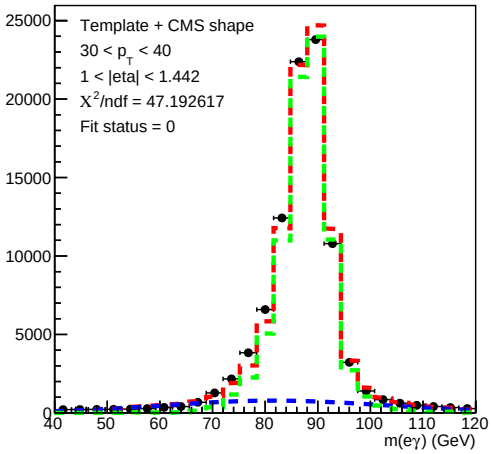
Data 2018



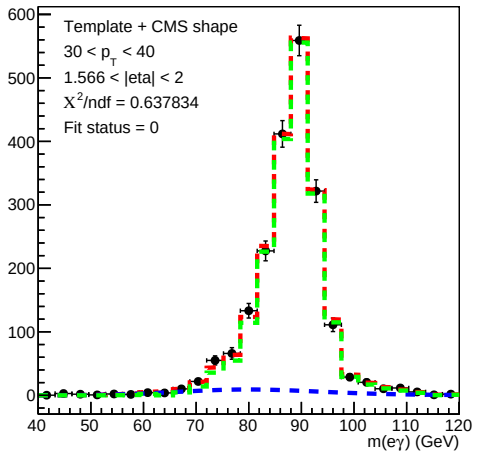
Monte Carlo 2018



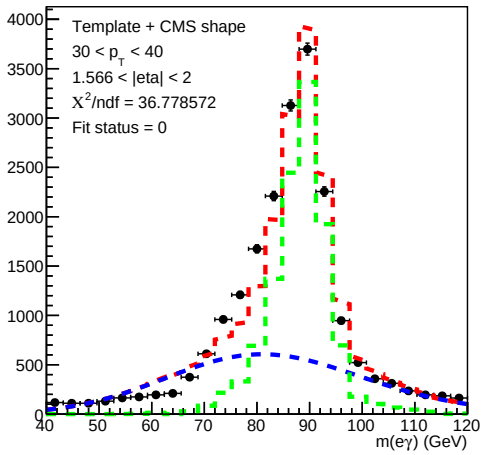
Data 2018



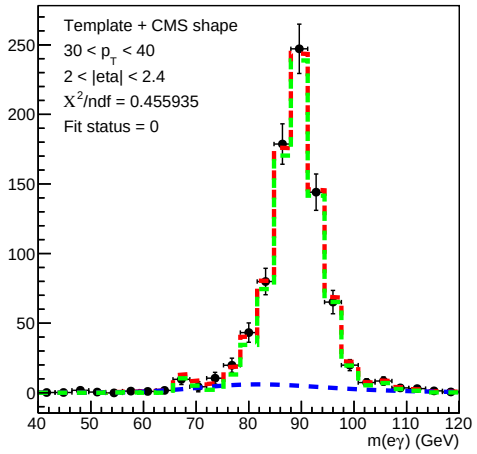
Monte Carlo 2018



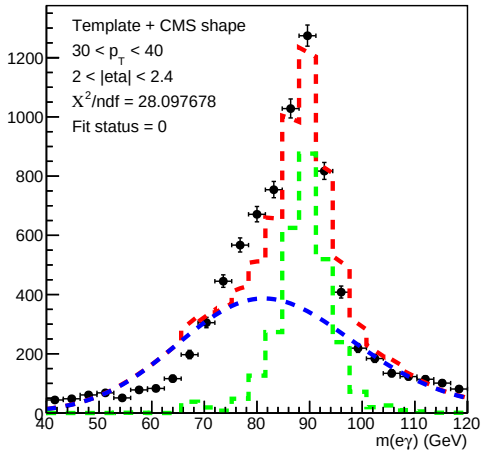
Data 2018



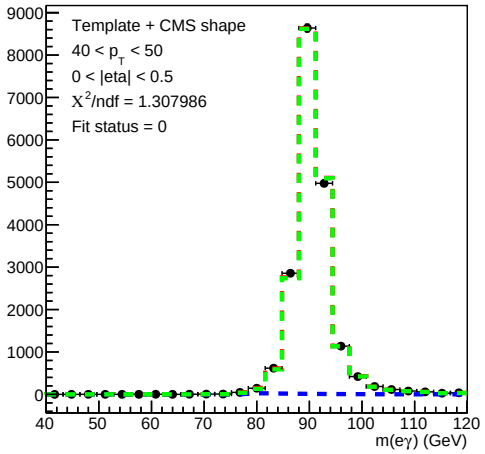
Monte Carlo 2018



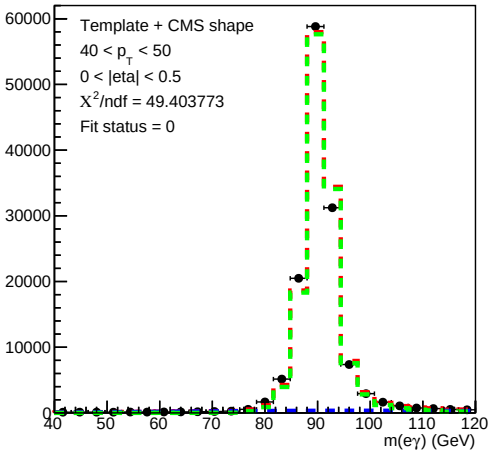
Data 2018



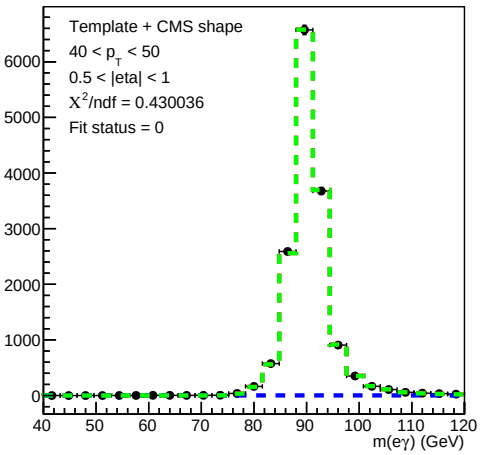
Monte Carlo 2018



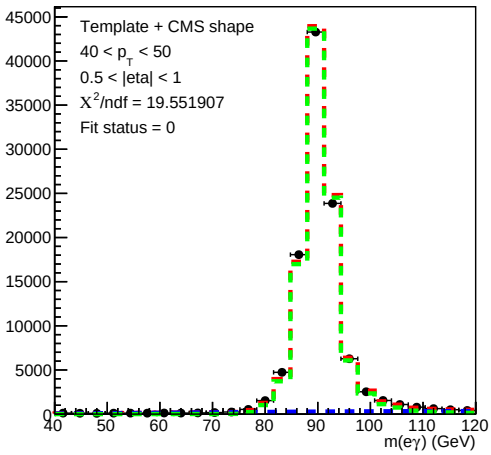
Data 2018



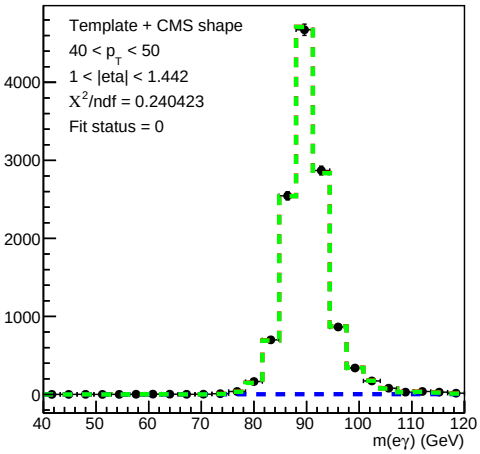
Monte Carlo 2018



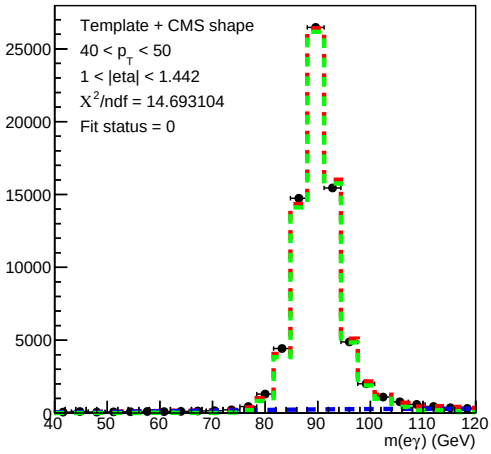
Data 2018



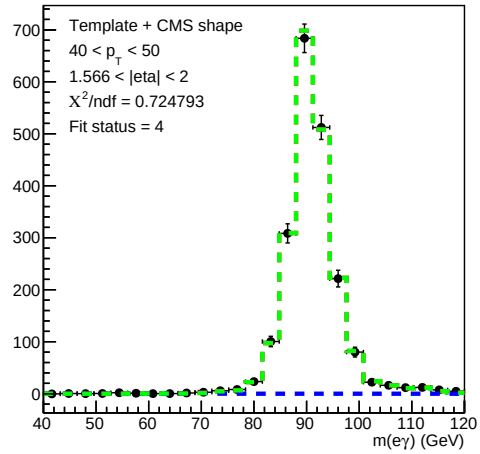
Monte Carlo 2018



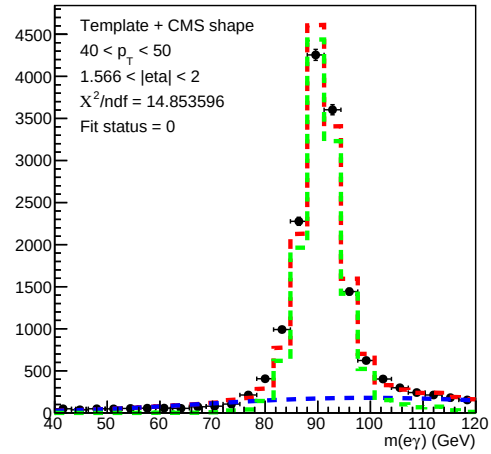
Data 2018



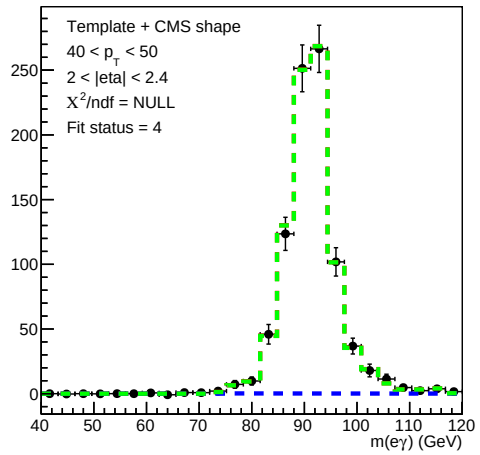
Monte Carlo 2018



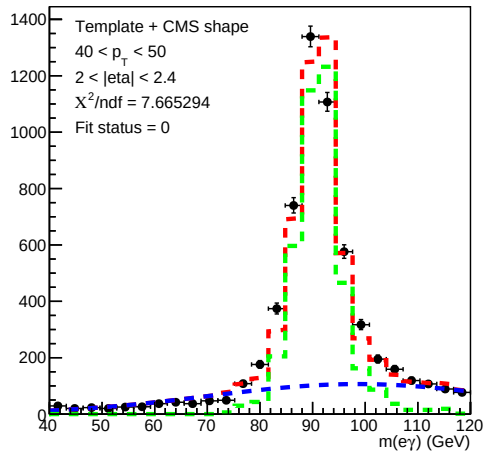
Data 2018



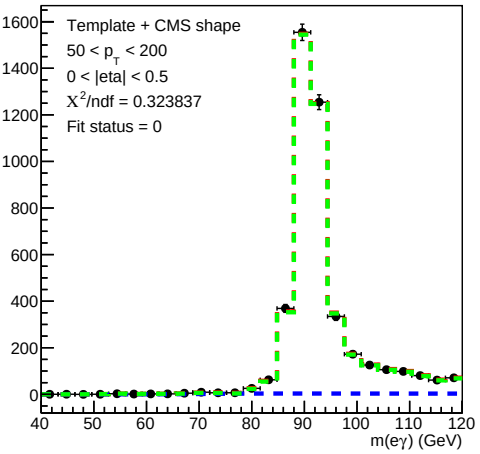
Monte Carlo 2018



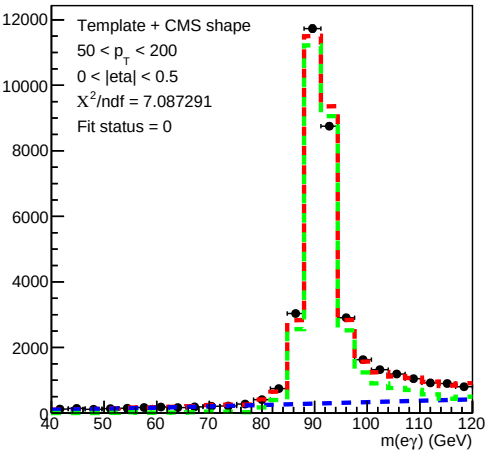
Data 2018



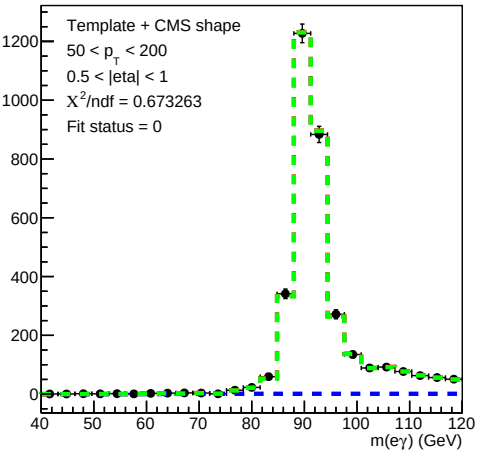
Monte Carlo 2018



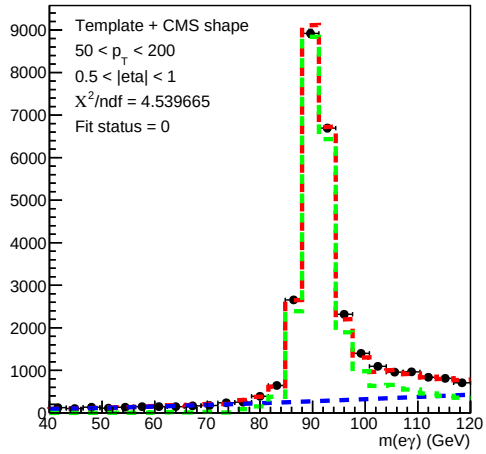
Data 2018



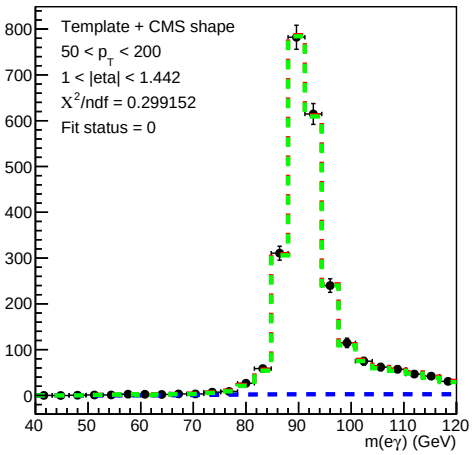
Monte Carlo 2018



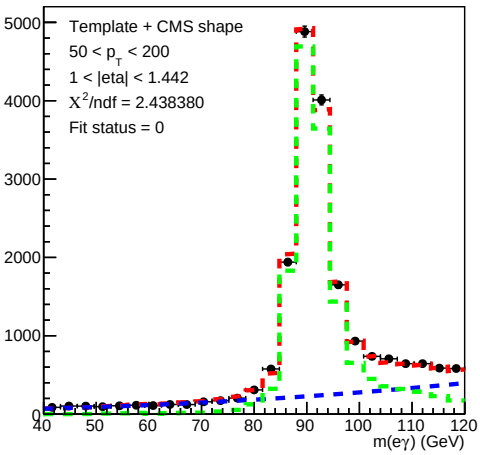
Data 2018



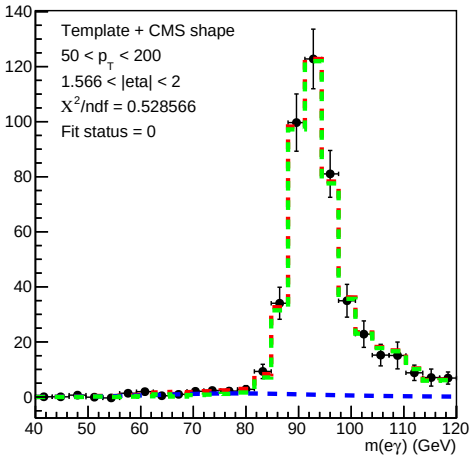
Monte Carlo 2018



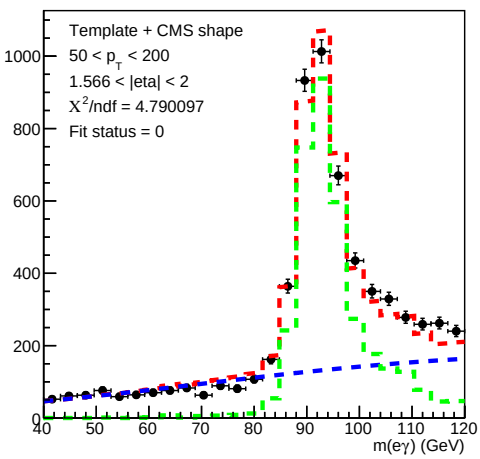
Data 2018



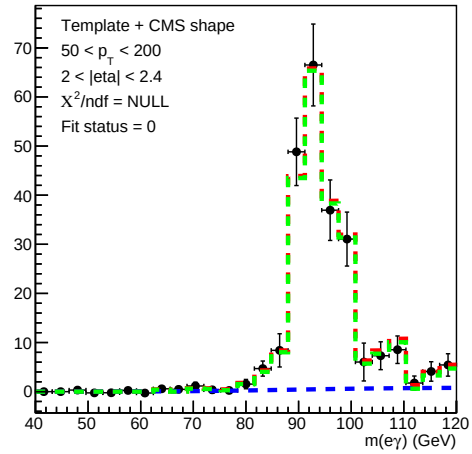
Monte Carlo 2018



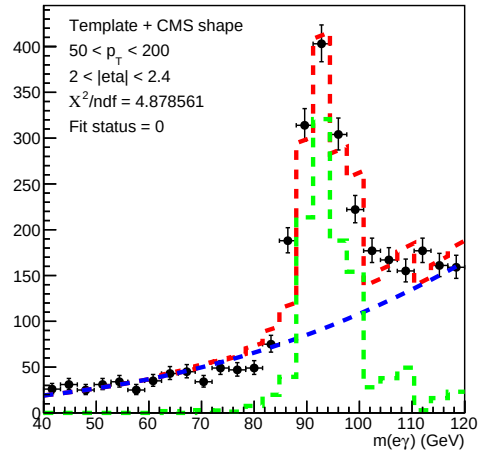
Data 2018



Monte Carlo 2018



Data 2018



BIBLIOGRAPHY

[1] The ALEPH Collaboration. 'Constraints on anomalous QGCs in e^+e^- interactions from 183 to 209 GeV'. In: *Physics Letters B* 602.1 (2004), pp. 31–40.

[2] The DELPHI Collaboration. 'Measurement of the $e^+e^- \to W^+W^-\gamma$ cross-section and limits on anomalous quartic gauge couplings with DELPHI'. In: *The European Physical Journal C* 31.2 (2003), pp. 139–147.

[3] The CMS Collaboration. 'Measurements of the pp $\to W\gamma\gamma$ and pp $\to Z\gamma\gamma$ cross sections and limits on anomalous quartic gauge couplings at $\sqrt{s} = 8$ TeV'. In: *Journal of High Energy Physics* 10 (2017), p. 072.

[4] The ATLAS Collaboration. 'Evidence of $W\gamma\gamma$ Production in pp Collisions at $\sqrt{s} = 8$ TeV and Limits on Anomalous Quartic Gauge Couplings with the ATLAS Detector'. In: *Physical Review Letters* 115 (2015), p. 031802.

[5] O. Hahn and L. Meitner. 'Uber die Absorption der β-Strahlen einiger Radioelemente'. In: *Physikalische Zeitschrift* 9 (1908), p. 321.

[6] O. Hahn and L. Meitner. 'Uber die β-Strahlen des Aktiniums'. In: *Physikalische Zeitschrift* 9 (1908), p. 697.

[7] O. Hahn and L. Meitner. 'Uber eine typische β-Strahlung des eigentlicher Radiums'. In: *Physikalische Zeitschrift* 10 (1909), p. 741.

[8] J. Chadwick. 'Intensitätavertailung im magnetischen Spektren der β-Strahlen von Radium B+C'. In: *Verhandlungen der Deutschen Physikalischen Gesellschaft* 16 (1914), p. 383.

[9] N. Bohr, H. Kramers and J. Slater. 'The quantum theory of radiation'. In: *Philosophical Magazine* 47 (1924), p. 785.

[10] W. Pauli. 'Zur älteren und neureren Geschichte des Neutrinos'. In: *Wolfgang Pauli. Collected Scientific Papers* (1964), p. 1313.

[11] J. Chadwick. 'Possible Existence of a Neutron'. In: *Nature* 129.3252 (1932), pp. 312–312.

[12] C. L. Cowan, F. Reines, F. B. Harrison, H. W. Kruse and A. D. McGuire. 'Detection of the Free Neutrino: a Confirmation'. In: *Science* 124.3212 (1956), pp. 103–104.

[13] E. Fermi. 'Tentativo di una Teoria Dei Raggi β'. In: *Il Nuovo Cimento (1924-1942)* 11.1 (2008), p. 1.

[14] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes and R. P. Hudson. 'Experimental Test of Parity Conservation in Beta Decay'. In: *Physical Review* 105 (4 1957), pp. 1413–1415.

[15] N. Cabibbo. 'Unitary Symmetry and Leptonic Decays'. In: *Physical Review Letters* 10 (12 1963), pp. 531–533.

[16] Y. Nambu. 'Axial Vector Current Conservation in Weak Interactions'. In: *Physical Review Letters* 4 (7 1960), pp. 380–382.

[17] M. Gell-Mann. 'The symmetry group of vector and axial vector currents'. In: *Physics Physique Fizika* 1 (1 1964), pp. 63–75.

[18] M. Fierz. 'Über die künstliche Umwandlung des Protons in ein Neutron'. In: *Helvetica Physica Acta* 9 (1936), p. 245.

[19] V. A. Fock. 'Über die invariante Form der Wellen- und der Bewegungsgleichungen für einen geladenen Massenpunkt'. In: *Zeitschrift für Physik* 39.2 (1926), pp. 226–232.

[20] C. N. Yang and R. L. Mills. 'Conservation of Isotopic Spin and Isotopic Gauge Invariance'. In: *Physical Review* 96 (1 1954), pp. 191–195.

[21] J. S. Schwinger. 'A Theory of the Fundamental Interactions'. In: *Annals of Physics* 2 (1957), pp. 407–434.

[22] S. L. Glashow. 'Partial Symmetries of Weak Interactions'. In: *Nuclear Physics* 22 (1961), pp. 579–588.

[23] S. Weinberg. 'A model of leptons'. In: *Physical Review Letters* 19 (1967), p. 1264.

[24] A. Salam and J.C. Ward. 'Electromagnetic and weak interactions'. In: *Physics Letters* 13.2 (1964), pp. 168 –171.

[25] S. L. Glashow, J. Iliopoulos and L. Maiani. 'Weak Interactions with Lepton-Hadron Symmetry'. In: *Physical Review D* 2 (7 1970), pp. 1285–1292.

[26] G. 'tHooft. 'Renormalization of massless Yang-Mills fields'. In: *Nuclear Physics B* 33.1 (1971), pp. 173 –199.

[27] G.'t Hooft. 'Renormalizable Lagrangians for massive Yang-Mills fields'. In: *Nuclear Physics B* 35.1 (1971), pp. 167 –188.

[28] Y. Fukuda et al. 'Measurement of the Flux and Zenith-Angle Distribution of Upward Throughgoing Muons by Super-Kamiokande'. In: *Physical Review Letters* 82.13 (1999), pp. 2644–2648.

[29] Q. R. Ahmad et al. 'Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory'. In: *Physical Review Letters* 89.1 (2002).

[30] K. Eguchi et al. 'First Results from KamLAND: Evidence for Reactor Antineutrino Disappearance'. In: *Physical Review Letters* 90.2 (2003).

[31] M. Kobayashi and T. Maskawa. 'CP Violation in the Renormalizable Theory of Weak Interaction'. In: *Chinese Physics* 49 (1973), pp. 652–657.

[32] C. Patrignani et al. (Particle Data Group). 'Review of particle physics'. In: *Chinese Physics C* (40, 100001 (2016)).

[33] The CMS collaboration. 'Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC'. In: *Physics Letters B* 716 (2012), p. 30.

[34] The ATLAS collaboration. 'Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC'. In: *Physics Letters B* 716 (2012), p. 1.

[35] J. Goldstone. 'Field theories with "Superconductor" solutions'. In: *Il Nuovo Cimento* 19.1 (1961), pp. 154–164.

[36] J. Goldstone, A. Salam and S. Weinberg. 'Broken Symmetries'. In: *Physical Review* 127 (3 1962), pp. 965–970.

[37] J. Ellis, M. K. Gaillard and D. V. Nanopoulos. 'A historical profile of the Higgs boson'. In: *The Standard Theory of Particle Physics* (2016), pp. 255–274.

[38] F. Englert and R. Brout. 'Broken symmetries and the mass of gauge vector mesons'. In: *Physical Review Letters* 13 (1964), p. 321.

[39] P. W. Higgs. 'Broken symmetries and the masses of gauge bosons'. In: *Physical Review Letters* 13 (1964), p. 508.

[40] G. S. Guralnik, C. R. Hagen and T. W. B. Kibble. 'Global Conservation Laws and Massless Particles'. In: *Physical Review Letters* 13 (20 1964), pp. 585–587.

[41] F.J. Hasert et al. 'Observation of Neutrino Like Interactions Without Muon Or Electron in the Gargamelle Neutrino Experiment'. In: *Physics Letters B* 46 (1973), pp. 138–140.

[42] The UA1 collaboration. 'Experimental observation of isolated large transverse energy electrons with associated missing energy at $s = 540\,GeV$'. In: *Physics Letters B* 122 (1983), p. 103.

[43] S. van der Meer. 'Stochastic damping of betatron oscillations in the ISR'. In: (Aug. 1972). Ed. by C. Pellegrini and A.M. Sessler, pp. 261–269.

[44] The ALEPH, DELPHI, L3, OPAL Collaborations, the LEP Electroweak Working Group. 'Electroweak Measurements in Electron-Positron Collisions at W-Boson-Pair Energies at LEP'. In: *Physics Reports* 532 (2013), p. 119.

[45] The ALEPH, DELPHI, L3, OPAL, SLD Collaborations, the LEP Electroweak Working Group, the SLD Electroweak and Heavy Flavour Groups. 'Precision Electroweak Measurements on the Z Resonance'. In: *Physics Reports* 427 (2006), p. 257.

[46] P. Abreu et al. 'Search for new phenomena using single photon events in the DELPHI detector at LEP'. In: *Zeitschrift fur Physik C* 74 (1997), pp. 577–586.

[47] M. Acciarri et al. 'Determination of the number of light neutrino species from single photon production at LEP'. In: *Physics Letters B* 431 (1998), pp. 199–208.

[48] R. Akers et al. 'Measurement of single photon production in e+ e- collisions near the Zo resonance'. In: *Zeitschrift fur Physik C* 65 (1995), pp. 47–66.

[49] J. E. Augustin et al. 'Discovery of a Narrow Resonance in e^+e^- Annihilation'. In: *Physical Review Letters* 33 (23 1974), pp. 1406–1408.

[50] J. J. Aubert et al. 'Experimental Observation of a Heavy Particle J'. In: *Physical Review Letters* 33 (23 1974), pp. 1404–1406.

[51] S. W. Herb et al. 'Observation of a Dimuon Resonance at 9.5 GeV in 400-GeV Proton-Nucleus Collisions'. In: *Physical Review Letters* 39 (5 1977), pp. 252–255.

[52] The CDF collaboration. 'Observation of top quark production in $\overline{\mathrm{p}}\mathrm{p}$ collisions with the Collider Detector at Fermilab'. In: *Physical Review Letters* 74 (1995), p. 2626.

[53] The D0 collaboration. 'Search for high mass top quark production in $\overline{\mathrm{p}}\mathrm{p}$ collisions at $\sqrt{s} = 1.8\,\mathrm{TeV}$'. In: *Physical Review Letters* 74 (1995), p. 2422.

[54] The CMS Collaboration. 'Public plots for Higgs $\rightarrow \mu\mu$'. CMS Collection. 2020.

[55] E. Corbelli and P. Salucci. 'The extended rotation curve and the dark matter halo of M33'. In: *Monthly Notices of the Royal Astronomical Society* 311.2 (2000), pp. 441–447.

[56] X. Wu, T. Chiueh, L. Fang and Y. Xue. 'A comparison of different cluster mass estimates: consistency or discrepancy?' In: *Monthly Notices of the Royal Astronomical Society* 301.3 (1998), pp. 861–871.

[57] W. J. Percival, S. Cole, D. J. Eisenstein, R. C. Nichol, J. A. Peacock, A. C. Pope and A. S. Szalay. 'Measuring the Baryon Acoustic Oscillation scale using the Sloan Digital Sky Survey and 2dF Galaxy Redshift Survey'. In: *Monthly Notices of the Royal Astronomical Society* 381.3 (2007), pp. 1053–1066.

[58] Q. R. Ahmad et al. 'Measurement of the Rate of $\nu_e + d \rightarrow \pi\pi e^-$ Interactions Produced by ^{8}B Solar Neutrinos at the Sudbury Neutrino Observatory'. In: *Physical Review Letters* 87.7 (2001).

[59] S. Fukuda et al. 'Determination of solar neutrino oscillation parameters using 1496 days of Super-Kamiokande-I data'. In: *Physics Letters B* 539 (2002), p. 179.

[60] Y. Abe et al. 'Indication for the disappearance of reactor electron antineutrinos in the Double Chooz experiment'. In: *Physical Review Letters* 108.13 (2012).

[61] F. P. An et al. 'Observation of Electron-Antineutrino Disappearance at Daya Bay'. In: *Physical Review Letters* 108.17 (2012).

[62] Y. Fukuda et al. 'Evidence for Oscillation of Atmospheric Neutrinos'. In: *Physical Review Letters* 81.8 (1998), pp. 1562–1567.

[63] B. Pontecorvo. 'Neutrino experiments and the problem of conservation of leptonic charge'. In: *Soviet Physics JETP* 26 (1968), p. 984.

[64] I. J. R. Aitchison A. J. G. Hey. *Gauge theories in particle physics, Vol. 1*. Fourth. CRC Press, 2013.

[65] J. D. Bjorken. 'Asymptotic sum rules at infinite momentum'. In: *Physical Review* 179 (1969), p. 1547.

[66] R. P. Feynman. 'Very high-energy collisions of hadrons'. In: *Physical Review Letters* 23 (1969), p. 1415.

[67] M. Riordan. 'The discovery of quarks'. In: *Science* (Vol. 256, Issue 5061, pp. 1287-1293).

[68] The CHARM Collaboration. 'Experimental Study of Neutral Current and Charged Current Neutrino Cross-Sections'. In: *Physics Letters* (99B (1981) 265).

[69] I. J. R. Aitchison A. J. G. Hey. *Gauge theories in particle physics, Vol. 2*. Institute of physics publishing, 2004.

[70] Jr. Callan C. G. and D. J. Gross. 'High-energy electroproduction and the constitution of the electric current'. In: *Physical Review Letters* 22 (1969), pp. 156–159.

[71] Y. L. Dokshitzer. 'Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics.' In: *Soviet Physics JETP* 46 (1977), pp. 641–653.

[72] V.N. Gribov and L.N. Lipatov. 'Deep inelastic e p scattering in perturbation theory'. In: *Soviet JournalProgress of Theoretical Physics* 15 (1972), pp. 438–450.

[73] G. Altarelli G. Parisi. 'Asymptotic freedom in parton language'. In: *Nuclear Physics B* (Volume 126, Issue 2, 8 August 1977, Pages 298-318).

[74] H. Abramowicz et al. 'Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data'. In: *The European Physical Journal C* 75.12 (2015), p. 580.

[75] R. D. Ball et al. 'Parton distributions from high-precision collider data. Parton distributions from high-precision collider data'. In: *European Physical Journal C* 77.CAVENDISH-HEP-17-06. 10 (2017), 663. 95 p.

[76] C. Degrande, N. Greiner, W. Kilian, O. Mattelaer, H. Mebane, T. Stelzer, S. Willenbrock and C. Zhang. 'Effective field theory: A modern approach to anomalous couplings'. In: *Annals of Physics* 335 (2013), pp. 21–32.

[77] S. Weinberg. 'Baryon- and Lepton-Nonconserving Processes'. In: *Physical Review Letters* 43 (21 1979), pp. 1566–1570.

[78] G. Perez, M. Sekulla and D. Zeppenfeld. 'Anomalous quartic gauge couplings and unitarization for the vector boson scattering process $pp \to W^+W^+jjX \to \ell^+\nu_\ell\ell^+\nu_\ell jjX$'. In: *The European Physical Journal C* 78.9 (2018).

[79] O. J. P. Éboli, M. C. Gonzalez-Garcia and J. K. Mizukoshi. '$pp \to jje\pm\mu\pm\nu\nu$ and $jje\pm\mu\mp\nu\nu$ at $O(\alpha_{em}^6)$ and $O(\alpha_{em}^4\alpha_s^2)$ for the study of the quartic electroweak gauge boson vertex at CERN LHC'. In: *Physical Review D* 74.7 (2006).

[80] L. Evans and P. Bryant. 'LHC Machine'. In: *Journal of Instrumentation* 3.08 (2008), S08001–S08001.

[81] Apollinari G., Bejar Alonso I., Bruning O., Fessia P., Lamont M., Rossi L. and Tavian L. *High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1*. CERN Yellow Reports: Monographs. Geneva: CERN, 2017.

[82] The CMS Collaboration. 'The CMS experiment at the CERN LHC'. In: *Journal of Instrumentation* 3.08 (2008), S08004.

[83] M. Brice and C. Marcelloni. 'The central part of CMS is lowered. Descente dans la caverne de la partie centrale du detecteur de particules CMS (Compact Muon Solenoid).' 2007.

[84] M. Brice. 'First half of CMS inner tracker barrel'. 2006.

[85] A. Dominguez et al. *CMS Technical Design Report for the Pixel Detector Upgrade*. Tech. rep. CERN-LHCC-2012-016. CMS-TDR-11. 2012.

[86] P. Schleper, G. Steinbrueck and M. Stoye. *Software Alignment of the CMS Tracker using MILLEPEDE II*. Tech. rep. CMS-NOTE-2006-011. Geneva: CERN, 2006.

[87] E. Widl, R. Fruhwirth and W. Adam. *A Kalman Filter for Track-based Alignment*. Tech. rep. CMS-NOTE-2006-022. Geneva: CERN, 2006.

[88] M. Brice. 'Images of the CMS ECAL Barrel (EB)'. CMS Collection. 2008.

[89] M. Hoch. 'Early Construction: Muon Chambers'. CMS Collection. 2007.

[90] The CMS collaboration. 'Particle-flow reconstruction and global event description with the CMS detector'. In: *Journal of Instrumentation* 12.10 (2017), P10003–P10003.

[91] D. Buskulic et al. 'Performance of the ALEPH detector at LEP'. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 360.3 (1995), pp. 481 –506.

[92] W. Adam, B. Mangano, T. Speer and T. Todorov. *Track Reconstruction in the CMS tracker*. Tech. rep. CMS-NOTE-2006-041. Geneva: CERN, 2006.

[93] W. Adam, R. Fraoehwirth, A. Strandlie and T. Todorov. 'Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC'. In: *Journal of Physics G: Nuclear and Particle Physics* 31.9 (2005), N9–N20.

[94] J. Allison et al. 'Geant4 developments and applications'. In: *IEEE Transactions on Nuclear Science* 53.1 (2006), pp. 270–278.

[95] The CMS collaboration. 'Performance of CMS muon reconstruction in pp collision events at $\sqrt{s}$ = 7 TeV'. In: *Journal of Instrumentation* 7.10 (2012), P10002–P10002.

[96] S. Frixione, Z. Kunszt and A. Signer. 'Three-jet cross sections to next-to-leading order'. In: *Nuclear Physics B* 467.3 (1996), pp. 399–442.

[97] S. Catani and M.H. Seymour. 'A general algorithm for calculating jet cross sections in NLO QCD'. In: *Nuclear Physics B* 485.1-2 (1997), pp. 291–419.

[98] S. Frixione. 'A general approach to jet cross sections in QCD'. In: *Nuclear Physics B* 507.1-2 (1997), pp. 295–314.

[99] D. A. Kosower. 'Antenna factorization of gauge-theory amplitudes'. In: *Physical Review D* 57.9 (1998), pp. 5410–5416.

[100] J.M. Campbell, M.A. Cullen and E.W.N. Glover. 'Four jet event shapes in electron-positron annihilation'. In: *The European Physical Journal C* 9.2 (1999), pp. 245–265.

[101] Z. Bern, L. Dixon, D. C. Dunbar and D. A. Kosower. 'One-loop n-point gauge theory amplitudes, unitarity and collinear limits'. In: *Nuclear Physics B* 425.1-2 (1994), pp. 217–260.

[102] F. del Aguila and R Pittau. 'Recursive numerical calculus of one-loop tensor integrals'. In: *Journal of High Energy Physics* 2004.07 (2004), pp. 017–017.

[103] Z. Bern, L. J. Dixon and D.A. Kosower. 'Bootstrapping multiparton loop amplitudes in QCD'. In: *Physical Review D* 73.6 (2006).

[104] G. Ossola, C. G. Papadopoulos and R. Pittau. 'Reducing full one-loop amplitudes to scalar integrals at the integrand level'. In: *Nuclear Physics B* 763.1-2 (2007), pp. 147–169.

[105] C. Anastasiou, R. Britto, B. Feng, Z. Kunszt and P. Mastrolia. 'd-dimensional unitarity cut method'. In: *Physics Letters B* 645.2-3 (2007), pp. 213–216.

[106] C. Anastasiou, R. Britto, B. Feng, Z. Kunszt and P. Mastrolia. 'Unitarity cuts and reduction to master integrals inddimensions for one-loop amplitudes'. In: *Journal of High Energy Physics* 2007.03 (2007), pp. 111–111.

[107] R.K. Ellis, W.T. Giele and Z. Kunszt. 'A numerical unitarity formalism for evaluating one-loop amplitudes'. In: *Journal of High Energy Physics* 2008.03 (2008), pp. 003–003.

[108] R. K. Ellis, W. T. Giele, Z. Kunszt and K. Melnikov. 'Masses, fermions and generalized D-dimensional unitarity'. In: *Nuclear Physics B* 822.1-2 (2009), pp. 270–282.

[109] W. T Giele, Z. Kunszt and K. Melnikov. 'Full one-loop amplitudes from tree amplitudes'. In: *Journal of High Energy Physics* 2008.04 (2008), pp. 049–049.

[110] F. Cascioli, P. Maierhofer and S. Pozzorini. 'Scattering Amplitudes with Open Loops'. In: *Physical Review Letters* 108.11 (2012).

[111] P. Mastrolia, E. Mirabella and T. Peraro. 'Integrand reduction of one-loop scattering amplitudes through Laurent series expansion'. In: *Journal of High Energy Physics* 2012.6 (2012).

[112] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer and Tim Stelzer. 'MadGraph 5: going beyond'. In: *Journal of High Energy Physics* 2011.6 (2011).

[113] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli and M. Zaro. 'The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations'. In: *Journal of High Energy Physics* 2014.7 (2014).

[114] V. Hirschi, R. Frederix, S. Frixione, M. Vittoria Garzelli, F. Maltoni and R. Pittau. 'Automation of one-loop QCD computations'. In: *Journal of High Energy Physics* 2011.5 (2011).

[115] C. Degrande, C. Duhr, B. Fuks, D. Grellscheid, O. Mattelaer and T. Reiter. 'UFO - The Universal FeynRules Output'. In: *Computer Physics Communications* 183.6 (2012), pp. 1201–1214.

[116] P. de Aquino, W. Link, F. Maltoni, O. Mattelaer and T. Stelzer. 'ALOHA: Automatic libraries of helicity amplitudes for Feynman diagram computations'. In: *Computer Physics Communications* 183.10 (2012), pp. 2254–2263.

[117] J. M. Butterworth et al. *The tools and Monte Carlo working group: Summary Report from the Les Houches 2009 Workshop on TeV Colliders*. 2010.

[118] J. Alwall et al. 'A standard format for Les Houches Event Files'. In: *Computer Physics Communications* 176.4 (2007), pp. 300–304.

[119] M. L Mangano, M. Moretti, F. Piccinini and M. Treccani. 'Matching matrix elements and shower evolution for top-pair production in hadronic collisions'. In: *Journal of High Energy Physics* 2007.01 (2007), pp. 013–013.

[120] R. Frederix and S. Frixione. 'Merging meets matching in MC@NLO'. In: *Journal of High Energy Physics* 2012.12 (2012).

[121] A. S. Belyaev, O. J. P. Èboli, M. C. Gonzalez-Garcia, J. K. Mizukoshi, S. F. Novaes and I. Zacharov. 'Strongly interacting vector bosons at the CERN LHC: Quartic anomalous couplings'. In: *Physical Review D* 59.1 (1998).

[122] O. J. P. Éboli, M. C. Gonzalez-Garcia, S. M. Lietti and S. F. Novaes. 'Anomalous quartic gauge boson couplings at hadron colliders'. In: *Physical Review D* 63.7 (2001).

[123] O. J. P. Éboli, M. C. Gonzalez-Garcia and S. M. Lietti. 'Bosonic quartic couplings at CERN LHC'. In: *Physical Review D* 69.9 (2004).

[124] O.J.P. Éboli and M.C. Gonzalez-Garcia. 'Classifying the bosonic quartic couplings'. In: *Physical Review D* 93.9 (2016).

[125] Törbjorn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen and Peter Z. Skands. 'An introduction to PYTHIA 8.2'. In: *Computer Physics Communications* 191 (2015), pp. 159–177.

[126] Bo Andersson, G. Gustafson, G. Ingelman and T. Sjöstrand. 'Parton Fragmentation and String Dynamics'. In: *Physics Reports* 97 (1983), pp. 31–145.

[127] T. Sjöstrand. 'Jet fragmentation of multiparton configurations in a string framework'. In: Nuclear Physics, Section B 248.2 (Dec. 1984), pp. 469–502.

[128] T. Sjöstrand and M. van Zijl. 'A multiple-interaction model for the event structure in hadron collisions'. In: *Physical Review D* 36 (7 1987), pp. 2019–2041.

[129] The CMS Collaboration. 'Extraction and validation of a new set of CMS pythia8 tunes from underlying-event measurements'. In: *The European Physical Journal C* 80.1 (2020).

[130] T. Gleisberg, S. Hoeche, F. Krauss, A. Schaelicke, S. Schumann and J. Winter. 'SHERPA 1.$_7$ a proof-of-concept version'. In: *Journal of High Energy Physics* 2004.02 (2004), pp. 056–056.

[131] T. Gleisberg, S. Hoeche, F. Krauss, M. Schoenherr, S. Schumann, F. Siegert and J. Winter. 'Event generation with SHERPA 1.1'. In: *Journal of High Energy Physics* 2009.02 (2009), pp. 007–007.

[132] P. Nason. 'A New method for combining NLO QCD with shower Monte Carlo algorithms'. In: *Journal of High Energy Physics* 11 (2004), p. 040.

[133] S. Frixione, P. Nason and C. Oleari. 'Matching NLO QCD computations with Parton Shower simulations: the POWHEG method'. In: *Journal of High Energy Physics* 11 (2007), p. 070.

[134] S. Alioli, P. Nason, C. Oleari and E. Re. 'A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX'. In: *Journal of High Energy Physics* 06 (2010), p. 043.

[135] S. Agostinelli et al. 'Geant4 - a simulation toolkit'. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250 –303.

[136] S. Alioli, P. Nason, C. Oleari and E. Re. 'A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX'. In: *Journal of High Energy Physics* 2010.6 (2010).

[137] The CMS Collaboration. 'Event generator tunes obtained from underlying event and multiparton scattering measurements'. In: *The European Physical Journal C* 76.3 (2016).

[138] The NNPDF collaboration. 'Parton distributions for the LHC run II'. In: *Journal of High Energy Physics* 2015.4 (2015), p. 40.

[139] Y. Li and F. Petriello. 'Combining QCD and electroweak corrections to dilepton production in the framework of the FEWZ simulation code'. In: *Physical Review D* 86 (9 2012), p. 094034.

[140] M. Grazzini, S. Kallweit and M. Wiesemann. 'Fully differential NNLO computations with MATRIX'. In: *The European Physical Journal C* 78.7 (2018).

[141] A. Bodek, A. van Dyne, J. Y. Han, W. Sakumoto and A. Strelnikov. 'Extracting muon momentum scale corrections for hadron collider experiments'. In: *The European Physical Journal C* 72.10 (2012).

[142] 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.

[143] G. Cowan, K. Cranmer, E. Gross and O. Vitells. 'Asymptotic formulae for likelihood-based tests of new physics'. In: *The European Physical Journal C* 71.2 (2011).

[144] G. Watt and R. S. Thorne. 'Study of Monte Carlo approach to experimental uncertainty propagation with MSTW 2008 PDFs'. In: *Journal of High Energy Physics* 2012.8 (2012).

[145] A. Buckley, J. Butterworth, D. Grellscheid, H. Hoeth, L. Lönnblad, J. Monk, H. Schulz and F. Siegert. 'Rivet user manual'. In: *Computer Physics Communications* 184.12 (2013), pp. 2803–2819.

[146] G. Sorrentino. 'Characterization of Z $\gamma\gamma$ and W $\gamma\gamma$ events in pp collisions at the LHC and limits on anomalous couplings'. 2019.

[147] S. S. Wilks. 'The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses'. In: *Annals of Mathematical Statistics* 9.1 (Mar. 1938), pp. 60–62.

[148] The CMS Collaboration. 'Measurement of the cross section for electroweak production of a Z boson, a photon and two jets in proton-proton collisions at $\sqrt{s}$ = 13 TeV and constraints on anomalous quartic couplings'. In: *Journal of High Energy Physics* 2020.6 (2020).

[149] The CMS Collaboration. 'Measurement of electroweak WZ boson production and search for new physics in WZ + two jets events in pp collisions at s=13TeV'. In: *Physics Letters B* 795 (2019), pp. 281–307.

[150] The CMS Collaboration. 'Search for the production of $W^{\pm}W^{\pm}W^{\mp}$ events at $\sqrt{s} = 13$ TeV'. In: *Physical Review D* 100.1 (2019), p. 012004.

[151] The CMS Collaboration. 'Observation of electroweak production of Wγ with two jets in proton-proton collisions at $\sqrt{s} = 13$ TeV'. In: (Aug. 2020).

[152] The CMS Collaboration. 'Measurements of production cross sections of WZ and same-sign WW boson pairs in association with two jets in proton-proton collisions at $\sqrt{s} = 13$ TeV'. In: *Physics Letters B* 809 (2020), p. 135710.

[153] The CMS Collaboration. 'Evidence for electroweak production of four charged leptons and two jets in proton-proton collisions at $\sqrt{s} = 13$ TeV'. In: (Aug. 2020).

[154] The CMS Collaboration. 'Search for anomalous electroweak production of vector boson pairs in association with two jets in proton-proton collisions at 13 TeV'. In: *Physics Letters B* 798 ().

[155] The ATLAS Collaboration. 'Studies of Zγ production in association with a high-mass dijet system in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector'. In: *Journal of High Energy Physics* 07 (2017), p. 107.