SAPIENZA
UNIVERSITÀ DI ROMA

UNIVERSITÉ LIBRE DE BRUXELLES
CO-TUTELLE AVEC SAPIENZA-UNIVERSITÀ DI ROMA
FACULTÉ DES SCIENCES

Search for new physics in dielectron and diphoton final states at CMS

Thèse présentée par:

Giuseppe Fasanella

En vue de l'obtention du grade de:

Docteur en Sciences

Bruxelles, 15 Septembre 2017

Constitution du jury de thèse:

- Prof. Barbara Clerbaux (ULB, Bruxelles), promoteur
- Dr. Paolo Meridiani (Sapienza, Roma), co-promoteur
- Prof. Daniele Del Re (Sapienza, Roma)
- Prof. Louis Fayard (LAL, Orsay)
- Dr. Ivan Mikulec (HEPHY, Vienna)
- Prof. Michel Tytgat (ULB, Bruxelles)
- Prof. Pascal Vanlaer (ULB, Bruxelles)

Evaluateurs externes:

- Prof. Attilio Andreazza (UniMi, Milano)
- Prof. Luigi Rolandi (SNS, Pisa)

Abstract

This thesis describes searches for new massive resonances that decay in an electron-positron or photon pair in the final state. Different datasets, coming from proton-proton collisions at a center-of-mass energy $\sqrt{s} = 13$ TeV at the Large Hadron Collider (LHC) and collected by the CMS experiment in 2015 and 2016, have been analyzed. After a chapter devoted to the description of the standard model of elementary particle physics, the motivation for the introduction of new theories that go beyond the standard model are introduced and some classes of models are described. The techniques used in order to reconstruct the particles produced in the collisions are discussed afterwards, with a special emphasis on electron/positron and photon reconstructions. Two separate analyses are presented.

The first one is the search for new heavy resonances decaying in an electron-positron pair in the final state. Such resonances are predicted by a variety of models such as grand unified theories or theories that introduce extra space-like dimensions. Their signature would appear as a localised excess of events in the electron-positron invariant mass spectrum. The event selection is optimized in order to be highly efficient for high-energy electrons/positrons and to avoid loosing potential signal events. The analysis relies on simulated samples for the estimation of the main source of background, which is the standard model Drell-Yan process. Data-driven approaches are pursued for both validating the simulation of the subleading background processes with prompt electrons in the final state and the determination of the background coming from processes of quantum chromo dynamics. After having inspected the electron-positron invariant mass, no excess over the standard model expectation is observed, and 95% confidence level upper limits are set on the ratio of production cross-section times branching ratio of a new resonance to the one at the Z boson peak, using the data collected in 2016 (35.9 fb^{-1}).

The second analysis presented in this thesis is the search for new heavy resonances decaying in the diphoton final state, the existence of which is predicted by models with non-minimal scalar sectors or by theories postulating the existence of additional space-like dimensions. Their signature would appear as a localised excess of events in the diphoton invariant mass spectrum. As for the case of the dielectron analysis, the event selection has been optimized in order to be highly efficient for high-energy photons. The background estimation is completely data-driven and achieved via a parametrization of the observed diphoton invariant mass spectrum. After the inspection of the diphoton invariant mass, no excess over the standard model expectation is observed, and 95% confidence level upper limits are set on the production cross-section times branching ratio, using the data collected in the first half of 2016 (12.9 fb^{-1}). Results have also been combined with those obtained with the same analysis techniques but with different datasets collected in 2012 and 2015 by the CMS experiment.

Résumé

Cette thèse décrit les recherches de nouvelles résonances massives se décomposant en une paire électron-positron ou une paire de photons dans l'état final. Différents ensembles de données, provenant des collisions proton-proton à une énergie dans le centre de masse $\sqrt{s} = 13$ TeV au Large Hadron Collider (LHC) et enregistrés par l'expérience CMS en 2015 et 2016, ont été analysés. Après un chapitre consacré à la description du modèle standard de la physique des particules élémentaires, les motivations pour postuler de nouvelles théories qui dépassent le modèle standard sont introduites et certaines classes de modèles sont décrites. Les techniques utilisées pour reconstituer les particules produites dans les collisions sont discutées ensuite, avec un accent particulier sur la reconstruction des électrons/positrons et des photons. Deux analyses séparées sont présentées.

La première est la recherche d'une nouvelle résonance massive se désintégrant en une paire électron-positron dans l'état final. De telles résonances sont prédites par une variété de modèles tels que les théories de grande unification ou les théories introduisant des dimensions supplémentaires. Leur signature apparaîtrait comme un excès localisé d'événements dans le spectre de masse invariante des paires électron-positron. La sélection des événements est optimisée pour être très efficace pour les électrons/positron à haute énergie et éviter de perdre des événements de signal potentiels. L'analyse repose sur des échantillons simulés pour l'estimation de la principale source de bruit de fond, qui est le processus Drell-Yan du modèle standard. Des approches basées sur les données sont poursuivies pour la validation de la simulation du bruit de fond avec des électrons/positrons produits dans l'état final et pour la détermination du bruit de fond provenant de processus de *quantum chromodynamics*. Après avoir inspecté la distribution de masse invariante des paires électron-positron, aucun excès sur la prédiction du modèle standard n'est observée, et des limites supérieures à 95% de niveau de confiance sont placées sur le rapport entre la section efficace de production multipliée par le rapport de branchement d'une nouvelle résonance et cette même quantité au pic du boson Z, avec les données prises en 2016 (35.9 fb^{-1}).

La deuxième analyse présentée dans cette thèse est la recherche d'une nouvelle résonance lourde se désintégrant en une paire de photons dans l'état final, dont l'existence est prédite par des modèles avec des secteurs scalaires non minimaux ou par des théories qui postulent l'existence de dimensions spatiales supplémentaires. Leur signature apparaîtrait comme un excès localisé d'événements dans le spectre de masse invariante des paires de photons. La sélection des événements a été optimisée afin d'être très efficace pour les photons à haute énergie et d'éviter la perte de signaux potentiels. L'estimation du bruit de fond est entièrement basée sur les données et réalisée avec une paramétrisation du spectre de masse invariante observé. Après l'inspection de la masse invariante des paires de photons, aucun excès par rapport à la prédiction du modèle standard n'est observé, et des limites supérieures à 95% de niveau de confiance sont placées sur la section efficace de production multipliée par le rapport de branchement d'une nouvelle résonance, avec les données prises jusqu'à Juillet 2016 (12.9 fb^{-1}). Les résultats ont également été combinés avec ceux obtenus par les mêmes techniques d'analyse, mais avec des ensembles de données différents enregistrées par l'expérience CMS en 2012 et 2015.

Sommario

Questa tesi descrive le ricerche di nuove risonanze massive che decadono in coppie elettrone-positrone o in coppie di fotoni nello stato finale. I dati analizzati, provenienti da collisioni protone-protone a una energia del centro di massa $\sqrt{s} = 13$ TeV sono stati raccolti nel 2015 e nel 2016 dall'esperimento CMS presso il Large Hadron Collider (LHC). Dopo un capitolo dedicato alla descrizione del modello standard delle particelle elementari, vengono analizzate le motivazioni per l'introduzione di nuove teorie che vanno al di là del modello standard e alcune classi di modelli sono descritte. Le tecniche utilizzate per ricostruire le particelle prodotte nelle collisioni sono in seguito discusse, con una particolare enfasi sulla ricostruzione di elettroni e fotoni. Due distinte analisi sono presentate.
La prima analisi verte sulla ricerca di nuove resonanze massive che decadono in coppie elettrone-positrone nello stato finale. Tali risonanze sono previste da numerosi modelli di nuova fisica come le *grand unified theories* o da teorie che introducono dimensioni spaziali aggiuntive. Una evidenza sperimentale dell'esistenza di tali particelle nuove apparirebbe come un eccesso localizzato di eventi nello spettro di massa invariante delle coppie elettrone-positrone. La selezione degli eventi è stata ottimizzata per essere altamente efficiente per elettroni di alta energia ed evitare di perdere eventuali eventi di segnale. L 'analisi si basa su campioni simulati per la stima della principale fonte di fondo, che è costituito dal processo Drell-Yan del modello standard. Un approccio basato sui dati viene invece perseguito sia per convalidare la simulazione del fondo sottodominante fatto di processi con elettroni nello stato finale e anche per la determinazione del fondo proveniente da processi di *quantum chromodynamics*. Dopo aver esaminato lo spettro di massa invariante delle coppie elettrone-positrone selezionate, nessun eccesso significativo è stato rilevato rispetto alle predizioni del modello standard e, di conseguenza, sono stati calcolati, utilizzando i dati raccolti nel 2016 (35.9 fb^{-1}), limiti superiori al 95% di livello di confidenza sul rapporto tra la sezione d'urto di produzione moltiplicata per il rapporto di decadimento di una nuova risonanza rispetto a quella del bosone Z.
La seconda analisi presentata in questa tesi è la ricerca di una nuova risonanza pesante che decade in coppie di fotoni nello stato finale, la cui esistenza è prevista da modelli con settori Higgs non minimali o da teorie che postulano l'esistenza di dimensioni spaziali aggiuntive. In questo caso, una loro evidenza sperimentale apparirebbe come un eccesso localizzato di eventi nello spettro di massa invariante delle coppie di fotoni selezionate. Di conseguenza, la selezione degli eventi è stata ottimizzata per essere altamente efficiente per fotoni di alta energia ed evitare perdite di eventuali eventi di segnale. La stima del fondo segue un approccio basato sui dati stessi e viene ottenuta tramite una parametrizzazione dello spettro di massa invariante osservato. Dopo l'ispezione della massa invariante delle coppie di fotoni selezionate, non si osserva nessun eccesso significativo rispetto alle predizioni del modello standard e pertanto, utilizzando i dati raccolti nella prima metà del 2016 (12.9 fb^{-1}), sono posti limiti superiori al 95% di livello di confidenza sulla sezione d'urto di produzione moltiplicata per il rapporto di decadimento di una tale risonanza. I risultati sono stati anche combinati con quelli ottenuti con le stesse tecniche di analisi ma per diversi set di dati, raccolti dall'esperimento CMS nel 2012 e nel 2015.

Acknowledgements

This thesis would have not been possible without the contribution of many people that helped me during the four years of doctoral studies.
First, I would like to thank Barbara Clerbaux and Paolo Meridiani who supervised to the work presented here with their guidance, support and countless advices during this extreme adventure of the spirit and the body which is named *PhD*.
I am grateful to all the jury members, Daniele Del Re, Louis Fayard, Ivan Mikulec, Michel Tytgat and Pascal Vanlaer, for reading this thesis and giving me their valuable comments.
I also want to thank the external evaluators Attilio Andreazza and Luigi Rolandi for their remarks and positive evaluations of this manuscript.
Major improvements to the quality of the analyses came over the years thanks to many people of the CMS Collaboration during internal meetings, approval talks and conferences. In particular, I want to thank Sam Harper and Pasquale Musella for the time they devoted to me in discussions and explanations and the many things I learnt from them in these occasions.
A special acknowlegment goes to Shahram Rahatlou and the INFN-Roma1 section for the invaluable opportunity to work at CERN during a rather intense year. The 4-th flour of "building 32" became increasingly crowded during that period, with notable spikes during the coffee breaks.
For lots of laughs and good moments I have to thank Vittorio, Martina (collega ad honorem), Livia, acm, marco, federico, jellysim, giulia e checco.
My work at the IIHE would not have been that fruitful and pleasant without my colleagues and friends. It is impossible to name everyone here, but special thanks go to my office mates, Laurent, Thomas and Aidan who have shared the office with me during my first years and the present ones, Diego and David to which I wish all the best for their future.
In ultimo, un ringraziamento a parte va a babbo Alfonso per i prosciutti di montagna *"che solo l'odore ti ubriacava"*, a mamma Giovanna, studentessa modello e oramai specialista nell'imbastire valigie ricolme di vettovaglie e con un libro di Thomas Mann in un angolo, e a Teresa e Antonietta, le mie fonti di fiducia sugli antichi segreti romantici e rusticani di Lucania.

Contents

Introduction

The standard model of elementary particle physics is an elegant and well established theory able to make predictions which managed to explain the experimental observations with outstanding precision over the years. Among them, a major prediction of the standard model is the existence of an additional elementary scalar boson introduced in 1964 by Brout, Englert and Higgs in order to explain why the elementary particles have non-zero masses. The experimental proof of the existence of this scalar boson, known as the Higgs boson, which is the manifestation of a mechanism introduced in Chapter 1, took several decades to be achieved and was one of the main motivations that lead to the construction of the Large Hadron Collider (LHC).

The LHC is an accelerator and collider installed in the same circular underground tunnel occupied until the year 2000 by the Large Electron Positron collider (LEP) with a circumference of 26.7 km. The LHC is able to provide proton-proton collisions with the highest center-of-mass energy ever achieved in a collider (13 TeV in 2015). Four interaction points are located in the ring, where four experiments are placed. One of them is the Compact Muon Solenoid (CMS), which is, as the ATLAS experiment, a general purpose detector. Thanks to the joint efforts of the ATLAS and CMS collaborations, the Higgs boson has been discovered in 2012 (48 years after its prediction) at a mass of around 125 GeV, thus proving the existence of the missing piece of the standard model. This discovery, which is certainly one of the most important achievements of modern particle physics research, accomplished one of the main goals of the LHC program.

Nevertheless, despite its tremendous success, the standard model is not able to describe the full picture of Nature (it shows no dark matter candidate, it doesn't explain the matter-antimatter asymmetry, it does not explain neutrino masses, it doesn't include the gravitational force in its framework,...) and presents some issues of internal consistency (hierarchy problem, number of free parameters,...). The standard model is then generally considered as a low-energy effective model of a more fundamental theory. In order to address some of its shortcomings, several models that go beyond the standard model have been proposed during the last decades, such as Grand Unified Theories (GUT) or models introducing extra space-like dimensions (ED). In particular, with the aim of unifying the electroweak and the strong forces, GUT introduce new neutral bosons heavier than the standard model Z boson, which are generically called $Z^{'}$ bosons. On the other hand, ED require the existence of additional spatial dimensions, which should manifest themselves in our ordinary 4 dimensional space-time through as a series of excited states of SM particles, called "Kaluza-Klein modes". The ED models predict the existence of Kaluza-Klein excited and massive states for the graviton (G), the mediator of the gravitational force, and possibly for the gauge bosons.

After the discovery of the Higgs boson in 2012, the LHC underwent a long shutdown period which lasted until spring 2015, when the machine was restarted and operated at a center-of-mass energy of 13 TeV. The data-taking period from 2015 to 2018 is called RunII, while the previous one, corresponding to a maximal center-of-mass energy of 8

TeV, is called RunI and will be considered in this thesis only when the corresponding datasets are combined with the more recent ones collected in RunII. The main goal of the RunII period is the quest for new physics beyond the standard model.

In this context, the results showed in this thesis deal with the searches for new massive particles decaying in the electron-positron or diphoton final state. Such new particles would manifest themselves as a localised excess of events in the observed invariant mass spectra. Both channels have the advantage that their reconstruction can be very well understood, which leads to a low background contamination coming from misreconstructed electron/photon candidates. Before the LHC RunII era, the most stringent limits for these final states have been provided by the CMS and ATLAS experiments by analyzing the RunI dataset: for example the existence of a new massive Z boson with SM-like couplings (Z'_{SSM}) decaying into an electron-positron pair was excluded at 95% confidence level for masses below ∼2.7 TeV, while the search for a new particle decaying in photon pairs excluded the invariant mass region below 1.4 - 2.7 TeV depending on the specific model considered (see Chapter 2 for a more detailed description). The increase in the center-of-mass energy from 8 TeV at the end of RunI to 13 TeV of RunII implies a potential dramatic increase of the production cross-section of heavy resonances (see Chapter 3). This fact is a strong motivation for the analyses of these final states at RunII, given that a discovery could be possible with a relatively small amount of data. Both searches have been redefined with respect to their RunI version, in order to challenge the different experimental conditions while retaining high efficiency for high-energy electron/photon candidates respectively. In the description of both analyses a particular attention is devoted to the procedures concerning the energy calibration of the electromagnetic calorimeter of the CMS detector and the determination of the mass scale and resolution. The resolution of a (narrow) resonance is in fact completely dominated by the experimental resolution of the detector which has then to be precisely measured in order to be able to determine the expected signal model. This task is achieved by starting from the Z peak region exploiting the SM decay of the Z boson in electron-positron pairs. This provides an excellent handle for the energy calibration in the invariant mass range around 90 GeV, which can then be extended using boosted Z events up to higher invariant mass regions. Two different approaches have been pursued by the analyses (see Sections 6.4 and 7.4) which are proven to be consistent with each other at the level of 0.1%.

In the first dataset collected by the CMS experiment in 2015 resulting in an integrated luminosity of 3.3 fb^{-1}, a mild excess of events ($\sim$ 2.9 standard deviations in terms of local p-value) in the diphoton final state has been reported at a mass of around 750 GeV. A similar excess has also been reported by the ATLAS collaboration at the same mass. This provided a certain amount of excitement in the community and strongly motivated the analysis of the 2016 dataset, with a much larger amount of integrated luminosity with respect to the 2015 one.

I participated to the searches for new physics in the electron-positron final state and in the diphoton final state since the beginning of the RunII data-taking. The analyses described in this thesis lead to several publications over the years: for the electron-positron final state two public analysis summaries (PAS) have been published and another one is going to be published in the immediate future, while the results obtained analyzing the 2015 dataset in combination with the results obtained at RunI have been published as a paper of the CMS collaboration; for the diphoton final state, three PAS documents have been published and two papers. Only the latest available results are reported in this thesis: in particular for the electron-positron final state the full 2016 dataset is considered, while for the diphoton final state the first half of the 2016 dataset is considered, in combination with the full 2012 and 2015 datasets.

The thesis is organised as follows. The standard model of elementary particle physics is introduced in Chapter 1, including a discussion of the symmetries of the model, the scalar sector and the main properties of the standard model Drell-Yan process. Chapter 2 lists the shortcomings of the standard model and reports how various theories beyond the standard model propose to solve them. In particular, the models that predict additional massive resonances decaying in electron-positron or diphoton final state are introduced. The design and operational parameters of the LHC are discussed in Chapter 3 as well as the phenomenological aspects of the proton-proton interactions. Chapter 4 describes the main features of the CMS detector, while the reconstruction techniques of the particle candidates coming from the registered signals in the CMS detector are explained in Chapter 5. Chapter 6 describes in detail the results of the search for new resonances decaying in the electron-positron final state, covering all the aspects of the analysis. The results of the search for new resonances decaying in the diphoton final state are shown in Chapter 7, covering as well all the aspects of the analysis. Finally, Chapter 8 exposes the conclusions coming from both searches.

Chapter 1

The standard model of particle physics

This chapter introduces the standard model (SM) of elementary particle physics, which describes in its conceptual framework the elementary particles and three of the four fundamental forces of Nature. Its gauge structure, based on the $SU(3) \times SU(2) \times U(1)$ symmetry group, is exposed in general terms. Finally, the cross section of a specific process, the Drell-Yan production, is derived using the SM formalism.

1.1 Elementary particles and forces: a general picture

Despite the greek meaning of the word, the atom is not indivisible. It consists, in fact, of smaller components: electrons, protons and eventually neutrons. While in the SM the electron is considered to be an elementary particle with no substructure, protons and neutrons are made of quarks, which are in turn believed to be elementary. In particular, the proton is made of two up quarks with an electric charge of $= 2/3\ e$, where e is the Coulomb charge $e = 1.6 \times 10^{-19}$ C, and one down quark with an electric charge of $-1/3\ e$. For the neutron, the numbers of up and down quarks are reversed. Quarks and leptons have spin $\frac{1}{2}$ and follow the Fermi-Dirac statistic, hence they are fermions.

During the last century, the knowledge of the elementary particles have been increased thanks to both theoretical and experimental efforts of many individuals: our most up-to-date insight of the elementary constituents of nature organizes both leptons and quarks into three families of particles. Starting from the leptons, the electron e belongs to the first generation, together with the electron neutrino ν_e[1]. The muon μ and the muon neutrino ν_μ constitute the second generation of leptons, whereas the tau τ and the tau neutrino ν_τ form the third generation. The masses of the charged leptons differ by four orders of magnitude between the first and third generations. Table 1.1 summarizes the leptons and their properties[2].

Notably, the same organization in three families can be found also for quarks. In addition to the up and down quarks, which constitute the first generation, two further generations of quarks have been found: the charm and strange quarks are placed in the second generation and the top and bottom quarks in the third generation. The charm and the top quark have the same electric charge as the up quark, while the strange and the bottom quark have the same electric charge as the down quark. Beside their electric charge, the six quarks carry also color charge, hence they can interact via strong interaction (described later in this section). The quarks and their properties are shown in Table 1.2. In particular, the quark masses span 5 orders of magnitude from the ≈ 2

[1] The ν_e has been introduced by Wolfgang Pauli in order to allow energy-momentum conservation in the nuclear beta decay

[2] In this thesis all masses and energies are expressed in natural units, where the speed of light c and $\hbar$ are taken as equal to 1.

Generation	Lepton	Charge	Mass
First	electron (e)	-e	511 MeV
	electron neutrino (ν_e)	0	< 2 eV
Second	muon (μ)	-e	105.67 MeV
	muon neutrino (ν_μ)	0	< 2 eV
Third	tau (τ)	-e	1776.99 MeV
	tau neutrino (ν_τ)	0	< 2 eV

Table 1.1: Properties of the leptons in the three generations. Neutrinos are known to have a tiny mass compared to the other SM particles, but non-zero, so to allow oscillations between neutrino families, which have been experimentally observed [1].

MeV of mass of the up quark, up to the top quark, which was discovered in 1995 at the Tevatron, and has a mass close to 173.2 GeV. The top quark is in fact the heaviest SM particle carrying a mass which is close to the one of a gold nucleus.

Generation	Quark	Charge	Mass
First	up quark (u)	2/3 e	$2.3^{+0.7}_{-0.5}$ MeV
	down quark (d)	-1/3 e	$4.8^{+0.5}_{-0.3}$ MeV
Second	charm quark (c)	2/3 e	1.275 ± 0.025 GeV
	strange quark (s)	-1/3 e	95 ± 5 MeV
Third	top quark (t)	2/3 e	$173.21 \pm 0.51 \pm 0.71$ GeV
	bottom quark (b)	-1/3 e	4.66 ± 0.03 GeV

Table 1.2: Quarks and their properties [1].

Hence, in the SM there are 12 elementary particles, 6 leptons and 6 quarks, while ordinary matter on earth is essentially composed of particles belonging to the first generations: up and down quarks in the nucleus and electrons in the electron cloud. In addition to this, it is important to notice that every particle has its own antiparticle, which is characterised by having the same mass but opposite quantum numbers, as predicted by the Dirac's theory. The question whether the neutrinos are their own antiparticles (Majorana particles) or not (Dirac particles) remains open to date.

In order for the elementary particles to interact with each other, forces have to act between them. Forces are mediated by particles with an integer spin that follow the Bose-Einstein statistic and are hence called bosons. There are four fundamental interactions in Nature which can act or not on a specific particle depending if this particle carries or not the corresponding charge.

- Electromagnetic force

All electrically charged particles are subject to electromagnetic interactions. The massless and chargeless photon (γ) is the carrier particle of the electromagnetic force and, owing to its masslessness, the electromagnetic force is a long range force with a $1/r$ potential. The theory describing the electromagnetic interactions is called quantum electrodynamics (QED).

- Weak force

Weak interactions are usually not relevant at energies well below ≈ 100 GeV, since the strong or electromagnetic interactions have couplings that are orders of magnitude larger. This is not true anymore for processes where the electromagnetic interactions and the strong interactions are forbidden, for example because of a quantum number conservation

law taking place. Weak interactions can involve neutrinos which have neither electric nor strong charge and, therefore, do not interact strongly or electromagnetically. Processes involving the changing of quark flavour must happen via weak interaction since this is not allowed with the other interactions. Contrary to the photon (and the gluons), the bosons which carry the weak interaction are heavy compared to most elementary particles in SM. There exists a charged ($W^{\pm}$) and a neutral version (Z) of the vector bosons that mediate the weak force. The range of the weak interactions is small (between 10^{-18} and 10^{-16} m) because of the mass of the bosons involved.

- Strong force

The strong interactions are responsible for the attracting force between quarks. It is mediated by the gluons which are massless spin one particles. Gluons exist as a colour octet: they follow the "8" representation (adjoint) of the $SU(3)$ group. Gluons also carry color charge, hence they can interact with themselves, unlike the electrical neutral photon. Due to the property known as confinement, it is not possible to observe quarks as free particles: they would instead undergo the hadronization process, which allows to re-arrange the color structure from colored quarks to colorless hadrons. The top quark is a special case since, due to its mass, the decay into a b quark (which then hadronizes) and a W boson happens before the hadronization process itself could start. The theory describing the strong interactions is called quantum chromodynamics (QCD). A more detailed description of the QCD properties can be found in Chapter 3.

- Gravitational force

All particles are affected by the gravitational force. However, even though it is the dominant force on the astronomical scale, it is negligible on a microscopic scale and, in general, when the energies involved are lower than the Planck scale (1.22×10^{19} GeV). The graviton G, a hypothetical, massless and chargeless elementary particle of spin two, would be the carrier of the gravitational force in a quantum field theory that involves gravity. However, such a theory is extremely difficult to construct, as will be explained in Chapter 2 and is not part of the SM.

The forces and some of their characteristics are summarized in Table 1.3.

Interaction	Range	Relative strength	Mediators
Strong	10^{-15} m	1	8 gluons (g)
Electromagnetic	∞	10^{-3}	photon (γ)
Weak	10^{-18} m	10^{-14}	W^+, W^-, Z
Gravitational	∞	10^{-43}	graviton (G) ?

Table 1.3: Range, relative strength with respect to the strong force, and mediators of the four fundamental interactions. The gravitational force is not included in the SM, and gravitons are hypothetical particles.

Due to reasons connected to the internal symmetry of the $SU(3) \times SU(2) \times U(1)$ group (that will be treated in section 1.3) to account for the observed masses of the elementary particles, a special role has to be assigned to at least an additional particle, which has to be a scalar boson. Its role is intimately connected to the spontaneus symmetry breaking mechanism which gives rise to non-zero masses for the other elementary particles (see again section 1.3). A new particle, whose couplings to the massive fermions have to be proportional to the mass of the fermions themselves, and proportional to the square root of the bosons' masses, was proposed in 1964 independently by Brout and Englert [2] and Higgs [3]. Such a particle have been observed in 2012 by the ATLAS [4] and CMS [5] experiments and it is known as the Brout-Englert-Higgs scalar boson (denoted in literature

as H or H^0), sometimes also referred to as the Higgs boson. This discovery constitues an outstanding milestone of the modern particle physics and a tremendous success of the SM theory.

A visual summary of the elementary particles and their organization in the SM framework, as well as the force carriers and the H boson is shown in Figure 1.1.

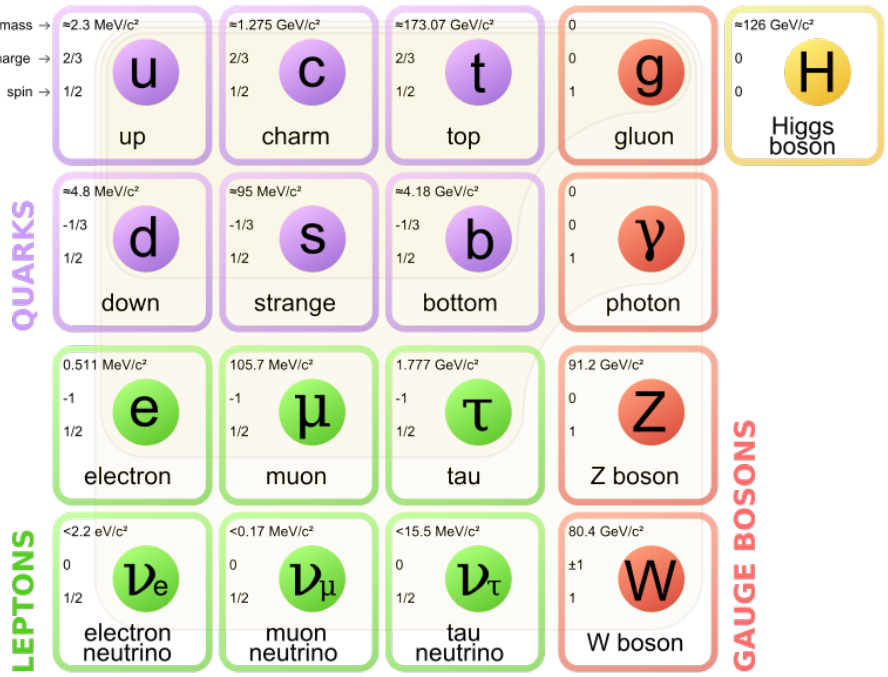

Figure 1.1: Overview of the Standard Model constituents.

1.2 Role of symmetries

In quantum field theories [6], the equations of motion of the different fields are derived, exploting the *least action principle*, from the Lagrangian density $\mathcal{L}$, which after integration over space and time defines the action S. For a spin 1/2 particle with mass m, described by a spinor field ψ, the action S is written, in the simplest case of absence of interaction, as:

$$S = \int dx^4 \mathcal{L} = \int dx^4 (i\bar{\psi}(x)\gamma^\mu \partial_\mu \psi(x) - m\bar{\psi}(x)\psi(x)) \tag{1.1}$$

where γ^μ are the Dirac's matrices. The term in parenthesis contains only the fermionic filed $\psi(x)$ and hence can be labeled as $\mathcal{L}_f = i\bar{\psi}(x)\gamma^\mu \partial_\mu \psi(x) - m\bar{\psi}(x)\psi(x)$. The equation of motion leads to the Dirac equation:

$$i\gamma^\mu \partial_\mu \psi(x) = m\psi(x) \tag{1.2}$$

As shown by Emmy Noether in her famous theorem, symmetries play a central role in physics: every continuous symmetry in the Lagrangian gives rise to a conserved quantity during the motion. Starting from the free spinor action defined in Equation 1.1, one sees that it is invariant under the transformation:

$$\psi(x) \to e^{ie\alpha}\psi(x) \tag{1.3}$$

where α is a global, i.e. not space-time dependent, phase rotation of the field and e is the already introduced Coulomb charge.

The leading "building principle" of the SM Lagrangian is that it is unnatural, in a relativistic theory, to allow a global phase transformation over the entire 4-d space, since it would be like if all the observers in the entire 4-d space agree on shifting their clocks by a certain fixed phase. In this sense, it is much more natural to allow local phase transformations, instead of global ones. The global symmetry has then to be forced to become a local one: this concept is called *symmetry gauging*[3]. If one wants to impose the invariance of the Lagrangian under the local transformation where $\alpha = \alpha(x)$:

$$\psi(x) \to e^{ie\alpha(x)}\psi(x) \tag{1.4}$$

one needs to introduce a new field $A^\mu(x)$, which must transform in the proper way under local phase transformation: the transformation of the ψ field and the A_μ field must act in a coordinated way as described by the transformation rule 1.5, so that the extra term coming from the derivates of the local phase $\alpha(x)$ compensate each other.

$$\psi(x) \to e^{ie\alpha(x)}\psi(x), \qquad A_\mu \to A_\mu - \partial_\mu\alpha(x) \tag{1.5}$$

The insertion of the new field A^μ can be elegantly achieved by simply replacing the partial derivative operator ∂ with the covariant derivative $\mathcal{D}$ in the Lagrangian:

$$\mathcal{D}_\mu = \partial_\mu + ieA_\mu(x) \tag{1.6}$$

The new Lagrangian is then given by:

$$\mathcal{L} = i\bar{\psi}(x)\gamma^\mu\partial_\mu\psi(x) - m\bar{\psi}(x)\psi(x) + ie\bar{\psi}(x)\gamma^\mu A_\mu\psi(x) - \frac{1}{4}(F_{\mu\nu}(x)F^{\mu\nu}(x)) = \mathcal{L}_f + \mathcal{L}_{gauge} \tag{1.7}$$

where:

$$F_{\mu\nu} = \partial_\mu A_\nu(x) - \partial_\nu A_\mu(x) \tag{1.8}$$

It can be seen that the Lagrangian (1.7) is composed of the two terms: $\mathcal{L}_f$ containing only the matter field $\psi(x)$ and $\mathcal{L}_{gauge}$ being the additional contributions containing the gauge field A_μ. It can be also shown that the Lagrangian is now invariant under coordinate local transformations of $\psi(x)$ and $A_\mu(x)$, as dictated in (1.5). In this sense, the introduction of a new bosonic field that couples to matter fields is a necessity to achieve the local gauge invariance of the action: in general, in the SM all interactions are consequences to this request. Ultimately, applying the Noether theorem, this symmetry leads to the law of conservation of the electric charge Q. The QED component of the SM Lagrangian is then a U(1) subgroup.[4].

Exploiting this "building principle", all the fermionic fields of Tables 1.1 and 1.2 are added by hand in the SM action, while the existence of the bosonic fields is a direct consequence of the required invariance properties of the action. Up to now the example dealt with the simplest $U(1)$ group of symmetry. The complete SM theory is based instead on the bigger group: $SU(3)_c \times SU(2)_L \times U(1)_Y$, obtained as direct product of smaller symmetry groups[5]. The Lagrangian of the electroweak theory (which treats the electromagnetic

[3] This argument gives credibility to the principle but it is not a rigorous demonstration (despite the fact that the symmetry gauging is one of the deepest and most profound concepts of the SM).

[4] Actually the U(1) group of the SM Lagrangian is not directly related to electric charge. The SM U(1) subgroup is instead related to the hypercharge $U(1)_Y$. The electric charge conservation rule is achieved in a more complex way: (spontaneously) breaking the bigger $SU(2)_L \times U(1)_Y$ group in the $U(1)_{em}$ group.

[5] The subscript c stands for color, L for left and Y for hypercharge. Their meaning will be explained later in the section

Field	$SU(3)_c$ representation	$SU(2)_L$ representation	Y	I_w^3	Q
u_{iL}	3	2	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{2}{3}$
d_{iL}	3	2	$\frac{1}{6}$	$-\frac{1}{2}$	$-\frac{1}{3}$
ℓ_{iL}	1	2	$-\frac{1}{2}$	$-\frac{1}{2}$	-1
ν_{iL}	1	2	$-\frac{1}{2}$	$\frac{1}{2}$	0
u_{iR}	3	1	$\frac{2}{3}$	0	$\frac{2}{3}$
d_{iR}	3	1	$-\frac{1}{3}$	0	$-\frac{1}{3}$
ℓ_{iR}	1	1	-1	0	-1
ν_{iR}	1	1	0	0	0

Table 1.4: Fermion content of the SM, with representations under $SU(3)_c$ and $SU(2)_L$, hypercharge Y, isospin I_w^3 and electric charge Q. The index i refers to the fermion generation, while the indices L and R represent the left-handed or right-handed nature of the particle [1].

and the weak forces into a unified framework) is invariant under transformations in the $SU(2)_L \times U(1)_Y$ subgroup, while the QCD Lagrangian is invariant under $SU(3)_c$.

In order to explain the subscript L, one can group the fermions of the SM by their chirality, either left or right. The chirality is defined by the projection operators $\mathcal{P}_{left} = (1-\gamma^5)/2$ and $\mathcal{P}_{right} = (1+\gamma^5)/2$, where $\gamma^5 = i\gamma^0\gamma^2\gamma^2\gamma^3$ and γ^i being the Dirac matrices. Left-handed fermions form doublets, while right-handed fermions form singlets. The left-handed quark doublets consist of up-type and down-type quarks and the left-handed lepton doublets consist of a neutrino and the associated charged lepton (see diagram 1.9). Up to now right-handed neutrinos have never been observed. The weak interaction only acts on left-handed particles and, thus, violates parity, which is the invariance under mirror operation at the origin in space: the subscript L of the $SU(2)_L$ group reflects this fact.

$$\ell_L = \begin{pmatrix} \nu_L \\ e_L \end{pmatrix},\ e_R, q_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix},\ u_R,\ d_R. \tag{1.9}$$

The quantum number associated to the $SU(2)_L$ symmetry is the weak isposin I_w, while the $U(1)_Y$ subgroup is related to the hypercharge Y. The weak hypercharge Y carried by the matter fields is related to the electric charge Q and the third component of weak isospin I_w^3 by:

$$Y = Q - I_w^3. \tag{1.10}$$

Finally, the subscript c of the $SU(3)_c$ group stands for *color*, which is the additional quantum number carried by the quarks[6]. Under $SU(3)_c$, quarks are color triplets while leptons are color singlets; quarks therefore carry a color index ranging between one and three, whereas leptons do not take part in strong interactions.

The fermion content of the SM is summarized in Table 1.4, together with its representation under the different groups of symmetry.

As explained earlier in the chapter, to each group of symmetry there should be an associated gauge field. The gauge field associated to the symmetry group $U(1)_Y$ is usually denoted as B_μ, with the hypercharge Y as generator of the group. Three gauge fields, W_μ^1, W_μ^2 and W_μ^3 are associated to $SU(2)_L$ group, with three generators that can be expressed as half of the Pauli matrices:

$$T_1 = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},\ T_2 = \frac{1}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix},\ \text{and } T_3 = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{1.11}$$

[6]The introduction of the color as an additional quantum number was done so to realize the Pauli's exclusion principles in particles like the Δ^{++} that would have apparently violated it.

The generators T^a satisfy the Lie algebra:

$$[T^a, T^b] = i\epsilon^{abc}T_c \quad \text{and} \quad [T^a, Y] = 0, \tag{1.12}$$

where ϵ^{abc} is the antisymmetric tensor. Finally, in the $SU(3)_c$ group, the eight generators (that can be expressed in term of the Gell-Mann matrices) correspond to the eight gluon fields $G_\mu^{1...8}$.

When "gauging" the $SU(3)$ and the $SU(2)$ symmetries, the same principle depicted in the simpler $U(1)$ symmetry gauging has to be implemented. A more complex algorithm is needed due to the fact that, unlike the $U(1)$ group, $SU(2)$ and $SU(3)$ are non-Abelian, i.e. non-commutating groups[7]. Apart for the complexity, the SM Lagrangian will be organized as in (1.7) so to include all the SM fermions in $\mathcal{L}_f$ while $\mathcal{L}_{gauge}$ will include the gauge fields G_μ, the W_μ and the B_μ.

Unlike the simplest $U(1)$ scenario, the striking consequence of the invariance under the $SU(3)_c \times SU(2)_L \times U(1)_Y$ group is that the SM Lagrangian cannot include any explicit mass term, neither for the fermionic field nor the bosonic fields: this is because a mass term like $m\bar{\psi}\psi$ can be simply re-written[8] as $m\bar{\psi}_R\psi_L$. Therefore, a mass term would couple right- and left-handed fields and, since the left-handed fermions transform as doublets and right-handed fermions as singlets under the SM gauge group, a term like $m\bar{\psi}_R\psi_L$ would break the invariance of the Lagrangian in an irreparable way. On the other hand, the impossibility to have a mass term would lead to the conclusion, in clear disagreement with the experimental facts, that all particles should be massless, both the fermions and the bosons. In reality, some of the gauge bosons, photons and gluons, are indeed massless, while the weak gauge bosons must have non zero masses in order to explain the "weakness" of the weak interaction. In addition to this, all fermions have masses, including the neutrinos. The solution to this problem is conceptually different for bosons and fermions and it's detailed in the next section.

In general terms:

- The fact that the weak gauge bosons have a mass different from zero indicates that the vacuum of the theory, i.e. the fundamental state does not share (is not invariant under) the same symmetries of the SM Lagrangian: the electroweak group $SU(2)_L \times U(1)_Y$ is not a symmetry of the vacuum. On the other hand, since gluons and photons are massless the SM symmetry group of the Lagrangian must (spontaneously) break in:

  $$SU(3)_c \times SU(2)_L \times U(1)_Y \rightarrow SU(3)_c \times U(1)_{em} \tag{1.13}$$

  where the subscript em in $U(1)_{em}$ stands for "electromagnetic" and is used to clearly indicate its difference with respect to the $U(1)_Y$ related to the hypercharge. This breaking pattern can be achieved introducing a new complex scalar doublet field ϕ in the theory. Three of the four degrees of freedom coming from the complex doublet ϕ are "used" to give masses to the W and the Z bosons, leaving an additional degree of freedom which would correspond to a new spin-0 particle.

- Once the ϕ field is introduced, the mass of the fermions can be achieved with the interaction terms between ths field ϕ and the matter field, via the so-called Yukawa terms.

[7] Ultimately, it's the covariant derivative which assumes a more complex form.
[8] Any field can be written in terms of chiral fields $\psi = \psi_R + \psi_L$.

1.3 Scalar sector

Given the "building principle" behind the SM formulation, explicit mass terms are forbidden inside the SM Lagrangian. Let's now imagine to introduce a certain field[9] ϕ and let's consider its expectation value $\langle 0|\phi|0\rangle$ calculated on the fundamental state of the theory $|0\rangle$, also called the vacuum of the theory. Suppose the theory is invariant (i.e. the Lagrangian describing the theory is invariant) under a certain symmetry of the fields, for example a simple $U(1)$ symmetry $\phi \to e^{i\alpha}\phi$. If the vacuum state itself were invariant under the same symmetry, $U(1)$ in this example, meaning

$$|0\rangle = e^{i\alpha}|0\rangle$$

then the expectation value of the ϕ field must be equal to zero[10]

$$\langle 0|\phi|0\rangle = 0$$

As a consequence if $\langle 0|\phi|0\rangle \neq 0$ then the vacuum state $|0\rangle$ is not invariant under the same symmetry of the Lagrangian. This situation defines what is called spontaneous symmetry breaking in group theory.

In the SM case, an additional complex scalar (i.e. having spin 0) doublet ϕ is simply introduced by hand in the Lagrangian and, making its components explicit, can be written as:

$$\phi = \frac{1}{\sqrt{2}}\begin{pmatrix}\varphi_1 + i\varphi_2 \\ \varphi_3 + i\varphi_4\end{pmatrix}. \tag{1.14}$$

The corresponding additional Lagrangian density $\mathcal{L}_\phi$, associated to the scalar sector, can be written as a kinematic term and a potential one:

$$\mathcal{L}_\phi = (D^\mu\phi)^\dagger D_\mu\phi - V(\phi) \tag{1.15}$$

Where the kinetic part includes the gauge covariant derivatives for the $SU(2)_L \times U(1)_Y$ group, which is defined as:

$$D_\mu\phi = \left(\partial_\mu + ig\vec{T}.\vec{W}_\mu + \frac{ig'}{2}B_\mu\right)\phi \tag{1.16}$$

The potential $V(\phi)$ has the most general renormalizable[11] form invariant under $SU(2)_L \times U(1)_Y$ in:

$$V(\phi) = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2 \tag{1.17}$$

where, in order to obtain a non-zero expectation value $\langle 0|\phi|0\rangle$ and so realize the required spontaneous breaking of the SM group of symmetry, the factor μ^2 has to be negative. Given the sign requirement for the μ^2 factor, the λ parameter has to be a positive real number in order to preserve the vacuum stability[12]. This choice for the sign of the parameters μ^2 and λ gives to the potential $V(\phi)$ the shape of a "Mexican hat", as shown in Figure 1.2. While a local maximum of the potential is found for $\phi = 0$, there is

[9] In principle it's not mandatory for this field to be added by hand as a new field in the Lagrangian, as it goes in the SM case. There are cases, like the Cooper pair in superconductivity, i.e. a pair of electrons (or other fermions) bounded together at low temperatures, where the new field (the bound state of electron pairs) is "dynamically created" and not explicitely present in the Lagrangian [7]. The SM approach is then the simplest possible.

[10] This is because $\langle 0|\phi|0\rangle = \langle 0|(e^{-i\alpha}e^{i\alpha})\phi|0\rangle = e^{i\alpha}\langle 0|\phi|0\rangle$, given that $\langle 0|e^{-i\alpha} = \langle 0|$ because of the vacuum invariance hypothesis. Hence $\langle 0|\phi|0\rangle = e^{i\alpha}\langle 0|\phi|0\rangle$ which means $\langle 0|\phi|0\rangle = 0$.

[11] In quantum field theories, divergencies tend to arise in calculations because all particles can contribute to a process as virtual particles in loops: then a cut-off is often needed in the computation of the physical quantities. If the cut-off disappears from the final results by its absorption in a finite number of measured constants, the theory is called renormalizable (see Section 1.5).

[12] To avoid that the minimum of the potential goes at $-\infty$.

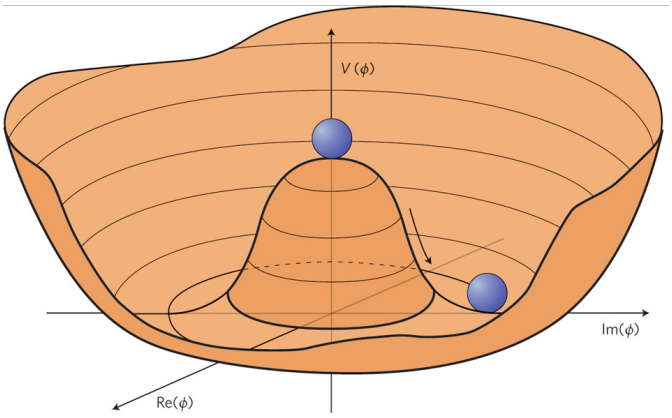

Figure 1.2: The Higgs potential in the simpler case where the ϕ field has only 2 degrees of freedom, instead of the 4 ones of a complex scalar doublet field.

a closed surface of minima corresponding to a non-zero field. Using the approach of the "small oscillations" in polar coordinates (perturbative expansion), the field ϕ can be developed around one of its degenerate minima. Clearly, the second derivative in the radial direction is positive, since the curvature of the potential is positive in that direction, while the second derivative in the angular coordinate is zero, since there is an entire closed cinconference of degenerate minima. The radial excitations give rise to a non-zero mass particle, while the angular excitations give rise to massless particles, called Goldstone bosons, one for each broken degree of symmetry, hence, in principle, three massless particles in the SM case as $SU(2)_L \times U(1)_Y \rightarrow U(1)_{em}$. On the other hand, the presence of those additional massless particles would be the scenario in case of the spontaneous breaking of a global symmetry of the Lagrangian, as stated by the Goldstone theorem. The SM case appears more complex: the broken symmetry $SU(2) \times U(1)$ is not a global one, since the SM Lagrangian is defined using the leading principle of the symmetry gauging: 4 gauge fields (3 W_μ e 1 B_μ) are in fact present. The 3 massless goldstone bosons that would appear in a global symmetry breaking are instead "re-absorbed", using the gauge invariance of the Lagrangian, by a re-definition of the gauge fields themselves, which in exchange gain mass different from zero [8]. The additional massive field corresponding to the radial excitation of the potential cannot instead be reabsorbed via a redifinition of the gauge fields and hence is a remarkable prediction of the SM theory. The associated massive particle is the already introduced H boson.

More specifally, the potential $V(\phi)$ has (degenerate) minima corresponding to the radial position $|\phi| = \sqrt{\dfrac{-\mu^2}{2\lambda}} \neq 0$. This non-zero quantity can be interpreted as the vacuum expectation value v of the field ϕ, measured to be about 246 GeV. It can be demonstrated that any scalar doublet ϕ can be written in the canonic form "down and real" using an appropriate matrix $U(x) \in SU(2) \times U(1)$, of the form $U(x) = e^{i\gamma \frac{T_2}{2}} e^{i(\alpha T_3 + \beta Y)}$. Using $U(x)$, the field ϕ can be re-written so that its non zero expectation value v is exposed:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \varphi_1 + i\varphi_2 \\ \varphi_3 + i\varphi_4 \end{pmatrix} = U(x) \begin{pmatrix} 0 \\ v + \frac{h(x)}{\sqrt{2}} \end{pmatrix} \tag{1.18}$$

where $h(x)$ is a field with zero expectation value over the vacuum. Given the local gauge invariance, the $U^{-1}(x)$ transformation can be applied to all fields without affecting in any way the form of the Lagrangian: the three degrees of freedom of $U(x)$ disappear, simply "absorbed" by the gauge fields, thus eliminating the degrees of freedom associated to the

massless Goldstone bosons. In particular, applying U^{-1} transforms the ϕ field as:

$$\phi \to U^{-1}\phi = \begin{pmatrix} 0 \\ v + \frac{h(x)}{\sqrt{2}} \end{pmatrix} \tag{1.19}$$

Using (1.19) and the perturbative approach around one arbitrary chosen minimum, the potential

$$V(\phi) = \mu^2|\phi|^2 + \lambda|\phi|^4$$

can be written in the suitable form:

$$\begin{aligned} V(h) &\simeq \frac{\mu^2}{2}h^2 + \lambda[v^2 + \frac{h^2}{2} + \sqrt{2}vh]^2 \\ &\simeq \frac{\mu^2}{2}h^2 + 2\lambda v^2 h^2 + \lambda v^2 h^2 + O(h^3) + O(h^4) \\ &\text{and given that } v = \sqrt{\frac{-\mu^2}{2\lambda}} \\ &= -\mu^2 h^2 + O(h^3) + O(h^4) \end{aligned} \tag{1.20}$$

where:

- No linear terms in the h field are present (since the potential is being expanding around a minimum).
- The constant terms are discarded since the equation of motions are obtained by deriving the Lagrangian.
- The cubic and quartic terms in the h fields are not detailed, but simply indicated as $O(h^3)$ and $O(h^4)$.
- The last equality gives rise to the mass of the h field:

$$m_H^2 = -2\mu^2 > 0. \tag{1.21}$$

The kinetic term of the lagrangian $\mathcal{L}_\phi$

$$(D_\mu\phi)^\dagger D_\mu\phi = |D_\mu\phi|^2 \tag{1.22}$$

again using the perturbative expansion, leads to:

$$|D_\mu\phi|^2 \simeq \frac{1}{2}\partial_\mu h\partial^\mu h + \frac{g^2v^2}{4}\left((W_\mu^1)^2 + (W_\mu^2)^2 + (W_\mu^3 - \frac{g'}{g}B_\mu)^2\right) \tag{1.23}$$

The charged vector bosons W^1 and W^2 therefore acquire a mass, given by:

$$m_{W^1}^2 = m_{W^2}^2 = \frac{g^2v^2}{2} \tag{1.24}$$

The third term of equation (1.23) is a linear combination of W_μ^3 and B_μ. The relation between the W_μ and B_μ fields and the physical fields Z_μ and A_μ, associated respectively to the neutral boson Z and the photon, is expressed by the set of relations:

$$\begin{cases} Z_\mu = W_\mu^3 \cos\theta_W - B_\mu \sin\theta_W \\ A_\mu = W_\mu^3 \sin\theta_W + B_\mu \cos\theta_W \\ \tan\theta_W = \frac{g'}{g} \end{cases} \tag{1.25}$$

Equation (1.23) can be therefore written as:

$$|D_\mu \phi|^2 \simeq \frac{1}{2}\partial_\mu h \partial^\mu h + \frac{g^2 v^2}{4}\left((W_\mu^1)^2 + (W_\mu^2)^2 + (\frac{Z_\mu}{\cos\theta_W})^2\right) \tag{1.26}$$

The mass of the Z boson is thus related to the mass of the charged W bosons via the Weinberg angle θ_W, which can be determined experimentally[13]:

$$\frac{m_{\mathrm{W}}}{m_{\mathrm{Z}}} = \cos\theta_W \tag{1.27}$$

There is no mass term for the A_μ field, hence the photon remains massless. Equation 1.26 elegantly solves the problem of the masses of the vector bosons and shows a kinetic term related to a new scalar field h, having a non-zero mass specified in (1.21).

To account for the fermions' masses a different approach can be pursued: given the existence of the field doublet ϕ, one can write coupling terms, known as the Yukawa couplings, between fermions and the ϕ field. Considering the electron e, for example, and indicating with l^e the $SU(2) \times U(1)$ doublet containing the left-handed electron and (electron) neutrino, one can write the following Lagrangian term

$$\mathcal{L}^e_{Yuk} = -\lambda_e \bar{l^e}_L \phi e_R + h.c. \tag{1.28}$$

which after the electroweak symmetry breaking (EWSB) procedure described earlier becomes

$$\mathcal{L}^e_{Yuk} \xrightarrow[EWSB]{} -\lambda_e v \bar{e}_{1L} e_{1R} + h.c. \tag{1.29}$$

A mass term can now be given to the electron e, where in particular:

$$m_e = \lambda_e v \tag{1.30}$$

A similar reasoning can be re-done for the down-type quark which can, like the electron, acquire a mass through Yukawa couplings to the ϕ doublet:

$$\mathcal{L}^d_{Yuk} = -\lambda_d \bar{q}_{1L} \phi d_{1R} + h.c. \xrightarrow[EWSB]{} -\lambda_d \frac{v}{\sqrt{2}} \bar{d}_{1L} d_{1R} + h.c. \tag{1.31}$$

On the other hand, the up-type quark and in general all the particles that are located in the "up-side" of the doublets detailed in (1.9), both quarks and neutrinos, cannot acquire a mass with a simple Yukawa term like the one written in (1.31), because of the canonical "down and real" form chosen for the ϕ doublet, the non-zero component of which, after EWSB, is always in the "down-side" of the doublet. The most economical solution is to make the "up" fields couple to the other possible $SU(2) \times U(1)$ term of interaction:

$$(l^i \phi^j)\epsilon_{ij} \tag{1.32}$$

where the indices $i, j = 1, 2$ run over the "up-side" and "down-side" components of the objects and the antisymmetric tensor ϵ_{ij} is used such that the "up" fields are now the ones gaining mass. Aesthetically the interaction term for the "up" fields can be seen as a coupling term with a transformed (rotated) ϕ field, $\tilde{\phi}$ defined as:

$$\tilde{\phi} = i\sigma_2(\phi^\dagger)^t, \tag{1.33}$$

where the the σ_2 matrix is one of the Pauli matrices $\sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$

[13] $\sin^2\theta_W \simeq 0.231$

For quarks and neutrinos an additional complication is also necessary: the eigenstates produced by weak interactions are flavour eigenstates but not, at the same time, mass eigenstates. The two basis (interaction and mass) are related by the rotation matrices, known as the CKM matrix (Cabibbo-Kobayashi-Maskawa) for quarks and the PMNS matrix (Pontecorvo–Maki–Nakagawa–Sakata) for neutrinos.

1.4 Standard model Lagrangian

Finally, the SM Lagrangian density can be decomposed as a sum of four different terms:

$$\mathcal{L}_{SM} = \mathcal{L}_f + \mathcal{L}_{gauge} + \mathcal{L}_\phi + \mathcal{L}_{Yuk} \tag{1.34}$$

which are related respectively to the fermion, gauge, scalar and Yukawa sectors. The four Lagrangian terms are detailed below.

- The fermionic part of the Lagrangian density consists of kinetic energy terms for quarks and leptons, namely:

$$\mathcal{L}_f = i\bar{q}_{iL}\not{D}q_{iL} + i\bar{u}_{iR}\not{D}u_{iR} + i\bar{d}_{iR}\not{D}d_{iR} + i\bar{\ell}_{iL}\not{D}\ell_{iL} + i\bar{e}_{iR}\not{D}e_{iR} \tag{1.35}$$

  The gauge-covariant derivatives contain the gauge tensors:

$$G^i_{\mu\nu} = \partial_\mu G^i_\nu - \partial_\nu G^i_\mu - g_s f_{ijk} G^j_\mu G^k_\nu, \text{ with } i,j,k = 1,...,8;$$

$$W^i_{\mu\nu} = \partial_\mu W^i_\nu - \partial_\nu W^i_\mu - g\epsilon_{ijk} W^j_\mu W^k_\nu, \text{ with } i,j,k = 1,...,3;$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$$

  where g_S and g are the coupling constants associated to the $SU(3)_c$ and $SU(2)_L$ symmetry groups respectively.

  The covariant derivatives are specified as:

$$D^\mu q_{iL} = (\partial^\mu + \frac{i}{2} g_S G^\mu_a \lambda_a + \frac{i}{2} g W^\mu_b \sigma_b + \frac{i}{6} g' B^\mu) q_{iL},$$

$$D^\mu u_{iR} = (\partial^\mu + \frac{i}{2} g_S G^\mu_a \lambda_a + \frac{2i}{3} g' B^\mu) u_{iR},$$

$$D^\mu d_{iR} = (\partial^\mu + \frac{i}{2} g_S G^\mu_a \lambda_a - \frac{i}{3} g' B^\mu) d_{iR},$$

$$D^\mu \ell_{iL} = (\partial^\mu + \frac{i}{2} g W^\mu_a \sigma_a - \frac{i}{2} g' B^\mu) \ell_{iL},$$

$$D^\mu e_{iR} = (\partial^\mu - i g' B^\mu) e_{iR}$$

  where g' is the coupling constant associated to the $U(1)_Y$ symmetry group.

- The gauge Lagrangian density $\mathcal{L}_{gauge}$ regroups the gauge fields of all three symmetry groups:

$$\mathcal{L}_{gauge} = -\frac{1}{4}G^i_{\mu\nu}G^{\mu\nu i} - \frac{1}{4}W^i_{\mu\nu}W^{\mu\nu i} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} \tag{1.36}$$

- The scalar sector is composed of the kinetic term and the potential one:

$$\mathcal{L}_\phi = (D^\mu\phi)^\dagger D_\mu\phi - \mu^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2 \tag{1.37}$$

  where the kinetic part includes the gauge covariant derivative defined as:

$$D_\mu\phi = \left(\partial_\mu + ig\vec{T}.\vec{W}_\mu + \frac{ig'}{2}B_\mu\right)\phi$$

- The last piece is the Yukawa Lagrangian density which describes the interactions between the fermions and the scalar doublet ϕ. If one notes Y^u, Y^d and Y^e three general complex 3×3 matrices of dimensionless couplings, the Yukawa Lagrangian density can be written as:

$$\mathcal{L}_{Yuk} = -Y^u_{ij}\bar{q_{iL}}u_{jR}\tilde{\phi} - Y^d_{ij}\bar{q_{iL}}d_{jR}\phi - Y^e_{ij}\bar{\ell_{iL}}e_{jR}\phi + h.c. \tag{1.38}$$

  where $\tilde{\phi}$ is defined as:

$$\tilde{\phi} = i\sigma_2(\phi^\dagger)^t \tag{1.39}$$

1.5 Radiative corrections and renormalization

The SM Lagrangian in Equation (1.34) contains all the information needed to compute physical quantities such as decay rates or cross-sections. In quantum field theory, the probability of a state $|a\rangle$ to evolve after some time to a state $|b\rangle$ is proportional to the square of the amplitude $\langle b|\hat{S}|a\rangle$, where $\hat{S}$ is the S-matrix which consists of a time ordered exponential of the interacting Hamiltonian. It is usually treated perturbatively, that is, the exponential is decomposed into a sequence of terms of increasing powers of the coupling constants. Each of its terms can be described by one or several Feynman diagrams from which an amplitude can be calculated using a finite set of rules. The leading order represents the classical amplitude and the higher orders are quantum corrections. A common issue when calculating the quantum corrections is the appearance of divergencies. In the SM though, these divergencies can be reabsorbed in the definition of the coupling constants at a given scale through a procedure named renormalization [9].

As an example one can consider the action in Equation (1.7) where, for the sake of simplicity, the spinor field is taken massless. The leading order and the 1-loop Feynman diagrams for the process $\psi\bar{\psi} \to \gamma \to \psi\bar{\psi}$ are shown in Figure 1.3. The leading order amplitude, computed applying the Feynman rules and naming q the exchanged momentum, is:

$$\mathcal{M}_{LO} = ie^2\gamma^\mu\frac{\eta_{\mu\nu}}{q^2}\gamma^\nu$$

For the 1-loop contribution, one must integrate over the fermion 4-momentum in the loop, k. The integral is proportional to $\int d^4k/k^4 \simeq \int dk/k \simeq ln(k)$. In order to obtain a finite result, one introduces a cut-off Λ inside the integral: $\int^\Lambda dk/k$ and the total amplitude including the 1-loop correction becomes:

$$\mathcal{M}_{1-loop} = ie^2\gamma^\mu\frac{\eta_{\mu\nu}}{q^2}\gamma^\nu(1 + \frac{e^2}{12\pi^2}\log\frac{q^2}{\Lambda^2}) \tag{1.40}$$

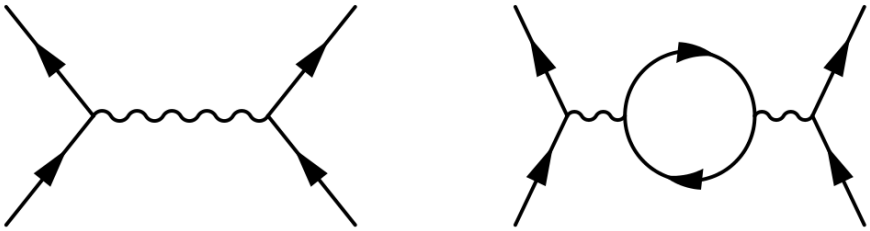

Figure 1.3: The leading order (left) and 1-loop correction (right) Feynman diagrams for the process $\psi\bar{\psi} \to \gamma \to \psi\bar{\psi}$ in QED.

One can get rid of this cut-off by trading the coupling present in the action, e, for the effective coupling e_{eff} at a given scale μ. The equation then becomes:

$$\mathcal{M}_{1-loop} = ie_{eff}^2 \gamma^\mu \frac{\eta_{\mu\nu}}{q^2} \gamma^\nu (1 + \frac{e_{eff}^2}{12\pi^2} \log \frac{q^2}{\mu^2}) \tag{1.41}$$

As a consequence, the coupling that should be used to calculate a physical process depends on its scale. The variation of the coupling with the scale q is described by the renormalization group equation which in the present case is given, for 1-loop corrections by:

$$\frac{d}{d\log q} e(q) = \frac{e^3(q)}{12\pi^2} \tag{1.42}$$

A similar equation holds for any parameter present in the action.

In the SM, the effective couplings g_i ($i = 1, 2, 3$) associated respectively to U(1), SU(2) and SU(3) are constrained by the equations:

$$\frac{d}{d\log q} g_i(q) = -\frac{b_i g_i^3(q)}{(4\pi)^2} \tag{1.43}$$

with [10]

$$b_1 = -\frac{4}{3} n_g - \frac{1}{10} n_h$$

$$b_2 = \frac{22}{3} - \frac{4}{3} n_g - \frac{1}{6} n_h$$

$$b_3 = 11 - \frac{4}{3} n_g$$

where n_g is the number of generations (3) and n_h the number of scalar bosons (1). One of the important consequences of these equations is the asymptotic freedom of QCD [11, 12]: because of the sign of b_3, g_3 weakens at high energy and the quarks can then be treated as free particles. This feature plays an essential role in the calculation of cross-sections in hadron collisions.

Equation (1.43) can be also seen as a powerful guide-line for the construction of models of new physics, since it happens to be very easy to build a theory which is not renormalizable which hence, in principle, has to be discarded.

1.6 Drell-Yan process

Given its importance in this thesis, the Drell-Yan process will be treated in details using the SM formalism. The Drell-Yan process is defined as the annihilation of a quark-antiquark pair into a lepton-antilepton pair. This process is described at leading order by the two Feynman diagrams drawn in Figure 1.4. These two amplitudes are proportional

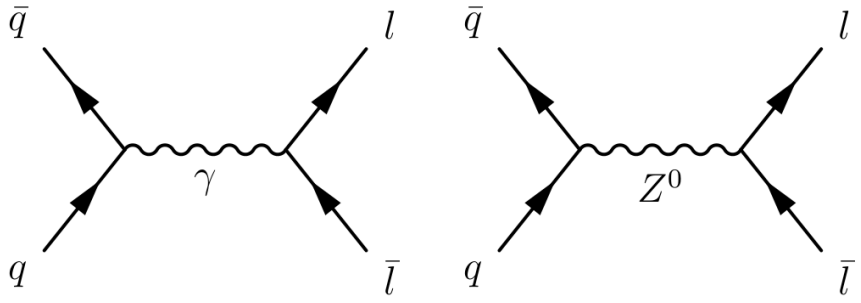

Figure 1.4: Feynman diagrams contributing to the Drell-Yan process at leading order. The left (right) diagram corresponds to the annihilation of a $q\bar{q}$ pair into a photon (a Z boson).

to the fine structure constant $\alpha \approx 1/137$.

In the Standard Model, there is actually a third diagram contributing, namely the production of the Higgs boson (H^0) predicted by the electroweak symmetry breaking mechanism (see diagram in Figure 1.5). However, as described in Section 1.3, the fermion

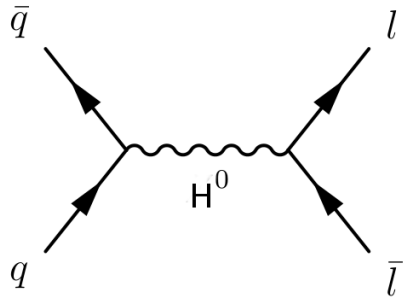

Figure 1.5: Feynman diagram of the process $q + \bar{q} \rightarrow H^0 \rightarrow l + \bar{l}$ at leading order.

coupling to the Higgs boson is proportional to m_f/v where m_f is the mass of the fermion and v is the vacuum expected value of the scalar field ($\approx$ 246 GeV). This leads to an amplitude proportional to $\frac{m_q(GeV)}{246}.\frac{m_l(GeV)}{246}$ for this diagram.
Since the production rate of each quark depends on its parton density function inside the hadrons (see Chapter 3):

- the contribution of the proton valence implies that up and down quarks ($m < 10$ MeV) are produced in larger amounts than the others;
- heavier quarks are harder to be produced because of their mass;
- a top-antitop initial state cannot occur below an invariant mass of 2 $\times m_t \approx 350$ GeV;

the two first diagrams are dominant by several orders of magnitude and the latter is usually neglected in the calculation.

1.7 Partonic cross-section

The diagrams in Figure 1.4 give rise to two matrices terms $\mathcal{M}_\gamma$ and $\mathcal{M}_Z$ and, hence, three amplitudes terms: $|\mathcal{M}_\gamma|^2$, $|\mathcal{M}_Z|^2$ and the interference $\mathcal{M}_\gamma \mathcal{M}_Z^* + c.c.$

The cross-section corresponding to the creation of a photon only can be written as[14]:

$$\frac{d\sigma_\gamma}{d\Omega} = \frac{e^4}{(4\pi)^2} Q_q^2 Q_l^2 \frac{1}{8s'} \left[(1+\cos\theta)^2 + (1-\cos\theta)^2\right] \tag{1.44}$$

where s' is the center-of-mass energy of the partonic process. It can be observed that this cross-section remains the same for $\theta \to \pi - \theta$.

The cross-section corresponding to the creation of a Z boson only can be written as:

$$\frac{d\sigma_Z}{d\Omega} = \frac{e^4}{(4\pi)^2} Q_q^2 Q_l^2 \frac{1}{8s'} |\mathcal{R}|^2 \left[c_{1,Z}(1+\cos\theta)^2 + c_{2,Z}(1-\cos\theta)^2\right] \tag{1.45}$$

where

$$c_{1,Z} = ((g_{V_l}^2 + g_{A_l}^2)(g_{V_q}^2 + g_{A_q}^2) + 4 g_{V_l} g_{A_l} g_{V_q} g_{A_q}) \tag{1.46}$$

$$c_{2,Z} = ((g_{V_l}^2 + g_{A_l}^2)(g_{V_q}^2 + g_{A_q}^2) - 4 g_{V_l} g_{A_l} g_{V_q} g_{A_q}) \tag{1.47}$$

and g_V and g_A are the vectorial and axial couplings associated to the lepton/quark. Because $c_{1,Z}$ and $c_{2,Z}$ are different, the cross-section is not invariant under $\theta \to \pi - \theta$. The Z boson alone can therefore introduce a forward-backward asymmetry.

The cross-section from the interference can be written as:

$$\frac{d\sigma_{int}}{d\Omega} = \frac{e^4}{(4\pi)^2} Q_q^2 Q_l^2 \frac{1}{8s'} \mathrm{Re}(\mathcal{R}) \left[c_{1,int}(1+\cos\theta)^2 + c_{2,int}(1-\cos\theta)^2\right] \tag{1.48}$$

$$c_{1,int} = 2(g_{V_l} g_{V_q} + g_{A_l} g_{A_q})$$

$$c_{2,int} = 2(g_{V_l} g_{V_q} - g_{A_l} g_{A_q})$$

$$\mathcal{R} = \left(\frac{1}{Q_l Q_q \sin^2 2\theta_W} \frac{1}{1 - M_Z^2/s' + i\Gamma_Z/M_Z}\right)$$

again not invariant under $\theta \to \pi - \theta$.

The total angular differential cross-section, which is the sum of equations (1.44), (1.45), (1.48), can be finally written as:

$$\frac{d\sigma_{\gamma/Z}}{d\Omega} = \frac{e^4}{(4\pi)^2} Q_q^2 Q_l^2 \frac{1}{4s'} \left[c_1(1+\cos^2\theta) + c_2\cos\theta\right] \tag{1.49}$$

where

$$c_1 = 1 + 2\mathrm{Re}(\mathcal{R}) g_{V_l} g_{V_q} + |\mathcal{R}|^2 (g_{V_l}^2 + g_{A_l}^2)(g_{V_q}^2 + g_{A_q}^2)$$

$$c_2 = 4\mathrm{Re}(\mathcal{R}) g_{A_l} g_{A_q} + 8|\mathcal{R}|^2 g_{V_l} g_{A_l} g_{V_q} g_{A_q}$$

The total cross-section is:

$$\sigma = \int_\Omega \frac{d\sigma_{\gamma+Z}}{d\Omega} d\Omega = \frac{4\pi}{3} \frac{\alpha^2}{s'} c_1 \tag{1.50}$$

where $\alpha = e^2/(4\pi)$.

[14] The angle θ is defined as the angle between the quark and the negative lepton directions of flight, in the qq center-of-mass frame.

1.8 Forward-backward asymmetry

The forward-backward asymmetry is obtained starting from the forward and backward cross-sections given by $\sigma_F = \sigma_{\theta<\pi/2}$ and $\sigma_B = \sigma_{\theta>\pi/2}$. The forward-backward asymmetry is then defined by :

$$A_{FB} = \frac{\sigma_F - \sigma_B}{\sigma_F + \sigma_B} = \frac{3}{8}\frac{c_2}{c_1} \tag{1.51}$$

An isotropic cross-section implies $A_{FB} = 0$ and the two extreme cases (+1 and -1) correspond to a purely forward/backward cross-section (the lepton is always closer to the quark/to the antiquark).

This asymmetry strongly depends on $\sqrt{s'}$:

- At low $\sqrt{s'}$ (< 10 GeV) the photon contribution is dominant and the forward-backward asymmetry is then zero since the photon couples equally to left-handed and right-handed particles
- In the intermediate region it goes negative (mostly driven by the effect of the globally negative interference term)
- It's $\approx$ zero around the Z peak
- At high energy (above 100 GeV) it becomes a positive constant

This variable is also of special interest in the searches for new physics: the angular differential cross-section of any spin 1 particle is proportional to $(1+\cos^2(\theta))+\frac{8}{3}A_{FB}\cos(\theta)$. The A_{FB} coefficient can therefore be used to distinguish signal from background or to identify a new signal in case of a discovery of a new particle [13].

Chapter 2

Theories beyond the standard model

While Chapter 1 introduced the SM of elementary particle physics, this chapter deals with the description of theories beyond the SM (BSM). Section 2.1 lists the main motivations that lead to the existence of BSM theories; Section 2.2 describes theories that lead to heavy resonances that decay to dilepton or diphoton final state[1]*.*

2.1 Motivation for new theories beyond the standard model

Despite its enormous success in accurately describing a vast amount of experimental data, spanning several orders of magnitudes in energy, it is commonly admitted that the Standard Model only constitutes a low energy approximation of a more fundamental theory. Indeed, there are some open questions or features which cannot be explained by the SM, as explained below. Some of them come from experimental observations (dark matter [14, 15], neutrino mass [16, 17]) some others are limitations of the current theory (lack of gravity description, convergence of the coupling constants [10]) or aspects dealing with the internal consistency of the theory (hierarchy problem, free parameters of the Lagrangian [1]). In more details:

- **Gravitational interaction:** the fourth fundamental interaction, gravity, is not included in the model. Gravity is, by many aspects, very different from the three other forces and the purpose to establish a common framework describing all of them has to face several difficulties. The Einstein's General Relativity (GR) theory shows that gravity is intimately connected to the space-time geometry which is, in turn, coupled to the particles energy-momentum tensor via the Einstein's fields equations: this makes its integration inside the SM framework subtler than simply adding a new interaction. To combine the quantum theory of the SM with the GR, a quantum theory of gravity is necessary; this would lead to a new field associated to gravity: a spin 2 particle, called graviton. It can be shown that such a theory is not renormalizable: loop corrections including gravitons induce ultraviolet divergencies that cannot be reabsorbed through the renomalization procedure, which is instead adopted in electroweak and chromodynamics theories (see Section 1.5). Finally, the strength of the gravitational force is much lower than the other ones. Whereas the strong, weak and electromagnetic forces have similar strengths at the electroweak scale (energies of $\mathcal{O}(100\text{ GeV})$), the energy at which gravitational interactions becomes relevant is at the order of the Planck scale of $E_{Pl} = 10^{19}$ GeV, which is defined by the Planck mass, $M_{Pl} = \sqrt{\hbar c/G}$, where G is the gravitational constant. The huge difference between

[1] In this thesis, the expression "dilepton final state" denotes the decay in electron-positron pairs (e^+e^-) or muon-antimuon pairs ($\mu\bar{\mu}$), while the expression "diphoton final state" denotes the decay in two photons ($\gamma\gamma$).

the electroweak scale and the Plack scale is also known as the hierarchy problem and is deeply connected to the problem of the fine-tuning of the Higgs boson mass (see the following point).

- **Hierarchy problem and fine-tuning of the Higgs boson mass:** after the discovery of the Higgs boson and the measurement of its mass of ≈ 125 GeV all the ingredients of the SM have been experimentally established. All particles in the standard model, including the Higgs boson, have a *bare mass* which is the mass obtained from the quantum propagator at the lowest order in perturbation theory. This is not anyway the *physical mass*, i.e. the mass that can be measured experimentally, because radiative corrections at higher orders coming from loops have to be considered. As explained in Section 1.5, the renormalization process relates the properties of the physical quantities (mass, charge, ...) to those of the bare particles, introducing suitable cut-off parameters in considering higher order correcttions. It is known [18] that the renormalization procedure corrects the squared bare mass (m_0) of the Higgs boson (H) with an extra term, including higher order corrections, δm_{H}^2 to obtain the physical mass m_{H}:

$$m_{\mathrm{H}}^2 = m_0^2 - \delta m_{\mathrm{H}}^2 \tag{2.1}$$

where δm_{H}^2 includes all contributions from radiative corrections to the Higgs propagator. The main ones[2] are those involving top quarks, vector bosons and the scalar boson itself. The corresponding Feynman diagrams are shown in Figure 2.1, where the Higgs boson is denoted as h [1].

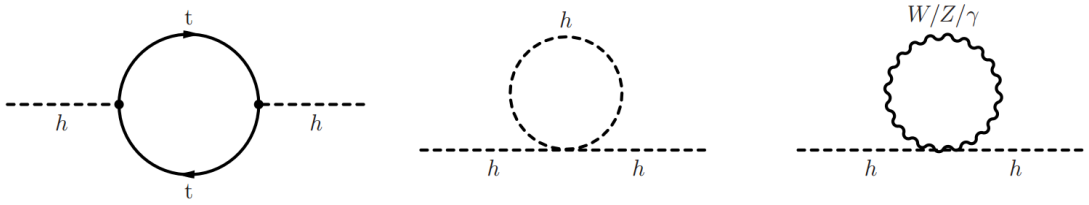

Figure 2.1: Main divergent contributions to the scalar boson mass predicted by the SM

The integrals corresponding to the amplitude of these processes are divergent, so a cut-off parameter Λ is introduced. This parameter represents the energy up to which the standard model can be still considered valid. In principle, one can assume that the SM is valid up to the Planck scale at which gravitational effects cannot be neglected. With this assumption Λ would be of the order of $\approx 10^{19}$ GeV. The full calculation gives that δm_{H}^2 is proportional to Λ^2:

$$\delta m_{\mathrm{H}}^2 \propto \Lambda^2 \approx 10^{38}\ \mathrm{GeV}^2 \tag{2.2}$$

Since m_{H} is ≈ 125 GeV ($\approx 10^2$), Equation (2.1) can be rewritten as:

$$10^4\ \mathrm{GeV}^2 \approx m_0^2 - \Lambda^2 \approx m_0^2 - 10^{38}\ \mathrm{GeV}^2$$

which means that m_0^2 is of the same order of Λ^2 (10^{38}) and that these two terms cancel with a very high precision to obtain the value of the physical mass. This mathematical problem, known as *fine-tuning*, does not invalidate the theory, which is still consistent. Anyway it seems an unnatural and implausible coincidence that

[2]This statement is true for any elementary spin 0 boson.

m_0^2 cancels all the loop contributions up to this astonishing, or even miraculous, precision.

The choice of Λ made in the previous calculation is somehow arbitrary because it is based on the assumption that the Standard Model is still valid up to the greatest possible energy, the Planck scale. If a lower Λ is chosen, the cancellation is tuned to an acceptable level. If, for instance, $\Lambda \approx$1 TeV is chosen, the hierarchy problem is completely solved since the cancellation is of the order of one over ten, which seems a natural and acceptable value. For this reason, if one accepts the fine-tuning argument, new physics phenomena at the TeV scale are expected, since at energy higher than $\Lambda =$1 TeV, the SM is not valid anymore.

- **Matter content of the universe**: astronomical observations show that the visible content of matter can only be approximately 5% of the total matter and energy content of our universe. This statement results from several recent cosmological observations. First, the measured orbital velocities of stars around their galaxy center [14, 15] is incompatible with the observed matter density in space: the stars are moving too fast. In order to reconciliate the experimental data with the theory, the existence of another kind of matter that does not interact via electromagnetic or strong interactions, the dark matter, has therefore been postulated. A second major result in cosmology is the discovery that the Universe is in accelerated expansion: in average galaxies recede from each other and their escape rate increases with the distance [19, 20]. Putting together these two cosmological results, one can conclude that the matter/energy content of the universe is made of 5% ordinary matter, 25% dark matter and 70% dark energy, which is thought to be responsible for the observed accelerated expansion of the universe, by introducing a repellent force (negative pressure). The SM does not offer good candidates or explainations for the dark matter and dark energy problems.

- **Neutrino masses**: originally, SM neutrinos were assumed massless. The fact that neutrinos can change from one flavour to another implies that they must have non-zero mass differences [16, 17] and that their mass eigenstates are different from their flavour eigenstates. A mass term for the neutrinos can, in principle, be added to the SM as described in Chapter 1, but it is not clear if the small masses that the neutrinos must have can arise from the same electroweak symmetry breaking mechanism that is in act for the other SM particles. In fact, even if the addition in the SM Lagrangain of the term in (1.39) accomplish the purpose to generate, through the electroweak symmetry breaking mechanism, a mass term for the neutrinos, it appears difficult to attribute to the same underlying mechanism the generation of masses so different as the electron and the neutrino ones, which have a difference of several orders of magnitude.
  In particular, the addition of the term (1.39) gives rise to a so-called Dirac mass term for the neutrino of the form $m_D(\bar{\nu}_R\nu_L - \bar{\nu}\nu_R)$, but there exists another possible contribution to the Lagrangian, which is also invariant under $SU(2) \times U(1)$ and is known as the Majorana mass term:

$$M_M(\nu_R\gamma^0\nu_R)$$

  The full Lagrangian involving the neutrino fields would then be of the form:

$$\mathcal{L}_{\nu-full} = M_M(\nu_R\gamma^0\nu_R) + m_D(\bar{\nu}_R\nu_L - \bar{\nu}\nu_R) \tag{2.3}$$

  In the limit where $M_M \gg m_D$ and after diagionalization of the mass matrix one gets, for each neutrino family, a "light neutrino" of mass $\frac{m_D^2}{M_M}$ and a "heavy neutrino" of mass

M_M. Since there are no restriction on the value of M_M, this mechanism, known as the "see-saw mechanism", achieves the suppression of the mass of the "light neutrino" even for values of the Dirac mass m_D of the order of the electroweak "natural" scale, i.e. the vacuum expectation value of the Higgs field.

- **Free parameters of the SM Lagrangian**: the SM contains 19 free parameters, that have to be measured. The parameters include the charged fermion masses, the mixing angles and the charge-parity (CP) violating phase of the Cabibbo-Kobayashi-Maskawa (CKM) matrix, the coupling constants of the three forces, and the mass and vacuum expectation value of the Higgs boson. However, it is widely believed that at least some of these parameters should be related to each other from a mechanism that is not described by the SM. As an example one could consider the different masses of the quark and lepton generations as arising from a common generation mechanism in a BSM theory, that has a spontaneously broken symmetry at the SM scale. The list of parameters is summarized in Table 2.1.

Quantity	Symbol	Value
Electron mass	m_{e}	511 keV
Muon mass	m_{μ}	105.7 MeV
Tau mass	m_{τ}	1.78 GeV
Up quak mass	m_{u}	2.3 MeV ($\mu_{\overline{\mathrm{MS}}}$=2 GeV)
Down quak mass	m_{d}	4.8 MeV ($\mu_{\overline{\mathrm{MS}}}$=2 GeV)
Strange quak mass	m_{s}	95 MeV ($\mu_{\overline{\mathrm{MS}}}$=2 GeV)
Charm quak mass	m_{c}	1.28 GeV ($\mu_{\overline{\mathrm{MS}}} = m_{\mathrm{s}}$)
Bottom quak mass	m_{b}	4.18 GeV ($\mu_{\overline{\mathrm{MS}}} = m_{\mathrm{b}}$)
Top quak mass	m_{t}	173.5 GeV
CKM 12-mixing angle	θ_{12}	12.9°
CKM 23-mixing angle	θ_{23}	2.4°
CKM 13-mixing angle	θ_{13}	0.2°
CKM CP violating phase	δ_{13}	69°
W boson mass	m_{W}	80.4 GeV
Z boson mass	m_{Z}	91.2 GeV
Strong coupling constant	α_s	0.119 ($\mu_{\overline{\mathrm{MS}}} = m_{\mathrm{Z}}$)
QCD vacuum angle	θ_{QCD}	~ 0
Higgs boson vacuum expectation value	v	246 GeV
Higgs boson mass	m_{H}	125.09 GeV [21]

Table 2.1: SM parameters. The quark masses are presented in the renormalization scheme known as $\overline{\mathrm{MS}}$ [1].

- **Convergence of the coupling constants**: the SM coupling constants of the electromagnetic interaction, the weak interaction and the strong interaction have a similar value at an energy scale of $\mathcal{O}(10^{16}$ GeV). However, they do not converge to a single value as shown in Figure 2.2. In order to unify the coupling constants, an extension of the SM would be necessary in order to modify their evolution above the electroweak scale.

All these aspects indicate that there must be new physics at a scale beyond the electroweak scale. What is unknown, however, is the energy scale at which this new physics will manifest itself. Driven by the arguments given while treating the hierarchy problem, it is believed that there should be new physics at the TeV scale, at which a discovery with direct searches at the Large Hadron Collider (LHC) could be possible.

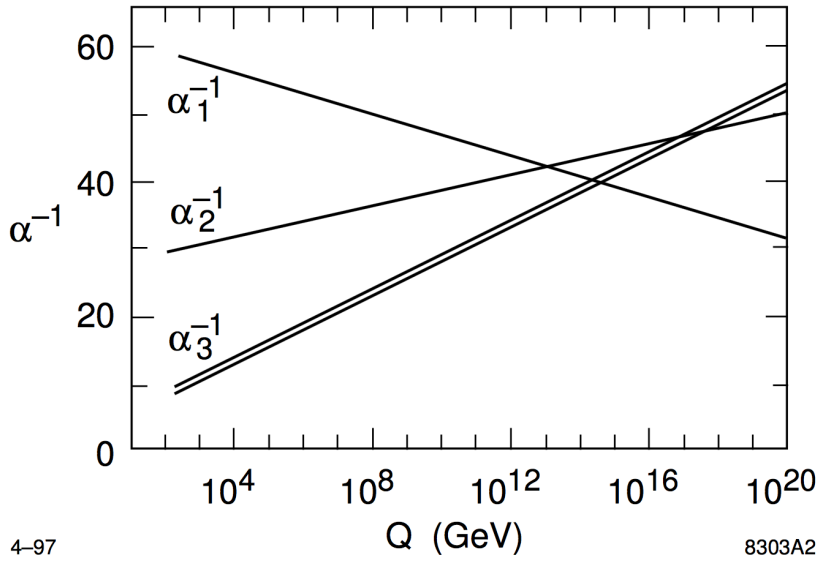

Figure 2.2: Evolution of the SM couplings $\alpha_i = \frac{g_i^2}{4\pi}$ as a function of the energy scale [10].

2.2 New massive resonances decaying into dilepton or diphoton final state

One of the most straightforward ways to observe physics beyond the SM at a hadron collider would be through an alteration (either a localized excess of events or a distortion) in the dilepton or diphoton[3] mass spectrum at high mass. The study of these final states is both theoretically and experimentally motivated: from a theoretical point of view, many BSM models of new physics give rise to high energy lepton/photon pairs in the final state (see next sections); moreover, at a hadron collider, the background for such final states is relatively low and this aspect, in combination with the high accuracy of the lepton/photon reconstruction provided by the CMS detector (see Chapter 5) make these channels experimentally well under control. In this section, some BSM models are described to account for the theoretical motivations for the studies presented in this thesis, while the experimental challenges and results will be trated in details in Chapters 6 and 7. Given the large amount of models, the studies presented in this thesis have been conceived as model independent searches for any possible signal of a new resonance which decays to two electrons/photons with a mass above 400/500 GeV. In both cases, the analysis strategy and the choice of a cut-based event selection (as described in Sections 6.3 and 7.3) rather than a MultiVariate one have been guided by the need to be sensitive to whichever new particle with different spin hypotheses.

From a theoretical point of view, a new massive resonance that decays in the dilepton/diphoton final state arises in many different types of BSM models such as supersymmetric models [22] (see Section 2.2.1), extensions of the SM gauge group in the framework of Grand Unification [23, 24, 25] (see Section 2.2.2), models introducing extra dimensions [26, 27] (see Section 2.2.3). Generically, in the searches for new physics, all particles that can give rise to a resonance in the dilepton mass spectrum are called Z′, while a spin 2 particle would be generically called "graviton". There isn't a widespread agreement in literature on how to generically indicate a new spin 0 particle: in this thesis, such a new

[3]In this thesis, two separate analyses will be presented. In particular, Chapter 6 describes in details the analysis targeting the dielectron final state, which is then combined with the dimuon final state (the analysis of this final state is not treated). Chapter 7 describes in the details the analysis targeting the diphoton final state.

particle will be simply referred as a scalar particle X.

2.2.1 Supersymmetry

Supersymmetry [18] is a new type of symmetry of the lagrangian where bosons and fermions are considered as components of a common super-multiplet. In this theory, any fermionic (bosonic) field is associated to a supersymmetric bosonic (fermionic) partner: any rotation in the supermultiplet space, i.e. any mixing of boson and fermion partners, should leave the lagrangian (hence the action) invariant.

There exist many supersymmetric models: the most simple one, the Minimal Supersymmetric Standard Model (MSSM) associates a superpartner (SUSY partner) to each SM particle: every fermion has a SUSY partner particle which is a boson, and vice versa. The superpartners of fermions, called *sfermions* are the *sleptons* $\tilde{\ell}$, the *sneutrinos* $\tilde{\nu}_\ell$ (ℓ=e,μ,τ), and the *squarks* $\tilde{q}$, all of them having spin 0. The gauge bosons superpartners are the *gauginos* $\tilde{\gamma}$, $\tilde{g}$, $\tilde{W}^{\pm}$ and $\tilde{Z}$, all of them having spin 1/2. In this model, the electroweak symmetry breaking requires the existence of two scalar doublets that will generate 5 physical fields[4]. Their superpartners are called *higgsinos*. In the SM the baryon number B and the lepton number L are (separetely) conserved quantities defined as

$$B = \frac{1}{3}(n_q - n_{\bar{q}}), \quad L = (n_l - n_{\bar{l}}) \tag{2.4}$$

where n_q, $n_{\bar{q}}$, n_l, and $n_{\bar{l}}$ are, respectively, the numbers of quarks, antiquarks, leptons and antileptons. This is generally not the case in a SUSY model, and in particular not true for the MSSM model, where terms violating B and L can easily appear in the SUSY action. To preserve these conservation laws, which are experimentally well verified, a new conserved quantum number called R-parity (R_p) is introduced as:

$$R_p = (-1)^{B-L+2s} \tag{2.5}$$

where s is the spin of the specific particle considered. With this definition all SM particles have $R_p = +1$ and all superpartners have $R_p = -1$.

Supersymmetry is extremely attractive for several aspects. First, it provides an explaination to the problem of the fine-tuning of the Higgs boson mass since the cut-off scale Λ up to which the SM would be valid is the scale of the superpartner masses. For energies compatible with the superpartner masses, the loop corrections to the Higgs boson mass must also take into account the superpartner corrections. Notably, the corrections of the bosonic and fermionic superpartners are of different sign with respect to the SM ones and turn out to exacly cancel out the SM divergent contributions. In order to avoid issue of naturalness for the Higgs boson mass, the supersymmetric particles cannot be much heavier than 1 TeV. Moreover, it could be pointed out it is somehow unnatural that the SM symmetry group is described by the direct product of three different groups: it would be indeed more natural a descriptions in terms of a bigger group reducing to the SM group after a certain symmetry breaking mechanism. In this respect, supersymmetric theories motivate extensions of the SM gauge group (discussed in the next section), in order to include particles and SUSY partners in a super-multiplet. Finally, supersymmetry provides a natural dark matter candidate: if R-parity is conserved then the lightest suuperparticle must be stable, since decays of superparticle into ordinary SM particles would not be allowed.

[4]This can be simply understood by counting the degrees of freedom. Two complex doublets have 8 degrees of freedom: three of them are absorbed by the gauge fields W and Z that acquire mass, leaving with 5 degrees of freedom associated to 5 "Higgs bosons" fields. This 5 "Higgs bosons" are usually parametrized in terms of two CP-even neutral scalars, one CP-odd neutral pseudoscalar and a pair of charged Higgs bosons. In general, any model with more than one complex doublet is said to have "non-minimal Higgs sector".

However, more complicated scenarios of supersymmetry allow violation of the R-parity [22] and make possible the decay of the "sneutrinos" into two leptons, with the interesting feature that the couplings to the lepton generations are, in general, not identical.

2.2.2 Grand Unified Theories

As pointed out in Section 1.2, the "building principle" that leads to the final SM action is the request of invariance under local transformations belonging to a certain symetry group. The SM gauge group G_{SM}, however, has a very peculiar structure coming from the direct product of three different gauge groups, being $G_{SM} = SU(3)_c \times SU(2)_L \times U(1)_Y$, which is felt as unnatural. Various BSM scenarios re-interpret the SM group G_{SM} as the "residuum" of a larger gauge group that broke down[5] at low energy, in a way that can be thought analogous to what happens already in the SM for the $SU(2)_L \times U(1)_Y \to U(1)_{em}$. Many of these models have the tantalizing feature of unifying the three gauge couplings g_1, g_2 and g_3 into a single fundamental coupling[6]: conceptually this would imply that all fundamental interactions included in the model have one common root. In the SM, the running of the coupling constants of the three forces makes them "almost to converge" at one value at high energy, without really crossing each other in one single point (see Figure 2.2). It is thought therefore, that the introduction of a larger symmetry group G modifies the couplings in such a way to unify them in one point at a certain scale, called GUT (from grand unified theories) scale: at E_{GUT} the three running gauge coupling constants of the SM gauge group become equal, while for energies $E \ll E_{GUT}$ the fundamental gauge group G is broken to retain the SM symmetry G_{SM}. During the process of the fundamental group symmetry breaking, extra $U(1)$ groups can easily appear, introducing new neutral gauge bosons. If their mass lies in the TeV range and they couple to quarks and/or leptons, it should be possible to produce them and detect them at the LHC.

On the other hand, the Achille's heel of these theories is that they predict the decay into a positron and a neutral pion of the proton[7], which is instead a stable particle or, more precisely, its lifetime is quantified to be $\tau_p > 10^{33}$ years [1]. In particular, the proton decay is mediated by the exchange of gauge bosons with a mass $\mathcal{O}(E_{GUT})$. Hence, to be consistent with the present experimental limits on τ_p, the grand unified scale must be $E_{GUT} > 10^{15}$ GeV. Such an energy is much larger than electroweak scale ($\mathcal{O}(100$ GeV)) and in meantime 3-4 orders of magnitude smaller than the Planck mass M_{Pl}, where gravity is expected to become as strong as the other interactions: the effects of gravity can be neglected in GUTs.

In the following the first GUT model, based on the $SU(5)$ symmetry group, will be shortly described focusing the attention on its limitations and how this limitations can be solved in a more sophisticated $SO(10)$ model; the more exotic E_6 models will then be described.

The $SU(5)$ model and its extension to $SO(10)$

As shown by H. Georgi and S.L. Glashow in 1974, the smallest gauge group which can contain the SM is $G = SU(5)$ [28]. It embeds $SU(3)_c \times SU(2)_L \times U(1)_Y$ and the matter particles representations in the following way:

- For each generation, the three color states of the right-handed down quarks and the charged and neutral left-handed leptons are grouped into a multiplet transforming

[5] The specific symmetry breaking mechanism depends on the specific BSM model and won't be treated in details in this thesis.

[6] Note that, even in these models, gravity is never taken into account.

[7] It is sort of the analogue of the β decay of the neutron described in the electroweak theory.

according to the $\bar{5}$ representation of $SU(5)$.

- For each generation, the three color states of the right-handed and left-handed up quarks and left-handed down quarks plus the right-handed charged leptons are grouped into a multiplet transforming according to the 10 representation of $SU(5)$.

- The gauge sector is composed of 24 bosons: the 12 SM gauge bosons plus 12 exotic gauge bosons. This feature can be easily predicted by counting the degrees of freedom (dof) characterizing the symmetry group: a $SU(n)$ group has (n^2-1) dof. The SM has hence 12 dof: 8 gluons coming the $SU(3)_c$ group, 3 Ws coming from $SU(2)_L$ and 1 B coming from $U(1)_Y$. A $SU(5)$ model will have instead $5^2-1=24$ dof, manifesting as 24 gauge bosons.

- The fundamental scale, where the three SM couplings are unified into a single coupling is $\mathcal{O}(10^{16}$ GeV).

In particular, the last statement, $E_{GUT} = \mathcal{O}(10^{16}$ GeV), is a consequence of the renormalization equations of the three couping constants, once the $SU(5)$ prediction of $\sin^2\theta_W$ =0.375 (at the unification scale) is taken into account. In 1974, when the model was proposed, the prediction $\sin^2\theta_W$ =0.375 was compatible with the experimental results and this model became very popular among physicists. Unfortunately, the latest measurements have significantly lowered the uncertainties on the θ_W value and $\sin^2\theta_W$ =0.375 is now ruled out. This, in turn, implies that the simplest $SU(5)$ model do not show the convergence of g_1, g_2 and g_3 to a single value[8], finally presenting a similar "misbehaviour" with respect to the SM one. The desidered convergence of the couplings can still be achieved in supersymmetric extensions of the $SU(5)$ model, such as the MSSM already introduced in the previous section (see Figure 2.3).

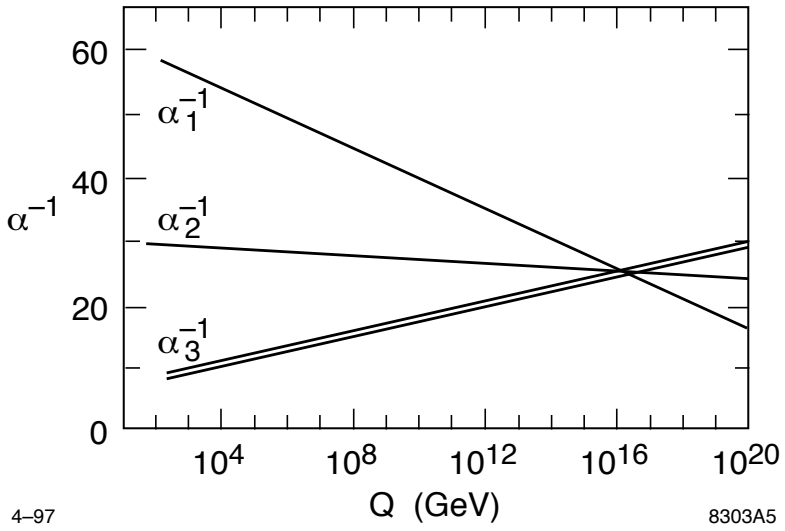

Figure 2.3: Evolution of the coupling constants $\alpha_1 = \frac{5}{3}\frac{g_1^2}{4\pi}$, $\alpha_2 = \frac{g_2^2}{4\pi}$ and $\alpha_3 = \frac{g_3^2}{4\pi}$ versus the energy scale of the process Q obtained with the renormalization group equations for the supersymmetric generalization of the SM (MSSM model) [10].

Coming to the Achille's heel of the theory, the requirement of the stability of the proton is not fulfilled neither by the standard $SU(5)$ model nor by its supersymmetric extensions.

[8]In the context of $SU(5)$ theories the SM g_1 gets re-defined as $g_1 = \sqrt{\frac{5}{3}}g_Y$.

In both cases, the exotic bosons imply the proton decay into a positron and a neutral pion and, even if this decay is suppressed, the predicted half time is $\approx 10^{29}$ years, several orders of magnitude smaller than the experimental lower limits.

Nevertheless, in the last decades, the $SU(5)$ model has been used as the starting point of a series of new models that could instead handle the experimantal observations about the proton stability. The most famous example is the $SO(10)$ model, proposed by H. Fritzsch and P. Minkowski in 1975 [29]. The $SO(10)$ model groups all matter particles belonging to the same generation into a single multiplet and has the nice feature to predict an half time for the proton decay which is not in contradiction with the experimental results. Moreover, in the $SU(5)$ model the new physics scale is $\mathcal{O}(10^{16}$ GeV), and so is the mass of the new gauge bosons: there is therefore no hope to produce them at the LHC. For a larger group, the situation would be different. For istance, the breaking scheme for $SO(10)$ is:

$$SO(10) \rightarrow SU(5) \times U(1)_\chi \rightarrow G_{SM} \times U(1)_\chi \tag{2.6}$$

where χ denotes the charge associated to the extra $U(1)_\chi$ group, leading to a new "photon-like" field, analogous to the SM B field associated to the SM $U(1)_Y$ group. While in the previously discussed $SU(5)$ model, the mass of the new bosons must be of the order $O(10^{16}$ GeV), there is no contraint on the breaking scale for this $U(1)_\chi$ group and it might happen in the TeV range. Therefore, particles coming from $SO(10)$ theories could be discovered at the LHC.

The E_6 models

Other frequent choices for gauge groups able to embed $SU(5)$ lie on the exceptional E_6 group. The symmetry breaking scheme is the following one:

$$E_6 \rightarrow SO(10) \times U(1)_\psi \rightarrow SU(5) \times U(1)_\chi \times U(1)_\psi \rightarrow G_{SM} \times U(1)_\chi \times U(1)_\psi \tag{2.7}$$

It is generally assumed that only one linear combination of the $U(1)_\chi$ and $U(1)_\psi$ groups is relevant at the TeV scale:

$$\mathrm{Z}'(\theta_{E_6}) = \mathrm{Z}'_\psi \cos\theta_{E_6} - \mathrm{Z}'_\chi \sin\theta_{E_6} \tag{2.8}$$

Each value of the mixing angle θ_{E_6} (simply referred as θ from now on), free parameter of the theory, corresponds to a $U(1)_\theta$ group and leads to a different Z' phenomenology. Some specific models are:

- Z'_ψ (θ=0) that only interacts through axial-vector couplings with the fermions (see Table 2.2) is predicted by superstring theories [30].
- Z'_χ ($\theta = -\pi/2$), corresponding to a pure $U(1)_\chi$ group.
- Z'_η ($\theta = \arccos\sqrt{5/8}$), also suggested by superstring theories [30]
- Z'_I ($\theta = \arccos(\sqrt{5/8}) - \pi/2$), that does not couple to up quarks and only couples to left-handed down quarks and right-handed leptons (see again Table 2.2).

Sequential standard model

Another model which is taken under consideration is the so-called Sequential Standard Model (SSM) [23]. This model predicts a new boson Z'_{SSM} heavier than the Z boson,

but with the same SM couplings to fermions and gauge bosons. This is often used as a benchmark model for experimental Z′ searches and not considered as a realistic scenario.

The Z′ couplings to up quarks, down quarks and charged leptons are given in Table 2.2 for the presented models, where notably any Z′ deriving from E_6 does not present vectorial coupling with the up quark.

Model	c_V^u	c_A^u	c_V^d	c_A^d	c_V^l	c_A^l
Z'_ψ	0	0.300547	0	0.300547	0	0.300547
Z'_η	0	0.380165	-0.285124	0.095041	0.285124	0.095041
Z'_χ	0	0.073458	-0.416249	-0.342792	0.416249	-0.342792
Z'_I	0	0	0.620752	-0.620752	-0.620752	-0.620752
Z'_{SSM}	-0.227388	0.592979	0.410183	-0.592979	0.044592	-0.592979

Table 2.2: Vector (c_V) and axial (c_A) couplings of the Z′ boson to up quarks (u), down quarks (d) and the charged leptons (l) for various E_6 models. For comparison, the Z'_{SSM} couplings, which are identical to the Z boson couplings are also given.

In case of the discovery of a new particle, in order the characterize its properties, one can note that each set of couplings induces a particular forward-backward asymmetry, as illustrated in Figure 2.4 [13], which shows the A_{FB} as a function of the dilepton invariant mass for various Z′ models of 500 GeV of mass. As expected, for a Z'_ψ which couples the same way to left-handed and right-handed particles, the asymmetry is null at the resonance mass.

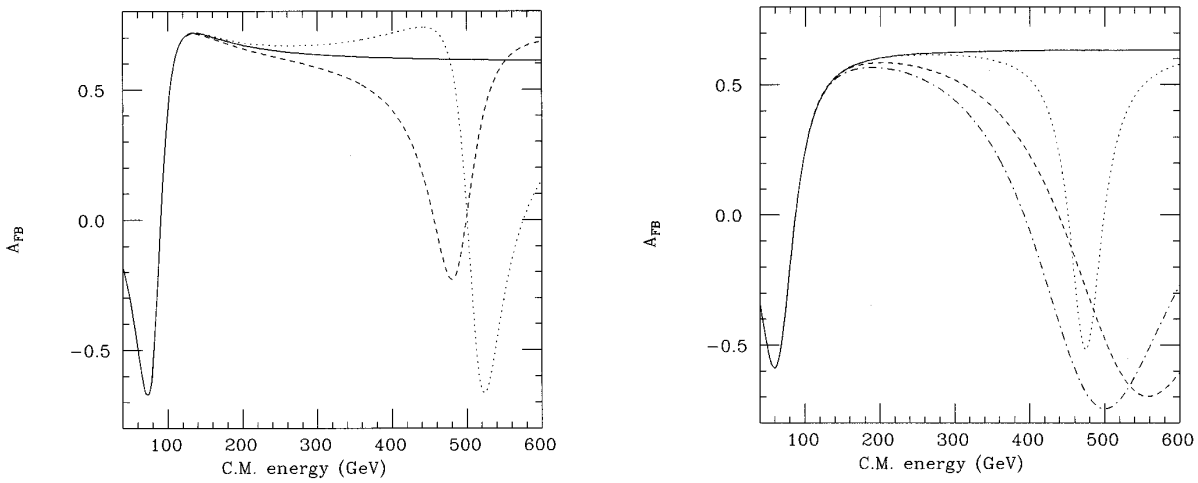

Figure 2.4: The forward-backward asymmetry for the $u\bar{u} \to l\bar{l}$ (left) and $d\bar{d} \to l\bar{l}$ (right) processes, as a function of the center-of-mass energy for the SM (solid line) and various models of Z′ at $M = 500$ GeV: Z'_ψ (dotted line), Z'_χ (dashed line) and Z'_I (dot-dashed line). The latter does not couple to up quarks and is therefore absent in the left plot [13].

2.2.3 Extra dimensions

Theories involving additional spatial dimensions represent a different class of models. They all emerged as solutions to the gauge hierarchy problem explaining the weakness of the gravitational force compared to the other forces by allowing the spin 2 graviton, as the carrier of the gravitational interaction, to propagate in the extra dimensions, while the other fields of the SM must remain "confined" in the usual 4-dimensional spacetime.

Gravity is only "perceived" as weak in the 4-dimensional world since the overlap of the wave functions of the SM particles with the graviton is small.

Large extra dimensions: ADD model

One of the first solutions of the hierarchy problem involving extra spatial dimensions was proposed by Arkani-Hamed, Dimopulous and Dvali [31], and it is known as ADD model. In this section, it will be shown that this model leads to the existence in the universe of large extra dimensions (LED) and it is then also known as the LED model. The central idea behind this model is the observation that, while the electroweak scale has been experimentally tested at distances of the order of $1/M_{EW}$, the gravitational force is far from being explored at distances of $1/M_{Pl} \sim 10^{-33}$ cm, while in reality it has never been tested for distances below ≈ 100 μm [32]. In particular, the fact that the gravitational strength between two masses decreases as the inverse of their distance squared is a direct consequence of the existence of only three spatial dimensions: the ADD theory hinges on the existence on n extra dimensions, with a radius smaller than 100 μm to account for the experimental tests. At this point, to account for the observed weakness of gravity compared to electroweak interactions, gravity has to be the only fundamental interaction able to propagate in the extra dimensions: the SM fields are "trapped" inside the 4-dimensional world, while gravitons are the only particles freely propagating in the whole space (called *bulk*). This reasoning leads to the idea of abandoning the interpretation of M_{Pl} as a fundamental energy scale, but rather as an "effective" one, while the real fundamental scale is the one in the $4+n$ dimensions $M_{Pl(4+n)}$. The two scales are related by the simple expression:

$$M_{Pl}^2 \sim M_{Pl(4+n)}^{n+2} R^n \tag{2.9}$$

where R is the radius of the extra dimensions (assuming all the extra dimensions share the same radius).

If this assumption is true, the gravitational potential for two test masses m_1 and m_2 placed at a distance $r \ll R$ can be calculated writing Gauss' low in (4+n) dimensions as:

$$V(r) \sim \frac{m_1 m_2}{M_{Pl(4+n)}^{n+2}} \frac{1}{r^{n+1}} \tag{2.10}$$

while for distances $\gg R$, the $1/r$ law holds, since the gravitational flux lines no longer propagate in the extra dimensions, but the usual potential given by:

$$V(r) \sim \frac{m_1 m_2}{M_{Pl}^2} \frac{1}{r} \tag{2.11}$$

is replaced by:

$$V(r) \sim \frac{m_1 m_2}{M_{Pl(4+n)}^{n+2} R^n} \frac{1}{r} \tag{2.12}$$

applying the Equation (2.9).

Moreover, this approach could be able to elude the hierarchy problem by requiring:

$$M_{Pl(4+n)} \sim 1 \text{ TeV}$$

Combining this request with the necessity to reproduce the observed strength of gravity leads to a condition for the size of the extra dimensions:

$$R \sim 10^{\frac{30}{n}-17} \text{cm} \times (\frac{\text{TeV}}{M_{EW}})^{1+\frac{2}{n}} \tag{2.13}$$

The case with $n = 1$ corresponds to $R \sim 10^{11}$ m, which would imply deviations from Newton's law at distances on the scale of the Solar System (the Earth's orbit has a radius of $\sim 1.5 \times 10^8$ km), at which it is well established. For $n = 2$, the value of R is $\sim$100 μm, just at the verdge of the experimental limits. Hence the number of extra dimensions must be $n \geq 2$. Table 2.3 shows the corresponding extra dimension size for 1 to 6 extra dimensions.

n	R(m)	R(eV^{-1})
1	10^{11} m	5×10^{17} eV^{-1}
2	100 μm	0.5 meV^{-1}
3	1 nm	5 keV^{-1}
4	3 pm	15 MeV^{-1}
5	100 fm	0.5 MeV^{-1}
6	10 fm	0.05 MeV^{-1}

Table 2.3: Radius of the extra dimensions in the ADD model to solve the hierarchy problem as a function of the number of extra dimensions. The radius R is expressed in terms of distance and energy units.

ADD dimensions are said "large" because their radius is in the range meV^{-1} – MeV^{-1} for $n \leq 6$, well above the TeV length $l_{TeV} = 1\text{TeV}^{-1}$. One observable effect of LEDs is due to a virtual graviton acting as a propagator the processes shown in Figure 2.5, which result in the production of a photon pair in the final state[9]. In this case, the graviton-induced

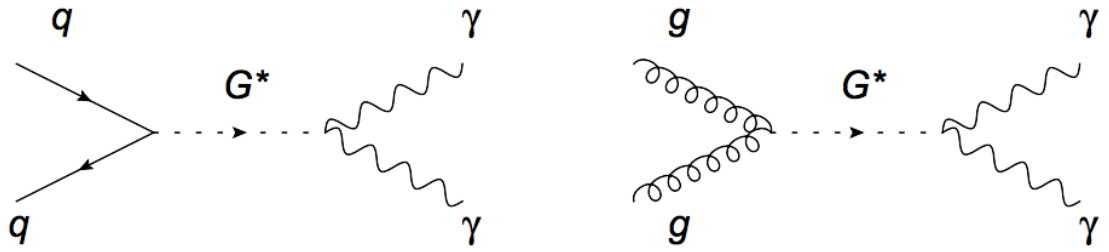

Figure 2.5: Feynman diagrams for virtual graviton production through $q\bar{q}$ in s-channel (left) and gluon-gluon fusion (right). In both cases, the graviton decays in the diphoton final state.

diagram interferes with the analogous SM diagrams and results in an enhancement of the invariant mass spectrum of the diphoton system, particularly at high masses. Since the graviton can propagate into the extra dimensions, it can also have momentum components in the extra dimensions. On the other hand, from the 4-dimensional point of view, the 4-dimensional components of such $(4 + n)$-momentum appear as "mass". More precisely, since the momentum is quantized, in the 4-dimensional world there will appear a series of excited states called Kaluza-Klein (KK) towers. The mass shift Δm between two consecutive states is:

$$\Delta m \sim 1/R$$

For $M_{Pl(4+n)}$ masses in the TeV range and two extra dimensions, this mass difference between the KK excitations is so small that physics following this model would not appear as single resonances but as a continuous distortion of the measured diphoton/dilepton invariant mass spectrum.

[9]Phenomenologically, gravitons preferentially decay into two gauge bosons, such as photons, rather than into two leptons, because the graviton has spin 2, and so fermions cannot be produced in an s wave function.

Kaluza Klein excitations of the gauge bosons. In addition to the large extra dimensions of the ADD model, it is possible to add other smaller extra dimensions, with $R \sim 1$ TeV^{-1}, where the graviton but also the SM gauge bosons are allowed to propagate. This scenario would lead to a complete different phenomenology: the KK modes in fact would be spaced every few TeV in mass and the first would be the only one likely to be visible at the LHC.

Warped extra dimensions: the Randall-Sundrum model

A closer look at the geometric solution for the hierarchy problem proposed by ADD reveals one flaw: the discrepancy between M_{Pl} and M_{EW} is not really removed, it is moved elsewhere. In fact, a result of compactifying the extra dimensions on a circle with a radius of R is a new hierarchy between the electroweak scale, and the compactification scale $1/R$. This inspired a different approach, proposed by Randall and Sundrum [27], with only one extra dimension[10] with its inverse radius[11] $1/r_c = 10^{19} \sim M_{Pl}$ GeV (in order to eliminate the "flaw" of the ADD model) and non-trivial geometry. The authors propose a set-up with the extra dimension, denoted by the coordinate $\phi \in [-\pi, \pi]$, which is compactified on an orbifold (circle with the additional condition (x,ϕ)=(x,$-\phi$)). The two fixed points of the orbifold ($\phi = 0, \pi$) hold the three "standard" spatial dimensions. As for the ADD model, the full space is often called the *bulk*, while a subset with p dimensions is denoted as p-brane; the usual 3 spatial dimensions are then referred to as the 3-brane, or even the $\phi = \pi$-brane[12]. The 5-dimensional action for the gravity sector is given by:

$$S_{gravity} = \int d^4x \int_{-\pi}^{+\pi} d\phi \sqrt{-\mathcal{G}} \left\{ -\Lambda + 2M^3_{Pl(5d)} R \right\} \tag{2.14}$$

where $\mathcal{G}$ is the determinant of the 5-dimensional metric $G^{\mu\nu}$, Λ is the cosmological constant, $M_{Pl(5d)}$ is the 5-dimensional Planck mass, and R the 5-dimensional Ricci scalar.

The solution of the Einstein's equations in such 5-dimensional space, if one requires to satisfy the 4-dimensional Poincaré invariance, is an anti-de Sitter (AdS_5) space with non-factorizable geometry, given by the metric:

$$ds^2 = e^{-2kr_c|\phi|} \eta_{\mu\nu} dx^\mu dx^\nu + r_c^2 d\phi^2 \tag{2.15}$$

where k is the AdS_5 curvature parameter, r_c is the compactification radius, x^μ are the standard 4-dimensional coordinates and $\eta_{\mu\nu}$ is the 4-dimensional metric tensor. This metric has two important consequences:

- The 5-dimensional and 4-dimensional Planck masses are related by:

$$M^2_{Pl} = \frac{M^3_{Pl(5d)}}{k} (1 - e^{-2kr_c\pi}) \tag{2.16}$$

  M_{Pl} and $M_{Pl(5d)}$ are therefore close to each other for $k \sim A/r_c$, with A a positive number "not too big", of order $\mathcal{O}(1-10)$.
  This request is compatible with the fact that, in order to solve the SM hierarchy problem via imposing its scale of validity $\Lambda_\pi \sim 1$ TeV, one can infer from

$$\Lambda_\pi = M_{Pl} e^{-kr_c\pi}$$

  a value of $kr_c = A \sim 10$.

[10] There are indeed other versions of the model with more than 1 extra dimension, but they won't be treated in this section.

[11] The radius is called r_c in this section to avoid confusion with the Ricci scalar R also present in the equations.

[12] As explained before, given the periodic conditions of the space, $\phi = \pi$ and $\phi = 0$ are indistinguishable.

- Assuming that the SM fields are confined to the brane located at $\phi = \pi$, the electroweak energy scale v, or any physical mass m in the $\phi = \pi$-brane is exponentially suppressed compared to the 5-dimensional mass m_0:

$$m = e^{-kr_c\pi} m_0 \tag{2.17}$$

Hence, in the Randall-Sundrum approach, there is only one fundamental scale, that being the Planck scale, since:

$$M_{Pl} \sim M_{Pl(5d)} \sim m_0 \sim 1/r_c \sim 10^{19} \text{ GeV}$$

KK graviton excitations in this model have significantly different properties than in the case with large extra dimensions. The interactions of matter fields with the graviton towers are given by the lagrangian [33]

$$L = -\frac{1}{M_{Pl}} T^{\alpha\beta} h^{(0)}_{\alpha\beta} - \frac{1}{\Lambda_\pi} T^{\alpha\beta} \sum_{n=1}^{\infty} h^{(n)}_{\alpha\beta} \tag{2.18}$$

where $h_{\alpha\beta}$ are the graviton fields, $T^{\alpha\beta}$ is the energy-momentum tensor of the matter fields, and Λ_π is the coupling parameter, introduced earlier, being ~ 1 TeV. As can be seen from the lagrangian, while the massless zero-mode of the graviton couples with $1/M_{Pl}$ strength (as for the ADD case), the coupling of the KK excitations is comparable to $1/\Lambda_\pi$. The masses of the KK modes are:

$$m_n = k x_n e^{-kr_c\pi} \tag{2.19}$$

where x_n is the n-th root of the J_1 Bessel function. Such KK gravitons would show up in collider experiments as individual resonances in the high tail of the diphoton/dilepton mass spectrum[13], as the mass difference between two consecutive states is expected to be around 1 TeV. The Randall-Sundrum model phenomenology is governed by two free parameters: the mass of the first graviton excitation m_1 and the normalized curvature parameter c, defined as $c = k/M_{Pl}$, sometimes also denoted as $\widetilde{k}$.
A different approach used in order to define the graviton model under consideration uses the ratio between the natural width of the graviton and its mass, denoted as Γ/M. This is the approach followed in Chapter 7.

2.2.4 Status of the searches at the end of LHC RunI era

This thesis will only deal with the results coming from the analyses of the second period of data taking (RunII) of the LHC collider, which started in the second half of 2015, after a long shut-down period of approximately 2 years (see Chapter 3).

However, the hunt for new heavy particles that could decay in the dielectron/diphoton final state did not start at the LHC. Direct searches in the high mass tail of both dielectron and diphoton mass spectra have already been conducted at the Tevatron. The CDF and D0 collaborations excluded a Z'_{SSM} in the dielectron channel below 963 GeV and 1023 GeV respectively [34, 35]. A search for KK graviton decaying in two photons has been also performed by the CDF [36] experiment, excluding graviton masses up to 459 GeV and 963 GeV depending of the parameters of the models. Morevoer, both CDF [37] and D0 [38] set a lower limit of ≈ 1050 GeV on the mass of a Randall-Sundrum graviton (in the scenario $c = 0.1$) combining the dielectron and diphoton channels.

Finally, given the amount of publications, the results from the first period of data taking (RunI) at the LHC are presented divided per decay channel:

[13] As exposed earlier, a graviton preferentially decays in the diphoton final state.

- **Dielectron final state:** The ATLAS and CMS collaborations published results from pp collisions at $\sqrt{s} = 7$ TeV [39, 40], at a combination of $\sqrt{s} = 7$ TeV and early 2012 $\sqrt{s} = 8$ TeV [41], and at $\sqrt{s} = 8$ TeV [42, 43]. Searches targeting signals from large extra dimensions have also been performed both by ATLAS [44] and CMS [45, 46]. Table 2.4 lists the lower limits obtained separately in the dielectron and the dimuon final states, and the combined dielectron and dimuon lower limits for the mass of various heavy resonances, obtained from searches at the LHC with a center-of-mass energy of 8 TeV and ≈ 20 fb^{-1} of integrated luminosity.

Lower mass limit [TeV] (Results obtained for $\sqrt{s} = 8$ TeV)				
ATLAS				
	dielectron	dimuon	combined	
Z′ model	observed	observed	expected	observed
Z'_{χ}			2.60	2.62
Z'_{ψ}			2.46	2.51
Z'_{SSM}	2.79	2.53	2.87	2.90
$G_{RS}(c/M_{Pl} = 0.01)$			1.28	1.25
$G_{RS}(c/M_{Pl} = 0.05)$			2.25	2.28
$G_{RS}(c/M_{Pl} = 0.1)$			2.67	2.68
CMS				
	dielectron	dimuon	combined	
Z′ model	observed	observed	expected	observed
Z'_{ψ}	2.34	2.39	2.57	2.57
Z'_{SSM}	2.67	2.73	2.90	2.90
$G_{RS}(c/M_{Pl} = 0.01)$	1.25	1.13	1.38	1.27
$G_{RS}(c/M_{Pl} = 0.05)$	2.13	2.12	2.35	2.35
$G_{RS}(c/M_{Pl} = 0.1)$	2.50	2.56	2.73	2.73

Table 2.4: Observed 95% confidence level (CL) limits in the dielectron and dimuon channels separately, and 95% CL limits in the combined final state, on the Z′ or a Randall-Sundrum graviton G_{RS} mass for various models, obtained from about 20 fb^{-1} of $\sqrt{s} = 8$ TeV LHC data, recorded in 2012.

- **Diphoton final state:** graviton searches in the diphoton final state have been published by the ATLAS and CMS collaborations at $\sqrt{s} = 7$ TeV [47, 48] and $\sqrt{s} = 8$ TeV [49, 50], where the last two referenced searches have been performed in the "low" energy range by both ATLAS (up to 650 GeV) and CMS (up to 850 GeV). There is also a more recent search by CMS exploring the high mass range (up to 3 TeV) at 8 TeV [51]. In this last case, specific models have been considered and so it is possible to quote the excluded mass region in a particular theoretical scenario, while in the previous searches no specific model was considered and it is only possible to state the the inspected mass region did not show any significant deviation from the background only hypothesis. For [51] 95% confidence level limits are set in the parameter space of the Randall-Sundrum model: masses below 1450 - 2780 GeV are excluded for the first excited state of the Randall-Sundrum graviton.

  Table 2.5 lists the mass range explored in the diphoton final state from searches at the LHC with a center-of-mass energy of 7 and 8 TeV.

With the increase of the available center-of-mass energy $\sqrt{s}$ from 8 TeV of RunI to 13 TeV of RunII, the phase space for a possible discovery have been dramatically increased (see Chapter 3), hence the possibility for the discovery of a new particle has been considerably improven. The results shown in this thesis are based on the proton-proton collision data collected by the CMS experiment (see Chapter 4), for an integrated luminosity of

Experiment	Year	$\sqrt{s}$ (TeV)	L (fb^{-1})	Mass range (GeV)
ATLAS	2013	7	4.9	142-3000
CMS	2013	7	2.2	200-2000
ATLAS	2014	8	20	65-650
CMS	2014	8	19.7	150-850
CMS	2015	8	19.7	500-3000

Table 2.5: Summary of diphoton searches at the LHC RunI, with the corresponding inspected mass region.

35.9 (12.9) fb^{-1} for the dilepton (diphoton) final state from the beginning of the RunII data taking period (2015) to the end of the 2016 data taking.

Chapter 3

Physics of proton-proton collisions at the Large Hadron Collider

In this chapter the design of the Large Hadron Collider will be described as well as the phenomenology of the proton-proton interactions. While the Large Hadron Collider can also support lead-lead or lead-proton collisions, the following chapter will describe only the phenomenology of the proton-proton collisions as they correspond to the dataset analyzed for the physics results exposed in this thesis.

3.1 The LHC collider

The Large Hadron Collider (LHC) [52] is a proton-proton (pp) accelerator and collider installed in the same circular underground tunnel occupied until the year 2000 by the Large Electron Positron collider (LEP)[1]. The tunnel is located at approximately 100 m below the level of the ground[2], under the Geneva area, and has a circumference of 26.7 km. Due to the need of achieving higher center-of-mass energy, pp proton beams replaced the electron and positron beams used for LEP. In fact, in a circular collider of radius R, the energy loss per turn due to the synchrotron radiation is proportional to $(E/m)^4/R$, where E and m are respectively the energy and mass of the accelerated particles: the use of protons, due to their higher mass compared to electrons, implies a smaller energy loss from synchrotron radiation. A pp collider was preferred to a $p\bar{p}$ collider because it allows to reach higher event rates. The low anti-proton production efficiency (10^5 protons are needed to create an anti-proton via the reaction $pp \rightarrow ppp\bar{p}$) and larger time needed to accumulate them would make almost impossible to achieve the required statistics of events needed in the search for new physics at the LHC. This in turn excludes the $p\bar{p}$ collider configuration, where a common vacuum and magnet system are used for both circulating p and $\bar{p}$ beams, as it was done for example at the Tevatron: to collide two counter-rotating proton beams opposite magnetic dipole fields are required with separate vacuum chambers. Because of the limited size of the tunnel inherited from the LEP era, the LHC uses twin bore magnets instead of two separate rings of magnets.

The LHC ring is divided in eight arcs and eight straight sections, four of them host various equipment needed for the accelerator while in the other four the two beams are brought into collision at the four interaction points (IPs), where the four main experiments of the LHC are located:

[1] In this latter case, electrons and positrons were accelerated and brought in collisions.
[2] The closest point to the surface is at 45 m, while the deepest one is at 170 m.

- CMS (Compact Muon Solenoid) [53], which will be described in greater details in Chapter 4, is a general-purpose detector with a large program of physics analyses;
- ATLAS (A Toroidal LHC ApparatuS) [54] covers the same physics objectives as CMS while exploiting different technical solutions, including a large toroidal magnet;
- LHCb (Large Hadron Collider beauty) [55] studies matter-antimatter asymmetry via CP violation, through studies involving b quarks;
- ALICE (A Large Ion Collider Experiment) [56] is designed to address the physics of the quark-gluon plasma at extreme values of energy density and temperature in nucleus-nucleus collisions.

ATLAS and CMS analyze completely independent datasets and, since they exploit completely different technical implementation, their respective results can cross-check each other and be combined to increase the precision of the measurements.

The existing CERN infrastructure, shown in Figure 3.1, is used for injecting the protons into the LHC.

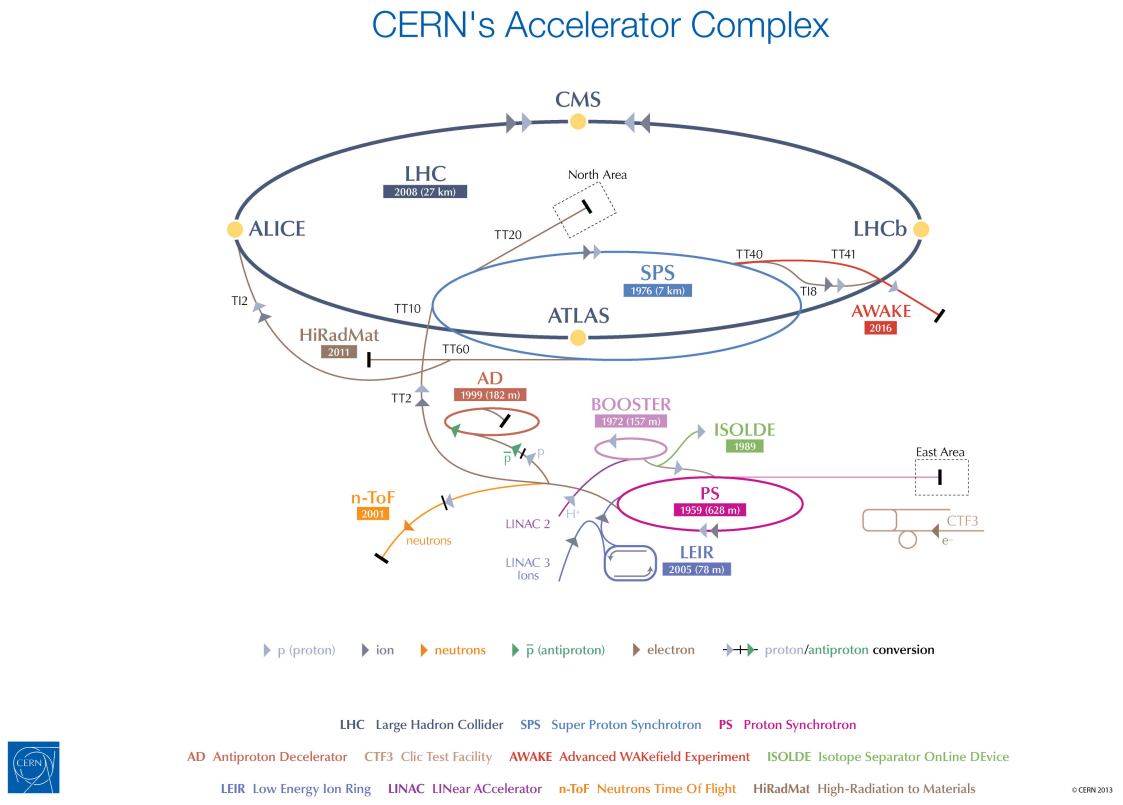

Figure 3.1: The CERN accelerator complex.

In particular, protons are produced starting from a duoplasmatron, which is an ion source where a heated cathode emits electrons, which in turn ionize a hydrogen gas, forming a plasma. An electric field confines the plasma, while an electrode extracts the protons from it. Protons are first accelerated in a linear accelerator, the LINAC2, until they reach an energy of about 50 MeV. They are then injected in a circular accelerator, the Proton Synchroton Booster (PSB), where they reach an energy of 1.4 GeV, before entering the Proton Synchrotron (PS). In the Super Proton Synchrotron (SPS), the proton energy

increases from 26 to 450 GeV, and the protons are then injected in the LHC, where they are finally accelerated to their final energy (3.5 TeV in 2011, 4 TeV in 2012, 6.5 TeV in 2015 and 2016). Three to four cycles of the PS synchrotron are needed to fill the SPS, whereas twelve cycles of the SPS are required to fill the LHC. The total injection time is about twenty minutes, and about twenty additional minutes are needed to increase the beam energy from 450 GeV to 6.5 TeV. When completely filled, the LHC nominally contains 2808 bunches of approximately 10^{11} protons.
Once inside the LHC, the protons are accelerated by sixteen radiofrequency cavities, while 1232 niobium-titanium superconducting dipole magnets ensure the deflection of the beams generating a magnetic field up to 8.3 T; quadrupole magnets are used to collimate the beams. The superconducting magnets used to bend the protons trajectory operate at a temperature below 2 K, obtained with a pressurized bath of superfluid helium at about 0.13 MPa. Three vacuum systems are also part of the LHC architecture: the beam vacuum (10^{-10} to 10^{-11} mbar at room temperature), the insulation vacuum for helium distribution (about 10^{-6} mbar) and the insulation vacuum for cryomagnets (about 10^{-6} mbar).
The machine is designed to achieve an energy per proton of 7 TeV, which would result in a design center-of-mass energy of $\sqrt{s} = 14$ TeV. The time between two bunch crossings in any IP is 25 ns, which means a bunch space ≈ 7.5 m along the beam axis. Neverthless, not all the design parameters has been reached in the operations: in particular the center-of-mass energy of 14 TeV has not been achieved yet. In the years 2010 and 2011 the LHC operated with proton beam energies of 3.5 TeV. In 2012, the beam energy of 4 TeV was reached, resulting in a proton-proton center-of-mass energy of 8 TeV and a bunch spacing of 50 ns. This first LHC running period is called RunI and won't be considered in this thesis.
Starting from spring 2013, the LHC went through a shut down period of about 2 years to allow consolidation and upgrade of numerous machine systems. In July 2015 LHC started to collide proton beams with a center-of-mass energy of 13 TeV (LHC running period called RunII). After a short period of 50 ns operation (Run2015B), the machine collected data with a bunch spacing of 25 ns in both 2015 and 2016. The LHC will continue operate at 13 or 14 TeV center-of-mass energy until 2018. A second long shutdown (LS2) is planned in 2019 and 2020, while RunIII will extend until 2023. After the Phase-1, which includes RunI, RunII and RunIII, the Phase-2 should extend the data taking up to approximately 2037. The integrated luminosity (see next section) collected in Phase-1 is expected to reach 300 fb^{-1}, while 3000 fb^{-1} should be collected by the end of Phase-2. An overview of the LHC schedule up to 2021 is presented in Figure 3.2.

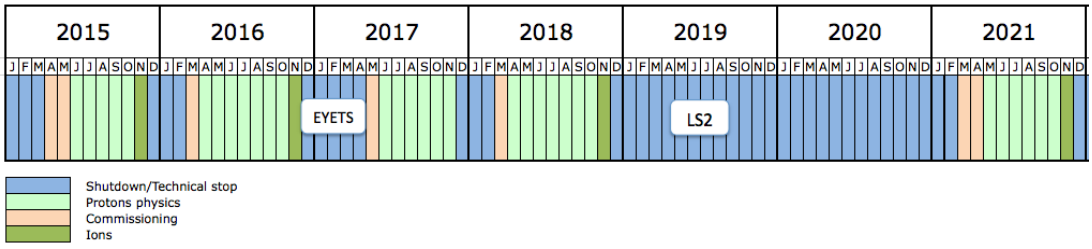

Figure 3.2: The LHC schedule up to year 2021.

Luminosity

An important parameter used to quantify the performances of a collider is the instantaneous luminosity $\mathcal{L}$. Given the event rate R_i of the process i, defined as the number of events occurring per unit of time, one can write:

$$R_i = \frac{dN_i}{dt} = \sigma_i \times \mathcal{L} \tag{3.1}$$

where σ_i is the cross section of the process i and $\mathcal{L}$ is the machine instantaneous luminosity. The instantaneous luminosity depends only on machine parameters and can be written as:

$$\mathcal{L} = \frac{N_1 N_2 n_b f_{rev}}{A} \tag{3.2}$$

where N_1 and N_2 are the number of particles in the two colliding bunches, A is the overlap area of the two bunches in the transverse plane with respect to the beam direction, n_b is the number of bunches in each beam, and f_{rev} is the revolution frequency of the bunches (with a design value of 11245 Hz). At the LHC, $N_1 = N_2$ both equal to the number of protons per bunch N_p ($\approx 10^{11}$) and, since the area of overlap is difficult to be measured, for a Gaussian-shaped beam distribution $\mathcal{L}$ can be rewritten as :

$$\mathcal{L} = N_p^2 n_b f_{rev} \frac{\gamma}{4\pi\epsilon_n \beta^*} F \tag{3.3}$$

where γ is the relativistic Lorentz factor, ϵ_n is the normalized transverse beam emittance (with a design value of 3.75 μm), β^* is the so called betatron function at the IP, and F is the geometric luminosity reduction factor due to the crossing angle at the IP.

The maximum number of bunches per beam and the revolution frequency are defined by the circumference of the LHC. Hence, in order to get as many events of interest as possible, on can either increase the number of particles in a bunch or focus the two beams on a smaller area.

The summary of the the design values for the LHC machine parameters can be found in Table 3.1. During collisions the number of particles in the bunches, and thus also the

	Parameter	Design
Center-of-mass energy	$\sqrt{s}$	14 [TeV]
Luminosity	$\mathcal{L}$	10^{34} [cm^{-2}s^{-1}]
Bunch spacing		25 [ns]
Number of bunches	n_b	2808
Number of protons/bunch	N_p	1.15×10^{11}
Emittance	ϵ_n	3.75 [μm]
β^* at the IP	β^*	0.55 [m]

Table 3.1: Machine parameters (design values) of the LHC.

instantaneous luminosity, decreases exponentially from the initial peak luminosity and after several hours of data dating the instantaneous luminosity $\mathcal{L}$ tends to decrease so much that it is more efficient to abort the fill and refill the machine with new beams instead of taking data at very low luminosity. The peak luminosity of the LHC for both 2015 and 2016 is shown in Figure 3.3.

If $\mathcal{L}$ is the instantaneous luminosity, L is the integrated luminosity, where $\mathcal{L}$ is integrated over time. The integrated luminosity for all the pp fills taken during 2015 and 2016 is shown in Figure 3.4.

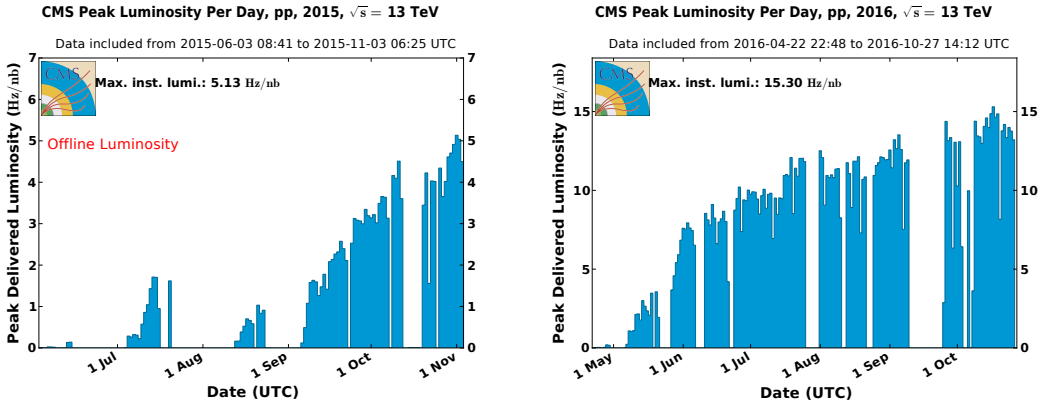

Figure 3.3: Peak luminosity delivered during day for 2015 (left) and 2016 (right) as measured by the CMS experiment [57].

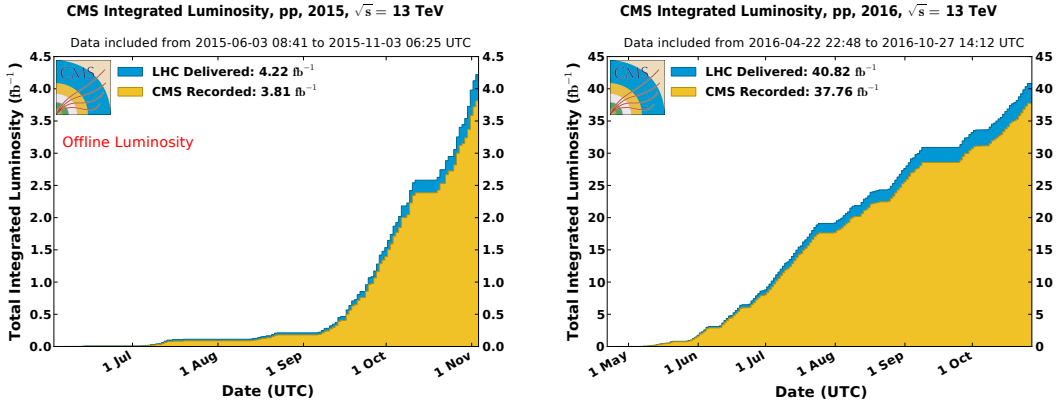

Figure 3.4: Cumulative offline luminosity versus day delivered to (blue), and recorded by CMS (orange) during stable beams for pp collisions at 13 TeV center-of-mass energy in 2015 (left) and 2016 (right) [57].

3.2 Phenomenology of proton-proton interactions

The phenomenology of proton-proton interactions have peculiar characteristics that have to be taken into account both in the machine design and at analysis level:

- The proton is a composite particle whose *valence* quarks are two up-type quarks and one down-type quark. The rest mass of the valence quarks added together makes up about 1% of the total proton mass of 938.3 MeV. The quarks are in fact held together inside the proton by gluons that are interacting with them and exchanging the colour charge. However, at higher loop order the gluon exchanged between two quarks can interact with other gluons in the proton or produce a quark-antiquark pair that annihilates shortly afterwards. Such quarks and antiquarks produced from gluons are called *sea* quarks. Both valence and sea quarks are generally referred to as *partons*, the inner constituents of the proton.

- There are two possible kinds of proton-proton interactions: soft and hard collisions. In soft collisions only a small momentum is transferred and particle scattering at large angle is suppressed. The final state particles have small transverse momentum ($\sim 10^2$ MeV), so that most of them escape down the beam pipe. The other possible scenario is when two of the proton partons take part in the interaction (referred to as hard interaction) with high transferred transverse[3] momentum. From now on, only hard collisions will be considered. The corresponding cross sections are several order of magnitudes smaller than those of the soft interactions, as depicted in Figure 3.5.

- The elementary cross section associated to a given hard interaction involving a parton a and a parton b is estimated using the QFT prescriptions described in Chapter 1. In order to estimate the total cross section of the process though, one needs to know also the probability of finding such partons in the colliding protons and their momentum distribution. For this purpose, see the description of the parton distribution functions in Section 3.2.1.

- Before or after the two partons interact with each other, they can radiate other gauge bosons. This radiation of particles is called *initial state radiation* (ISR) when it happens before the hard interaction, and *final state radiation* (FSR) if it occurs with the decay products of the hard interaction.

- All the observed hadrons are colorless, hence if the final state of a hard interaction contains particles that carry a colour charge (like e.g. quarks), they have to form new particles re-arranging their colour structure in order to become colour neutral. This process is called *hadronization* and results in showers of particles that form a cone along the initial particles direction which are called *jets*. The only exception is the top quark, which has a lifetime shorter than the timescale at which the hadronization takes place and, therefore, decays before it can hadronize. After the hard interaction, the remnants of the two protons are not colour neutral anymore and have to hadronize as well, forming jets that fly along the beam axis (underlying event, whose produced hadrons carry in general small transverse momenta). The hadronization process is currently described only phenomenologically through various models, among which the most commonly used is the Lund string model [59].

- Finally, the fact that the LHC bunches are made of many ($\approx 10^{11}$) protons leads to what is called *pile-up* in the event, due to the presence of several proton-proton interactions besides the one of interest (see Section 3.2.2).

[3] Everywhere in this thesis, transverse is defined with respect to the beam directions.

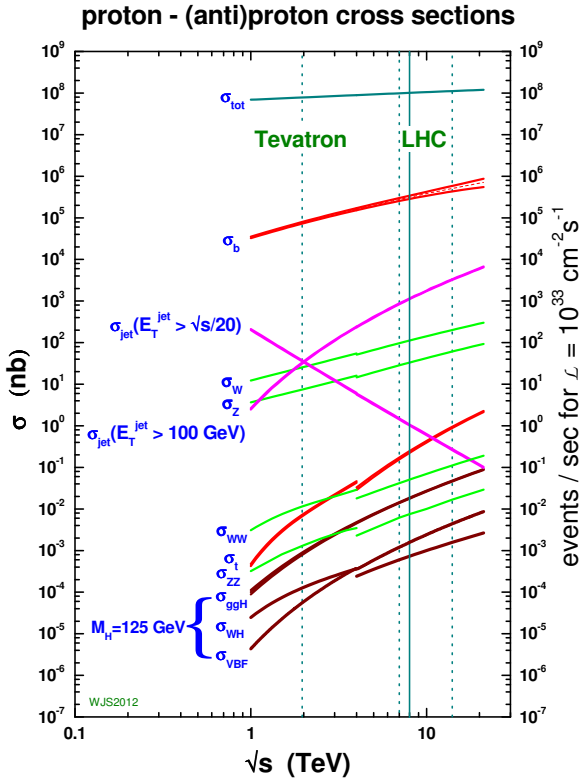

Figure 3.5: Cross section of SM processes as a function of the center-of-mass energy of proton proton collisions. The vertical lines mark the center-of-mass energies of the Tevatron (1.96 TeV) and the LHC (7,8 and 13 TeV) [58].

.

3.2.1 Parton distribution functions

The parton distribution function (PDF) is the probability density to find a parton p carrying a fraction x of the longitudinal proton momentum of a proton (see note made for Equation 3.4). The PDF depends on the squared four-momentum Q^2 transferred in the collision and is labeled as $f_p(x_p, Q^2)$. PDFs are different for gluons, valence quarks and low-momentum sea quark-antiquark pairs of all flavours and depend on the energy scale at which the interaction between the partons takes place: in particular for interactions with "high exchanged momenta" where a shorter distance scale is probed, the contributions of gluons and sea quarks become dominant.

PDFs don't come directly from first-principles criteria, hence must be experimentally measured. Currently they are essentially extracted from the study of lepton-hadron collisions such as those provided by HERA [60, 61]. Once the PDF have been measured at a given Q^2, it can be extrapolated at another energy using the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) evolution equation [62, 63, 64]. Different sets of PDF exist provided by different collaborations, such as CTEQ [65, 66], MSTW [67] or NNPDF [68]: As an example MSTW parton distribution functions are shown in Figure 3.6 for two different values of Q^2 and superimposing the PDFs of the different quark flavors and gluons. The bump around $x = 0.1$ visible for up and down quarks in both left and right plots of Figure 3.6 corresponds to the contribution of the valence quarks.

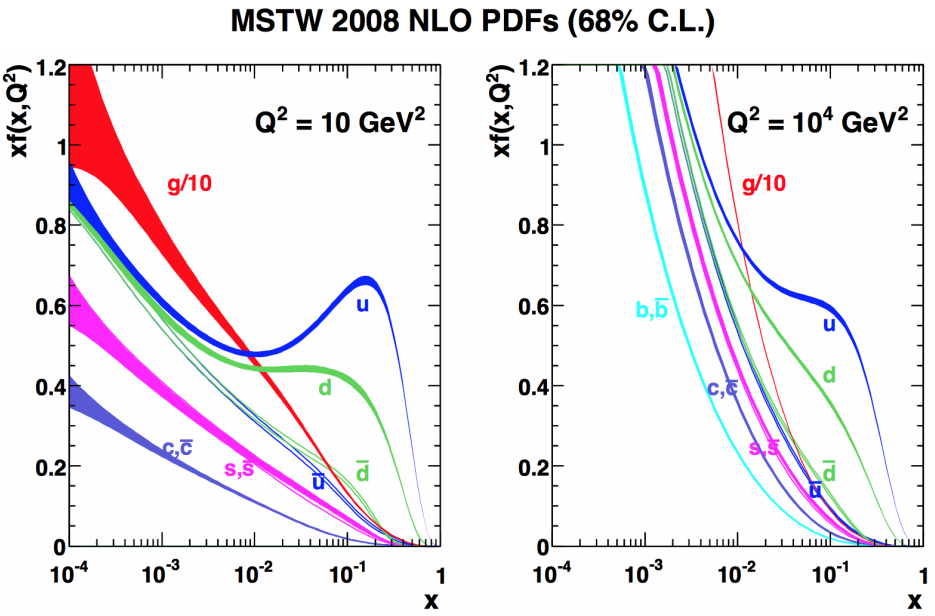

Figure 3.6: Parton density functions multiplied by the fraction of the total momentum carried by the parton ($xf(x, Q^2)$), including the 1 standard deviation uncertainty bands, for two different invariant momentum transferred Q^2=10 GeV^2 (left) and Q^2=10^4 GeV^2 (right) [67].

While the total center-of-mass energy carried by the pp system is $\sqrt{s}$, the effective center-of-mass energy of the hard scattering $\sqrt{\hat{s}}$ is given by the following relation:

$$\sqrt{\hat{s}} = \sqrt{x_a x_b s} \tag{3.4}$$

where x_a and x_b are the fractions of the longitudinal momentum of the proton carried by the interacting partons[4] a and b. To probe physics at a certain energy scale, the Q^2 value has to be in the range of the squared effective center-of-mass energy ($\hat{s}$) of the hard scattering, which corresponds to the squared invariant mass M^2 of the system of the interacting particles. From Equation (3.4) this means that to study physics at the TeV scale with a collider reaching $\sqrt{s}$ = 13 TeV, the average x of the partons has to be around 0.1. From the corresponding PDF in Figure 3.7 (Q^2=(1000 GeV)^2) it can be seen that at such values the up quark and down quark contents show an excess over the other quarks, which means that the interactions are dominated by the valence quarks and the gluons.

Once the PDFs are known at a given Q^2, the total proton-proton cross section is computed as the convolution of the elementary cross section and the PDFs, integrated over all the possible elementary processes (also known as *factorization theorem* [70]) giving rise to the considered final state. Figure 3.8 clearly shows that the RunII data taking at 13 TeV is expected to benefit from a significant increase in the cross section of the production of heavy resonances due to the fact that the ratio between the PDFs in the two scenarios $\sqrt{s}$ = 13 TeV and $\sqrt{s}$ = 8 TeV (end of RunI scenario) is greater than 1. For example, for a resonance of mass $\approx$ 2.5 TeV (from $q\bar{q}$ production mode) there is a factor $\approx$ 10 between the 13 TeV scenario and the 8 TeV scenario [71]. This fact can make the discovery of a new TeV scale particle possible with a few fb^{-1} of integrated luminosity and strongly motivates the searches for new physics at high mass.

[4]This is true in a fast moving frame, also referred to as "infinite momentum frame".

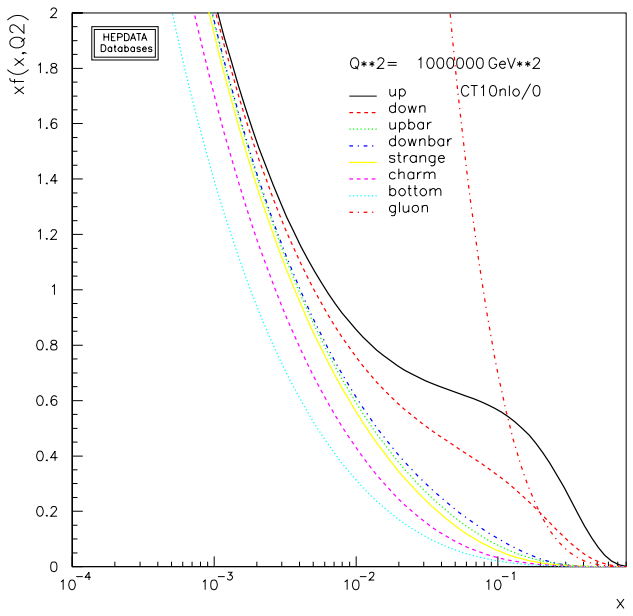

Figure 3.7: PDFs for different partons in a proton, obtained with the CT10 parametrisation [65] where the Q^2 is chosen for physics studies at the TeV scale. The plot was generated with the specific tool from the HepData project [69].

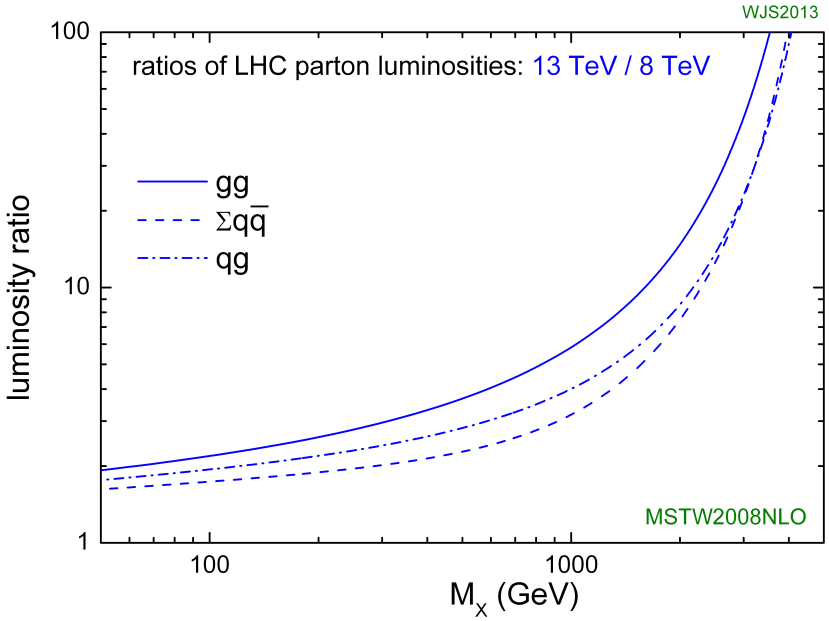

Figure 3.8: Ratio of LHC parton luminosities between 13 and 8 TeV [71].

3.2.2 Pile-up

More than one independent proton-proton interactions can take place simultanously in a bunch crossing at the interaction point. The interaction of two protons forms a primary vertex, from which the particles that were created in the interaction originate.

The number of primary vertices created on average depends on the beam parameters, e.g. the number of particles in a bunch and how small is the focusing area of the bunches. In 2012 this number has been measured by the CMS experiment and, on average, corresponded to 21 interactions per bunch crossing as shown in Figure 3.9 [57]. Similar plots for 2015 and 2016 have not been published yet by the CMS collaboration, but they are derived at analysis level and available in the Sections 6.1 and 7.1.

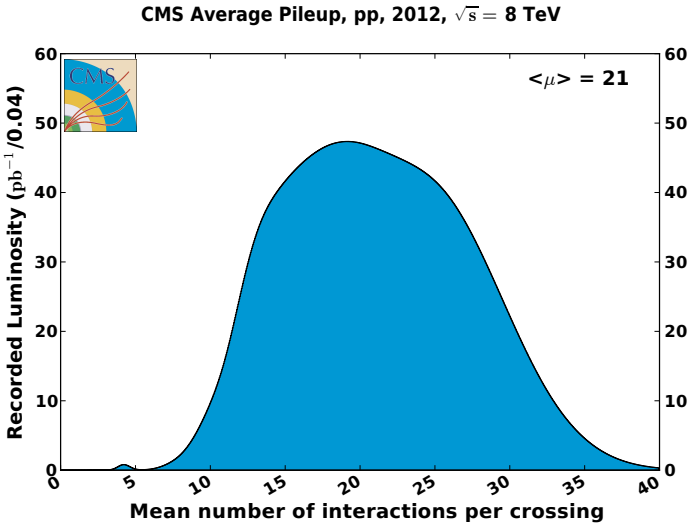

Figure 3.9: Luminosity recorded as a function of the mean number of interactions per bunch crossing in the CMS experiment in 2012 [57].

The presence of many primary vertices per bunch crossing presents a challenge for the event reconstruction, since the particles originating from different primary vertices are superimposed in the detector.

Chapter 4

The CMS detector

This chapter describes the general concepts of the Compact Muon Solenoid (CMS) experiment. Its design is briefly exposed starting form the innermost region to the outermost one. Given its importance in the context of this thesis, more details are given concerning the electromagnetic calorimeter (ECAL), while less information are provided for the other subsystems.

4.1 Overall concept

The Compact Muon Solenoid (CMS) is one of the two general purpose experiments taking data at the LHC, the other one being ATLAS (A Toroidal LHC ApparatuS). After the discovery of the Higgs boson (see Section 1.3) achieved in 2012 by the ATLAS [4] and CMS [5] experiments, the main goal of the two collaborations become the search for new physics beyond the standard model, either via direct discovery or via precision measurements of already known particles and phenomena [72]. The CMS detector has a cylindrical shape with an total length of 28.7 m, 21.6 m of which make the main cylinder with a diameter of 15 m, and the rest of the length coming from the forward calorimeter. The total mass is 14000 t. CMS is made of several layers of detectors, designed to identify and detect the different SM particles and cover most of the 4π solid angle. One of its peculiar feature is the presence of a 3.8 T superconducting solenoidal magnetic field. In order to achieve good momenta and energy resolution, the tracking detector (Tracker) and the two calorimeters, both the electromagnetic (ECAL) and the hadronic one (HCAL), are located inside[1] the volume of the magnetic field (13 m-long, with 5.9 m of inner diameter). The muon chambers are instead located outside the volume of the magnet, whose return field is large enough to saturate 1.5 m of iron, allowing four muon stations to be integrated so to ensure robustness and full geometric coverage. The overall layout of CMS is shown in Figure 4.1.

Before moving to the description of the CMS subdectors, the coordinate conventions is described in the following section.

4.2 Coordinate conventions

The origin of the coordinate system adopted by CMS is centered at the nominal collision point inside the experiment, while the y-axis points vertically upward, and the x-axis points radially inward, toward the center of the LHC ring. The coordinate system is shown in Figure 4.2. The azimuthal angle ϕ is measured from the x-axis in the x-y plane.

[1] Most of the HCAL is inside the magnet, with an outer detector (HO) located outside the coil (see dedicated section).

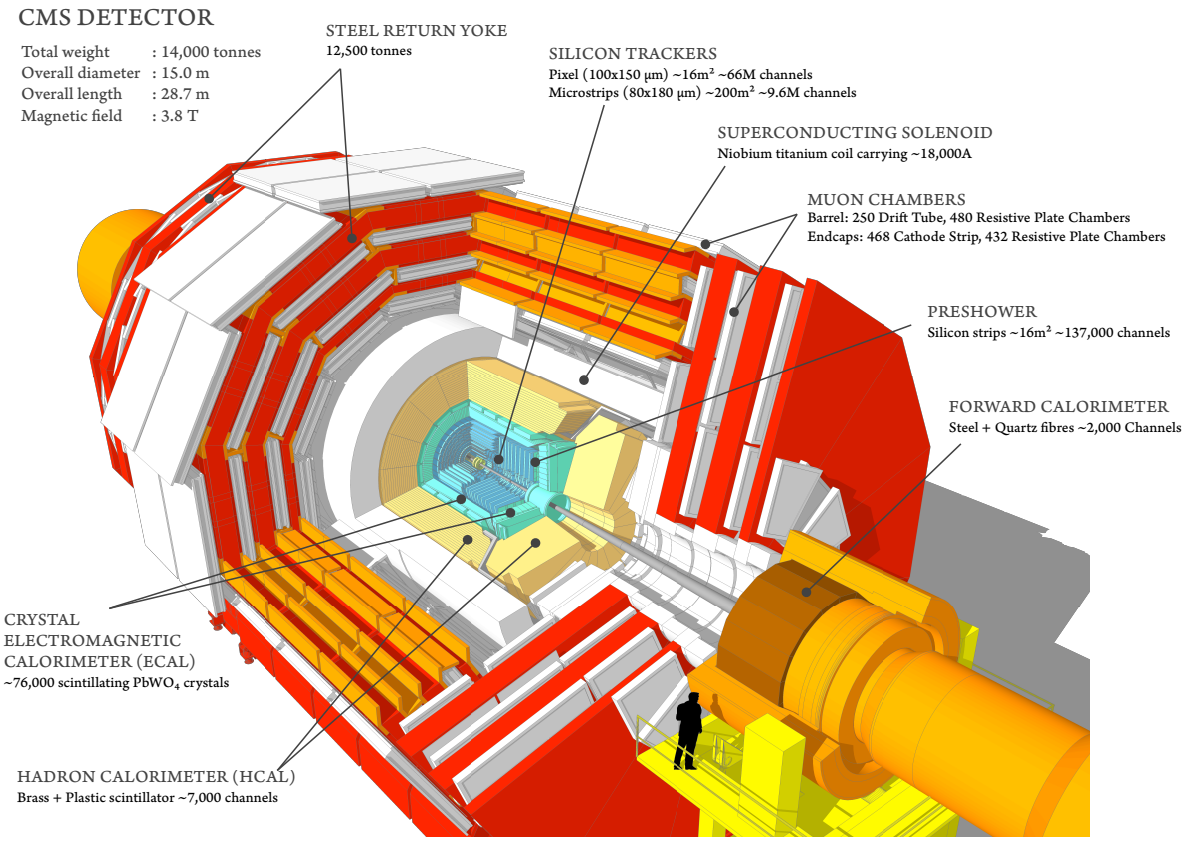

Figure 4.1: An exploded view of the CMS detector.

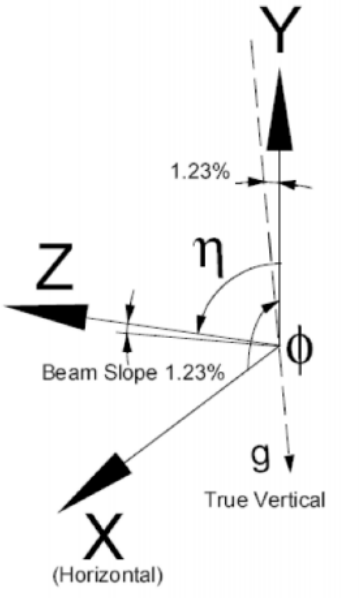

Figure 4.2: Coordinate system adopted by CMS.

The polar angle θ is measured from the z-axis.

Two important quantities for the physics analyses are defined in the transverse plane with respect to the beam direction: the so-called transverse momentum and transverse energy, denoted by p_T and E_T respectively. Their importance is given by the fact that the

interaction between the elementary partons happens with unknown center-of-mass energy (see Section 3.2) as well as unkown is the total initial energy of the interacting partons. For this reason the energy conservation rule can be applied only in the plane transverse to the beams direction. The transverse momentum and transverse energy are computed according to the following relations:

$$p_T = psin\theta \tag{4.1}$$

$$E_T = Esin\theta \tag{4.2}$$

It is also important to note that particles which escape the detection, for example neutrinos, leave an imbalance in the transverse plane total momentum, which is quantified as missing transverse energy, i.e. the negative vectorial sum of the transverse momenta of all the detected particles in the event:

$$E_T^{miss} = -\sum_i p_T^i \tag{4.3}$$

Moreover, the center-of-mass may be boosted along the beam direction. This is the reason why it is useful to use experimental quantities that are invariant under such boosts.

The rapidity is defined as:

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z} \tag{4.4}$$

and has the property of being additive under Lorentz boosts along the z direction, i.e. it is simply shifted by a constant when subjected to such transformations. For ultrarelativistic particles ($p \gg m$) the rapidity is approximated by the pseudorapidity:

$$\eta = -\ln \tan \frac{\theta}{2} \tag{4.5}$$

where θ is again the angle between the particle momentum and the z axis. The pseudorapidity can be reconstructed from the measurement of the θ angle and can be also defined for particles whose mass and momentum are not measured.

The value $\eta = 0$ corresponds to the direction perpendicular to the beams direction, while the limit $\eta = \infty$ gives the direction parallel to the beams direction. Usually, the subdetectors' regions are defined according the their η value, where the central part ($\eta < 1.5$) is called *barrel* and the outer parts ($\eta > 1.5$) are denoted as *endcaps*.

In the following sections, the CMS subdetectors are described from the innermost region (the closest one to the nominal interaction point) to the outermost region. The chapter ends with a short description of the trigger and data acquisition systems.

4.3 Inner tracking system

Placed within the magnetic field, the tracker [73] is the subdetector which is the closest one to the interaction point. Its purpose is to identify tracks and vertexes in a high track multiplicity environment. In order to provide good radiation hardness, high granularity and large hit redundancy to achieve good performances in pattern recognition, the silicon (Si) technology has been chosen for the whole volume of the tracker (given by a cylinder of 5.8 m length and 2.6 m diameter). A global view of the tracker layout is depicted in Figure 4.3.

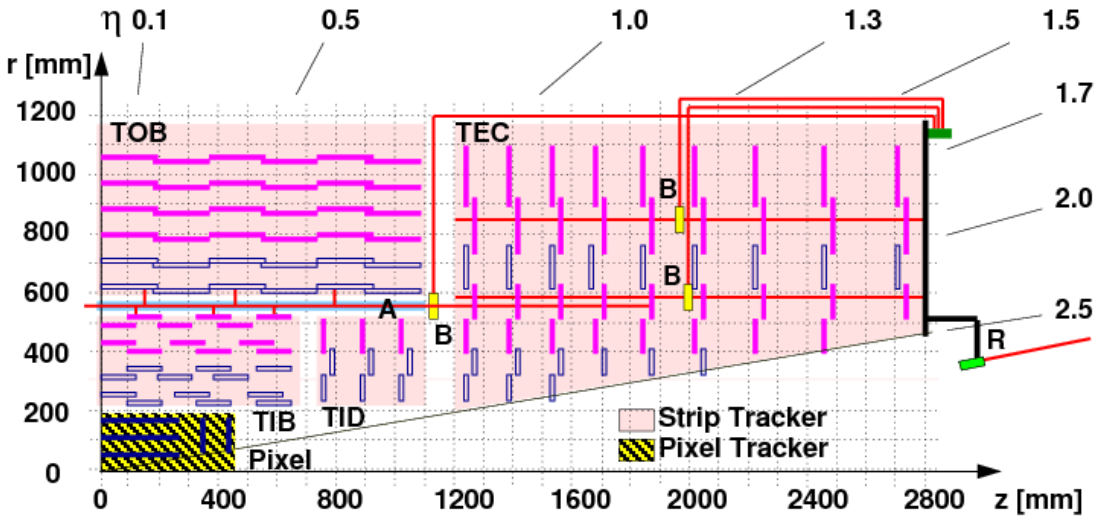

Figure 4.3: A global view of the tracker layout.

The innermost tracker is made of three layers of silicon pixel detectors named Tracker Pixel Barrel (TPB), ranging from 4.4 cm to 10.2 cm diameters, and two wheels of Tracker Pixel Endcap (TPE), covering the pseudorapidity range up to $|\eta| = 2.5$. TPB and TPE contain 48 million and 18 million pixels, respectively. The pixels have a size of $100 \times 150\ \mu\text{m}^2$.

Thanks to the large Lorentz drift angle in the magnetic field, the measured hit resolution in the TPB is 9.4 μm in the $r - \phi$ plane and 20-40 μm in the longitudinal direction. The longitudinal resolution depends on the angle of the track relative to the sensor. For longer clusters, sharing of charge among pixels improves the resolution, with optimal resolution reached for interception angles of $\pm 30°$.

The silicon strip tracker is placed outside of the pixel tracker. The barrel part of the strip tracker is divided in the 4 layers of the Tracker Inner Barrel (TIB) and the 6 layers of the Tracker Outer Barrel (TOB). Coverage in the forward region is provided by the 3 Tracker Inner Discs (TID), and the 9 disks of the tracker endcap (TEC) on each side. The dimensions of the strips varies between 80 μm in the innermost layers of the TIB, and 183 μm in the outer layers of the TOB. In the disks the dimension varies between 97 μm and 184 μm. Some of the modules are composed by two detectors mounted back-to-back with the strips rotated by 100 mrad. These double-sided (stereo) modules will also provide a measurement in the coordinate orthogonal to the strips. The single point resolution that can be achieved depends strongly on the size of the cluster and on the pitch of the sensor and varies not only as a function of the cluster width, but also as a function of pseudorapidity, as the energy deposited by a charged particle in the silicon depends on the angle at which it crosses the sensor plane. The measured hit resolution in the barrel strip detector varies between $\sim$20 μm and $\sim$30 μm in the $r - \phi$ plane in the TIB and TOB. The total area of the Si detectors is around 200 m^2, providing a coverage up to $\eta = 2.5$. The material budget inside the active volume of the tracker increases from 0.4 X_0[2] at $\eta = 0$ to around 1 X_0 at $|\eta| = 1.6$, before decreasing to 0.6 X_0 at $|\eta| = 2.5$.

[2] X_0 defines the so-called *radiation length*, i.e. the typical length after which the energy of an incoming electron is reduced by a factor $1/e$ only by radiative processes (bremsstrahlung).

4.4 Electromagnetic calorimeter

The electromagnetic calorimeter [74] plays an essential role in the context of this thesis, since it is the subdetector devoted to the identification (in combination with the tracker) and the measure of the energies of electrons and photons. ECAL is a homogeneous calorimeter of almost 76000 Lead Tungstate $PbWO_4$ scintillating crystals divided into a barrel and two endcaps, with coverage in pseudorapidity up to $|\eta| < 3.0$. A preshower system is installed in front of the edges of ECAL for π^0 rejection[3]. A 3D view of the barrel and endcap electromagnetic calorimeter is shown in Figure 4.4.

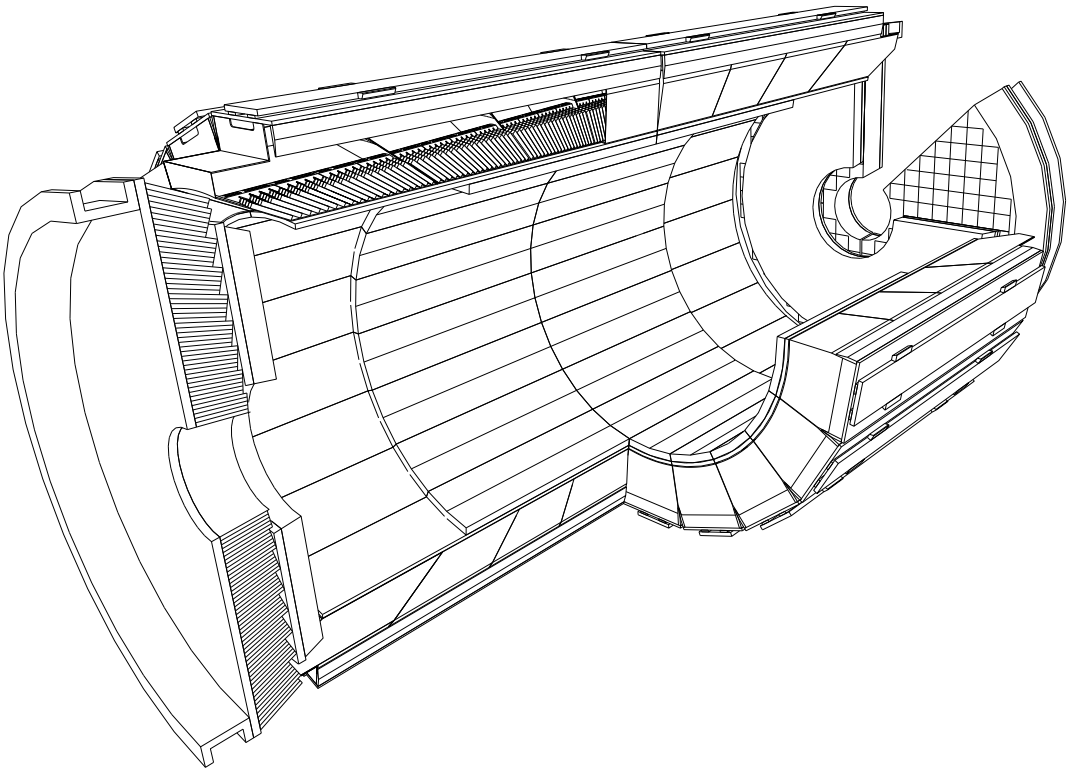

Figure 4.4: A 3D view of the electromagnetic calorimeter.

4.4.1 Barrel calorimeter

The barrel part of the ECAL covers the pseudorapidity range $|\eta| < 1.479$ (see Figure 4.5). The front face of the crystals is at a radius of 1.29 m from the nominal interaction point and each crystal face has an area of $22 \times 22\,\text{mm}^2$ and a length of 230 mm corresponding to $25.8\,X_0$. The truncated pyramid-shaped crystals are mounted in a geometry which is off-pointing with respect to the mean position of the primary interaction vertex, with a 3° tilt in both ϕ and in η directions, in order to avoid the scenario in which a particle could go right along the separation between two center-pointing crystals. In terms of $\eta \times \phi$ dimensions, the crystal size corresponds to $\Delta\eta \times \Delta\phi = 0.0175 \times 0.0175$[4]. The barrel granularity is 360-fold in ϕ and (85×2 half barrel)-fold in η, resulting in a total number of 61200 crystals. The crystal volume in the barrel amounts to $8.14\,\text{m}^3$ (67.4 t).

[3]The separation angle between two photons coming from the $\pi^0 \to \gamma\gamma$ decay, in case of a boosted π^0, would be small enough for the two photons to be misintified as only one photon, if both of them go in the same ECAL crystal. Since the energy release in the preshower is different in case of two collimated photons or only one photon, it offers a handle in order to solve this ambiguity.

[4]There are 360 crystals in the ϕ direction, which means that a crystal covers exactly 1° in ϕ.

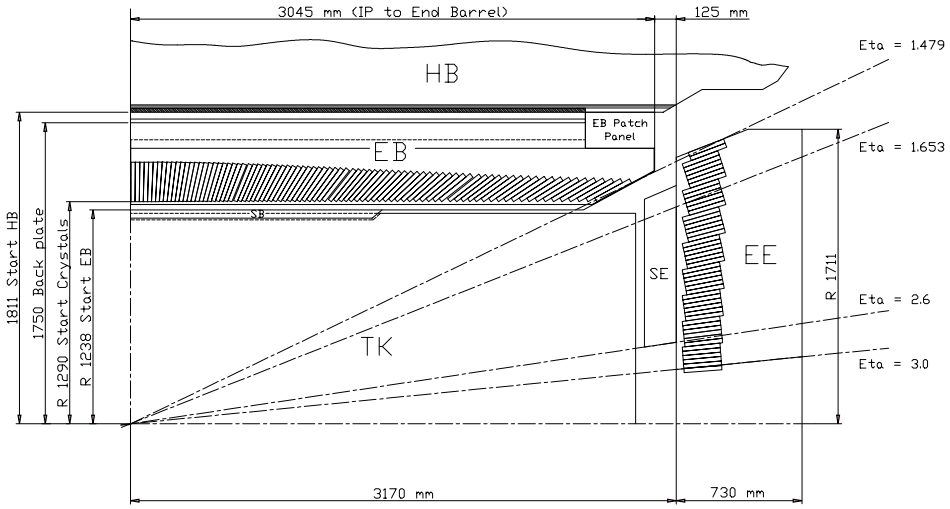

Figure 4.5: Longitudinal section of the electromagnetic calorimeter (one quadrant).

Crystals for each half-barrel are grouped in 18 supermodules, each subtending 20° in ϕ. Each supermodule consists of four modules: 500 crystals are located in the first module and 400 crystals in each of the remaining three ones. For the sake of simplicity in construction and assembly, crystals have been grouped in arrays of 2×5 crystals, contained in a very thin ($200\,\mu$m) alveolar structure. Thermal regulation is granted by two active systems: a specially regulated cooling circuit which keeps the operating temperature (ambient temperature) of the crystal array and of the APDs (avalanche photodiodes) within a tight temperature spread of $\pm0.05\,^{\circ}$C, ensuring adequate thermal stability; and the power cooling circuit which antagonises the heat generated by all power sources in the supermodule (each supermodule is designed as a separate thermal entity).

4.4.2 Endcap calorimeter

The endcap part of the ECAL covers a pseudorapidity range from 1.48 to 3.0. The design of the endcaps provides precision energy measurement up to $|\eta| = 2.5$. Crystals are however installed up to $|\eta| = 3$ in order to extend the energy-flow measurement in the forward direction. The mechanical design of the endcap calorimeter is based on an off-pointing pseudo-projective geometry using tapered crystals of the same shape and dimensions ($24.7\times24.7\times220\,\mathrm{mm}^3$) grouped together into units of 36 (called supercrystals). A total of 268 identical supercrystals is used to cover each endcap with a further 64 sectioned supercrystals used to complete the inner and outer perimeter. Each endcap contains 7324 crystals, corresponding to a volume of $1.52\,\mathrm{m}^3$ (for a total weight of 12.6 t). Both endcaps are identical and each endcap detector is constructed using Dee-shaped sections as seen in Figure 4.6.

Figure 4.7 shows the total thickness (in radiation lengths) of the ECAL as a function of pseudorapidity. The endcap part also includes the preshower detector (see the specific subsection). Because of the higher radiation levels in the forward region, the endcaps must tolerate very large doses and neutron fluences, hence all materials used in this region had to be chosen with higher radiation hardness. Also the electronics is different between barrel and endcap for the same reason: avalanche photodiodes (APDs) are employed in the barrel and vacuum phototriodes (VPTs) in the endcaps.

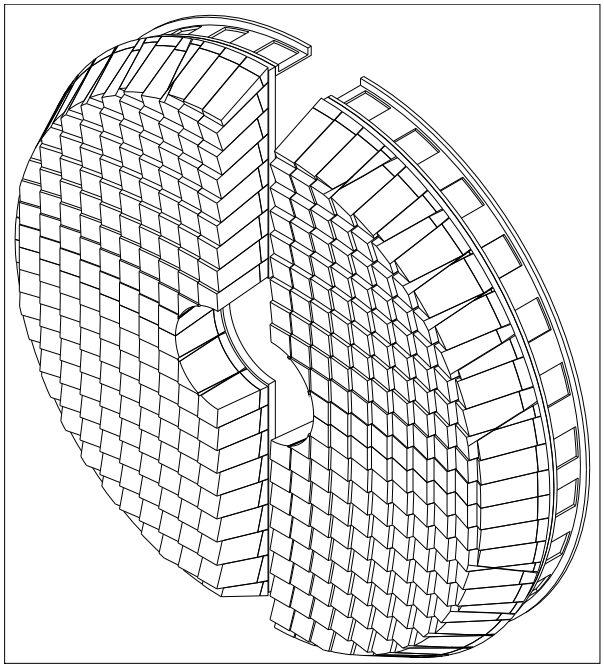

Figure 4.6: A single endcap with Dees apart.

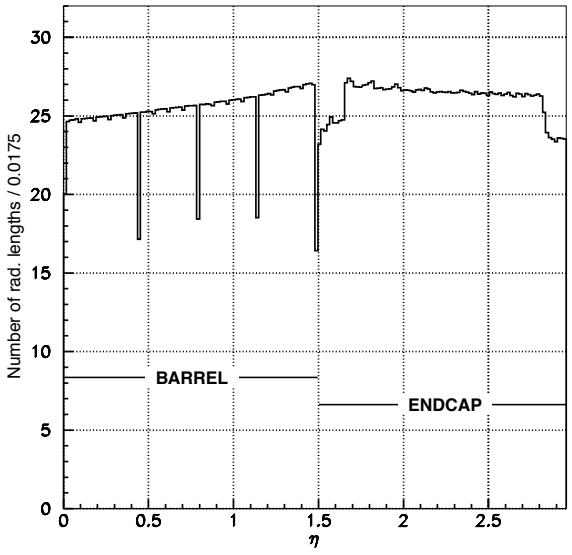

Figure 4.7: Total thickness in X_0 of the ECAL as a function of pseudorapidity, averaged over ϕ. In the endcap region, the preshower contribution is also considered [74].

4.4.3 Preshower detector

The endcap preshower covers a pseudorapidity range from $|\eta| = 1.65$ to 2.61. Its main function is to provide π^0-γ separation. The preshower detector, placed in front of the crystals, contains two lead converters of a total thickness of $2\,X_0$ and $1\,X_0$ respectively, followed by detector planes of silicon strips with a pitch of $< 2\,\text{mm}$. The impact position of the electromagnetic shower is determined by the barycenter of the deposited energy. The accuracy is typically $300\,\mu\text{m}$ at 50 GeV. In order to correct for the energy deposited in the lead converter, the energy measured in the silicon is used to apply corrections to the energy measurement in the crystal. The fraction of energy deposited in the preshower (typically 5% at 20 GeV) decreases with increasing incident energy. Figure 4.8 shows the

layout of the preshower.

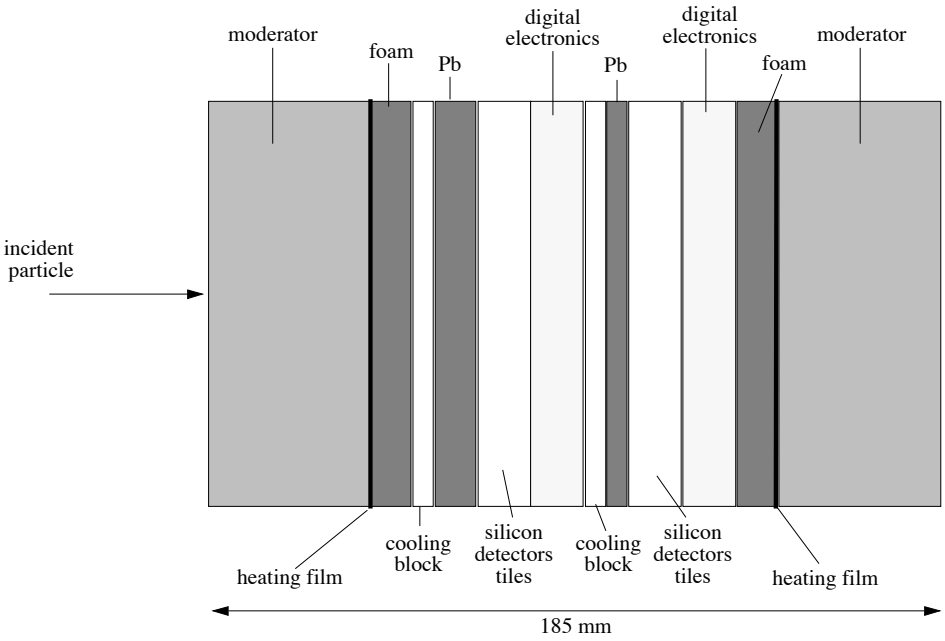

Figure 4.8: Schematic view of a section of the endcap preshower.

To maintain its performance during the lifetime of the experiment, heating films and insulating foam glued on the moderators guarantee that the external surfaces are kept at the ambient temperature of the neighboring detectors.

4.4.4 Lead tungstate crystals

The lead tungstate crystals ($PbWO_4$) were chosen for operation at LHC due to their peculiar characteristics [75]. In particular, the high density ($8.3\,g/cm^3$), short radiation length (0.89 cm) and small Molière radius[5] (2.2 cm) results in a fine granularity and a compact calorimeter (the compactness is a very important parameter due to the fact that the calorimeters are placed inside the magnetic field and that it's technically challenging to achieve a magnetic field, both 3.8 T intense and constant inside the all internal volume). The scintillation decay time is of the same order of magnitude of the LHC bunch crossing time: about 80% of the light is emitted in 25 ns, while the light output is relatively low: about 100 scintillating photons per MeV. The electronics collect about 4.5 photoelectrons per MeV in both the avalanche photodiodes (APDs) and the vacuum phototriodes (VPTs), where the higher APD quantum efficiency[6] (75% vs 20%) is balanced by their smaller surface coverage on the back face of the crystal. The crystals emit blue-green scintillation light with a broad maximum at 420 nm [75].

4.4.5 Energy resolution

The ECAL energy resolution has been parameterized as:

$$\frac{\sigma(E)}{E} = \frac{a_{stoc}}{\sqrt{E}} \oplus \frac{b_n}{E} \oplus c \quad \text{(E in GeV)} \tag{4.6}$$

[5]The Molière radius (R_M) defines the lateral development of an electromagnetic shower: a cylinder of infinite length and radius R_M contains $\approx 90\%$ of the shower's energy.

[6]The device efficiency to convert a photon in photoelectrons.

where a_{stoc} is the stochastic term, b_n the noise, and c the constant term. Figure 4.9 summarizes the different contributions expected for the energy resolution. Terms representing the degradation of the energy resolution at extremely high energies have not been included. The stochastic term includes fluctuations in the shower containment as well as a contribution from photostatistics. The noise term contains the contributions from electronic noise and pile-up energy; the former is quite important at low energy, the latter is negligible at low luminosity. The curve labeled *intrinsic* in Figure 4.9 includes the shower containment and a constant term of 0.55%. The constant term must, in fact, be kept down to this level in order to profit from the excellent stochastic term of $PbWO_4$ in the energy range relevant for the searches for new physics. To achieve this goal, in situ calibration/monitoring using isolated high p_T electrons is performed. The parameters, measured in an electron test beam, for incident electrons of different energies from 20 to 250 GeV, with a 3x3 crystal configuration, considering E in GeV, correspond to a_{stoc}=0.028 $\sqrt{(GeV)}$, b_n=0.12 GeV and c =0.003.

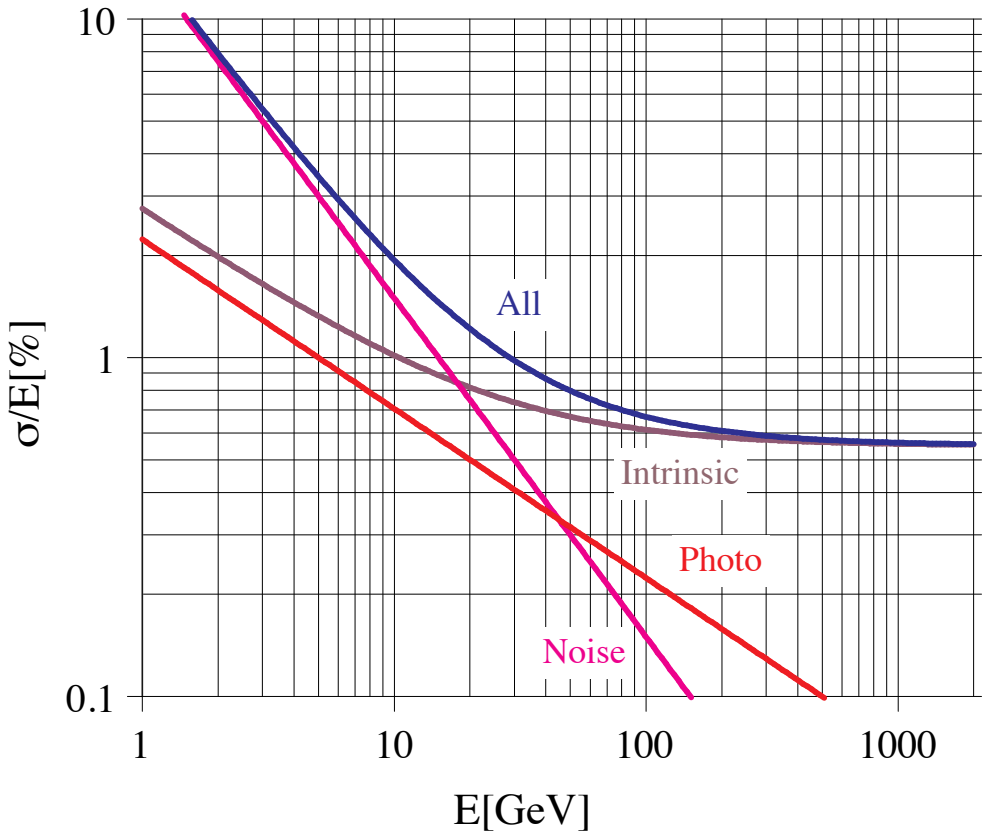

Figure 4.9: Different contributions to the energy resolution of the $PbWO_4$ calorimeter. The noise term contains the contributions from electronic noise and pile-up energy. The curve labelled "photo" describes the contribution from photostatistics, while the curve labelled "intrinsic" includes the shower containment and a costant term of 0.55%. The parametrization given in Equation (4.6) re-arrange the 3 contributions in a slightly different way, including the shower containment and the photostatistics contributions in the stochastic term a_{stoc} [75].

A detailed study of the energy resolution behaviour both at low and high energy region is crucial in order to describe a reliable signal model in the searches for new physics beyond the standard model. These studies will be described later in this thesis, in Chapters 6 and 7, where the focus will be the importance in the context of the analyses.

4.5 Hadronic calorimeter

The design of the hadronic calorimeter (HCAL) [76] is strongly influenced by the choice of the magnet parameters since most of the CMS calorimetry is located inside the magnet coil (see Figure 4.11). An important requirement of HCAL is to minimize the non-Gaussian tails in the energy resolution and to provide good containment and hermeticity. Hence, the HCAL design maximizes material inside the magnet coil in terms of interaction lengths[7]. This is complemented by an additional layer of scintillators, referred to as the hadron outer (HO) detector, which is installed outside the magnet. Brass has been chosen as absorber material as it has a reasonably short interaction length (5.15 interaction length in 79 cm), it is relatively easy to mold and it is non-magnetic. Maximizing the amount of absorber before the magnet requires minimizing the amount of space devoted to the active medium. The tile/fiber technology has hence been chosen: it consists of plastic scintillator tiles read out with embedded wavelength-shifting (WLS) fibers. The photodetection readout is based on multi-channel hybrid photodiodes (HPDs). The absorber structure is assembled by bolting together precisely segmented and overlapping brass plates so as to leave space to insert the scintillator plates, which have a thickness of 3.7 mm. The overall assembly enables the HCAL to be built with essentially no uninstrumented cracks or dead areas in ϕ. The gap between the barrel and the endcap of HCAL, through which the services of the ECAL and the inner tracker pass, is inclined at 53° and points away from the center of the detector. Since the identification of forward jets is very important for the rejection of many backgrounds, the barrel and the endcap parts, which cover up to $|\eta| < 3.0$, are complemented by a very forward calorimeter (HF), placed at ± 11.2 m from the interaction point, which extends the pseudorapidity range of the calorimetry up to $|\eta| < 5.2$. As the particle flux in this very forward region is extremely high, a radiation hard technology, using Cherenkov light in quartz fibers was chosen, using steel as an absorber. The HF detector is also used as a real-time monitor for the luminosity on a bunch-by-bunch basis.

The HCAL baseline single-particle energy resolution is:

$$\frac{\sigma_E}{E} = \frac{65\%}{\sqrt{E}} \oplus 5\% \tag{4.7}$$

in the barrel,

$$\frac{\sigma_E}{E} = \frac{83\%}{\sqrt{E}} \oplus 5\% \tag{4.8}$$

in the endcaps, and

$$\frac{\sigma_E}{E} = \frac{100\%}{\sqrt{E}} \oplus 5\% \tag{4.9}$$

in the forward calorimeter (where E is expressed in GeV).

4.6 Magnet

The design of the CMS magnet, in particular its bending power, was driven by the required performance of the muon system. The momentum determination of charged particles is performed by measuring the particles trajectories inside the solenoid. In particular, the momentum resolution is given by the formula:

$$\frac{\Delta p}{p} = \Delta s \frac{8p}{0.3BR^2} \tag{4.10}$$

[7]The interaction length is the equivalent of the radiation length (X_0) for hadronic showers, modulo the fact that it's generally much bigger, with a factor between 1 to 30 depending on atomic mass number Z of the specific material considered.

where $p = \gamma mv$ is the particle momentum, B is the magnetic induction, s is the sagitta and R is the tracker radius. To achieve the goal of the unambiguous determination of the curvature (hence the electric charge) for muons with momentum of 1 TeV, the requirement on the momentum resolution is: $\Delta p/p \sim 10\%$ at $p = 1$ TeV. Therefore strong field and large radius are an efficient approach to reach optimal momentum resolution: CMS preferred indeed a higher field within a relatively compact space. The solenoid of the CMS detector produces uniform field in the axial direction, while the flux return is assured by an external iron yoke with three layers, in between which the muon system is installed. The superconducting magnet has a length of 12.5 m and a diameter of the cold bore of 6.3 m. It is made from a 4-layer winding of NbTi (Niobium-Titanium) cable reinforced with aluminium, weighting a total of 220 t, and kept at a temperature of 4.5 K with liquid helium. It was designed to produce a field of 4 T but operate at a lower field of 3.8 T. The magnetic field is generated by a 18 kA current circulation in the cables. The magnet system stores an energy of 2.5 GJ.

The parameters of the CMS superconducting solenoid are summarized in Table 4.1.

Parameter	Value
Field	3.8 T
Inner bore	5.9 m
Length	12.9 m
Number of turns	2168
Current	19.5 kA
Stored energy	2.7 GJ

Table 4.1: Parameters of the CMS superconducting solenoid.

4.7 Muon system

The muon system is the outermost of the CMS subdetectors. Its main goals are the identification of muons, thanks to their high penetrating power[8], and a precise measurement of their momentum, with the help of the information coming from the tracker. The muon system also works as trigger for events which involve muons and it provides a precise time measurement of the bunch crossing. The CMS muon system [77] relies on three kinds of gaseous detectors: drift tubes (DT), cathode strip chambers (CSC) and resistive plate chambers (RPC). The DT and the CSC provide an excellent spatial resolution for the measurement of charged particle momentum; the RPC are used for trigger issues because of the very good timing. The active parts of the muon system are hosted into stations which are interleaved by the iron layers of the return yoke of the magnet. The longitudinal view of a quarter of the muon system is given in Figure 4.10. The barrel extends up to $|\eta| < 1.4$, the endcaps up to $|\eta| < 2.4$.

Before moving to report in short details the trigger and data acquisition systems, a schematic view of a transverse slice of the CMS barrel is provided in Figure 4.11, as a visual summary of what has been presented in the previous sections.

4.8 Trigger and data acquisition

The bunch crossing frequency at the CMS interaction point is 40 MHz (bunch spacing of 25 ns) while technical difficulties in handling, storing and processing extremely large

[8]Muons are, in fact, minimum ionizing particles (MIP).

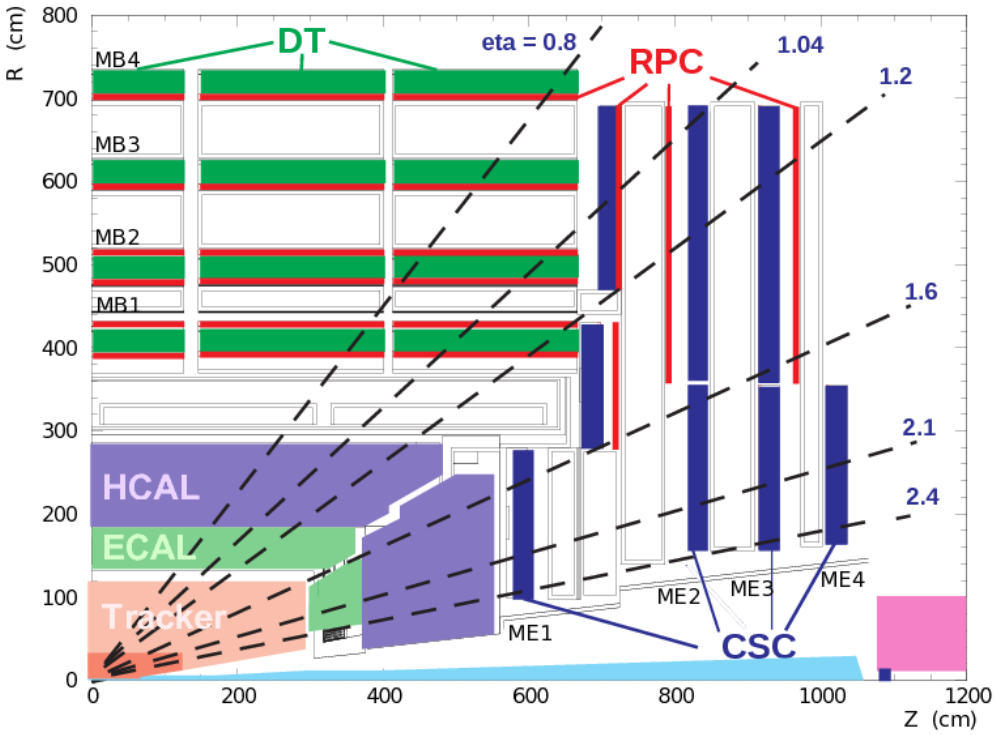

Figure 4.10: Longitudinal view of one quarter of the CMS muon system.

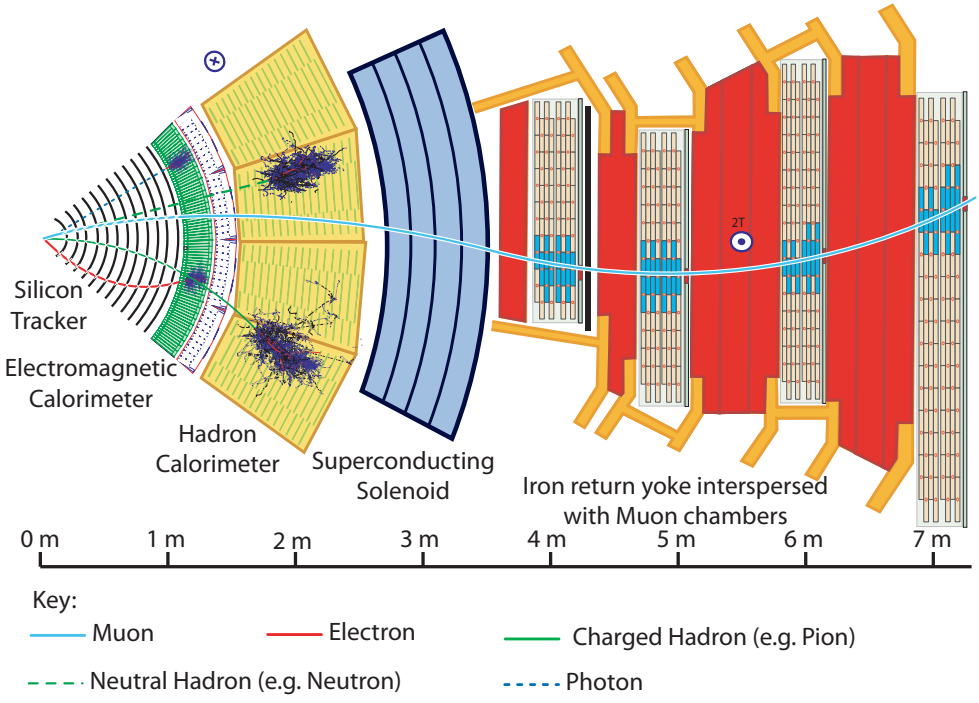

Figure 4.11: Schematic view of a transverse slice of the central part of the CMS detector.

amounts of data impose a limit of about 1 kHz on the rate of events that can be written to permanent storage, as the average event size is about 1 MB. At the LHC nominal luminosity the total event rate for inelastic interactions is expected to be of the order of 10^9 Hz while the rate of interesting events is rather small in comparison (see Figure 4.12).

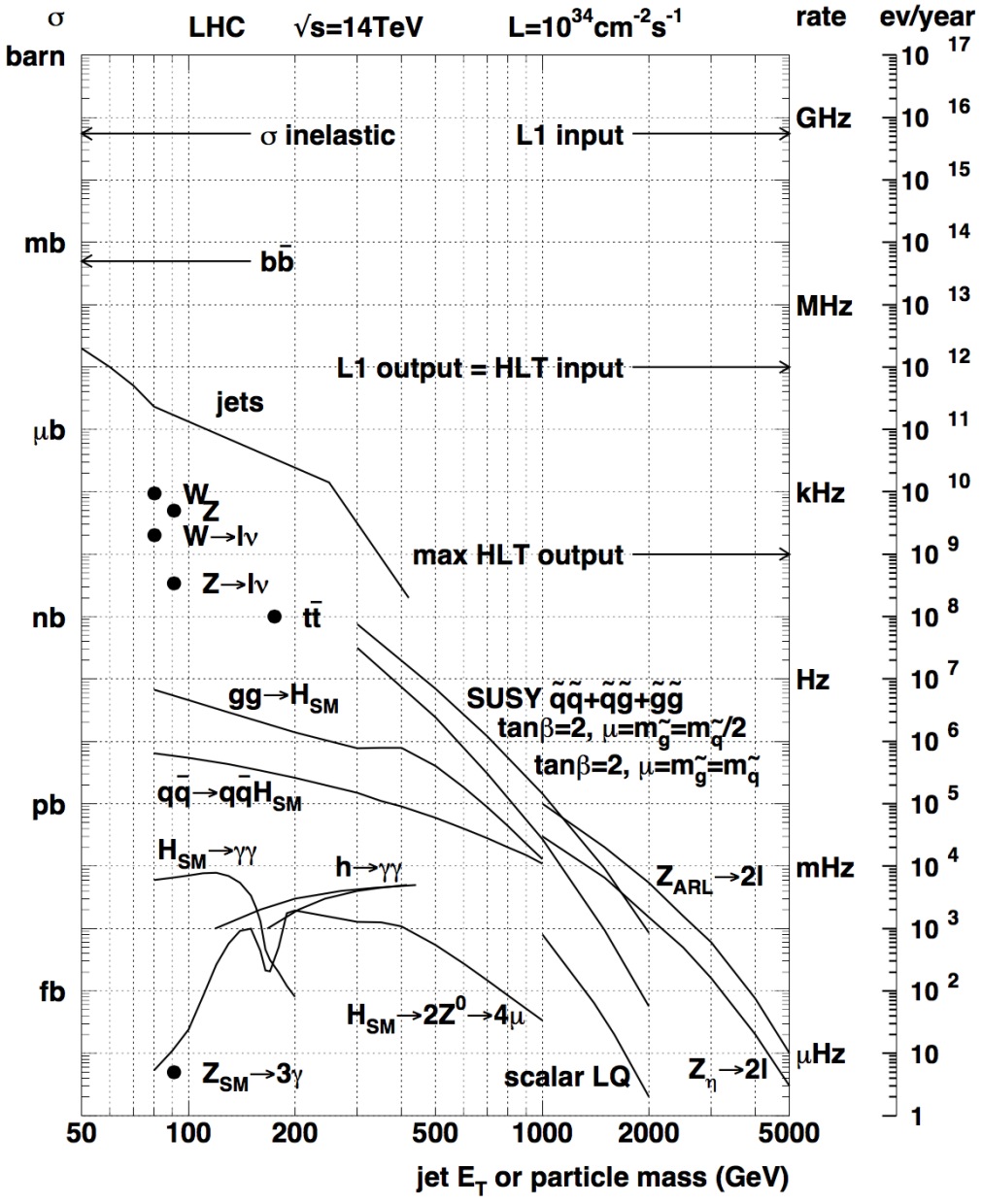

Figure 4.12: Event cross-sections and rates of selected processes for the LHC design luminosity of 10^{34} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$ as a function of the mass of produced objects [72].

A sophisticated trigger system is then necessary to select events of interest, within a limited amount of time available for the selection since the bunch crossing time is 25 ns. This interval of time would not be enough to read out all raw data from the detectors, and for this reason CMS uses a multi-level trigger design, where each step of the selection uses only part of the available data. In this way higher trigger levels have to process fewer events and have more time available: they can go into finer detail and use more refined algorithms. The two steps of the CMS selection chain are: the first level (L1) trigger, built from custom hardware, which reduces the rate to a maximum of 100 kHz, and the high level trigger (HLT), running the CMS reconstruction software on a processor farm, which performs higher level reconstruction and reduces the rate of events selected by the

L1 trigger to about 1 kHz before the events are stored on disk.

4.8.1 L1 trigger

The L1 system is built from custom designed, programmable electronics, and is located underground, both in the service and the experimental caverns. Within a time budget of 3.2 μs, it has to decide if an event is kept or discarded, and transfer this decision back to the subdetectors, which keep the high resolution data in memory in the meantime. The L1 is divided in a muon trigger and a calorimeter trigger, which classify and rank interesting event candidates, reconstructed from low resolution data read out from the subdetectors. The rank of a candidate is determined by energy or momentum and data quality. The calorimeter and muon triggers do not perform any selection themselves. They identify "trigger objects" of different types: e/γ (isolated and not), jets and muons. Based on the input from the muon trigger and the calorimeter trigger, the global trigger calculates the final trigger decision. Up to 511 trigger algorithms[9] can be executed in parallel to generate a decision. The simplest triggers are in general based on the presence of one object with an E_T or p_T above a predefined threshold (single-object triggers) or based on the presence of two objects of the same type (di-object triggers) with either symmetric or asymmetric thresholds. Other requirements are those for multiple objects of the same or different types ("mixed" and multiple-object triggers). The high resolution data from the inner tracker are not used to generate the L1 decision, which means that there is no information about the vertices and no distinction between electrons and photons available at this level.

4.8.2 High level trigger

The goal of the HLT is to reduce the event rate from the maximum L1 output ($\approx$ 100 kHz) to 600 Hz which is the maximum rate for mass storage. Once the L1 trigger has accepted an event, the data of this event are transferred from the buffer memory to the surface, where they are reconstructed in the HLT. The HLT is a special part of the CMS software and runs on a farm of several thousand processors. Each processor works on the reconstruction of one event at a time, to get to a trigger decision within 100 ms on average. Since the time budget for one event is much larger than at the L1 trigger, more complicated algorithms, including tracking, can be executed at the HLT level. Once an event is accepted, it is stored on disk and fully reconstructed offline at a later time.

The use of standard software techniques and languages makes it possible to benefit from the continuous improvements in the reconstruction software. In particular the algorithms used in the HLT, which access data with full resolution and granularity from any part of the detector, is identical to those used in the off-line reconstruction. However, in order to discard uninteresting events as soon as possible, the selection is organized in a sequence of logical steps: the Level-2 and Level-3. The Level-2 uses the full information from calorimeters and muon detectors (but not tracker) and reduces the event rate by roughly one order of magnitude. The data from the silicon tracker represent almost 80% of the event size and require complex and time consuming algorithms for the reconstruction. For this reason this information is used only during the Level-3 selection.

The HLT consists of approximately 400 trigger paths[10], which, starting from the seed of the L1 trigger, look for different objects and signatures in the event. One trigger path

[9] The maximum number of L1 algorithms evolved in time. At the beginning of the RunII data-taking this number was 128, the same as in the RunI period, while during the 2016 data-taking, it ranged from 244 up to 305. Further improvements, deployed at the beginning of 2017, allowed to reach 511 L1 bits.

[10] If an event is not accepted by a path, it can still be accepted by a different path, or even be accepted by several paths.

is built from reconstruction modules and filter modules. After some parts of the data are reconstructed, a filter module decides either the reconstructed objects pass the thresholds and the next step in reconstruction is started, or the event is not accepted by the path. In the latter case, the execution of the path is stopped and the following reconstruction steps and filter steps are not performed to save computation time.

If the acceptance rate is too high (for example in case of a trigger path with very low thresholds), the trigger path can be prescaled to lower the rate. A prescale value of 10, for example, means that the path is executed only for 1 over 10 events (randomly chosen to avoid biases) that were accepted by the L1 trigger and, consequently, the trigger rate for that path is 10 times smaller. The prescale value for one trigger path has several predefined levels, depending on the instantaneous luminosity of the LHC machine. During an LHC fill, the instantaneous luminosity decreases, and the prescale values can be changed during a CMS run to keep the global trigger rate at an optimal level.

The complete description of the HLT system is beyond the purpose of this thesis and won't then be treated in more details.

Chapter 5

Event reconstruction

The aim of this chapter is to describe the way the information coming from the CMS subdetectors is combined together to reconstruct physical objects. The electron and photon reconstruction is described in Section 5.1, the muon reconstruction in Section 5.2, the tau reconstruction in Section 5.4, the jet and b-tagged jet reconstruction in Section 5.5 and finally the missing transverse energy reconstruction in Section 5.6. Given the importance of the electron and photon reconstruction in this thesis, the specific section is treated in more details.

5.1 Electron and photon reconstruction

The strategies used in CMS to reconstruct electrons and photons have large similarities due to the similarities of the objects' signatures in the calorimeters (almost the entire energy is expected to be deposited in the ECAL in the form of an electromagnetic shower with a small lateral extension; virtually no deposit in HCAL is expected). The presence of a track matching or not the energy deposit in ECAL allows one to disentangle between the two objects. The fact that a photon can convert into electron pairs in the tracker and that electrons can radiate bremsstrahlung photons, which in turn can convert in electron pairs, implies however that in both cases the energy of the incident particle can be distributed between several energy deposits largely spread in the ϕ direction, in reason of the bending of the electrons due to the magnetic field of CMS. In some cases, this feature also makes the electron/photon distinction more difficult and various algorithms have been developped to identify the primary particle.

This section is organized in the following way: first the common aspects of electrons and photons reconstruction are presented then the peculiarities of the individual reconstruction methods are presented.

5.1.1 Energy measurement in ECAL

The electrical signal coming from the photodetectors (APDs and VPTs) is amplified and shaped by a multi-gain preamplifier (MGPA) [78], which uses three parallel amplification stages. The output is digitized by a 12 bit Analog-to-Digital Converter (ADC) running at 40 MHz, which records ten consecutive samples and selects the gain with the highest non-saturated signal. This provides a dynamic range of about 5×10^4 from the least significant bit of 35 MeV to a saturation energy of about 1.7 (3.5) TeV in EB (EE). The data consist of a series of consecutive digitizations, corresponding to a sequence of samplings of the signal at 40 MHz. A set of 10 consecutive samplings (taken with a time distance of 25 ns) is readout and used to reconstruct the signal amplitude. Figure 5.1 shows an example of

the time sampling for a signal pulse as a function of the time difference $(T - T_{max})$, where T and T_{max} indicate the time of the generic ADC sample and the time corresponding to the maximum of the pulse shape respectively. The signal pulse is expected to start from the fourth sample and the baseline pedestal value is estimated from the first three samples.

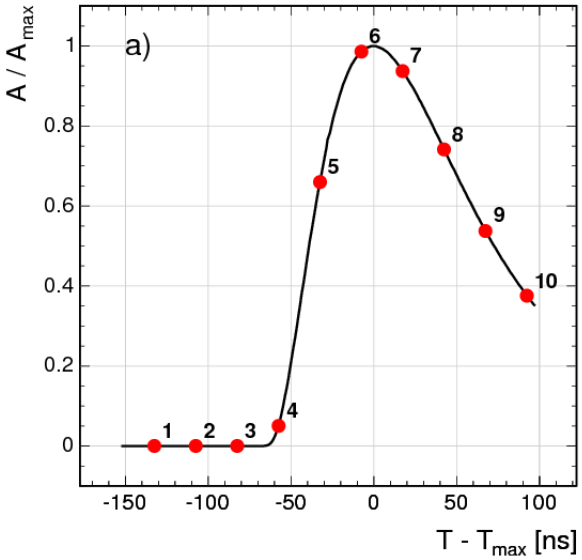

Figure 5.1: Pulse shape measured in the ECAL as a function of $(T - T_{max})$ [79].

5.1.1.1 Signal pulse reconstruction

During LHC RunI a digital filtering algorithm was used, where the signal amplitude $\mathcal{A}$ was estimated as the linear combination[1] of the $N = 10$ samples S_i:

$$\mathcal{A} = \sum_{i=1}^{N} w_i \times S_i \tag{5.1}$$

where the weights w_i were computed in order to minimize the variance of $\mathcal{A}$. A detailed description of this method (known as "weight method") can be found in [80].

For the LHC RunII a new ECAL pulse reconstruction algorithm was developed, due to the requirements of the new data taking conditions, in particular the unprecedented pile-up conditions (see Section 3.2). Several methods have been investigated to mitigate the effect of pile-up, maintaining optimal noise filtering [81] and a template fit with multiple components (known as "multifit method"[79]) was chosen for the RunII data taking.

The multifit algorithm estimates the signal amplitude and up to 9 out of time amplitudes by minimization of the χ^2, given by:

$$\chi^2 = \sum_{i=1}^{N} \frac{\left((\sum_{j=1}^{M} \mathcal{A}_j p_j)_i - S_i \right)^2}{\sigma_{S_i}^2} \tag{5.2}$$

where i runs over the $N = 10$ samples and j runs over the $M \leq 10$ pulse templates p_j. The pulse templates for each crystal were measured from dedicated low pile-up pp collisions

[1] The simplest method to take the sampling on the maximum as the amplitude measurement was exploited, during the RunI data-taking, for energies close to the saturation point of the three available amplifiers (see Section 5.1.2).

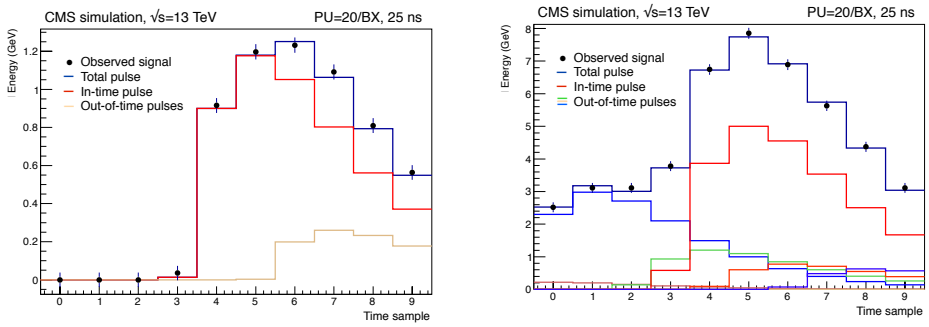

Figure 5.2: Two examples of fitted pulses for simulated events with 20 average pile-up interactions and 25 ns bunch spacing, for a signal in the barrel (left) and in the endcaps (right). Dots represent the 10 digitized samples, the red distributions (other light colors) represent the fitted in-time (out-of time) pulses with positive amplitude. The dark blue histograms represent the sum of all the fitted contributions [79].

data recorded by CMS. The total electronic noise is denoted as σ_{S_i} and measured from dedicated pedestal runs, which measured the noise in all three gains of the MGPA in the absence of signal pulses.

The technique of Non-Negative-Least-Squares [82] is used to perform the χ^2 minimization of Equation (5.2) with the constraint that the fitted amplitudes must be all positive. Examples of two fitted pulses in the barrel and in the endcaps are shown in Figure 5.2. The fit is performed in $\approx$ 10 ms/event, for events with an average pile-up of 40 and for 25 ns bunch spacing.

The residual contribution of out-of-time pile-up to the energy resolution has been estimated for the multifit algorithm using simulated samples of unconverted photons. It is observed to be highly suppressed for signal pulses in both the barrel and endcaps. The improvement in energy resolution with respect the RunI reconstruction algorithm for collisions with 25 ns bunch spacing is substantial especially for low p_T photons and electrons, given the larger contribution of pile-up to the total energy estimate, and is still significant for those at high p_T ($p_T > 50$ GeV).

With the amount of statistics collected in the first half of 2016 ($\sim$13 fb^{-1}) an important issue of the multifit algorithm, known as "slew rate issue", was exposed. It resulted in affecting the energy scale of electrons with high transverse momentum up to a maximal effect quantified to $\sim$4%. Given the importance of the high-p_T region in this thesis, the issue will be described in the next section, as well as the strategy that has been used in order to solve it.

5.1.2 Discussion of the ECAL slew rate issue

When converting the analog signals coming from the ECAL crystals to digital signals, three gains are available with the different gain factors of 12, 6 and 1. When the ADC (analog-to-digital converter) count reaches its maximum with a given gain, a lower gain factor is considered (gain switch). At the lowest gain factor (gain 1), the ADC saturates at an energy of $\sim$1.7 TeV in the barrel and $\sim$ 3.2 TeV in the endcaps.
The ECAL electronics showed a non-linear behaviour at the end of each gain range which results in a distortion of the pulse shape in the case where the ADC switches gains during the pulse. This happened for electrons and photons having an energy close to the gain thresholds, for which the signal is too steep for the electronics to be able to follow it (this

is known as "slew rate") and, in particular, the 4-th sample (see Figure 5.1) of the pulse shape was read using a higher gain factor than the correct one. The 5-th sample (see again Figure 5.1) is instead always read with the proper gain factor, given that the delay of 25 ns from the 4-th to the 5-th sample is sufficient in order to not be affected by the slew rate of the electronics. Hence, the events characterized by the presence of a gain switch from the 4-th sample to the 5-th sample of the pulse shape need a special care. The multifit algorithm (see previous section) did not take this effect correctly into account and, for those events only, it turned out to reconstruct a lower amplitude for the pulse shape, due to an over-estimation of the out-of-time pile-up components, with witch the fitting algorithm tried to compensate the distortion of the pulse shape. This issue, known as "slew rate issue" or "gain switch issue", was not simulated in the Monte Carlo samples, therefore only electrons/photons in data (and in the barrel) encountering a gain switch threshold are effected. The events with at least one electron/photon encountering gain 1 switches ($E_{\mathrm{T}} >\sim 300$ GeV) are of course the most problematic ones given the final states targeted in the analyses described in this thesis, but also the energy of electrons/photons affected by gain 6 switches ($E_{\mathrm{T}} >\sim 150$ GeV) needs to be treated.

Once the problem has been understood to be due to the multifit algorithm, it was possible to overcome it simply by reconstructing the energy of the crystals encountering a gain switch as it was done with the RunI "weight" method, i.e. by simply taking the maximum sample as the estimate of the amplitude. Figure 5.3 shows the Z peak for barrel-barrel electron pairs that pass the HEEP selection (see Section 6.3) with at least one electron having a switch to gain 6 (left) or gain 1 (right) in the 5×5 crystal matrix centred on the seed crystal, with the Z mass calculated with the electron energy before and after the software fix. As can be seen from the figure, in the case without the fix and expecially in the "gain 1 scenario", the Z peak is shifted from 91 GeV and a secondary tail where one of the electrons is grossly misreconstructed appears. With the fix applied, the Z peak position is recovered to 91 GeV, the resolution improves to expected values and the tail at 75 GeV disappears, demonstrating the effectiveness of the fix strategy.

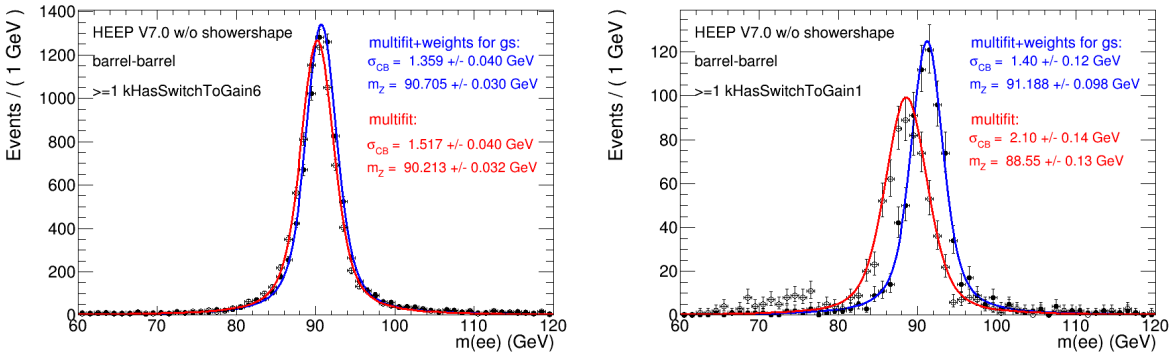

Figure 5.3: The Z peak region for barrel-barrel dielectron events where at least one electron has a gain 6 switch (left) or gain 1 switch (right). The electrons are required to pass the HEEP selection. When using the standard multifit algorithm for gain switched crystals the Z peak is shifted and the mass resolution is degraded. When using the modified multifit algorithm (standard multifit plus weight method in case of gain switches) the Z peak is restored to 91 GeV and the mass resolution is significantly improved.

In order to further assess the stability of the software fix as a function of E_{T}, Figure 5.4 shows the Z peak position for barrel-barrel pairs passing the HEEP selection as a function of the leading electron E_{T} with (blue points) and without (red points) the fix. No gain switch requirement is applied. It can be seen that there is a remaining drift effect in

the mass scale as a function of the leading electron E_{T} even when the fix is applied. On the other hand, the effect is only ≈ 0.5 GeV from 35 to 800 GeV, hence is covered by the assigned 1% (2%) energy scale systematic uncertainty considered by the diphoton (dielectron) analysis described in this thesis.

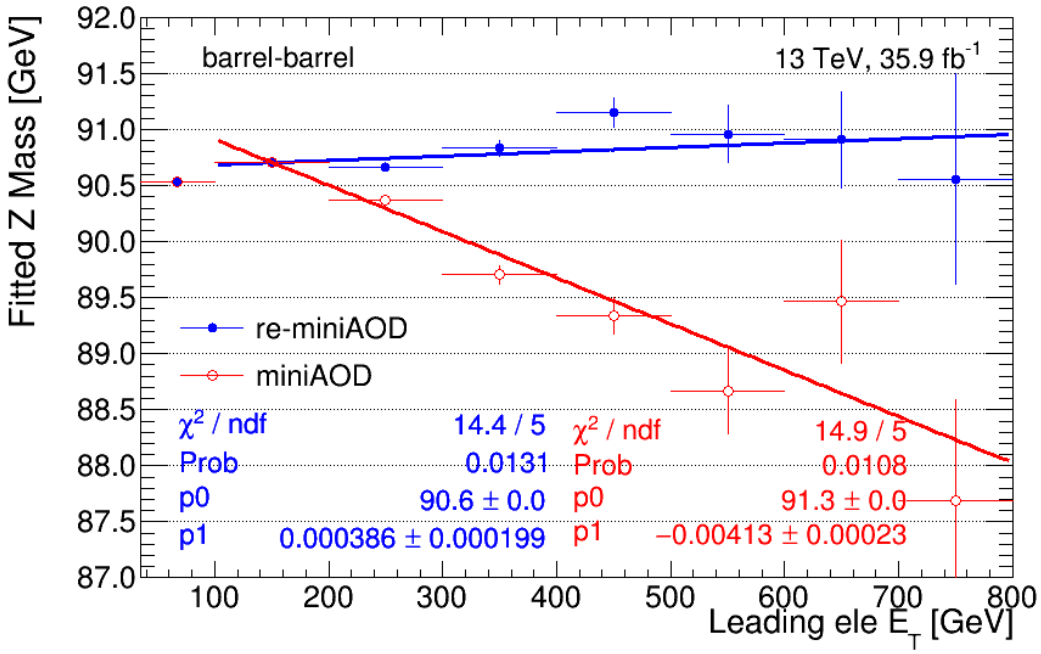

Figure 5.4: The Z peak position for barrel-barrel dielectron events as a function of leading electron E_{T} using the energy before and after the gain switch fix. The peak position after the fix (blue points and line) is clearly much more stable as a function of the leading electron E_{T} with respect to the scenario where the fix is not applied (red points and line).

5.1.3 Clustering algorithms

Electrons and photons deposit their energy over several ECAL crystals. In addition, the presence of material in front of ECAL causes conversions of photons and bremsstrahlung from electrons, and the radiated energy is spread along ϕ by the strong magnetic field (see Section 4.6). Clustering algorithms are used to collect the energy deposits in ECAL, including the contributions from the radiated energy. The clustering algorithm takes into account the correlation in the $\eta - \phi$ plane of the radiated energy due to the combined effect of the magnetic field and material budget distribution.

In fact, an electron/photon interacting in the ECAL, will deposit in average 94% (97%) of its energy in a matrix of 3×3 (5×5) crystals around the crystal with the largest energy deposit, which is denoted as the seed crystal. The material in front of the ECAL has a thickness between 0.4 and 2 radiation lengths. This means that at $\eta \sim 0$, on average 33% of the electron energy is radiated before reaching the ECAL[2] and in the direction of the largest material budget at $|\eta| \sim 1.4$ the loss is on average 86%. The energy lost via radiative processes has to be added to the energy deposited by the electron in the ECAL. This is done by the construction of a supercluster (SC) in the two ways defined below (two different algorithms are employed for barrel and endcaps). Since the trajectory

[2]For the photon case, the tipical "conversion length" is related to the radiation length for an electron by a multiplicative factor of $\frac{9}{7}$.

of the electron is curved in r$-\phi$ plane, the energy deposit in the ECAL barrel coming from radiated photons is mainly spread in the ϕ direction. In order to build the SC, dominoes of 5×1 crystals[3] in $\eta - \phi$ are produced, extending for 17 crystals (0.3 rad) in the ϕ direction around the seed crystal, which must have a transverse energy of at least 1 GeV. If the energy of the domino exceeds a threshold of 0.1 GeV it is grouped with nearby dominoes to form a cluster. The clusters themselves must have a seed domino that exceeds an energy of 0.35 GeV to be added to the SC. In the endcaps, the energy deposit from bremsstrahlung follows a trajectory in η and ϕ. The energy is collected in a 5×5 matrix around a seed crystal that has to exceed a transverse energy threshold of 0.18 GeV. Around the seed crystal the energy is collected in 5×5 matrices along roads in η and ϕ. These roads have a range of ± 0.07 in η and ± 0.3 rad in ϕ around the seed crystal, and the transverse energy of the 5×5 cluster has to exceed 1 GeV if it is to be added to the SC. The energy collected in the pre-shower detector situated in front of the 5×5 matrix is also added to the SC energy. The position of the supercluster is computed as the energy-weighted mean (barycenter) of the cluster positions, whereas its (raw) energy[4] E_{raw} is simply taken as the sum of the energy of all its constituent clusters.

The electron or photon energy is then estimated as:

$$E_{e,\gamma} = F_{e,\gamma} \left[G \times \sum_i (IC_i \times LC_i(t) \times \mathcal{A}_i) + E_{ES} \right] = F_{e,\gamma} E_{\text{raw}} \tag{5.3}$$

where the sum is performed over all the clustered crystals. The amplitude measured in the i-th crystal is labeled by $\mathcal{A}_i$, while $LC_i(t)$ is a time dependent correction that accounts for time variation of the channel response due to changes in crystal transparency (see Section 5.1.4.1). The IC_i parameter is a relative calibration constant that takes into account differences in the crystal light yields and photodetector response (see Section 5.1.4.2) and G is a scale factor converting the digital scale into GeV. For clusters in the endcap region the corresponding energy in the preshower (E_{ES}) is added. Finally $F_{e,\gamma}$ is a particle-dependent correction, also denoted as cluster correction, applied to the SC energy. It accounts for biases in the energy reconstruction related to the geometry of the detector, the upstream material, and the clustering of energy emitted by bremsstrahlung or photon conversions. The pile-up dependence of the energy scale and its correlation with the cluster shape and position is also taken into account by a multivariate algorithm trained on simulated samples of electrons and photons (giving raise to what is generally called "electron-tuned regression" and "photon-tuned regression").

In the following, the methods used to derive the parameters in Equation (5.3) are described. They are based on in-situ measurements of copious physics processes producing electrons and photons in the final state: $\pi^0, \eta \to \gamma\gamma$, W→ e$\nu$, Z→ee.

Also, a definition that will be useful for the future is the ratio of the energy in a fixed array of 3×3 crystals around the seed crystal over the SC energy without the $F_{e,\gamma}$ correction of Equation (5.3). This variable is called R_9 and defined as $R_9 = E_{3\times3}/E_{\text{raw}}$: the R_9 variable offers a convenient way to identify electrons with little radiation in the tracker or unconverted photons, for which a better energy resolution is expected.

5.1.4 Energy calibrations

The ECAL energy calibration workflow [83] exploits different methods and physics channels to measure the terms entering in the energy reconstructions expressed by Equation

[3] In order to be considered in the supercluster building procedure, the crystals must have a signal that is 2σ above the electronic noise level of 80 MeV in the barrel and up to 300 MeV in the endcaps.

[4] Also denoted as E_{SC} in literature.

(5.3). The first step consists of correcting for the single channel response time variation. Once the response of the channels is stable in time, collision events are used to derive the intercalibration corrections. As final step of the calibration procedure, the absolute energy scale G is tuned.

5.1.4.1 Corrections for time-dependent response changes

Ionizing radiation induce loss of transparency in the ECAL crystals reducing their measured response to the deposited energy. This loss in transparency depends on the dose rate, which varies with η, and a partial recovery of transparency is observed in the absence of radiation. The changes in transparency are measured and corrected using a dedicated "laser monitoring" system [84] which injects laser light ($\lambda \sim 440$ nm, close to the peak of the scintillation light spectrum for $PbWO_4$) into each crystal. One measurement point per crystal is typically recorded every 30 minutes. The change in transparency (R/R_0) does not directly measure the change in the amount of the scintillation light (S/S_0) since the two have different spectra and the optical photons travel different paths to reach the photodetectors, but they can be related by a power law:

$$\frac{S}{S_0} = \left(\frac{R}{R_0}\right)^{\alpha} \tag{5.4}$$

where $\alpha \sim 1.5$. The history of the relative response variation measured by the laser monitoring system in the period 2011-2016 is shown in Figure 5.5.

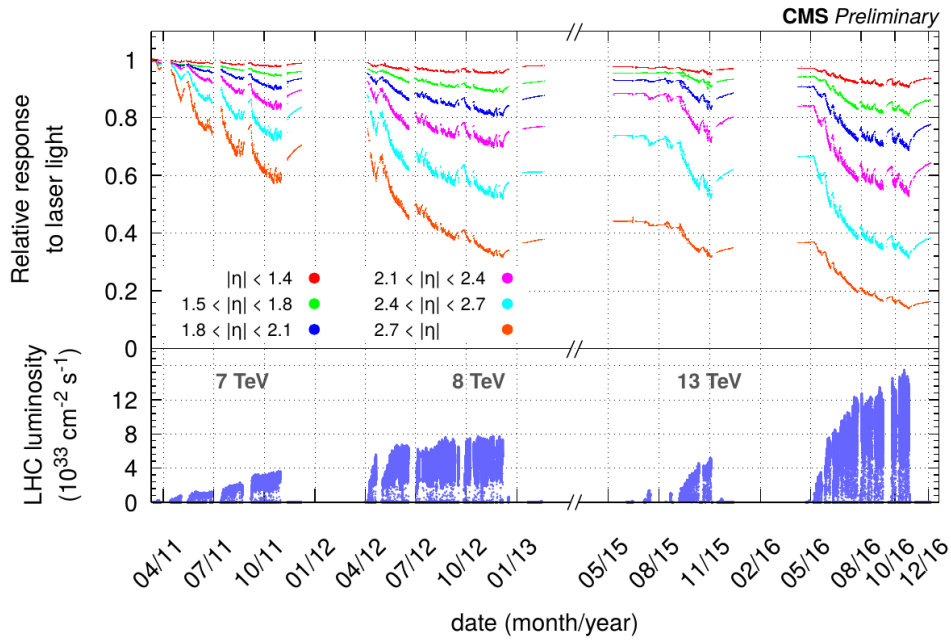

Figure 5.5: Relative response variation measured by the laser monitoring system ($\frac{S}{S_0}$) in the period 2011-2016. The response is averaged over the pseudorapidity ranges listed in the legend [79].

The corrections for transparency loss, the LC (laser correction) parameters of Equation (5.3), are then validated with collisions data, by examining the stability of the reconstructed invariant mass of π^0 decays and using the ratio of E/p for isolated electrons from W$\rightarrow$ eν and Z$\rightarrow$ ee decays. In the ratio, E is the energy measured in the calorimeter and p is the momentum measured in the tracker. Figure 5.6 shows the stability of the E/p ratio measured from 2015 data (in barrel) before and after the LC are applied. The stability of

the ratio after corrections is better than 0.1% in the barrel and 0.4% in the endcaps. Any residual imperfections in the LC (e.g. due to the dispersion in α values between crystals) are removed by applying additional time-dependent corrections, explained below.

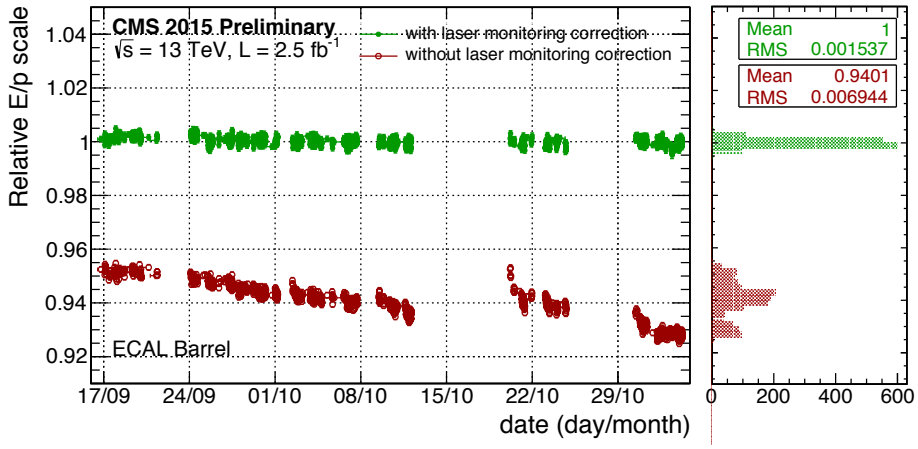

Figure 5.6: History plot for 2015 data of the ratio of electron energy E, measured in the ECAL barrel, to the electron momentum p, measured in the tracker. The electrons are selected from W$\to$ eν decays. The projections along the y-axis are shown in the histograms on the right side [79].

5.1.4.2 Intercalibrations and energy scale

Relative calibration of the crystals[5] is obtained from LHC collisions data. These intercalibration constants (IC) are calculated using several independent methods, and the resulting constants are combined to provide one number per crystal. The three methods used to achieve the task are detailed below.

- The ϕ-symmetry method is based on the equalization of the average energy measured in different channels located at a constant value of $|\eta|$. This method is in fact based on the assumption that, in minimum bias events, the physics processes are ϕ-symmetric and the CMS detector is mainly ϕ-symmetric too. It employs a dedicated data stream with reduced event information. The stream records a high rate of events ($\sim$1.5 kHz) and the method achieves a statistical precision of better than 0.2% (0.4%) in EB (EE) for a typical LHC fill. The accuracy of the method is limited to few percent by systematic uncertainties in the distribution of material in front of the ECAL. Since this uncertainty does not vary with time, the method can be used to track possible time variations in the IC values. Corrections that account for these variations are propagated to the final calibration, in time intervals of approximately one month.

- The π^0 mass method is based on the reconstruction of the peak in the spectrum of the invariant mass of unconverted photon pairs from π^0 and η decays. The photons are reconstructed using the energy sum in a 3×3 matrix of crystals centered on the crystal with the highest energy deposit, and an iterative method is used to determine the IC value of each crystal. It employs a dedicated data stream with high rate ($\sim$ 7 kHz) and reaches a precision of 0.5% in the central barrel ($|\eta| < 0.8$). This precision is dominated by systematic uncertainties.

- The E/p method is based on the comparison of the energy measured in the ECAL to the momentum measured by the tracker for isolated electrons, and an iterative

[5]The IC_i parameters in Equation (5.3).

procedure is used to extract the IC value for each crystal. The precision in the central barrel reaches the systematic limit of 0.5%, while for $|\eta| > 1$ the statistical contribution is the limiting factor.

The combined intercalibration was obtained from the mean of the individual intercalibration constants at a fixed value of $|\eta|$, weighted by their respective precisions. In the region $|\eta| > 2.5$, beyond the tracker acceptance, the E/p method can not be used and the high pile-up prevents the reconstruction of the invariant mass peak from π^0 or η, therefore only the IC estimate from the ϕ-symmetry are available. The precision of the different methods and of their combination is reported in Figure 5.7 for the barrel case; it is at the level of ∼1.5% in the endcaps.

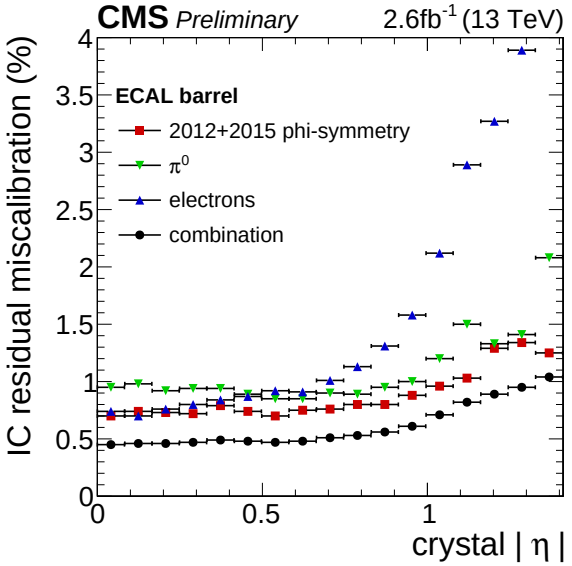

Figure 5.7: Precision of the calibration coefficients as a function of $|\eta|$, for the barrel. The precisions of the different methods and of their combination are reported [79].

The calibration of the η rings (η scale) is obtained from the invariant mass of Z→ee decays, selecting a sample of non-showering electrons. The η scale is set to match the expectation from Monte Carlo simulations.

Finally, the overall energy scale G is set, separately for EB and EE, such that the reconstructed Z peak in data matches that in the Monte Carlo.

5.1.5 Energy resolution and scale

The energy resolution for electrons and photons plays a crucial role in many CMS analyses and in particular for the searches of dielectron and diphoton resonances at high mass.

Figure 5.8 shows the measured energy resolution for electrons from Z→ee decays plotted as a function of $|\eta|$ for 2015 data. The resolution is affected by the amount of material in front of ECAL (which increases beyond $|\eta| > 1$) and the presence of cracks between modules (vertical lines in the plot). The resolution is shown for two sets of intercalibration constants: initial calibration constants derived from prompt reconstruction of 2015 data (black points) and with calibration constants derived from the re-reconstructed 2015 data (blue points). Figure 5.8 also shows the resolution expected from Monte Carlo simulations (red points). The residual discrepancy between data and Monte Carlo is accounted for by adding in quadrature a constant Gaussian smearing to the electron and photon energies

in Monte Carlo events. This last point, as well as the energy scale discrepancy between data and MC, will be treated in details in Chapters 6 and 7 given their major importance in the analyses presented in this thesis.

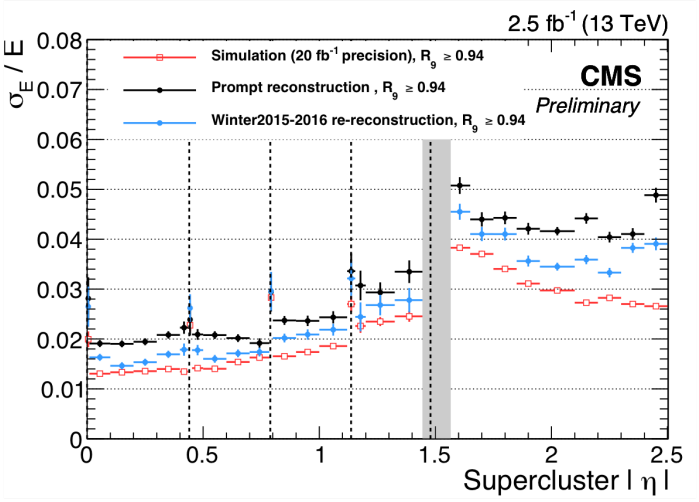

Figure 5.8: Relative electron energy resolution unfolded in bins of pseudo-rapidity $|\eta|$ for the barrel and the endcaps. Electrons from Z→ee decays are used. The resolution is shown for low bremsstrahlung electrons ($R_9 > 0.94$ with $R_9 = E_{3\times3}/E_{\text{raw}}$). All corrections from Equation (5.3) are applied [79].

Having described the common strategies of electrons and photons energy reconstruction, in the following the peculiarities of the two objects will be presented.

5.1.6 Electron reconstruction

The peculiarity of the electron reconstruction with respect to the photon recostruction consists in the required matching between an energy deposit in ECAL and a track in the tracker, which is of course not required for photons.
More specifically, four steps are involved: the track seed selection, the track building and the track fitting (the last two are usually referred to as "tracking") and the track-supercluster matching. They are described below.

Track seeding

Track seeds for electron tracks, which are the starting point for the electron track reconstruction, are built from doublets or triplets of hits in the pixel detector following two different approaches: the ECAL driven seeding and the tracker driven seeding [85].
For the ECAL driven seeding, the procedure starts from an ECAL SC, with at least 4 GeV of transverse energy and a veto of 0.15 on the ratio of hadronic energy to SC energy. The hadronic energy is calculated from the HCAL towers in a cone of $\Delta R = 0.15$ around the direction of the electron. Hits in the first pixel layer are searched by back propagating the trajectory from the barycenter of the SC, under both charge assumptions. If a pixel hit is found in a relatively wide window around the prediction from the back propagation, the track is refitted starting from the position of the hit and searching for a second hit in the next layers with a narrower window. If the first two hits are matched with the prediction from the SC, then the seed is selected.
On the other hand, tracker driven seeds are selected from tracks that were reconstructed

with the Kalman filter (KF) algorithm [86]. This algorithm is not suited for electrons that emit bremsstrahlung photons as the curvature of the track changes in that case. All seeds of KF tracks that match a supercluster in the ECAL and pass a matching criterion of the ratio between SC energy and track momentum $E/p > 0.65$ (0.75), having track momenta $2 < p < 6$ GeV ($p \geq 6$ GeV), are selected.

Depending on the analysis needs, a seeding strategy could be preferred with respect to the other. For example in the search for high mass resonances decaying in dielectron final state, the ECAL driven seeding is required at selection level.

Tracking

Once the track seed has been obtained, the tracking procedure can take place. The tracking procedure consists of the "track building" outward from the seed, for which the combinatorial track finder method (CTF) [85] is used (which is an extension of the standard KF method), followed by the "track fitting" which uses a Gaussian sum filter (GSF) method [87] in a backward fit. For the track building, starting from the seed, the combinatorial track finding algorithm iteratively adds successive layers, using the Bethe–Heitler (BH) model [88] for the modeling of the electron losses. Owing to the possibility of emitted bremsstrahlung photons, a very loose requirement between the predicted hits and the found hits is applied. Not more than one layer can have no compatible hit found, and in case of multiple hits found up to five candidate trajectories are generated per layer. Since the distribution of the energy loss after the BH model is non-Gaussian, fitting the track with the KF algorithm that uses Gaussian distributions does not give good results. For this reason, the GSF algorithm models the BH energy loss distribution as a sum of six Gaussian distributions with different means, widths and amplitudes. After passing through a layer, six new trajectory components are generated with the weight according to the weight of the initial trajectory multiplied by the weight of the Gaussian component in the BH energy loss distribution estimation. To limit the maximal number of followed trajectories to 12, the ones with low weight are dropped or merged if they are similar. Finally, the track parameters obtained have their uncertainty distributed according to the sum of Gaussian distributions from the trajectory components. For the value of the track parameter the mode of the distribution is used.

Track-supercluster matching

In order to build GSF electron candidates, a track has to be associated to a SC.
For ECAL driven tracks, the position of the SC is taken as the energy weighted position and the position of the track is the extrapolated position at the SC from the innermost track position. The difference should be smaller than 0.02 in the η direction and 0.15 rad in the ϕ direction.
For tracker driven tracks a multivariate technique, using a boosted decision tree (BDT), is used, that combines track observables and SC observables to get a global identification variable. For a successful matching, the track-SC combination should be higher than a threshold of this variable.

Charge and momentum measurement

The charge of the electron candidate is defined by the majority of the charge estimation by three different methods. The first method measures the charge from the curvature of the GSF track. The second estimate comes from the curvature of the KF track associated to the GSF track if they share at least one innermost hit. The third charge estimation comes

from the comparison of the ϕ direction of the SC position as measured from the beam spot with the ϕ direction as measured from the first hit of the GSF track. Simulations predict a charge mis-identification rate of 1.2% for electrons with a transverse momentum $p_T \sim 35$ GeV. The momentum of the electron candidates is measured by combining the momentum as measured by the tracking procedure with measurements from the ECAL. The weighting of the two measurements depends on the track parameters and the SC parameters. For electrons with high energies the precision of the energy measurement from the ECAL outweighs the one from the tracker, and the transverse momentum of the electron candidate is defined by the energy measurement from the ECAL SC E_{SC} and the polar angle of the track at the interaction point θ_{track}:

$$p_T = E_{SC} \sin \theta_{track} \tag{5.5}$$

Since the transverse momentum measurement for high energy electrons is based on the energy measurement from the ECAL, it is in the following called E_{T}.

A schematic view of the electron reconstruction procedure, considering also multiple bremsstrahlung emissions, is shown in Figure 5.9.

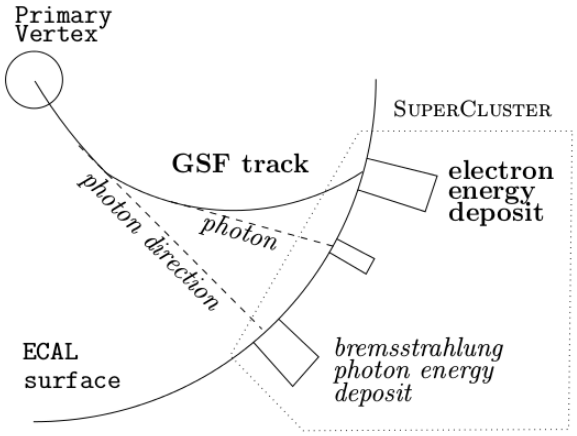

Figure 5.9: Schematic view of an electron reconstructed in CMS. The track is reconstructed by the GSF algorithm taking into account the trajectory kinks due to energetic bremsstrahlung photon emission. The energy deposits belonging to the emitted photons are collected together to the electron cluster by the clustering algorithm.

5.1.7 Photon reconstruction

The photon is the simplest electromagnetic object. Any reconstructed SC with $p_T >$ 10 GeV is considered as a photon candidate [89]. At this stage each electron is also reconstructed as a photon candidate since a veto with respect to reconstructed prompt electrons is applied only at final analysis level, with dedicated photon selections (see Section 7.3.2). Photon candidates which produce an electron pair (conversion) in the material upstream ECAL are tagged using an algorithm which searches for conversion tracks matching the ECAL SC, as will be described in the following subsection.

5.1.7.1 Photon Conversions

Conversion track pairs are reconstructed starting form standard tracks reconstructed as described before and preselected with basic quality criteria (number of hits in the tracker

layers greater than 4, $\chi^2_{track} < 10$). Hence opposite-charge pairs are combined together and asked to satisfy the photon conversion topology. Electron conversion tracks are also required to have $p_T > 1$ GeV and requested to be matched to a SC, by requiring that ΔR between the track direction and the SC position to be smaller than 0.1. About one quarter of the events have at least one of the photon reconstructed and selected as a photon producing electron pairs.

Before concluding the section about electron and photon reconstruction at CMS, it is worth noting that the final selection of electrons or photons applied at analysis level includes various other conditions (on the shower shape, on their isolation in the detector, ...) in order to efficiently distinguish them from jets. Several variables are helpful in this task and can be combined in a set of selection criteria used by the CMS analyses. Two specific sections are devoted to describe the selection criteria of high-energy electrons and photons, respectively in chapters 6 and 7.

5.2 Muon reconstruction

The muon reconstruction [77] is obtained using the information coming from the muon system and the inner tracker. Muon candidates that are reconstructed only by using the information coming from the muon system are referred to as "standalone muon candidates", while muon candidates that are reconstructed by combining the information coming from the muon system inwards to the inner tracker are referred to as "global muon candidates". Finally, muon candidates reconstructed by combining the information coming from the inner tracker outward to the muon station are referred to as "tracker muon candidates". Since the global muon reconstruction algorithm starts from standalone muon tracks going inwards, it is mostly efficient for muons leaving hits in several muon stations, while the tracker muon reconstruction is more efficient for low p_T muon candidates.

5.2.1 Standalone muon reconstruction

The standalone muon reconstruction aims to reconstruct tracks in the muon system. The reconstruction of muon tracks in the muon system starts from seeds, generated by the DT and the CSC (see Section 4.7), which are fitted to produce segments (track stubs). The track stubs obtained from the seeds are propagated to the innermost muon layer. From there a first KF pre-filter is used to find track segments in the outward direction. After the pre-filter, a filtering step, using the same technique, performs a fit in the inward direction, using the individual hits of the track segments and tighter criteria for the matching between the projected hit position[6] and the hit in the detector. A second tracking step is done in a similar way but using the hits coming from the RPC (see again Section 4.7) but without generating track segments in this case. The inclusion of the RPC measurements improves the reconstruction of low momentum muons and muons that pass a gap between the DT or CSC detectors.

5.2.2 Global muon reconstruction

The global muon reconstruction begins after the completion of the standalone reconstruction, by requiring that each standalone track is matched to a compatible track in the tracker. The process of identify the tracker track to combine with a given standalone

[6]For the propagation of the trajectory between the muon stations, energy loss from multiple scattering, ionisation and bremsstrahlung emissions are taken into account via dedicated parametrizations of these effects.

muon track is referred to as track matching and consists of two steps. The first step is the definition of a region of interest (ROI) in the track parameter space that roughly corresponds to the standalone muon track, and to select the subset of the tracker tracks inside this ROI. The determination of the ROI is based on the assumption that the standalone muon originates from the interaction point.

The second step is to iterate over this subset, applying more stringent spatial and momentum matching criteria to choose the best tracker track to combine with the standalone muon. The matching is performed by propagating the muon and the tracker tracks onto the same plane and looking for the best χ^2 value from the comparison of track parameters in the ROI. This outside-in algorithm is paired with an inside-out identification algorithm, where candidate tracker tracks are extrapolated to the muon system taking into account the magnetic field, the average expected energy losses, and multiple Coulomb scattering in the detector material.

If there is a suitable match between the tracker track and the standalone muon track, a final fit is performed all over the hits. However it is also possible to combine only a subset of the hits for the final fit. In particular, choosing a subset of the muon hits provides a better momentum resolution for high energy muons, when the measurements in the muon system are frequently contaminated by electromagnetic showers.

If the matching fails, the reconstruction is stopped and no global track is produced. If the matching algorithm selects more than one tracker track for a given standalone track, all matched tracks proceed in the reconstruction chain and the global track with the best χ^2 is chosen.

The global muon reconstruction ends with the matching of the global muon track and the energy deposits in the calorimeters.

5.2.3 Tracker muon reconstruction

Global muon track reconstruction starts from the muon system and combines standalone muon tracks with tracks reconstructed in the inner tracker. However, a large fraction of muons with transverse momentum below 6-7 GeV does not leave enough hits in the muon spectrometer to be reconstructed as standalone muons. Moreover, some muons can escape in the gap between the wheels.

A complementary approach, which starts from the tracker tracks, has therefore been designed [90] to identify off-line these muons and hence improve the muon reconstruction efficiency: it consists in considering all tracker tracks and identifying them as muons by looking for compatible signatures in the calorimeters and in the muon system.

The algorithm starts extrapolating each reconstructed tracker track outward to its most probable location within each detector of interest (ECAL, HCAL, HO, muon system). After collecting the associated signals from each detector, the algorithm determines compatibility variables corresponding to how well the observed signals fit with the hypothesis that the tracker track is produced by a muon. Based on the energy deposits in the calorimeters, a compatibility variable is determined, which describes how consistent the energies are with respect to what is expected for a muon.

If the extrapolated track matches at least one muon segment in the muon system, the corresponding tracker track qualifies as a "tracker muon".

The efficiency for reconstructing a muon as global or tracker muon is as high as 99%.

Momentum assignment

The momentum assigned to the muon candidate uses the estimates coming from the final track fit performed in the global and tracker muon reconstructions. If both estimates give $p_T > 200$ GeV and the ratio of charge and momentum agrees with each other within 2σ, the estimate from the global muon reconstruction is chosen. Otherwise, the tracker muon momentum is taken.

It is worth nothing that, in addition to these criteria, a specific selection has been developed for high p_T ($p_T > 200$ GeV) muons. In particular, the goodness of the global muon track fit selection, based on the χ^2 of the track is not requested, but an additional cut based on the $\sigma(p_T)/p_T$ for the track used for momentum determination is applied. The high-p_T muon selection is used to select the muon candidates in the search for high mass resonances decaying in dimuon final state, whose results are combined with the ones obtained looking at the dielectron final state (see Chapter 6).

5.3 Particle-flow algorithm

Before moving on with the reconstruction techniques for the remaining SM particles, few words about the so-called particle-flow (PF) algorithm [91, 92] are necessary. This algorithm allows to reconstruct stable particles combining, thanks to the high granularity of the CMS detector, information coming from all subdetectors under the form of calorimeter clusters and tracks (individually denoted as PF elements).

As a single physical particle can create multiple PF elements, such as a track and several calorimeter clusters, a link algorithm has been designed to fully reconstruct particles and to limit double counting, via computing the distance between the tracks and the calorimeter clusters to determine whether they correspond to the same physical objects. At the end of this procedure electrons, photons, muons, charged and neutral hadrons are reconstructed as individual PF particles, which are eventually combined to form more complex objects such as hadronically decaying taus, jets, or transverse missing energy.

5.4 Tau reconstruction

Tau leptons are characterized by a short lifetime (2.9×10^{-13} s) and their decay occurs inside the detector. They are therefore reconstructed from their visible decay products. In the case of a leptonic decay (about one third of the cases), the electron or the muon is the only reconstructed particle and the standard tools for electron and muon reconstruction, which have been described in the previous sections, are used. In the remaining two thirds of cases, taus decay hadronically with the production of a jet with very specific features. In particular, the jet is characterized by small track multiplicity: the hadronic decay products of a tau contain only one or three charged particles, mostly pions, which is much less than the usual track multiplicity for gluon or quark jets. The neutral part consists mostly of neutral pions which decay into two photons. Finally, jets coming from tau decays are on average more collimated and isolated than quark and gluon jets. In the case of a hadronic decay, taus are reconstructed with the Hadrons plus Strips (HPS) algorithm [93, 94]. The HPS algorithm builds taus from tracks, and energy deposits in ECAL. The efficiency in identifying hadronically decaying taus is typically between 45 and 70%, for misidentification rates of a jet as a tau of the order of 1%. The tau reconstruction is hard in several aspects: the background from jets is important and, since at least one neutrino is produced during the tau decay, only a fraction of its total momentum is measured. The

consequence is that in most analyses involving leptons in the final states, the electron and muon channel provide a much better sensitivity than the tau channel. In the search for heavy dilepton resonances, for istance, the use of the visible mass of the tau pair (defined as the invariant mass of the visible decay products of the tau with the missing transverse energy) leads to a width over mass ratio up to 50% for 2 TeV masses [95]. An important exception is the Higgs boson case where the large enhancement of the cross section in this channel, compared to the lighter leptons, makes it by far the most sensitive lepton channel.

5.5 Jet reconstruction

The cross section to produce jets is by far the largest one in pp collisions at the LHC to the point that jets represent a background for many analyses. Quark and gluon jets are built from PF objects which need to be "grouped together" into a jet. To achieve this task several algorithm can be used; the most commonly used one is the anti-k$_T$ algorithm [96]. This algorithm, in order to cluster individual PF objects into jets, proceeds by defining distances d_{ij} between two entities (particles, pseudojets) i and j, and distances d_{iB} between an entity i and the beam:

$$d_{ij} = \min(k_{ti}^{-2}, k_{tj}^{-2}) \frac{\Delta_{ij}^2}{R^2} \tag{5.6}$$

$$d_{iB} = k_{ti}^{-2} \tag{5.7}$$

where k_{ti} and k_{tj} are the transverse momenta of the i and j entities respectively, $\Delta_{ij}^2 = (\phi_i - \phi_j)^2 + (\eta_i - \eta_j)^2$, and R is the cone size parameter which can be chosen in the range from 0.4-0.7 depending on the analyses choices. If the smallest distance is of d_{ij}-type, the entities i and j are combined into a new single entity, while if it is of d_{iB}-type, the i entity is considered as a jet and removed from the list of entities. The procedure continues until the entity list is empty. Unlike other jet clustering algorithms, the anti-k$_T$ algorithm produces jets with a conical shape, clustered around the hardest particles and with boundaries resilient with respect to soft radiation (the algorithm is less sensitive on the low p_T objects while still including them).
Three different types of jets are reconstructed, depending on the way information from the subdetectors is combined [97].

- Calorimeter jets use information from the calorimeter towers in the HCAL and the corresponding ECAL crystals.
- Jet-plus-track (JPT) jets use the information from the tracker in addition to the calorimeter jets, to add tracks that are bent out of the cone defined by the calorimeter jets. The calorimeter jets parameters are corrected by taking into account these additional tracks.
- Particle flow jets are generated by the clustering of PF candidate particles and taking the vectorial sum of their four momenta.

The raw jet energies are corrected to ensure a uniform response in η and an absolute calibration in p_T. The target of the calibration is that the reconstructed jet energy matches the energy of the generated jet. The correction to the raw p_T of the jet can be decomposed in four multiplicative terms [96]:

- An offset correction, to remove the energy due to particles not involved in the hard-scattering process (pile-up, detector noise);

- A MC calibration factor, which corrects the reconstructed energy to match the generated MC particle jet energy, based on simulations;
- A residual calibration for the relative energy scale, to correct the energy response as a function of the pseudorapidity;
- A residual calibration for the absolute energy scale, to make the energy response uniform as a function of the transverse momentum.

The final energy resolution for a jet of 100 GeV of p_T is around 10%.

b-jet reconstruction

Except for the top quark, which decays defore hadronizing, all the quarks will generate a jet. However jets originating from b quark hadronization (b-jets), can be distinguished from other jets coming from gluons, light-flavor quarks (u, d, s) and c quark fragmentation using track, vertex and identified lepton information. Different algorithms to tag b-jets exist but only the Combined Secondary Vertex (CSV) algorithm is described here as it is the most commonly used in the CMS analyses. Since B hadrons typically have a lifetime of $c\tau \simeq 450$ μm, a powerful handle to discriminate between b-jets and other jets is the existence of a secondary vertex. A secondary vertex is defined as a vertex sharing less than 65% of its tracks with the primary vertex and separated radially from the primary vertex with a significance at least 3σ. In addition, if the radial distance exceeds 2.5 cm and if the mass is compatible with a K^0 or greater than 6.5 GeV, the secondary vertex is rejected. The last condition for secondary vertices is that the flight direction of each candidate is in a cone with $\Delta R = 0.5$ around the jet direction. When no secondary vertex is found, in about 35% of cases for real b-jets, the CSV algorithm can use so-called "pseudo-vertices", from tracks whose impact parameter is more than 2σ away. If no pseudo-vertex is found, the CSV algorithm proceeds from simple track variables. The information used by the CSV algorithm to identify b-jets are summarized as follows [98]:

- The presence of a secondary vertex, a pseudo-vertex or none of them;
- The flight distance significance between the primary and the secondary (or pseudo-) vertex in the transverse plane;
- The number of tracks at the secondary or pseudo-vertex;
- The ratio of the energy carried by tracks at the vertex with respect to all tracks in the jet;
- The pseudorapidities of the tracks at the vertex with respect to the jet axis;
- The number of tracks in the jet.

A likelihood ratio to reject c-jets and another one to reject light-parton jets are combined to form the final CSV discriminator. The efficiency of the CSV algorithm in data and simulations for the medium working point is close to 70% with a mistagging rate of about 1.5%.

5.6 Transverse missing energy

Neutrinos and other hypothetical neutral weakly interacting particles cannot be detected by CMS. However, some information about their presence can be gathered from the

detection of a momentum imbalance in the transverse plane to the beam axis. The missing transverse energy is noted $\vec{\not{E}}_T$, while its magnitude is referred to as $\not{E}_T$.

The most widely used type of $\vec{\not{E}}_T$ in CMS is the particle-flow (PF) $\vec{\not{E}}_T$, which is the negative vectorial sum over the transverse momenta of all PF particles. A bias in the $\not{E}_T$ measurement can be introduced for several reasons, such as the non-linearity of the response of the calorimeter for hadronic particles, or the minimum energy thresholds in the calorimeters. In addition, another bias comes from the pile-up interactions. This can be corrected by subtracting from the $\vec{\not{E}}_T$ a certain fraction $f(\vec{v})$ of $\vec{v}$, the vectorial p_T sum of charged particles associated to each pile-up vertex:

$$\vec{\not{E}}_T^{corr} = \vec{\not{E}}_T - \sum_{PU} f(\vec{v})\vec{v} \tag{5.8}$$

Finally, an asymmetry in the ϕ variable is observed for data and simulated events, and is found to be related to the number of reconstructed vertices N_{vtx}. The $\vec{\not{E}}_T$ projection in the x- and y-directions are independently corrected by some functions of N_{vtx}.

Another type of $\not{E}_T$ based on an multivariate (MVA) method is designed to reduce the influence of pile-up interactions, which do not have significant $\not{E}_T$ but degrade the $\not{E}_T$ measurement resolution by 3.3-3.6 GeV on average for each single pile-up vertex. The MVA $\vec{\not{E}}_T$ relies on the identification of jets originating from pile-up interactions with a MVA discriminator that takes as input jet shape variables and vertex information.
In Z boson decay events, the transverse energy can be decomposed in three components, as illustrated in Figure 5.10.

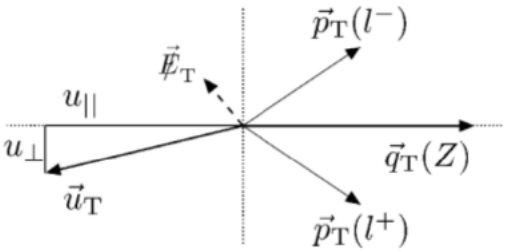

Figure 5.10: Schematic view of the Z boson transverse momentum $\vec{q}_T$, the hadronic recoil $\vec{u}_T$ with its parallel and perpendicular projections along $\vec{q}_T$, and the $\vec{\not{E}}_T$.

The following relation holds:

$$\vec{q}_T + \vec{u}_T + \vec{\not{E}}_T = \vec{0} \tag{5.9}$$

where $\vec{q_T} = \vec{p_T}(l^+) + \vec{p_T}(l^-)$ is a well-measured momentum scale (vectorial sum of the lepton transverse momenta), and $\vec{u_T}$, which is the vectorial sum of the transverse momenta for all reconstructed PF particles except the leptons originating from the Z boson decay is the hadronic recoil . The hadronic recoil can be decomposed in two components parallel or perpendicular to $\vec{q}_T$ direction: $\vec{u}_T = \vec{u}_\perp + \vec{u}_\parallel$. The MVA $\not{E}_T$ is obtained after recomputing the hadronic recoil $\vec{u_T}$ in order to avoid biases. A first BDT is trained to correct the direction of $\vec{u_T}$ to correspond to the generated direction in simulated Z+jets events, while a second BDT estimates its magnitude after direction corrections. The corrected hadronic recoil is added to $\vec{q}_T$ to give the negative MVA $\not{E}_T$. In comparison with the PF $\vec{\not{E}}_T$, the MVA $\vec{\not{E}}_T$ resolution is less sensitive to the number of interaction vertices.

As pointed out before, the $\vec{\not{E}}_T$ measurement [99] strongly relies on the reconstruction of all other physics objects, and is sensitive to a wide range of effects. Artificially large $\not{E}_T$ can be measured because of spurious detector signals. Sources of fake $\not{E}_T$ are:

- Dead cells in the ECAL;
- Beam-halo particles;
- Particles striking sensors in the ECAL barrel detector;
- Noise from HCAL hybrid photodiode;
- Direct particle interactions with light guides and photomultipliers tubes in the forward calorimeter;
- High-amplitude anomalous pulses in the ECAL endcaps;
- A misfire of the HCAL laser calibration system;
- A defective track reconstruction, from coherent noise in the silicon strip tracker.

Even if dedicated algorithms are used to identify and remove these events generating fake $\not{E}_T$, the problem is still present. Among others, this is the main reason why in the searches for high mass resonances decaying in dielectron or diphoton pairs no requirements are made, at final selection level, on the missing transverse energy (see Sections 6.3 and 7.3).

5.7 Summary

This chapter described the techniques used in CMS to reconstruct the final state objects. The electron and photon reconstruction has been described in details in Section 5.1, while the muon, tau, jet and missing transverse energy reconstructions are described in Sections 5.2-5.6.

My personal contributions on the determination of the energy scale and resolution of the ECAL detector are briefly introduced in Section 5.1.5 in order to place them in the general picture. A detailed description can be found in Section 7.4. In addition, a discussion of the so-called "slew rate issue" is reported in Section 5.1.2, which I investigated in the second half of 2016, when the amount of integrated luminosity in data exposed it.

Chapter 6

Search for new physics in the dielectron final state at CMS

One of the most striking signatures of physics beyond the standard model would be the observation of a narrow resonance in the invariant mass spectrum of lepton pairs. As described in details in Chapter 2, many BSM models predict the existence of such resonances at the TeV scale, generally denoted as Z′. Examples are discussed in Chapter 2 and cited here for completeness: they include a new heavy Z boson-like particle with SM-like couplings (Z'_{SSM}[1]) and the Z'_{ψ} boson inspired by superstring models. Another example is the Kaluza-Klein graviton of the Randall-Sundrum (RS) model of extra dimensions which would give a spin-2 dilepton resonance.
The analysis reported in this chapter was performed with the full 2016 dataset, corresponding to an integrated luminosity of 35.9 fb^{-1}.

6.1 Data and MC samples

The primary datasets used in this analysis are summarised in Table 6.1. The data from the "Moriond17" re-reconstruction campaign are used for eras from "2016B" to "2016G", while the prompt reconstruction is used for era "2016H", for an integrated luminosity of 35.9 fb^{-1} collected by the CMS experiment during the entire 2016. The analyzed sample fulfilled standard data quality criteria for all components of the CMS detector ("golden json"). Different datasets are used in the analysis. For the main results the "DoubleEG" dataset has been used, in which at trigger level two electromagnetic objects compatible with two electrons/photons are present. For trigger and selection efficiency measurements the "SingleElectron" dataset is used, in which at trigger level one electromagnetic object compatible with one electron is present. For the QCD background study ("fake rate" method) the "SinglePhoton" dataset has been used, in which at trigger level one electromagnetic object compatible with one photon is present. Finally, for the $t\bar{t}$-like background study ("eμ" method) the "SingleMuon" dataset has been used, in which at trigger level a muon candidate is present.

The different Monte Carlo (MC) simulated samples used in this analysis were centrally produced during the "RunIISummer16" campaign. Table 6.2 reports the samples related to the different background sources for the analysis (see Section 6.5) and it is organized as follows: in the top part the MC samples related to the SM Drell-Yan (DY) process are listed[2], followed by $t\bar{t}$ samples, diboson samples (WW, WZ and ZZ), samples containing a

[1] The Z'_{SSM} model is considered as a reference model, rather than a realistic one.
[2] The Drell-Yan samples are available in different mass bins and different precision for the cross section computation in quantum field theory.

Dataset	Integrated luminosity
/X/Run2016B-03Feb2017_ver2-v2/MINIAOD	5.8 fb^{-1}
/X/Run2016C-03Feb2017-v1/MINIAOD	2.6 fb^{-1}
/X/Run2016D-03Feb2017-v1/MINIAOD	4.3 fb^{-1}
/X/Run2016E-03Feb2017-v1/MINIAOD	4.0 fb^{-1}
/X/Run2016F-03Feb2017-v1/MINIAOD	3.1 fb^{-1}
/X/Run2016G-03Feb2017-v1/MINIAOD	7.5 fb^{-1}
/X/Run2016H-03Feb2017_ver2-v1/MINIAOD	8.4 fb^{-1}
/X/Run2016H-03Feb2017_ver3-v1/MINIAOD	0.2 fb^{-1}

Table 6.1: Datasets (X) used in the analysis. Depending on the different studies, specified in the text, X="DoubleEG", "SingleElectron", "SinglePhoton", "SingleMuon".

W boson and jets in the final state and finally samples with photons and jets in the final state ("GJets" in the Table). The DY background is generated with POWHEG v2 [100, 101, 102, 103, 104, 105] for the next-to-leading order (NLO) matrix elements using the NNPDF3.0 parton distribution functions (PDFs) and PYTHIA 8.2 [106] for parton showering and hadronization. The final state photon radiation is handled by PHOTOS [107]. The DY cross section from NLO is corrected to next-to-next-to-leading order (NNLO) in perturbative QCD by using a dilepton invariant mass dependent k-factor according to the predictions of the FEWZ 3.1.b2 program [108]. For the FEWZ predictions (see Section 6.5.1) a special PDF set [109] is used in which the QCD parton distribution functions, based on the PDF4LHC [110] set, are combined with the photon parton distribution functions, accounting for missing electro-weak corrections at NLO. Another irreducible non-resonant background arise from a $\gamma\gamma$ initial state via t- and u-channel processes. This photon-induced process would also produce two leptons in the final state but according to studies with recent PDFs containing a photon component [109] its contribution has no significant effect on the background yield. This contribution is included in the k-factor that corrects the DY NLO cross section.

The $t\bar{t}$ background is simulated with POWHEG v2. The corresponding cross section is calculated at NNLO with TOP++ assuming a mass of 172.5 GeV for the top quark. The tW and WW productions are simulated with POWHEG v2, with parton showering and hadronization described by PYTHIA 8.2.

The inclusive diboson processes WW, WZ and ZZ are simulated at leading order (LO) using the PYTHIA program along with the NNPDF2.3 [111] PDFs.

W+jets production is simulated at LO with the MADGRAPH5 [112] program.

The NNPDF3.0 [113] PDFs are used for samples generated at NLO. PDFs are evaluated using the LHAPDF library [114, 115, 116]. The detector response is simulated using the GEANT4 [117] package.

All the MC samples include the simulation of pile-up interactions (see Section 3.2) comparable with the pile-up distribution of the full 2016 dataset under study. In Figure 6.1, the data and MC pile-up distributions are compared. MC events are then re-weighted at analysis level to account for the pile-up difference between data and MC.

6.2 Trigger

Different high level triggers (HLT) have been used for the analysis. The main signal trigger asks for two electron candidates with the kinematic requirements of having $E_T>$ 33 GeV, loose calorimeter identification requirements and very loose matching criteria between the GSF track and the supercluster in ECAL.

Sample	Cross section [pb]
ZToEE_NNPDF30_13TeV-powheg_M_50_120	1975
ZToEE_NNPDF30_13TeV-powheg_M_120_200	19.32
ZToEE_NNPDF30_13TeV-powheg_M_200_400	2.73
ZToEE_NNPDF30_13TeV-powheg_M_400_800	0.241
ZToEE_NNPDF30_13TeV-powheg_M_800_1400	1.68E-2
ZToEE_NNPDF30_13TeV-powheg_M_14000_2300	1.39E-3
ZToEE_NNPDF30_13TeV-powheg_M_2300_3500	8.948E-5
ZToEE_NNPDF30_13TeV-powheg_M_3500_4500	4.135E-6
ZToEE_NNPDF30_13TeV-powheg_M_4500_6000	4.56E-7
ZToEE_NNPDF30_13TeV-powheg_M_6000_Inf	2.06E-8
DYToEE_NNPDF30_ 13TeV-powheg-pythia8	1921.8
DYJetsToLL_M-50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	5765.4
DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	5765.4
TTTo2L2Nu_noSC_TuneCUETP8M2T4_13TeV-powheg	87.31
TTToLL_MLL_500To800_TuneCUETP8M1_13TeV-powheg-pythia8	0.326
TTToLL_MLL_800To1200_TuneCUETP8M1_13TeV-powheg-pythia8	3.26E-2
TTToLL_MLL_1200To1800_TuneCUETP8M1_13TeV-powheg-pythia8	3.05E-3
TTToLL_MLL_1800ToInf_TuneCUETP8M1_13TeV-powheg-pythia8	1.74E-4
TT_TuneCUETP8M2T4_13TeV-powheg	831.76
ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8	35.6
ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8	35.6
WW_TuneCUETP8M1_13TeV-pythia8	118.7
WWTo2L2Nu_13TeV-powheg	12.178
WWTo2L2Nu_Mll_200To600_13TeV-powheg	0.1322
WWTo2L2Nu_Mll_600To1200_13TeV-powheg	5.404E-3
WWTo2L2Nu_Mll_1200To2500_13TeV-powheg	3.3931E-4
WWTo2L2Nu_Mll_2500ToInf_13TeV-powheg	5.1484E-6
WZ_TuneCUETP8M1_13TeV-pythia8	47.13
ZZ_TuneCUETP8M1_13TeV-pythia8	16.523
WJetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	61526.7
GJets_HT-40To100_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	20790
GJets_HT-100To200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	9238
GJets_HT-200To400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	2305
GJets_HT-400To600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	274.4
GJets_HT-600ToInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	93.46

Table 6.2: Complete list of the MC samples used in the analysis.

It is so labeled as "HLT_DoubleEle33_CaloIDL_GsfTrkIdVL". The first-level trigger (L1) seeding for this HLT evolved during the data taking: it is always seeded by the OR of a DoubleEG seed, a SingleEG seed and a SingleJet seed (respectively requiring two deposits in ECAL, one localised deposit in ECAL and an object compatible at L1 with a jet), with the exact threshold of each of those seeds changing in time. From run 275319 it is also seeded by a SingleTau seed (requiring the presence of an object compatible at L1 with a τ). The presence of the SingleJet and SingleTau seeds are meant to mitigate the case of a loss of efficiency in the EG seeds at high E_{T}. The lowest E_{T} threshold for the SingleEG was 40 GeV, with the corresponding E_{T} thresholds for the DoubleEG seed being 24 GeV and 17 GeV for the two ECAL deposits respectively.

The HLT efficiency has been factorized as the product of three terms: the efficiency of the L1 trigger, the HLT turn on curve with respect to the supercluster E_{T}, and the online electron identification efficiency of the trigger (HLT ID). Only the HLT turn on curve is applied in the main part of the analysis: the L1 efficiency will be shown in Section 6.2.1

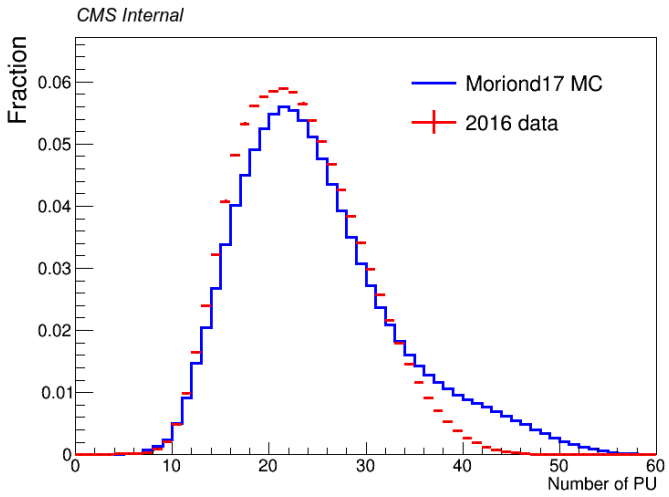

Figure 6.1: Pile-up distribution for 2016 data (red histogram) and MC samples (blue histogram) [118].

to be close to 100% while the HLT ID efficiency will be shown to be E_{T}-independent and thus cancels out in the way the final limits are presented, as a ratio of cross sections normalized to the Z peak (see Section 6.6).

Trigger efficiency: methodology

The tag-and-probe method [119] is used to measure efficiencies directly in data using pair produced objects, as the electrons from the DY process. A tight selection is applied to one object, called the tag, and the efficiency is then measured with a second object, called the probe, which is selected with a loose selection. The efficiency ϵ is defined as

$$\epsilon = \frac{N_{\text{passing probes}}}{N_{\text{all probes}}} \tag{6.1}$$

where $N_{\text{all probes}}$ is the total number of all selected probes and $N_{\text{passing probes}}$ is the total number of selected probes which pass the selection criteria whose efficiency is under evaluation. The tag does not enter the efficiency calculation and is used to ensure a sample of high purity, given its tight requirements. For all the efficiencies measurements presented in this thesis, electrons from the Z decays are used.

In order to measure the trigger efficiency, events are selected by requiring the tight working point (WPTight) of the single electron trigger labeled as "HLT_Ele27_eta2p1_WPTight", which requires an energy deposit in ECAL with $E_{\mathrm{T}}>$ 27 GeV with the η cut restricted up to the absolute value of 2.1. The tag is required to pass this trigger, pass the final event selection (denoted as HEEP selection and described in section 6.3), and be in the barrel[3]. To simplify the computation probes can not be also tags. In the case of the probe being in the barrel, the tag is required to have a smaller supercluster ϕ than the probe for even number events and a larger supercluster ϕ for odd number events. The probe must also pass the HEEP selection.

The L1 efficiency and HLT turn on curves are fitted with either a single or double turn on function (defined in terms of the "error function" erf), as appropriate. The double turn on function is shown in Equation (6.2), with the single turn on function being identical to the double turn on function except that the B terms are removed. The A_0 and B_0

[3]This requirement is asked in order to ensure a sample with high purity.

parameters can be interpreted as the efficiency at the plateau, the A_1 and B_1 as the value where the efficiency reaches half maximum and the A_2 and B_2 terms drive the rapidity of the turn on of the curve.

$$f(E_T) = 0.5 \cdot A_0 \cdot \left(1 + erf\left(\frac{E_T - A_1}{\sqrt{2} \cdot A_2}\right)\right) + 0.5 \cdot B_0 \cdot \left(1 + erf\left(\frac{E_T - B_1}{\sqrt{2} \cdot B_2}\right)\right) \tag{6.2}$$

6.2.1 Primary signal trigger: L1 efficiency

As mentioned before, the unprescaled L1 seeding requirements for the primary signal trigger evolved over time. The DoubleEG E_{T} thresholds ranged from 22.10 GeV to 24.17 GeV, while the SingleEG E_{T} thresholds ranged from 26 GeV to 40 GeV. Table 6.3 shows the percentage of the 2016 dataset having a given L1 seed as the lowest unprescaled one. Approximately 80% of the data had SingleEG30 as L1 seed, meaning that the corresponding E_{T} threshold is 30 GeV, with only 13% having SingleEG40 as the lowest unprescaled SingleEG seed.

Lowest unprescaled SingleEG seed	Fraction of 2016 dataset
SingleEG26	0.2%
SingleEG30	79.9%
SingleEG32	5.2%
SingleEG34	1.3%
SingleEG36	0.03%
SingleEG40	13.3%

Table 6.3: The fraction of the 2016 dataset with a given lowest unprescaled SingleEG seed.

The efficiency for an electron to pass the lowest unprescaled SingleEG seed, shown in Figure 6.2, has been measured in data using the tag-and-probe method as a function of the supercluster E_{T}[4]. Applying it to MC events, this translates to an efficiency of 99.5% to select barrel-barrel and 98.8% to select barrel-endcap events in a mass range of 60 to 120 GeV and a $\sim$ 100% efficiency above 120 GeV. This statement can be taken as a lower bound on the efficiency, given that there is also the DoubleEG L1 seed which further increases the efficiency. Hence, it is assumed that the L1 efficiency is effectively $\sim$ 100% with a 0.5% uncertainty in the barrel and 1.2% in the endcap.

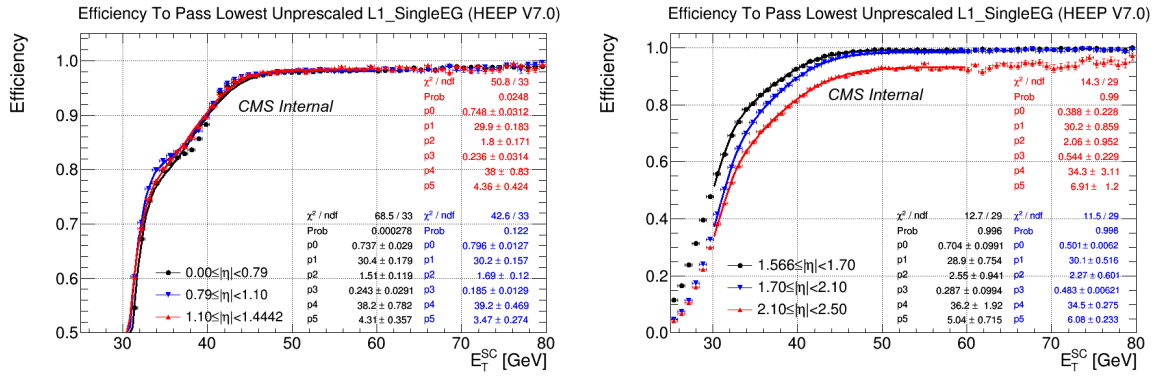

Figure 6.2: The efficiency (from tag-and-probe method) for an electron passing the HEEP selection to pass the lowest unprescaled L1 SingleEG threshold as a function of the supercluster E_{T} and $|\eta|$ for barrel electrons (left) and endcap electrons (right) [118].

[4]The supercluster E_{T} , differently with respect to E_{T}, is not measured with respect to the primary vertex, but with respect to the origin of the detector (0,0,0).

6.2.2 Primary signal trigger: HLT efficiency

The HLT efficiency is divided into two components, the efficiency of the E_T>33 GeV cut (HLT turn on curve) and the efficiency of the online electron identification criteria (HLT ID), i.e the shower shape criteria (CaloIdL) and the GSF track matching criteria with the SC (GsfTrkIdVL) asked by the HLT. The efficiencies are again measured using the tag-and-probe method.

Figure 6.3 shows the turn on curve of the E_T cut as a function of the offline reconstructed E_T for three $|\eta|$ regions. The trigger is almost fully efficient for electron E_T> 35 GeV in the barrel and mostly turned on for electron E_T> 35 GeV in the endcap. These curves are used to weight the Monte Carlo events, using the functional form depicted in Equation (6.2), in order to take into account the small inefficiency loss at low E_T introduced by the turn on effect.

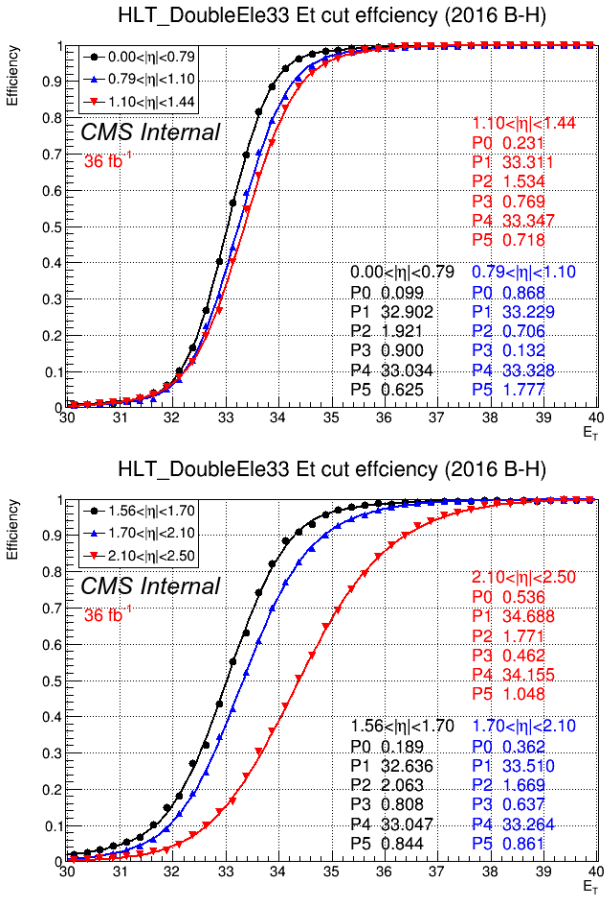

Figure 6.3: The efficiency (from tag-and-probe method) for electrons in the barrel (top) and endcaps (bottom) passing the HEEP selection to pass the online cut E_T> 33 GeV [118].

Finally, the efficiency of the online identification requirements of the HLT is shown in Figure 6.4. As the efficiency is flat vs E_T, there is no need to furtherly weight the MC with this extra factor, because it will automatically be included in the final limit results as they are presented as a ratio of cross sections with respect to the Z peak (see Section 6.6).

To conclude the section, the efficiency of the HLT_Ele27_eta2p1_WPTight trigger path is briefly discussed below.

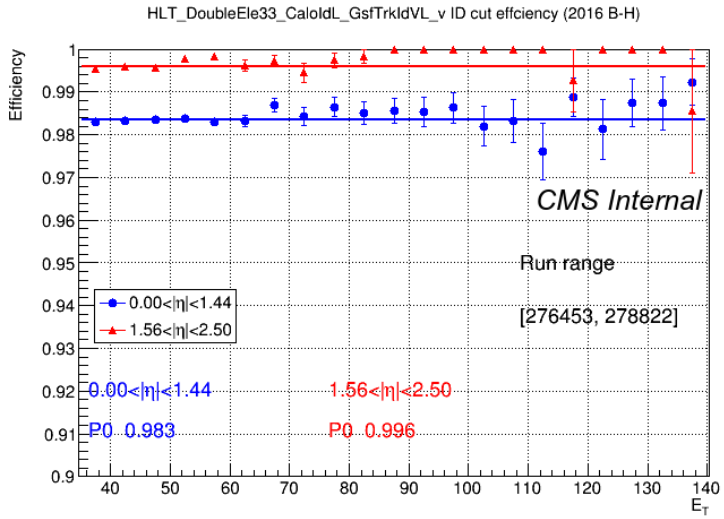

Figure 6.4: The efficiency (from tag-and-probe method) for electrons in barrel and endcaps passing the HEEP selection to pass the online identification requirements of the HLT as a function of E_{T} [118].

HLT Ele27 trigger efficiency

The trigger path HLT_Ele27_eta2p1_WPTight is used for the determination of the data-MC scale factors of the HEEP selection. Therefore, even if this is not the primary HLT of the analysis, its efficiency has been also studied in data using the tag-and-probe method and shown in Figure 6.5.

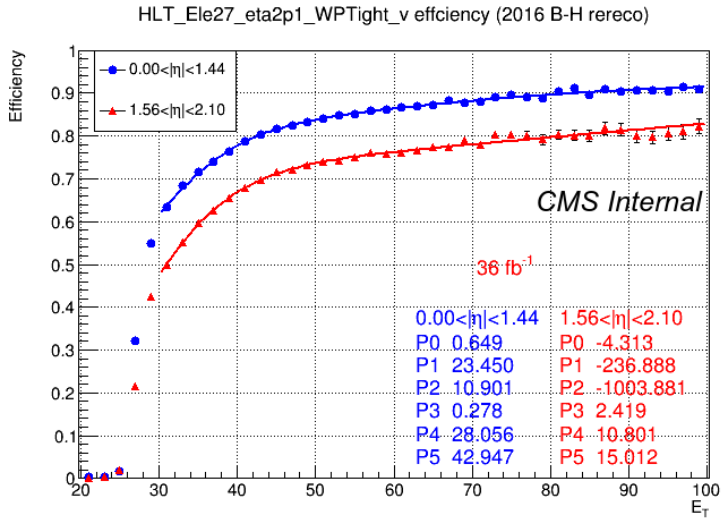

Figure 6.5: The efficiency for an electron passing the HEEP selection to pass the HLT_Ele27_eta2p1_WPTight trigger as a function of the supercluster E_{T} [118].

6.3 Event selection

A high selection efficiency for electron E_{T} above 100 GeV and a flat E_{T} dependance of the efficiency itself are crucial requirements for a search targeting new high mass resonances decaying into electron-positron pairs. It is also important that the selection performs

reasonably well for lower E_{T} values, down to 35 GeV, in order to use the Z peak in all the control region studies. Finally, given the fact that the validation in data of the selection at very high E_{T} is technically difficult due to lack of statistics, a simple and robust selection is preferred with respect to a more complex MultiVariate approach, even if in principle it could lead to a slightly more efficient selection.

6.3.1 Electron identification and isolation requirements

In order to fulfill these requirements, a cut based selection has been developped: it consists of a series of cuts on several variables that exploit the specificity of high E_{T} electrons. Only electron candidates[5] with $E_{\mathrm{T}}>$ 35 GeV and well reconstructed in the tracker and ECAL sensitive regions are selected. Candidates in the ECAL transition region ($1.4442 < |\eta| < 1.566$) and beyond the $|\eta|$ coverage ($|\eta| > 2.5$) of the tracker are therefore discarded. A different selection is applied for candidates reconstructed in the ECAL barrel ($|\eta| < 1.4442$) and endcaps ($1.566 < |\eta| < 2.5$). This cut-based selection is called HEEP (High Energy Electron Pairs): the V7.0 implementation of the cuts are listed in Table 6.4.

Variable	Barrel	Endcap
	Acceptance selections	
E_{T}	$E_{\mathrm{T}}>$ 35 GeV	$E_{\mathrm{T}}>$ 35 GeV
η	$\lvert\eta\rvert < 1.4442$	$1.566 < \lvert\eta\rvert < 2.5$
	Identification selections	
$\Delta\eta_{in}^{seed}$	$\lvert\Delta\eta_{in}^{seed}\rvert < 0.004$	$\lvert\Delta\eta_{in}^{seed}\rvert < 0.006$
$\Delta\phi_{in}$	$\lvert\Delta\phi_{in}\rvert < 0.06$	$\lvert\Delta\phi_{in}\rvert < 0.06$
H/E	$H/E < 1/E + 0.05$	$H/E < 5/E + 0.05$
$\sigma_{i\eta i\eta}$	-	$\sigma_{i\eta i\eta} < 0.03$
$\frac{E_{1\times5}}{E_{5\times5}}$ and $\frac{E_{2\times5}}{E_{5\times5}}$	$\frac{E_{1\times5}}{E_{5\times5}} > 0.83$ or $\frac{E_{2\times5}}{E_{5\times5}} > 0.94$	-
Inner lost layer hits	lost hits ≤ 1	lost hits ≤ 1
Impact parameter d_{xy}	$\lvert d_{xy}\rvert < 0.02$ cm	$\lvert d_{xy}\rvert < 0.05$ cm
	Isolation selections	
Calorimeter isolation Iso	$Iso < 2 + 0.03E_{\mathrm{T}}[\mathrm{GeV}] + 0.28\rho$	$Iso < 2.5 + 0.28\rho$ ($E_{\mathrm{T}}<$50 GeV) else $Iso < 2.5 + 0.03(E_{\mathrm{T}}[\mathrm{GeV}] - 50) + 0.28\rho$
p_T isolation $Isopt$	$Isopt <$ 5 GeV	$Isopt <$ 5 GeV

Table 6.4: Definitions of HEEP selection cuts.

The variables used in the HEEP definition are defined as follows:

- $\Delta\eta_{in}^{seed}$: the difference in η, between the track position as measured in the inner layers, extrapolated to the interaction vertex and then extrapolated to the calorimeter and the position of the supercluster's seed. The cut value in the barrel (0.004) is thighter than in the endcaps (0.006), since the tracker material budget is thicker in the latter case, so reducing the precision on the position measurement.

- $\Delta\phi_{in}$: is the analoguous of $\Delta\eta_{in}^{seed}$ for the ϕ coordinate[6]. Its distribution is however much broader than $\Delta\eta_{in}^{seed}$ because of the wider spread of the energy in ϕ with respect to η due to photon emissions from bremsstrahlung processes. Hence, the cut value (0.06 in both barrel and endcap regions) is around ten times looser than for $\Delta\eta_{in}^{seed}$.

- H/E: the ratio of the energy deposited in the HCAL tower behind the electron candidate and the energy of the electron candidate itself.

[5] An electron candidate is a GSF electron as described in Section 5.1.6.
[6] The only difference is that for $\Delta\eta_{in}^{seed}$ the position of the supercluster's seed is used.

- $\sigma_{i\eta i\eta}$: the (weighted) spatial second order moment of the electron candidate in the η direction, computed as

$$\sigma_{i\eta i\eta} = \sqrt{\frac{\sum_{i\in 5\times 5}(\eta_i - \bar{\eta})^2 w_i}{\sum_{i\in 5\times 5} w_i}}, \qquad w_i = \max\left(0, 4.7 + \log(E_i/E_{5\times 5})\right).$$

- $\frac{E_{1\times 5}}{E_{5\times 5}}$: the ratio of the energy contained in the 1×5 domino in $\eta \times \phi$ in the barrel ($x \times y$ in the endcaps) centered in ϕ on the seed crystal of the supercluster over the energy of the 5×5 matrix centered on the same seed crystal.

- $\frac{E_{2\times 5}}{E_{5\times 5}}$: the ratio of the energy contained in the most energetic 2×5 domino in $\eta \times \phi$ in the barrel ($x \times y$ in the endcaps) centered in ϕ on the seed crystal of the supercluster over the energy of the 5×5 matrix centered on the same seed crystal.

- Missing hits in the innermost layers of the trackers: defined as the number of missing hits in the innermost layers of the tracker (including the pixel) before the GSF track (see Section 5.1.6) first hit. It is mainly designed to reject photons that convert into a pair of electrons in the tracker.

- Impact parameter d_{xy}: the closest distance, in the transverse plane, between the primary vertex and the track of the GSF electron candidate. The distribution is wider in the endcaps due to the poorer resolution on the track momentum in that region. Similarly to the missing hit cut, the d_{xy} cut is mainly useful to reject converted photons.

- ECAL isolation: defined as the scalar sum of transverse energy of all the ECAL crystals with E_{T}> 80 MeV in the barrel (E_{T}> 100 MeV in the endcaps) in a cone of $\Delta R = 0.3$ centered on the GSF electron candidate position in the calorimeter, excluding those in an inner cone of radius 3 crystals and those in a $|\eta|$ strip of total width of 3 crystals (see left side of Figure 6.6). The inner cone veto removes the electron energy from the sum whereas the $|\eta|$ strip removes the energy from the bremsstrahlung photons.

- HCAL isolation: defined as the sum of transverse energy collected by all the towers of the first layer of the HCAL in a cone of $\Delta R = 0.3$ centered on the GSF electron candidate position in the calorimeter, excluding the towers in a cone of $\Delta R = 0.15$ (see right side of Figure 6.6).

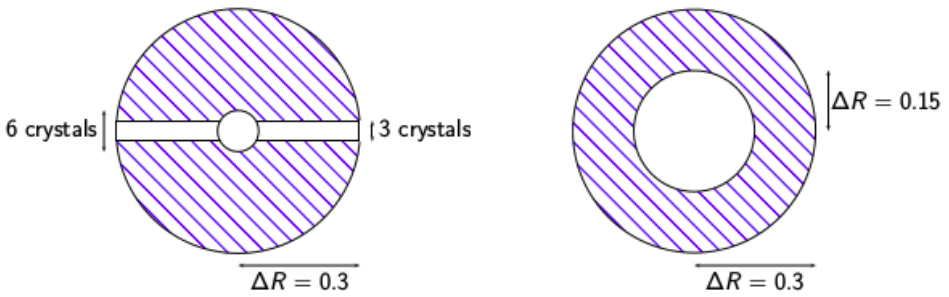

Figure 6.6: Definition of the regions used to compute the ECAL (left) and HCAL (right) isolation.

- Calorimeter isolation Iso: sum of the ECAL isolation and the HCAL isolation (defined in the previous points). The variable is strongly dependent on the electron energy and tends to increase for high E_{T} electrons due to the extension of the shower. A cut value of the form $Iso < a + b \cdot E_T$ is therefore applied.

- p_T isolation $Isopt$: defined as the sum of the p_T of all the reconstructed tracks between an inner and outer cone defined respectively as ΔR =0.04 and 0.3 and centered on the GSF electron candidate direction. Only tracks having $p_T > 700$ MeV and $|\Delta z|$ with the electron track < 0.2 cm are considered, where z is the minimum distance of the track to the nominal interaction point (0,0,0). The inner cone veto is present to remove duplicate tracks associated to the electron as well as tracks from converted bremsstrahlung photons. This variable substitutes the default tracker isolation variable in the GSF electron, for which the efficiency has a large pile-up dependence starting at $E_T \sim 100$ GeV, due to a misbehaviour of the reconstruction software that allows many low quality tracks to enter the isolation cone. The track isolation variable has then been redefined introducing ad-hoc "high purity" requirements which removed most of the problematic tracks that introduced the pile-up dependance of this variable.

Electron candidates are required to satisfy the HEEP selection criteria described above and events with at least 2 electrons passing HEEP are selected. At least one of the electron candidates has to be in barrel (events with both electron candidates in endcap regions are rejected). If more than one dielectron candidate is found in the event, only the pair with the two largest electron E_{T} is retained. No charge requirement for the electrons is asked to avoid efficiency losses at high mass for the main analysis. Charge requirements are introduced only in the estimation of the QCD contribution from data using same sign control region, in the computation of the HEEP scale factors (see Section 6.3.2) and for the $\mathrm{t\bar{t}}$ and $\mathrm{t\bar{t}}$-like background estimation with eμ studies (see Section 6.5.2).

6.3.2 Selection efficiency

The efficiency of the HEEP selection has been measured in data and MC using the tag-and-probe method (see Section 6.2) in $\mathrm{Z} \rightarrow$ ee events.

The following conditions are required:

- The tag must pass the HEEP selection.

- The tag must be a barrel electron.

- The tag must be matched to the HLT_Ele27_eta2p1_WPTight trigger in data (trigger turn on curve is applied to MC events)

- The tag is then paired with every other GSF electron candidate in the event that passes both the E_{T} and η acceptance cuts of the HEEP selection (probe).

- The invariant mass of the tag (T) and probe (P) system must satisfy the invariant mass requirement: $70 < m_{\mathrm{TP}} < 110$ GeV. If there are multiple tag-and-probe candidates per event, then all pairs are selected.

In this case, the efficiency is measured in both data and MC as a function of different variables of interest like electron E_{T}, η, ϕ.

In order to measure the HEEP selection efficiency in MC, the events are weighted using the trigger turn on curve related to the HLT_Ele27_eta2p1_WPTight path, shown in Figure 6.5 of Section 6.2.2, to emulate the efficiency of this trigger. In particular, the

curve in Figure 6.5 is used to simulate the effect of the trigger on the MC samples in the following way: for each electron in the MC, a random number is thrown from the uniform distribution between 0 and 1; if this number is lower than the E_{T},$|\eta|$-appropriate efficiency for that electron, the electron is consider to have passed the trigger.

In order to measure the HEEP selection efficiency in data, events are selected from the SingleElectron dataset requiring the HLT_Ele27_eta2p1_WPTight trigger. In order to have better data/MC agreement at the Z peak, electron energy in data is scaled using the mean values of the official electron/photon (EGM) scale corrections from the Physics Object Group (POG)[7], which are found to be in agreement with the dielectron mass scale corrections described in Section 6.4.

In Figure 6.7, the invariant mass distributions of selected tag-and-probe pairs are shown for all pairs (top), pairs in which the probe passes the HEEP selection cuts (middle and referred to as "pass-pass" region) and pairs in which the probe fails to pass the HEEP selection cuts (bottom and referred to as "pass-fail" region). Due to the different selection efficiencies in barrel and endcaps, the invariant mass distribution of selected TP pairs are separated for probe electrons in barrel (left plots) and endcaps (right plots). All standard model contributions are estimated from MC except for QCD processes. The QCD contribution is negligible in the pass-pass region compared to the DY. However, its contribution in the pass-fail region is important and should be taken into account in order to find the correct value of $N_{\text{all probes}}$ in Equation (6.1). The contribution of the QCD background, where selected electrons are misidentified jets, is extracted from data using the "same-sign" method, which exploits the fact that the probability of assigning positive or negative charge to the misidentified jet should be equal[8]. Therefore, the number of opposite sign and same sign tag-and-probe pairs are similar for QCD processes, as well as the invariant mass distribution shapes. The shape and normalization of QCD in the "pass-fail" region is estimated from data in the "pass-fail" same sign control region by subtracting all backgrounds from data using MC. It is also worth mentioning that the contribution of W+jets process is around three times more than the QCD one.

The HEEP selection efficiencies for data and MC are shown as functions of E_{T} and ϕ of the probe in Figures 6.8-6.9. In addition, data to MC scale factor per bin and the corresponding mean value are shown in the bottom part of the same plot. Data in the left part of the Figure 6.8 includes both DY and non-DY events and shows a dip of efficiency around $E_{\mathrm{T}} \approx 70$ GeV which is due to background contamination. In order to prove this last statement the efficiencies are compared with those obtained, in the right part of Figure 6.8 where the non-DY contributions are subtracted from data. In that case, the efficiences are flat as functions of E_{T} for $E_{\mathrm{T}} > 50$ GeV.

The main sources of systematic uncertainties on the scale factor originate from non-DY processes. The nominal value of the cross sections are varied by 10% and 50% for $t\bar{t}$ and W+jets processes, respectively. For the QCD estimation, a 50% uncertainty is considered. Besides, the uncertainty of the pile-up weights is also considered although it is negligible. The summary of the efficiencies and scale factors is shown in Table 6.5.

[7] The EGM corrections and the way they are derived will be described in Section 7.4.

[8] LHC is a proton-proton collider, hence positive charged jets should be produced with slightly bigger rate to which this study is not sensitive though.

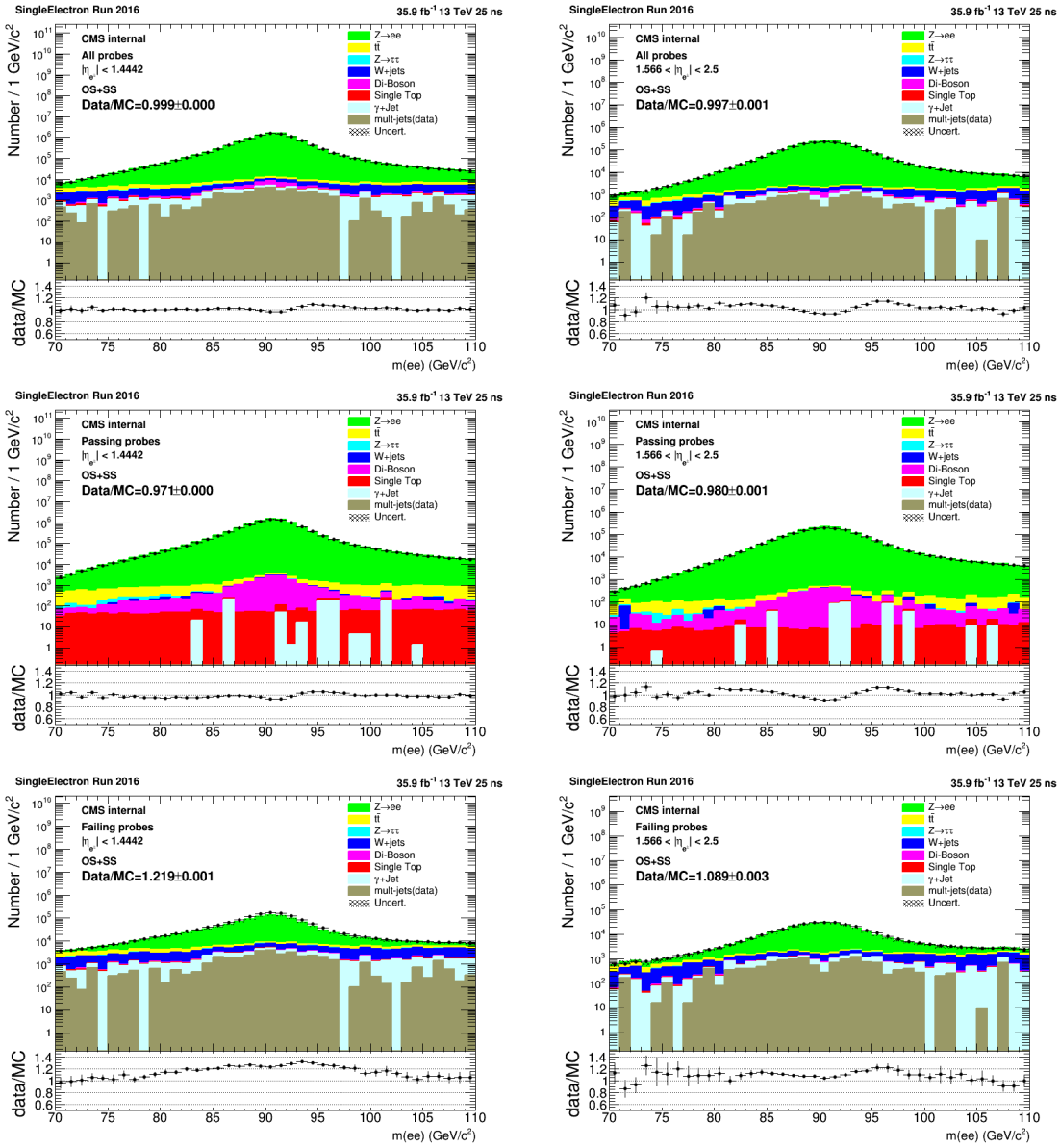

Figure 6.7: Invariant mass of the tag-and-probe system for probes in the barrel (left) and endcaps (right), where all the probes are included (top), only passing probes are included (in the middle) and only failed probes are included (in the bottom) [118].

	Barrel	Endcap
Data	$86.13\% \pm 0.01\%$(stat.)	$83.38\% \pm 0.03\%$(stat.)
DY + non-DY	$88.65\% \pm 0.03\%$(stat.)	$84.85\% \pm 0.09\%$(stat.)
Scale factor	0.972 ± 0.000(stat.) ± 0.006(syst.)	0.983 ± 0.001(stat.) ± 0.007(syst.)
Data - non-DY	$87.92\% \pm 0.03\%$(stat.)	$85.83\% \pm 0.09\%$(stat.)
DY	$90.50\% \pm 0.01\%$(stat.)	$87.35\% \pm 0.03\%$(stat.)
Scale factor	0.971 ± 0.000(stat.) ± 0.006(syst.)	0.983 ± 0.001(stat.) ± 0.007(syst.)

Table 6.5: Efficiencies and scale factors in MC and data in the barrel and endcap regions for non-DY processes included in the simulations (first block) or subtracted from the data (second block) [118].

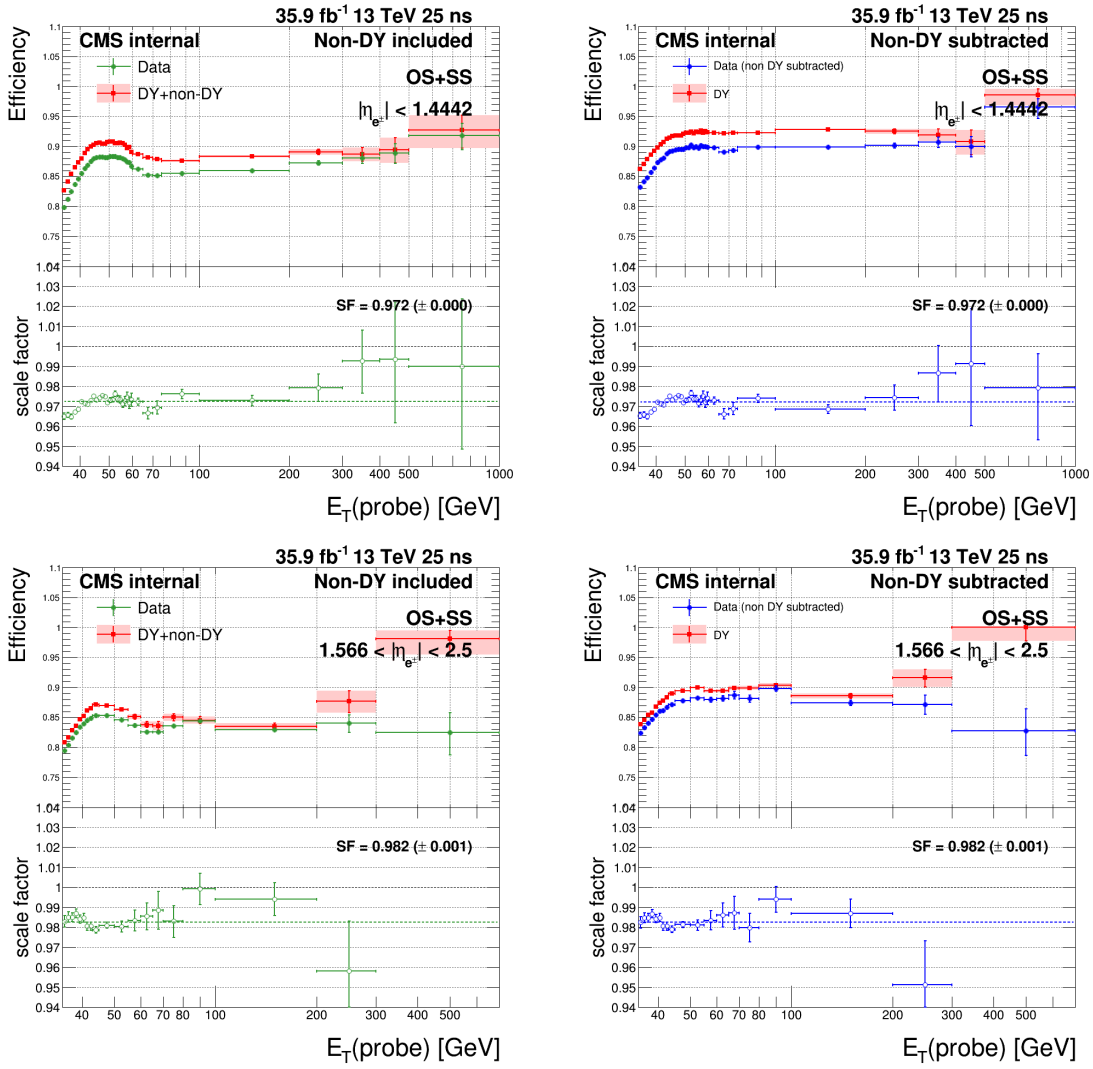

Figure 6.8: Efficiencies and scale factors in MC and data in the barrel (top) and endcap (bottom) where the non-DY processes are included in the simulations (left) or subtracted from the data (right) as functions of E_{T} of the probe [118].

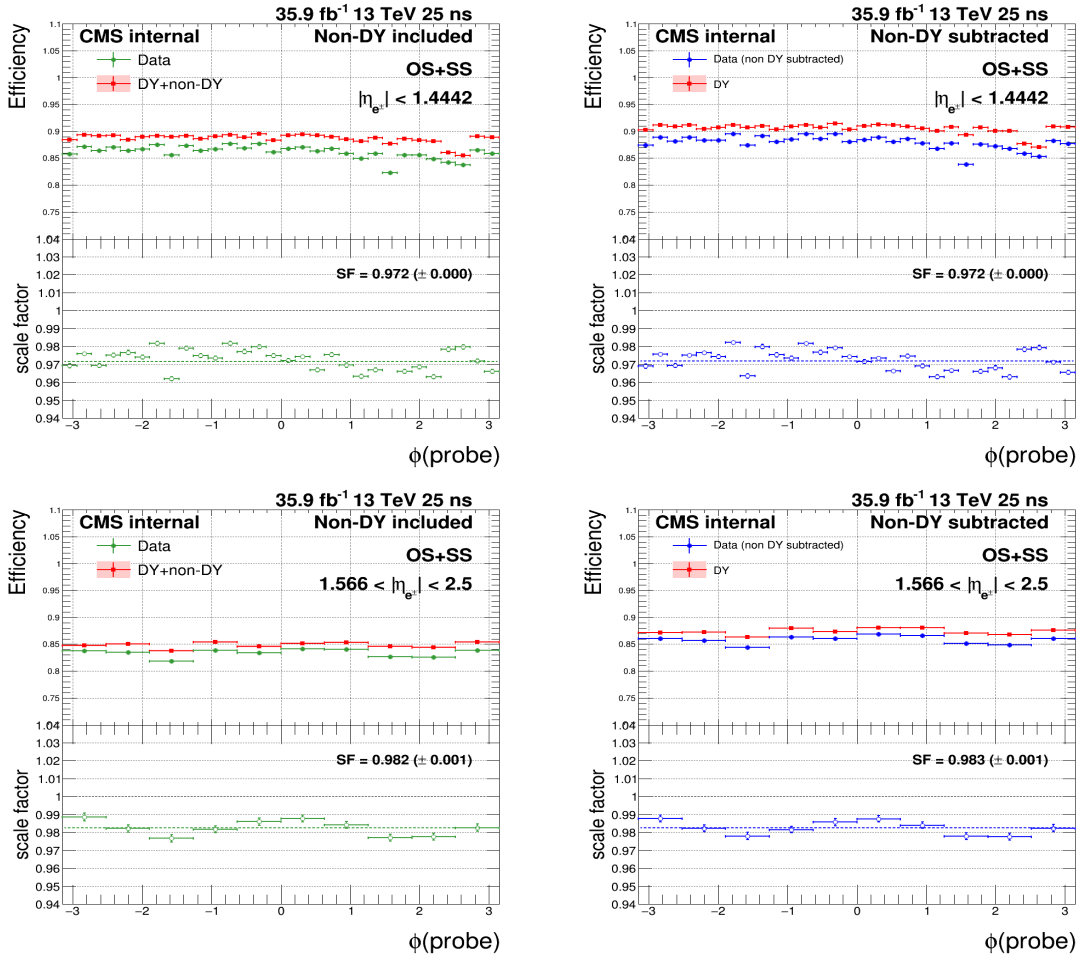

Figure 6.9: Efficiencies and scale factors in MC and data in the barrel (top) and endcap (bottom) where the non-DY processes are included in the simulations (left) or subtracted from the data (right) as functions of the ϕ of the probe [118].

6.4 Mass scale and resolution

The mass resolution function is a crucial point of the analysis, since its outcome enters in the signal model definition (see Section 6.7). Its estimation follows two steps: a data-MC comparison, and a MC-only study. The first step consists in the comparison between the broadness of the invariant mass distribution of the electron pairs m_{ee} selected requiring the HEEP selection at the Z peak ($60 < m_{ee} < 120$ GeV) between data and MC. Both distributions are fitted using a Breit-Wigner (BW) function, whose parameters are fixed to the Particle Data Group (PDG) value of the Z boson, convoluted with a double-sided crystal ball function (dCB), which is defined as a Gaussian core connected with two power-law functions on both sides. The second moment of the gaussian core of the dCB function (σ parameter) is then compared between data and MC in different $|\eta|$-categories. For the "BB category" both electrons are required to be in the ECAL barrel, while for the "BE category" one electron is required to be in the ECAL barrel, the other one in the ECAL endcap. A binned maximum likelihood fit is performed on both data and MC invariant mass distributions. Fit results for data and MC are shown in Figure 6.10.
One also observes that the Z peak is slightly broader in data compared to MC. The σ parameter σ_{MC} of the dCB which fits the MC distribution is then subtracted in quadrature from the σ_{data} coming from the fit to the data, thus defining the σ_{extra} parameter through the relation:

$$\sigma_{extra} = \sqrt{\sigma^2_{data} - \sigma^2_{MC}} \tag{6.3}$$

In Table 6.6 the results for the σ_{data}, σ_{MC} and σ_{extra} parameter are shown for the BB and BE categories. In addition to the σ parameters coming from the fit procedure described above, also an effective σ is quoted (σ^{eff}) in Table 6.7. The σ^{eff} quantity is defined as the minimum window around the mean of the distribution which contains 68% of the total events in the distribution. Note that the numbers in Table 6.6 are expressed in percentage [%] of the Z peak mass value (the PDG value of M_Z is taken as 91.1876 GeV), while the ones drawn on Figure 6.10 are the non-normalized fitted values. Also, the quoted mean values of the mass shift are the ones of the dCB function used to fit the mass spectra after convolution with the BW function.
The consistency of the numbers quoted in Table 6.6 with respect to the latest available official EGM scale and resolution (smearing) corrections from the electron/photon POG (described in next chapter's Section 7.4) has been checked. This was an important cross-check of the two methods, given that the weighted mean values of the EGM corrections are used for the HEEP selection scale factor determination (see Section 6.3.2) and the computation of the DY cross section (see Section 6.5.1), while the results described in this section are used in the signal model of the analysis.

In particular, the official electron/photon mean scale correction per electron candidate for EB (EE) is 1.0012 (1.0089), while the mean smearing correction for EB (EE) is 1.23 % (2.29 %), well in agreement, after computing the dielectron expectations for the analysis categories BB and BE, with the values reported in this section. The way to prove the last statement is the following: starting from the definition of the invariant mass itself $m_{ee} = \sqrt{2E_1E_2(1 - cos(\theta))}$, where E_1 and E_2 are the energies of the first and second electron considered and θ is the angle between the two electrons, one deduces the following relation between the mass scale and the energy scales per single electron:

$$\text{Mass scale} \approx \sqrt{scale_1 \times scale_2} \tag{6.4}$$

Using the mean values of the official EGM scale corrections quoted above (1.0012 for EB, 1.0089 for EE) one finds:

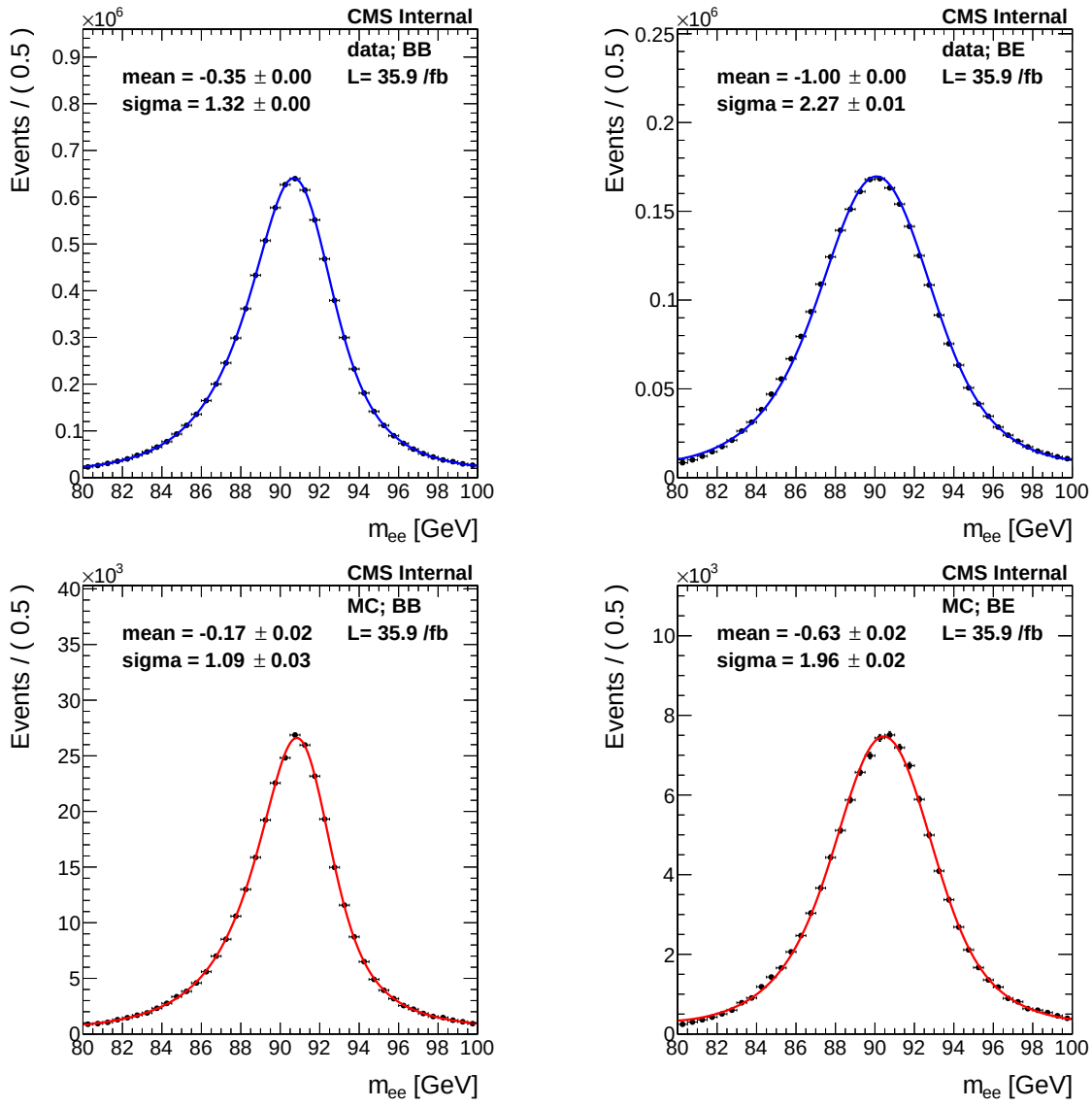

Figure 6.10: Invariant mass distribution at the Z peak in data (top) and MC (bottom) for the BB region (left) and BE (right) channel.

Category	$\frac{\Delta M}{M}$ [%]	σ_{data} [%]	σ_{MC} [%]	σ_{extra} [%]
BB	-0.19 ± 0.02	1.45 ± 0.00	1.20 ± 0.03	0.81 ± 0.04
BE	-0.40 ± 0.02	2.49 ± 0.01	2.15 ± 0.03	1.26 ± 0.05

Table 6.6: Results for the data-MC scale shift $\frac{\Delta M}{M}$, the σ_{data}, σ_{MC} and the σ_{extra} parameters for BB and BE categories relative to $M_Z = 91.19$ GeV.

Category	σ^{eff}_{data} [%]	σ^{eff}_{MC} [%]	σ^{eff}_{extra} [%]
BB	3.72	3.44	1.42
BE	4.63	4.07	2.19

Table 6.7: Results for the σ^{eff}_{data}, σ^{eff}_{MC} and σ^{eff}_{extra} parameters for BB and BE categories.

- Mass scale(BB, EGM) $\approx$ mean scale in EB = -0.12%.
  The last equality holds simply because the EGM corrections are meant to be applied to energy of the electron/photon in data in order to align the Z peak position in

data with respect to the peak position in the MC (see Section 7.4 for futher details). Hence if the mass scale in data has to be multiplied by the factor 1.0012, this means the the corresponding mass scale in the BB category is off by -0.12%.

- Mass scale(BE, EGM) $\approx \sqrt{\text{mean scale in EB} \times \text{mean scale in EE}}$ = -0.51% (following the same reasoning of the previous point).

Once that the mass scale corrections are derived using the expectations from the EGM single electron/photon corrections, one can compare them with the quantity $\frac{\Delta M}{M}$ of Table 6.6 which is defined as the (normalized) difference between the fitted Z peak positions in data and MC. The mass scale computed with the two methods compare as depicted in Table 6.8, showing a closure between the two methods at the level of per mil:

Mass scale [Category]	$\frac{\Delta M}{M}$	EGM expectations
BB	-0.19%	-0.12%
BE	-0.40%	-0.51%

Table 6.8: Comparison between the mass scale parameters computed in this section with respect to the expectations using the official EGM scale corrections.

The comparison between the σ_{extra} parameter derived in this section and the official EGM smearing coefficients (*smear* in the following equations) follows a different reasoning. The EGM smearing coefficients can be thought of as relative uncertainties on the electron/photon energies (see again Section 7.4 for further details):

$$smear = \frac{\Delta E}{E}$$

On the other hand, the σ_{extra} coefficient can be thought of as the relative uncertainty on the mass itself. Hence, propating the error on the invariant mass:

$$\sigma_{extra} = \frac{1}{2}\sqrt{(\frac{\Delta E_1}{E_1})^2 + (\frac{\Delta E_2}{E_2})^2} = \frac{1}{2}\sqrt{smear_1^2 + smear_2^2}$$

Which in turns can be expressed for the analysis categories as:

$$\sigma_{extra}(\text{BB, EGM}) = \frac{1}{2}\sqrt{2 \times (\text{mean smear in EB})^2} = \frac{\text{mean smear in EB}}{\sqrt{2}} = 0.87\%$$

$$\sigma_{extra}(\text{BE, EGM}) = \frac{1}{2}\sqrt{(\text{mean smear in EB})^2 + (\text{mean smear in EE})^2} = 1.29\%$$

At this point, once that the σ_{extra} parameters for BB and BE are derived using the expectations from the EGM single electron/photon smearing corrections, one can compare them with the fitted σ_{extra} quantity quoted in Table 6.6. The two methods compare as depicted in Table 6.9, showing a closure between the two methods at half per mil level:

σ_{extra} [Category]	fitted σ_{extra}	σ_{extra} from EGM expectations
BB	0.81%	0.87%
BE	1.26%	1.29%

Table 6.9: Comparison between the σ_{extra} parameters computed in this section with respect to the expectations using the official EGM smearing corrections for the BB and BE categories.

Both the closure tests for the mass scale and the σ_{extra} parameters give almost optimal agreement between the method described in this section and the expectation from the official EGM scale and smearing corrections. With this, one can conclude that the two approaches are equivalent.

The second step of the study is MC-only based. In particular, the mass resolution has been studied as a function of the generated invariant mass of the electron pair m_{gen}. In order to maximise the statistics in the m_{gen} range between $50 - 4500$ GeV, different DY samples are used (listed in Table 6.2).

For each bin of the generated invariant mass m_{gen}, the distribution of the difference between the reconstructed and the generated invariant mass, divided by the generated invariant mass is considered. The mass resolution, defined as $\frac{m_{reco}-m_{gen}}{m_{gen}}$, is obtained as a function of m_{gen} and a binned maximum-likelihood fit is performed using a "RooCruijff function"[9] (asymmetric Gaussian core, connected with two exponential tails, one on each side). While the definition of the RooCruijff function foresees 2 different σ parameters (one per each side of the Gaussian core), in this study the two σ parameters were constrained to be the same value, for sake of simplicity. The reconstructed invariant mass m_{reco} has been obtained requiring the HEEP selection. The fitted parameters of the RooCruijff functions were then studied as functions of the corresponding generated mass and an analytic parametrisation is provided and used as an input for the limit setting procedure (this will be treated in details in the next section). In particular, for the mass resolution, the σ parameter of the RooCruijff function, labeled as σ_{fit}, is added in quadrature with the σ_{extra} parameter derived in the first step of the procedure. Results for the BB region and BE region are shown in Figure 6.11. In this figure, the mass resolution plots are fitted using a standard "energy resolution function", described by the stochastic, noise and constant terms, to which an extra linear term is added, in the BB case only.

The binning of the x-axis is chosen in order to have a reasonable amount of statistics for each range of the invariant mass and ensure in this way a good quality, in terms of reduced χ^2, of the performed fits.

In the BB region there is a small linear rise in the mass resolution starting around $\approx$ 1 TeV. The effect has been studied in [120] and it is known to be due to leakage of the electromagnetic ECAL shower in the HCAL subdetector which worsen in this way the energy reconstruction driven by the ECAL detector. In fact, the effect of the leakage in the HCAL subdetector is visible as an increase in the H/E variable (which was defined in Section 6.3).

The mass scale, defined as $\frac{m_{reco}}{m_{gen}}$ has also been studied as a function of the generated invariant mass of the electron pair m_{gen}, using the same generated samples taken into account for the mass resolution determination. For each bin of the generated invariant mass m_{gen}, the distribution of the reconstructed invariant mass, divided by the generated invariant mass is analysed. The mean parameter of the RooCruijff function used to fit the resolution distribution is then taken as the mass scale simply adding the unity to it. The error on the parameter is taken accordingly.

Results for the BB region and BE region are shown in Figure 6.12. In the BB region there is a drop in the mass scale parameter starting around $\approx$ 1 TeV ($\approx$ 0.5 % effect), which is the counterpart of the rise observed in the mass resolution and due to the same reason.

Mass resolution parametric model

As described earlier, the quantity $\frac{(m_{reco}-m_{gen})}{m_{gen}}$ is fitted in bins of m_{gen} using a RooCruijff function (RC). The fits results are shown for the different mass points in Figure 6.13 for the BB category and in Figure 6.14 for the BE category. The 8 mass windows corresponds

[9]The choice of the RooCruijff function, instead of the dCB, is due to the fact that its fitted parameters showed a better (more continuous) behaviour as a function of the invariant mass, thus simplifying the determination of the mass resolution parametric model (see next section).

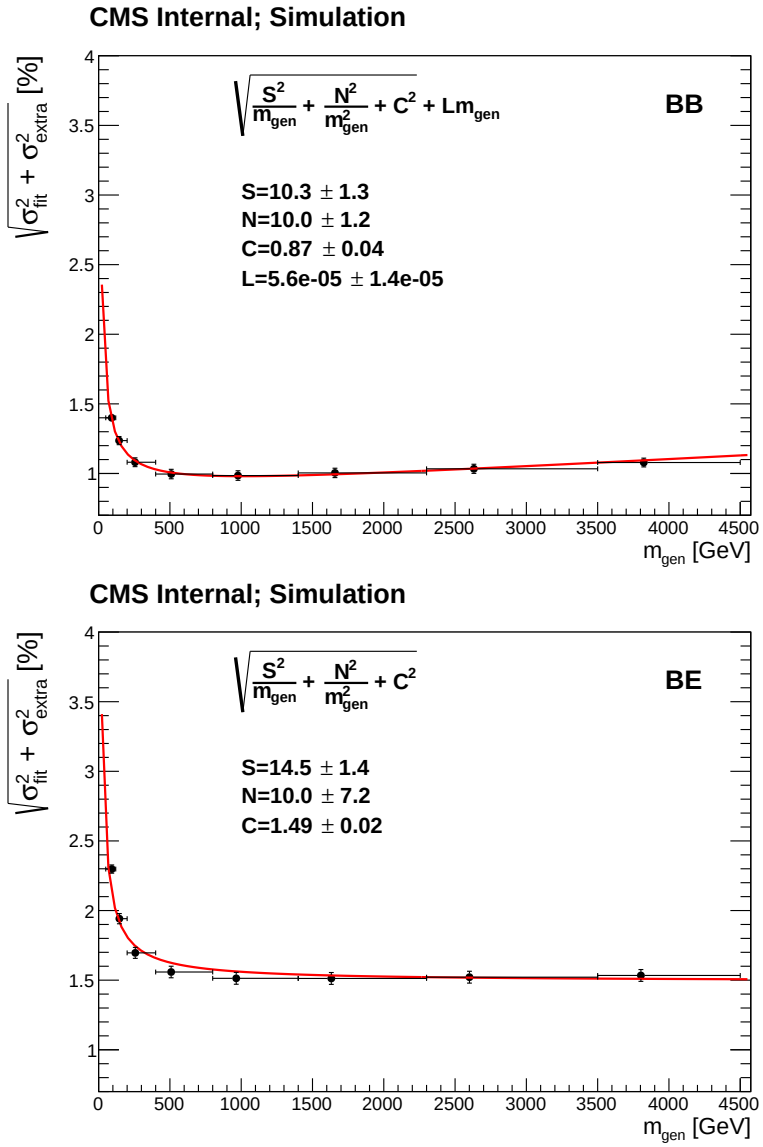

Figure 6.11: Mass resolution as a function of the generated invariant mass for the BB region (top) and BE (bottom) channel.

to those used in Figure 6.12 and are defined by the following bins: [50,120] GeV, [120,200] GeV, [200,400] GeV, [400,800] GeV, [800,1400] GeV, [1400, 2300] GeV, [2300,3500] GeV, [3500,4500] GeV, [4500,6000] GeV, >6000 GeV.

The parameters of the fitted RC functions are then drawn as functions of m_{gen} and a fit is superimposed in order to get an analytic description of their behaviour (see Figures 6.15 and 6.16).

As a consequence, in the limit setting procedure (see Section 6.7), the mass resolution is treated as a RC function whose parameters are described by the analytic functions derived from 6.15 and 6.16.

Finally, as a closure test of the procedure, the $\frac{(m_{reco}-m_{gen})}{m_{gen}}$ histograms are plotted superimposing the RC function that is used for the limit setting procedure, finding of good level of agreement for all the m_{gen} bins. The only bins where the agreement is not optimal are the first two (upper left in Figures 6.17 and 6.18) which correspond to m_{gen} <200 GeV. Given that the search region of the analysis starts above 200 GeV, this is not a reason of concern.

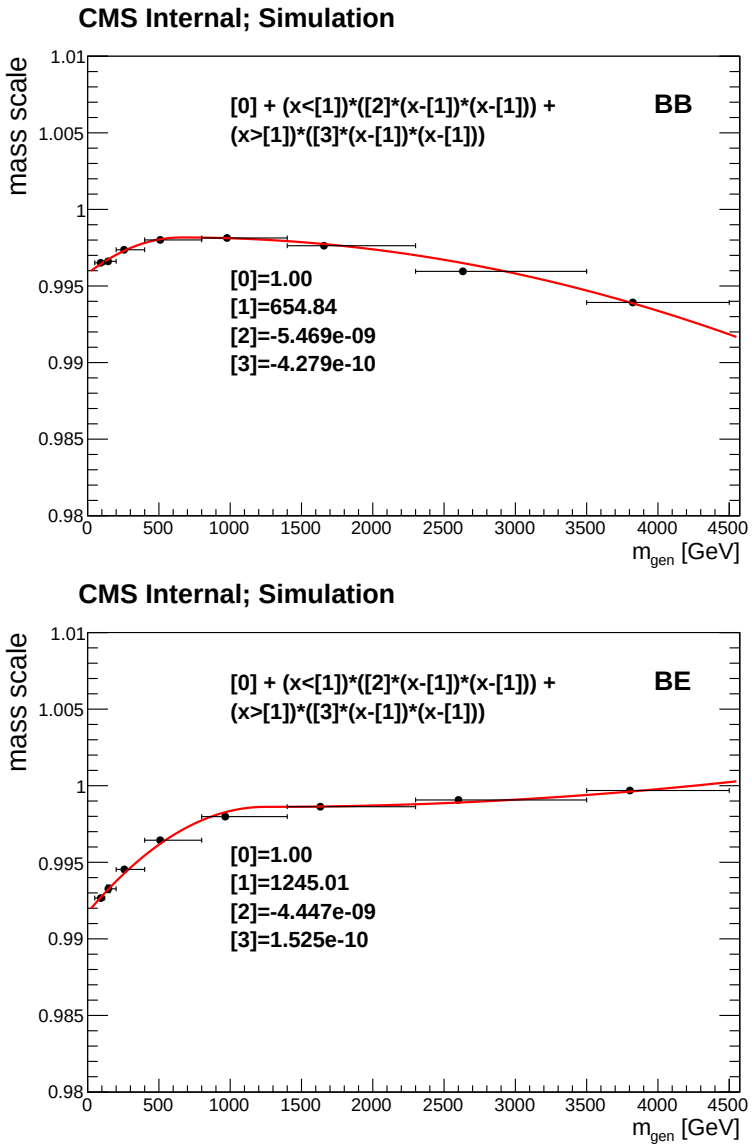

Figure 6.12: Mass scale as a function of the generated invariant mass for the BB region (top) and BE (bottom) channel.

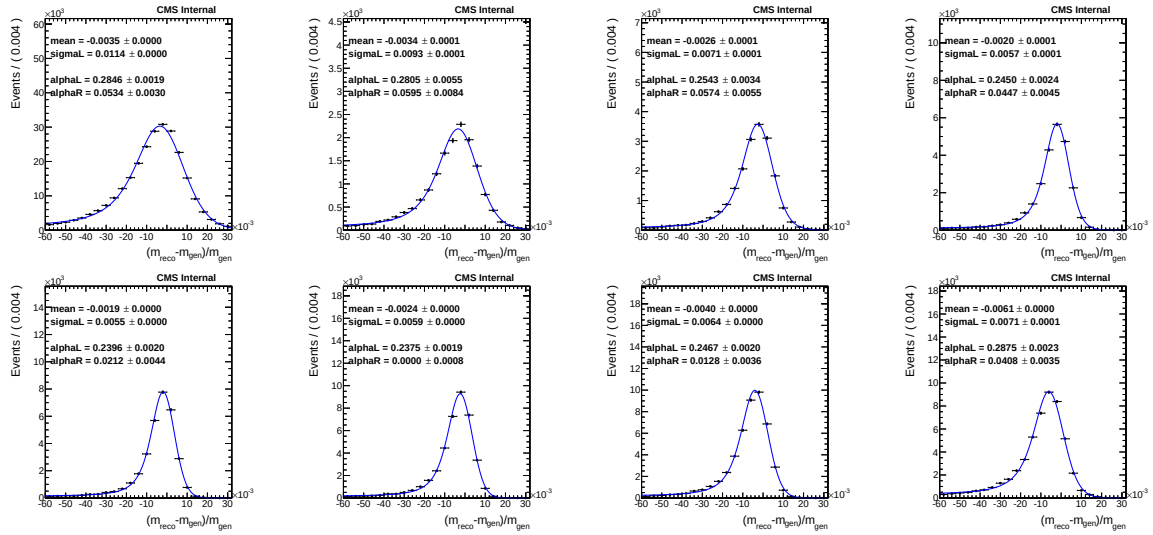

Figure 6.13: Fit results of the $\frac{(m_{reco}-m_{gen})}{m_{gen}}$ histograms in the BB category per m_{gen} bins. The 8 mass windows are [50,120] GeV, [120,200] GeV, [200,400] GeV, [400,800] GeV, [800,1400] GeV, [1400, 2300] GeV, [2300,3500] GeV, [3500,4500] GeV, [4500,6000] GeV and >6000 GeV.

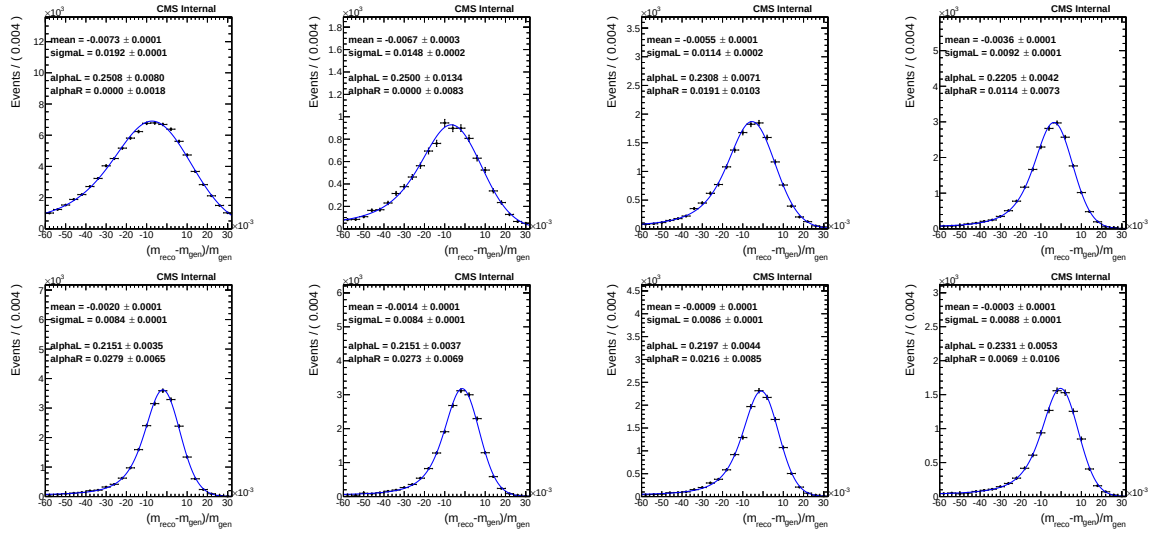

Figure 6.14: Fit results of the $\frac{(m_{reco}-m_{gen})}{m_{gen}}$ histograms in the BE category per m_{gen} bins. The 8 mass windows are [50,120] GeV, [120,200] GeV, [200,400] GeV, [400,800] GeV, [800,1400] GeV, [1400, 2300] GeV, [2300,3500] GeV, [3500,4500] GeV, [4500,6000] GeV and >6000 GeV.

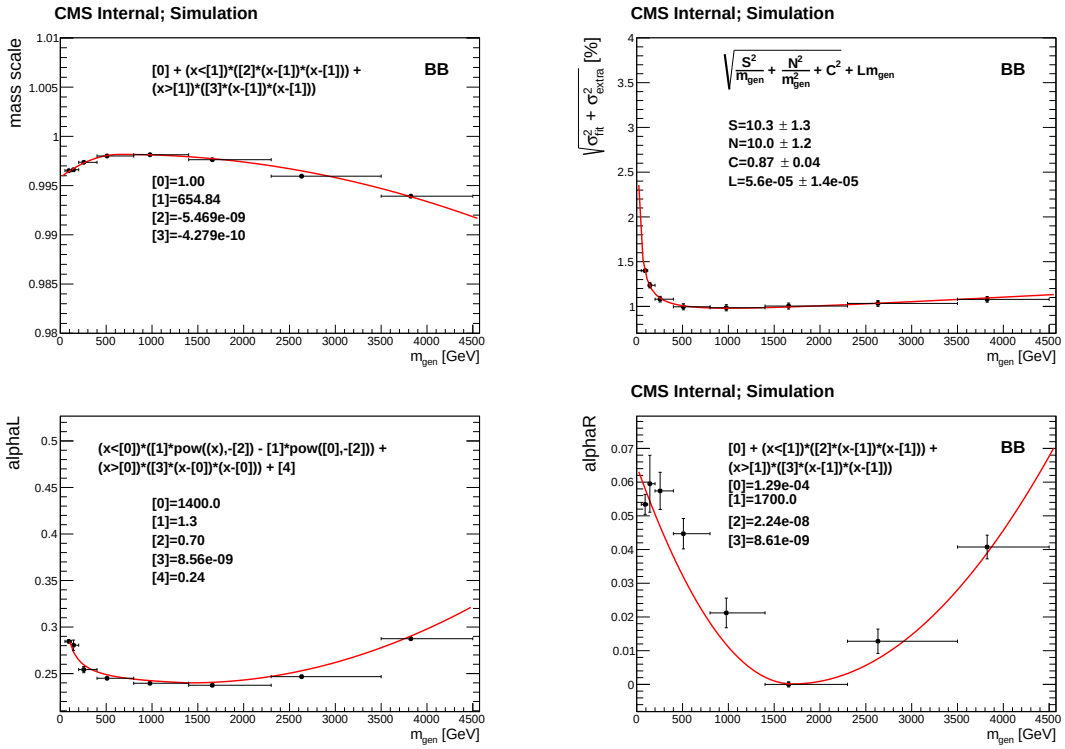

Figure 6.15: Parameters of the RC as a function of m_{gen} for BB category. From top, going clockwise: mean, resolution (with the σ_{extra} term added in quadrature), left and right coefficients of the exponential tails of the RC.

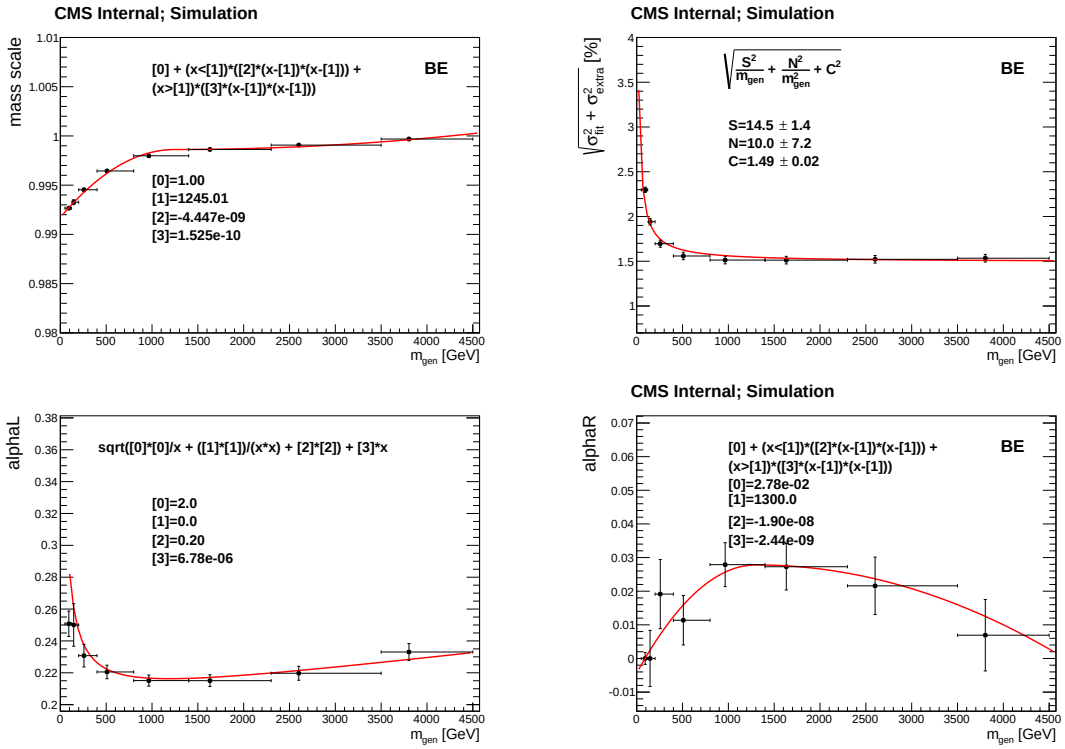

Figure 6.16: Parameters of the RC as a function of m_{gen} for BE category. From top, going clockwise: mean, resolution (with the σ_{extra} term added in quadrature), left and right coefficients of the exponential tails of the RC.

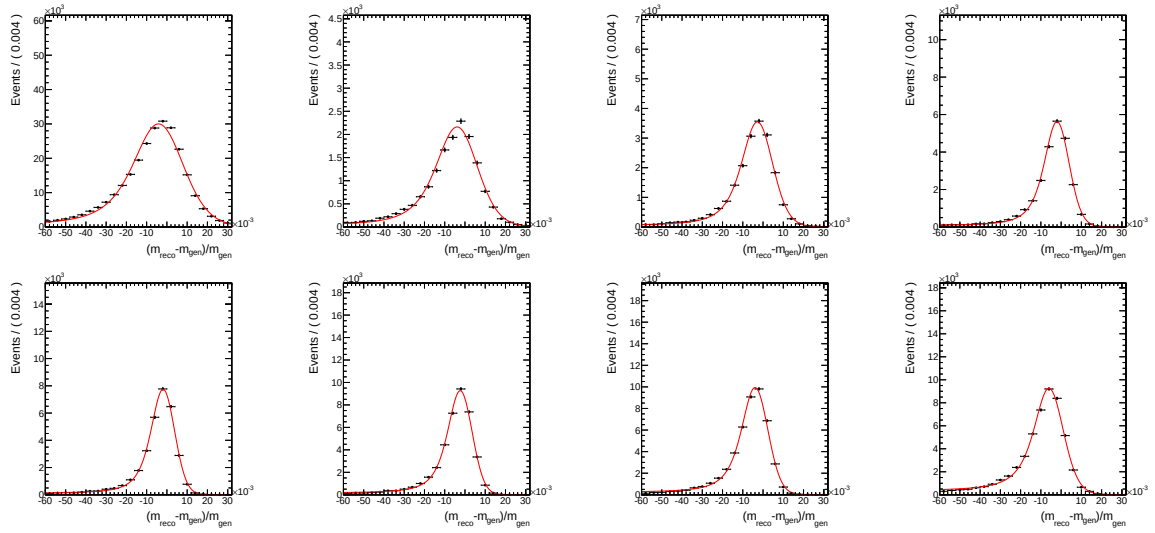

Figure 6.17: Closure test of the RC description in the BB category. The 8 mass windows are the same as in Figure 6.13.

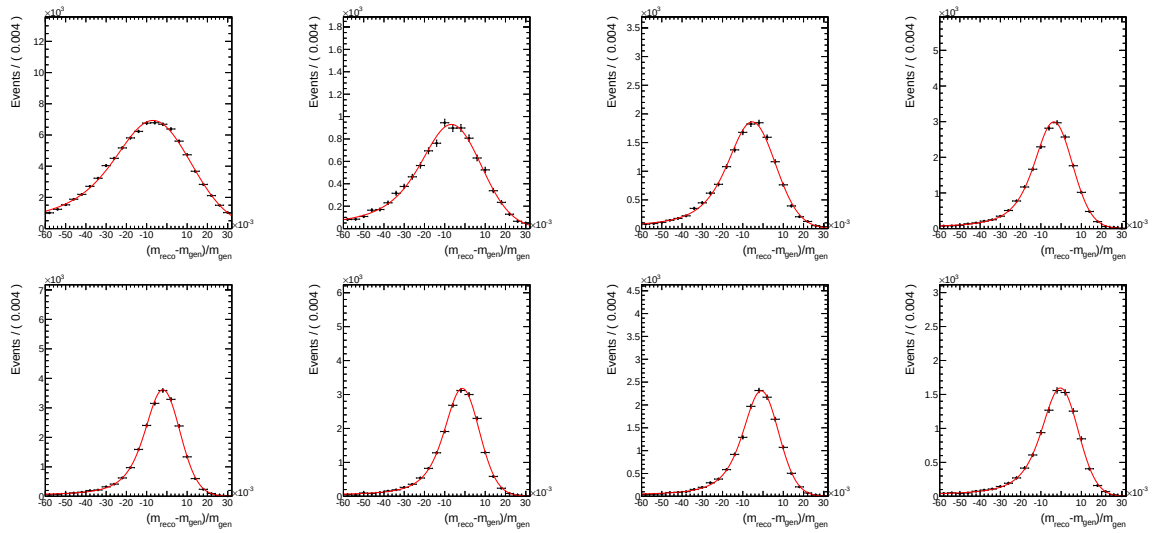

Figure 6.18: Closure test of the RC description in the BE category. The 8 mass windows are the same as in Figure 6.13.

6.5 Background composition

In the search for new physics in the dielectron final state, three main sources of background arise.

- The most significant source of background for the analysis is the irreducible DY process (see Section 6.5.1).
- The second most important type of background comes from real prompt electrons from $t\bar{t}$ processes[10], tW processes, diboson processes (WW, WZ and ZZ), as well as $Z/\gamma^* \to \tau^+\tau^-$ processes. This type of background is referred to as $t\bar{t}$ and $t\bar{t}$-like background as it is dominated by the $t\bar{t}$ process (see Section 6.5.2).
- The third type of background is the jet background, where one or more jets are misidentified as an electron, mainly arising from W+jets and multijets events (see Section 6.5.3).

Irreducible background yields are estimated with simulated samples and normalised to their relative cross-sections using the highest-order calculations available. The sum of these backgrounds are normalised to the observed yield in the dielectron invariant mass region of 60–120 GeV. The $t\bar{t}$ and $t\bar{t}$-like background is estimated as well using MC simulations, cross-checked using the e-μ control region in data. On the contrary, the jet background is estimated with data-driven methods, as described in Section 6.5.3.

6.5.1 Drell-Yan background

The SM Drell-Yan background is estimated using simulated events generated by the POWHEG v2 [100, 101, 102, 103, 104, 105] event generator interfaced with PYTHIA 8.2 [106], as described in Section 6.1. The Monte Carlo samples are normalised to the number of events in data in the Z peak region defined by the dielectron invariant mass region of 60-120 GeV and corrected with the trigger turn on curve, described in Section 6.2. Figure 6.19 shows the dielectron invariant mass distribution in data and MC at the Z peak for the BB and BE categories of the analysis. In this figure, MC samples are weighted for the quoted integrated luminosity of 35.9 fb^{-1}, in order to see the normalisation agreement between data and MC. In addition, electron energy corrections have been applied in data (scale) and MC (smearing corrections), using the (weighted) mean value of the official EGM corrections (see Section 6.3.2), in order to get better data-MC agreement.

As a cross-check of the level of understanding of the DY background, a cross-section measurement including the trigger efficiencies and HEEP efficiency scale factors have been provided and detailed in Table 6.10. The procedure is straightforward: starting from the total number of events in data, the number of expected background events is subtracted using the available MC samples, after taking into account the trigger efficiency scale factors (SF) (described in Section 6.2.2), the HEEP selection SF (described in Section 6.3.2) and the "acceptance×efficiency" of the detector (see Section 6.7). At the Z peak, with respect to the expected theory cross-section [121], scale factors of 0.971 for barrel electrons and 0.982 for endcap electrons are found. Also, a 2% difference between BB and BE categories is observed.

[10] The $t\bar{t}$ process becomes less important at high mass as the boost of the top quark causes the b-jet to enter the electron isolation cone and so fail the isolation requirements of the HEEP selection.

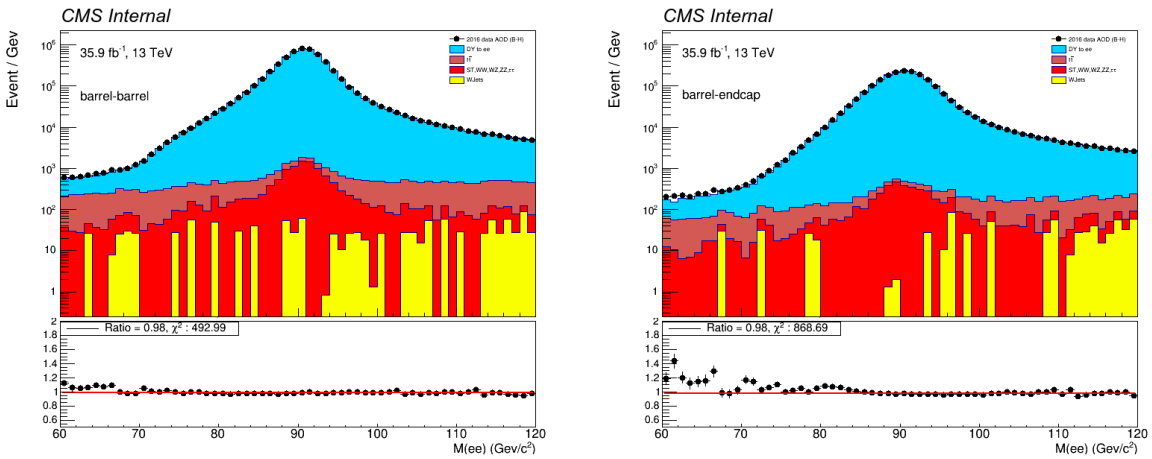

Figure 6.19: Dielectron invariant mass distribution in data and MC at the Z peak region, for the BB (left) and BE (right) categories. The MC samples are normalised for the integrated luminosity [118].

Variable	BB	BE
Data events	5760345±2400	2051759±1432
Background events	32805	11336
MC acceptance×efficiency	0.0880±0.001	0.0315±0.001
Trigger SF	0.979±0.001	0.985±0.001
HEEP SF	0.943±0.001	0.953±0.001
Luminosity	35867 pb^{-1}	35867 pb^{-1}
cross-section	1967 ± 3 (stat) ± 51 (lumi) pb	1922 ± 3 (stat) ± 50 (lumi) pb

Table 6.10: The cross-section measurement in the dielectron invariant mass range of 60-120 GeV. The HEEP efficiency scale factors are taken from Table 6.5. Note that in Table 6.5 the scale factors are for single electrons while here they are propagated for electron pairs. The luminosity is taken following the latest recommentations from the luminosity group.

DY background uncertainty

The main uncertainties on the Drell-Yan background originate from the PDF uncertainties and higher-order effects. For the different mass bins, the DY cross-sections have been evaluated using the FEWZ 3.1.b2 [108] program with NNLO accuracy in QCD and NLO for the electro-weak component and evaluated inside the analysis acceptance (E_{T}>35 GeV, $|\eta| < 2.5$, excluding the $|\eta|$ transition region of $1.4442 - 1.566$). The uncertainties on the ratios of the cross-sections with respect to the Z peak cross-section were estimated by taking into account possible correlations of the PDF uncertainties between the various mass bins [122]. The PDF uncertainties relative to the Z peak as a function of mass are quoted in Table 6.11.

The ratio of the FEWZ cross-sections to the ones predicted by the POWHEG samples used in the analysis is shown in Figure 6.20. It is clear that the POWHEG prediction is increasingly higher than the FEWZ prediction as the mass increases, hence a "FEWZ to POWHEG" k-factor has been applied to improve data-MC agreement at high mass. The functional form for the k-factor is shown in Figure 6.20.

6.5.2 $t\bar{t}$ and $t\bar{t}$-like background

The background from electrons arising from non-singularly produced W and Z bosons is estimated directly from Monte Carlo simulations (see Section 6.1 for further details on

Mass range (GeV)	PDF relative uncertainty
200-300	1.21%
400-500	1.54%
900-1000	2.16%
1400-1500	2.73%
1900-2000	3.24%
2400-2500	3.72%
2900-3000	4.27%
3400-3500	5.00%
3900-4000	5.94%
4400-4500	7.47%
4900-5000	10.2%
5400-5500	14.3%
5900-6000	19.9%

Table 6.11: The PDF uncertainties relative to the Z peak region as a function of mass [122].

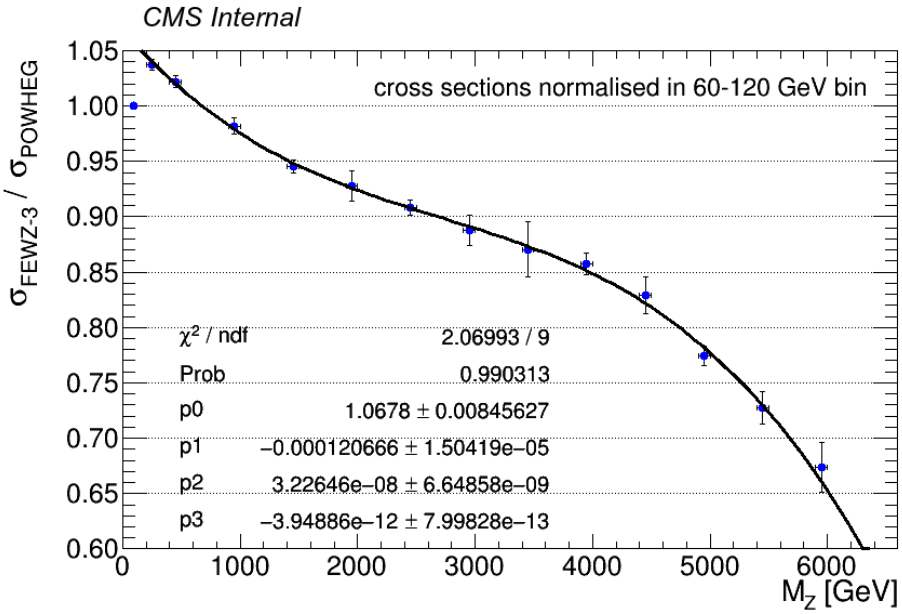

Figure 6.20: The ratio of the $Z/\gamma^* \to e^+e^-$ cross-sections as predicted by FEWZ 3.1.b2 and the ones predicted by POWHEG in various mass bins [118].

the simulated samples). Given that the leptonic decay branching ratio of the W boson is universal in the lepton flavour, all these processes, where two W bosons decays are involved, have a decay branching ratio to a pair of leptons of different flavour eμ twice as large as the branching ratio to electron pairs ee. For this reason, the eμ invariant mass spectrum can be exploited as a control region to validate the Monte Carlo predictions for these backgrounds. The validation technique of the $t\bar{t}$ and $t\bar{t}$-like background, using the eμ invariant mass spectrum, is described below, where in order to evalute the multijet background contribution two different methods have been employed: the "same sign" method and the "fake rate" method. The latter is described in details in Section 6.5.3, while here is used as an input for an alternative validation of the $t\bar{t}$ and $t\bar{t}$-like background samples.

eμ mass spectrum: same sign method

This study is derived using the SingleMuon dataset, where the electron-muon events are selected requiring that the first object is a global muon (see Section 5.2) passing a dedicated selection, designed for high E_{T} muons [123], and the second object is an electron passing the HEEP selection. Electrons are required to have $E_{\mathrm{T}}>$ 35 GeV while muons are required to have $E_{\mathrm{T}}>$ 53 GeV. Since high energetic muons can undergo bremsstrahlung processes, an associated supercluster in the ECAL (in the direction of the muon's inner track) could be present. Hence the selected muons could lead to fake electron candidates. Therefore, an electron veto is applied such that if there is a global muon with $E_{\mathrm{T}}>$ 5 GeV within a cone $\Delta R < 0.1$ built around the electron candidate direction, that electron is rejected. In order to validate the $t\bar{t}$ and $t\bar{t}$-like background samples the jet background component (QCD and W+jets) has to be estimated first. To accomplish this task the "same sign" $e\mu$ invariant mass spectrum is considered, where the selected electron and muon are requested to have the same electric charge[11]. Starting from the same sign eμ mass spectrum in data, the contributions of the SM processes, excluding of course the jet background itself[12], are subtracted from the data eμ mass spectrum. The spectrum resulting in the MC subtraction is taken as the jet background estimate. In the case that the data yield in one bin of the same sign invariant mass histogram is already smaller than the combined yield of all SM processes, the associated yield for the jet background is set to zero.
At this point, after having estimated the jet background directly from data, the distribution of all eμ pairs (both same sign and opposite sign) can be inspected, using as jet background estimate twice the estimate coming from the same sign eμ mass spectrum[13].

The final invariant mass spectrum of all eμ events (both same sign and opposite sign together) in data and MC is shown in Figure 6.21, where the jet background is computed as described above. The data-MC agreement is quite good up to 1 TeV of invariant mass.

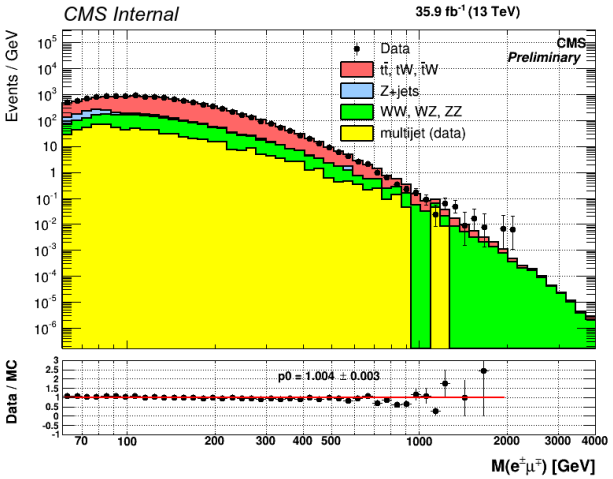

Figure 6.21: Invariant mass spectra of the eμ events (both same sign and opposite sign pairs are included) in data and MC with 35.9 fb^{-1} of integrated luminosity [118]. The jet background is estimated here from the invariant mass spectrum of the same sign eμ pairs.

[11] In alternative, the "fake rate" method could be also used to accomplish the same task. See next section.
[12] The MC simulations for the jet background are not reliable, given that the small misreconstruction rate for the jets translates in a low amount of statistics and huge associated uncertainties.
[13] This is because for the jet background the spectrum for the same sign should be the same as the opposite sign eμ pair.

eμ mass spectrum: fake rate method

An alternative validation of the $t\bar{t}$ and $t\bar{t}$-like background can be provided by using again the eμ mass spectrum but estimating the jet background using the "fake rate" method (described in Section 6.5.3) and used here simply as an input for the study.

In particular, a control region is defined, in which the muon is selected as described in the previous section while the electron is required to fail the HEEP selection but pass the pre-selection criteria detailed in Table 6.12 and match the first leg of the trigger HLT_DoubleEle33_WPTight. The events in the control region are then weighted with fake rate estimate applying the functional form detailed in Table 6.13 as a function of the electron E_{T} and $|\eta|$. Data in this control region is enriched with QCD and W+jets events. Therefore, the jet background itself is estimated subtracting all the other background components from the data eμ mass spectrum using the MC simulations.

In Figure 6.22 (top), the eμ mass distribution is shown in the control region used to determine the jet background, while in the bottom part, the final eμ mass distribution is shown in the signal region, after including the jet background estimated from the control region. Also in this case, the data-MC agreement is generally good in the eμ mass spectrum, hence the $t\bar{t}$ and $t\bar{t}$-like background sample are validated.

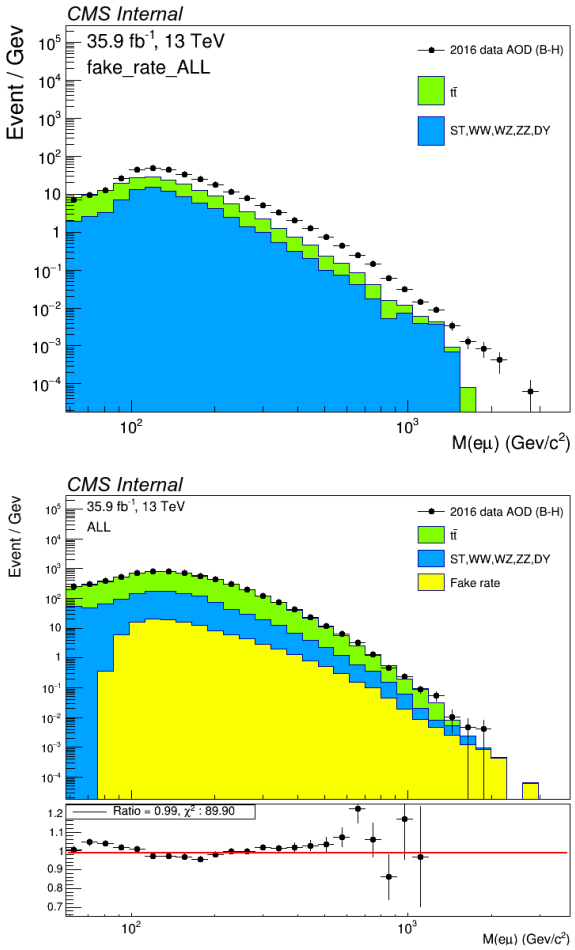

Figure 6.22: The eμ distribution in the control region (top), used to estimate the jet background. The bottom plot presents the eμ distribution in the signal region, where the jet background is taken from the estimation coming from the control region in the top plot [118].

6.5.3 Jet background

The jet background is the smallest of the significant SM backgrounds considered in this search. The primary components of the jet background are: di-jet events, where both jets have passed the HEEP selection, W+jets events, where the W decays to an $e\nu$ pair and at least a jet passes the HEEP selection and γ+jets events, where a photon and a jet pass the HEEP selection. The jet background is estimated using the "fake rate" method [120, 124] which will be described in the following sections.

Fake rate measurement

The fake rate (FR) is defined as the probability for a jet misreconstructed as an electron to pass the HEEP selection and is measured using events in data which passed the OR of the following triggers: HLT_Photon30_CaloIdVL, HLT_Photon50_CaloIdVL, HLT_Photon75_CaloIdVL and HLT_Photon90_CaloIdVL. These triggers ask for a localised deposit in ECAL compatible with a photon having $E_{\mathrm{T}}>$ 30,50,75 and 90 GeV respectively, having also passed very loose shower shape calorimeter requirements (CaloIdVL). The fake rate is estimated using a jet-enriched sample, built requiring that the electron candidates pass a dedicated loose pre-selection defined in Table 6.12 (they will referred to as "loose electrons" in the following). In addition, in the sample there can be only one ECAL driven reconstructed electron with $E_{\mathrm{T}}>$ 10 GeV and H/E$<$ 0.15 in the event, in order to reduce the contamination from $Z/\gamma^* \to e^+e^-$ events. The FR has been measured in bins of E_{T} and $|\eta|$ and computed using a jet template (whose accuracy is tested in the next section) normalised to the observed N-1 track isolation distribution[14] in the range of $10 < Isopt < 20$ GeV, where any real electron contamination is expected to be very small[15]. In order to measure the FR, the loose electrons (which passed the pre-selection) are required to pass the H/E and calorimeter isolation cuts of the HEEP selection, but fail at least one other cut of the HEEP (where the $Isopt$ cut has been removed, in order to build the track isolation template). The distribution of the N-1 track isolation distribution is then fitted with the jet template and the FR is defined as the jet template prediction in the signal region with $Isopt <$5 GeV, normalized to the background-only region $10 < Isopt < 20$ GeV (see next section for further details).

Variable	Barrel	Endcap
$\sigma_{i\eta i\eta}$	$<$0.013	$<$0.034
H/E	$<$0.15	$<$0.10
Missing hits	$<= 1$	$<= 1$
|dxy|	$<$ 0.02 cm	$<$ 0.05 cm

Table 6.12: The pre-selection requirements used for the fake rate calculation.

The fake rate measurement showed a relatively flat behaviour across the barrel while increasing with $|\eta|$ in the endcaps. Therefore, while the barrel was not splitted in $|\eta|$, the endcaps were splitted into two $|\eta|$ bins. The measured FR estimates are shown in Figure 6.23, together with arbitray chosen fitted functions to allow the FR estimate to be easier applied at analysis level. The fit parameters are summarised in Table 6.13. A 50% uncertainty on the method is assumed to cover deviations from the arbitrarily chosen fits.

Before moving to the estimation of the jet background contribution, the next section

[14] The N-1 track isolation distribution is simply the track isolation distribution $Isopt$ for electron candidates which pass the HEEP selection with the track isolation cut removed. The $Isopt$ variable is defined in Section 6.3.

[15] As a reminder for the reader: $Isopt <$ 5 GeV is the track isolation requirement of the HEEP selection.

Region	E_T range (GeV)	Functional form
Barrel	$35.0 \leq E_T < 131.6$ GeV	$0.106 - 0.0025 \times E_T + 2.26e-05 \times E_T^2 - 7.11e-08 \times E_T^3$
	$131.6 \leq E_T < 359.3$ GeV	$0.0139 - 0.000104 \times E_T + 3.6e-07 \times E_T^2 - 4.13e-10 \times E_T^3$
	$E_T \geq 359.3$ GeV	$0.00264 + 3.38e-06 \times E_T$
Barrel	$35.0 \leq E_T < 131.5$ GeV	$0.106 - 0.00252 \times E_T + 2.28 \times 10^{-5} \times E_T^2 - 7.21 \times 10^{-8} \times E_T^3$
	$131.5 \leq E_T < 355.0$ GeV	$0.0138 - 0.000103 \times E_T + 3.62 \times 10^{-7} \times E_T^2 - 4.25 \times 10^{-10} \times E_T^3$
	$E_T \geq 355.0$ GeV	$0.00279 + 2.43 \times 10^{-6} \times E_T$
Endcap $\lvert\eta\rvert < 2.0$	$35.0 \leq E_T < 122.0$ GeV	$0.117 - 0.0013 \times E_T + 4.67 \times 10^{-6} \times E_T^2$
	$122.0 \leq E_T < 226.3$ GeV	$0.0345 - 4.76 \times 10^{-5} \times E_T$
	$E_T \geq 226.3$ GeV	$0.0258 - 9.09 \times 10^{-6} \times E_T$
Endcap $\lvert\eta\rvert > 2.0$	$35.0 \leq E_T < 112.5$ GeV	$0.0809 - 0.000343 \times E_T$
	$E_T \geq 112.5$ GeV	0.0423

Table 6.13: The functional form approximation of the measured fake rate for HEEP electrons in the barrel and endcap regions, as a function of E_T.

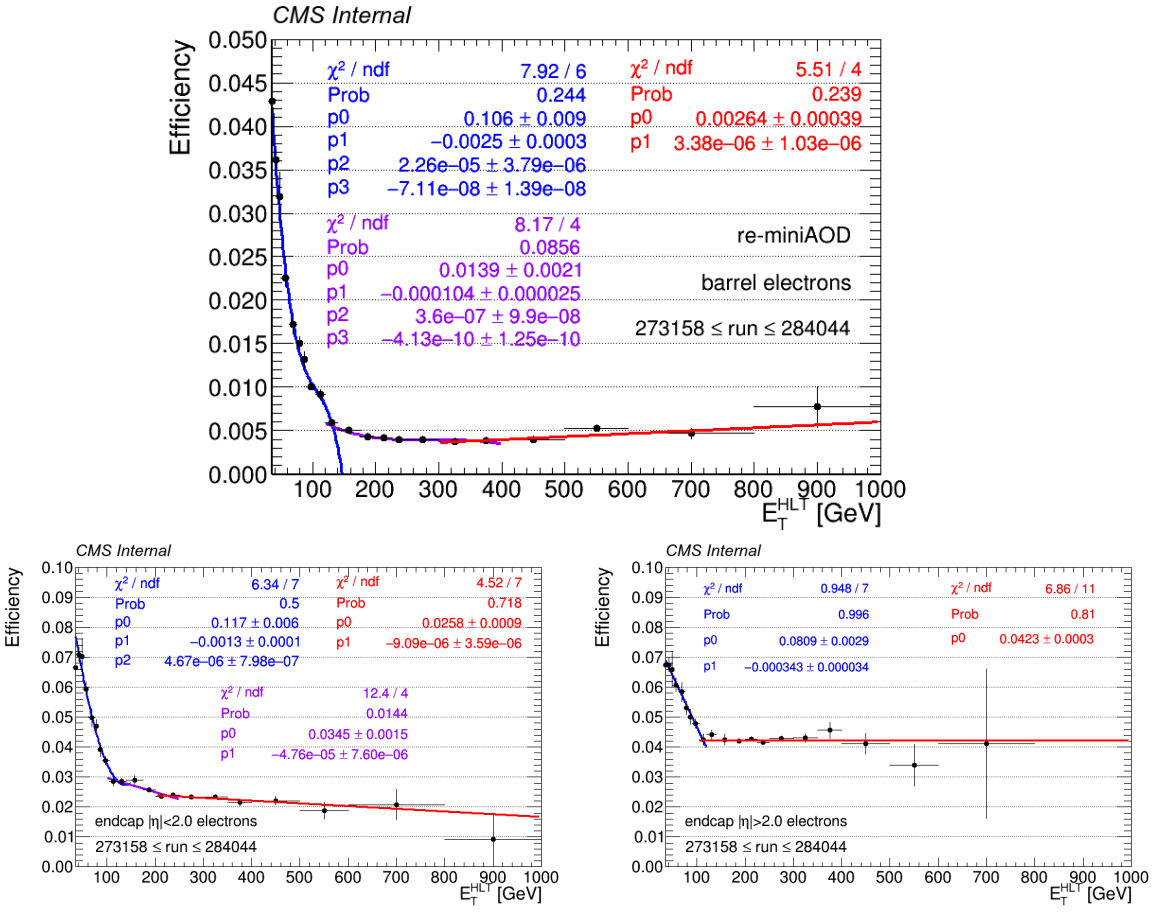

Figure 6.23: The measured FR as a function of E_T for the barrel region (top), the endcap low $|\eta|$ region (bottom left) and the endcap high $|\eta|$ region (bottom right) [118].

describes how the accuracy of the jet template used to fit the N-1 track isolation distribution is accessed.

Validation of the jet template

In order to scrutinise the accuracy of the jet template, an electron template is generated from the Drell-Yan Monte Carlo sample, defined as the N-1 track isolation distribution in the E_T, $|\eta|$ appropriate bin. At this point the *Isopt* distributions in data are fitted using

both templates[16]. An example of the tracker isolation distribution is shown in Figure 6.24 for the three $|\eta|$ regions (barrel and the two $|\eta|$ regions defined for the endcaps). As can be seen from Figure 6.24, there is excellent separation between the signal and background regions. Overall, there is an acceptable agreement between the data and the fit coming from the two templates together. Below 20 GeV, the maximum deviations are in the 10-20% range and there are no clear trends in the agreement across the different E_{T}, $|\eta|$ bins. However, as Figure 6.24 also shows, in higher E_{T} bins, the jet template falls slower than the data, leading to an over prediction for values of tracker isolation greater then 20 GeV. This effect is smaller in the endcaps. It implies that there is some small correlation between the tracker isolation variable and the other HEEP cuts, with jets having passed the other cuts being less likely to have a very high value of tracker isolation. However, the jet template is not required to model high values of tracker isolation correctly but the low values correctly, specifically the region from 0 to 5 GeV, and given the target accuracy of the method of 50%, this is the case. Hence, the jet template is validated.

Estimation of the jet background contribution

The jet background is estimated by selecting electron pairs passing the primary analysis trigger with one electron passing the HEEP selection and one electron passing the FR pre-selection defined in Table 6.12 but failing the HEEP selection. This is referred to as the "1FR" estimate. Since the loose electron candidate must fail the HEEP selection, the events must be weighted by a factor $\mathrm{FR}_{\mathrm{fail}}/(1-\mathrm{FR}_{\mathrm{fail}})$ where $\mathrm{FR}_{\mathrm{fail}}$ is the E_{T} and $|\eta|$ appropriate fake rate defined for the electron failing the HEEP selection. In the case of more than one electron pair in the same event satisfying these conditions, all valid pairs are allowed to enter the estimation. The residual contamination of $Z/\gamma^* \to e^+e^-$ events is corrected for, by subtracting the MC estimate.

The 1FR estimate includes the background estimation for W+jets, γ+jets and di-jets events but due to combinatorial effects, the 1FR estimate overestimates the di-jet contribution by a factor 2. To correct for this effect, a sample of events where both electron candidates pass the loose pre-selection only is derived and the FR is applied for both loose electrons. In this case, the events have to be weighted by a factor $\mathrm{FR}_1/(1-\mathrm{FR}_1) \times \mathrm{FR}_2/(1-\mathrm{FR}_2)$ where FR_1 (FR_2) is the E_{T} and $|\eta|$ appropriate fake rate for the first (second) electron, obtaining what is referred to as the "2FR" estimate. The 2FR contribution is finally subtracted from the 1FR estimate to evaluate the total jet background without any double counting.

Due to fake rate measurement uncertainties and statistical effects, the 2FR estimate can sometimes be greater than half the 1FR estimate. This implies that the entirety of the 1FR estimate is from di-jets and therefore the true estimate of the di-jet background is simply 50% of its value. Therefore the 1FR estimate can only be reduced to a minimum of 50% of its uncorrected value.

The fake rate estimates for the two electrons are taken to be 100% correlated. As the fake rate has an assigned 50% uncertainty, the uncertainty on the estimated jet background is taken to be 100%.

[16]Note that since there are no signal events in the region $10 < Isopt < 20$ GeV, in any fit with electron and jet templates, the jet template only, in order to give a reasonable estimate of the true background distribution, must be forced to match the data points in the background-only region $10 < Isopt < 20$ GeV. Therefore the electron template cannot contribute to the FR estimate and has in fact not been used for the purpose, but only as cross-check of the accuracy of the jet template itself.

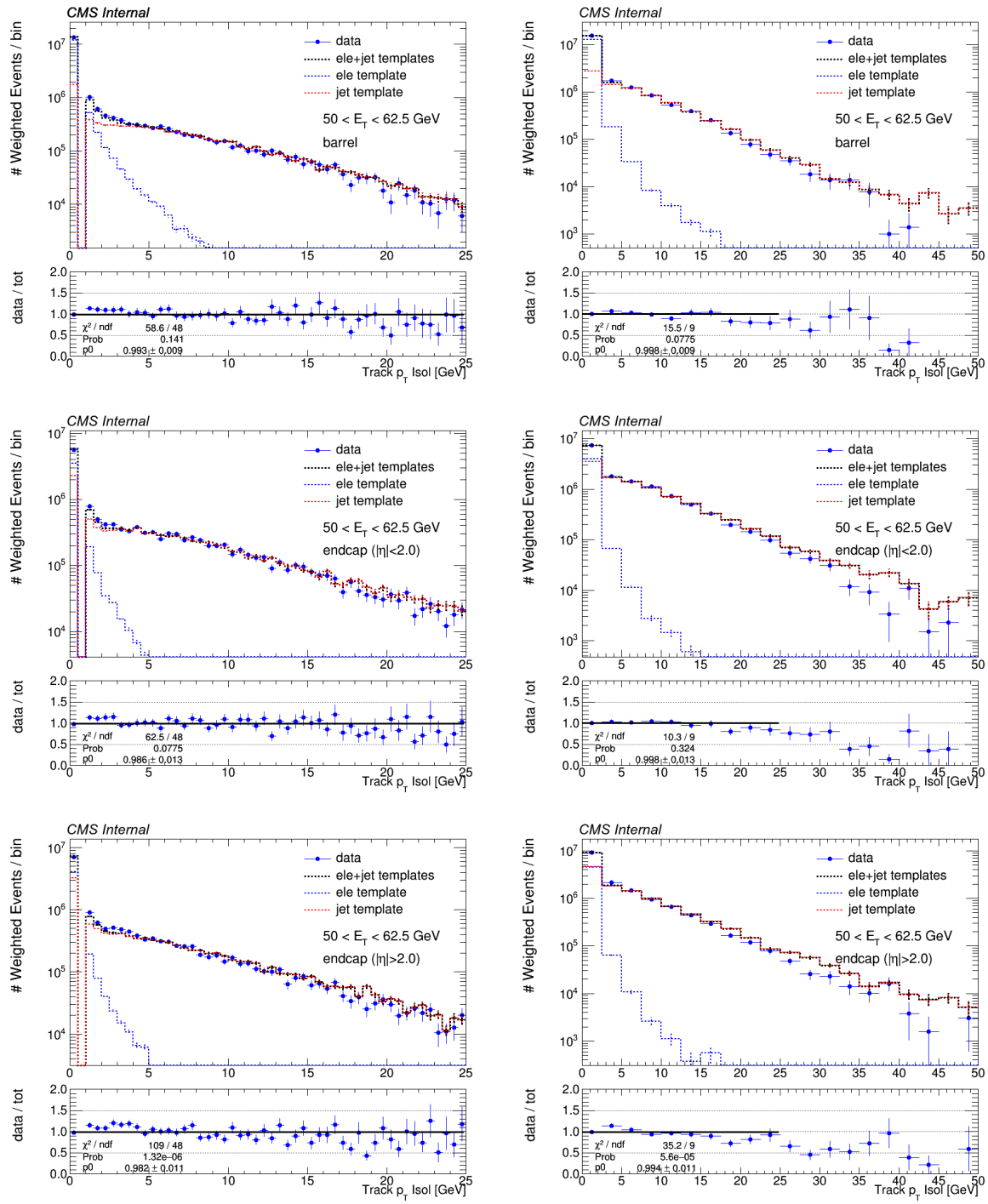

Figure 6.24: Example fits obtained using the electron and jet templates to the observed N-1 track isolation when binned in 0.5 GeV bins (left) and 2.5 GeV bins (right) for electron candidates passing all but the track isolation cut in barrel (top), inner endcap (middle) and outer endcap (bottom) [118].

6.6 Invariant mass spectra

The observed invariant mass spectra for dielectron events, selected as described in Section 6.3, are shown in Figure 6.25, where the binning of the x-axis varies according to the mass resolution, for barrel-barrel events, barrel-endcap events and all selected events. The SM background estimations are obtained using both MC simulations and data-driven methods

as described in Section 6.5.

Figure 6.26 shows the cumulated versions of the invariant mass spectra, with each bin containing the number of events with an invariant mass greater or equal than the mass of that bin.

Figure 6.27 shows the data minus background estimation over the background-only estimation in the signal region for the barrel-barrel categories, barrel-endcap categories and sum of barrel-barrel and barrel-endcap categories together. It can be seen that the distributions show no significant deviations from zero, hence a striking signature of new physics is not observed in data.

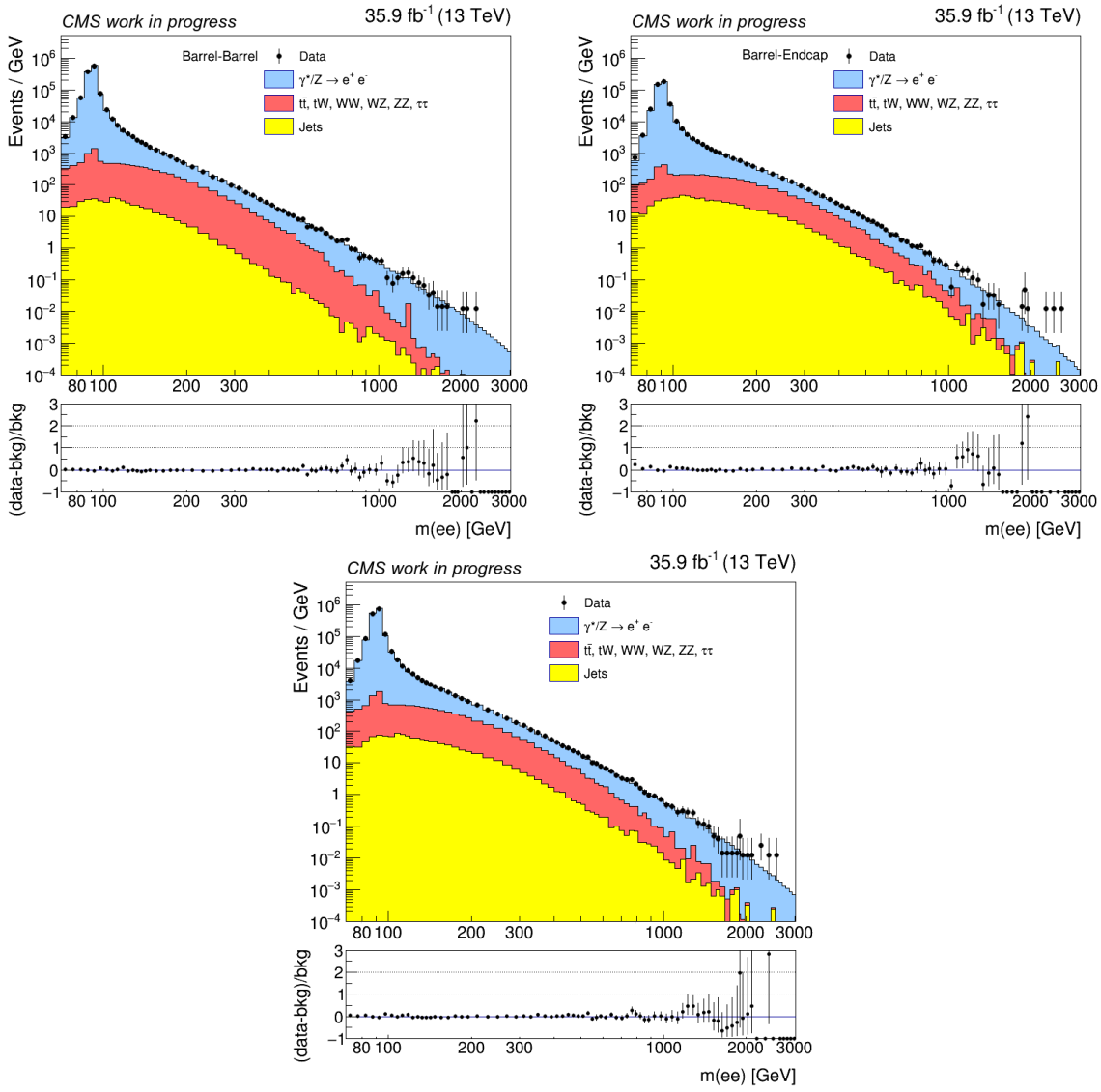

Figure 6.25: The observed dielectron mass spectrum for the barrel-barrel category (top left), the barrel-endcap category (top right) and for all selected events (bottom) together with the predicted standard model backgrounds [118].

In Table 6.14, the predicted SM backgrounds, normalized to data at the Z peak, and observed data yields are shown as a function of dielectron invariant mass.

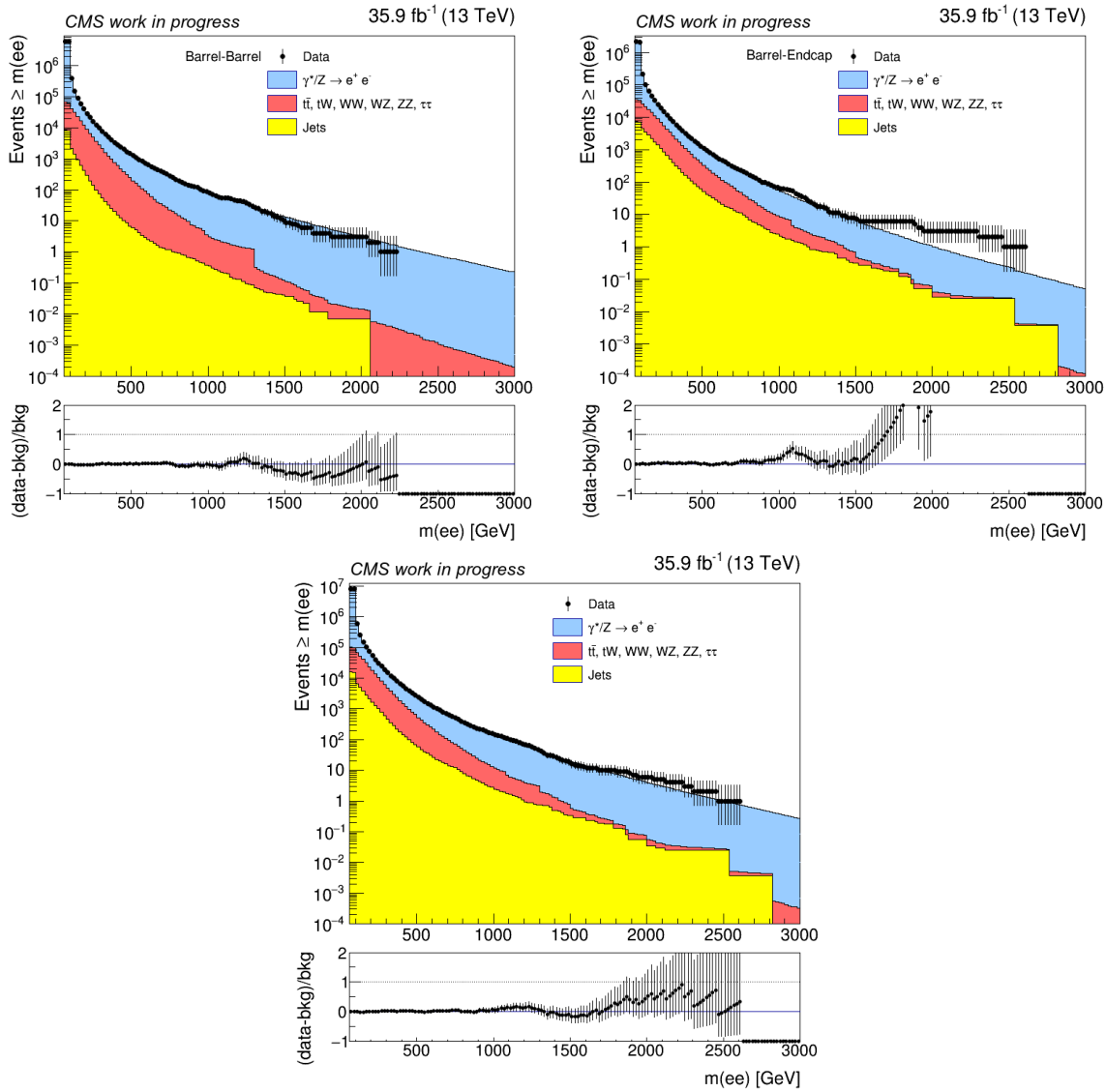

Figure 6.26: The integral of the measured dielectron mass spectrum for the barrel-barrel category (top left), the barrel-endcap category (top right) and for all selected events (bottom) together with the predicted standard model backgrounds [118].

Mass range (GeV)	Data	Total background	$\gamma^*/Z \rightarrow ee$	$t\bar{t}$ and $t\bar{t}$-like	Jets
60-120	7812104	7813715 ± 13568	7768204 ± 13566	41176 ± 175	4334
120-400	245101	250408 ± 1144	198687 ± 1136	46941 ± 130	4779
400-600	4297	4192 ± 35	2914 ± 30.49	1159 ± 18	118
600-900	943	934.16 ± 9.9	740 ± 7.31	170 ± 6.5	22.5
900-1300	182	175.45 ± 2.4	155 ± 0.96	15.89 ± 2.19	3.70
1300-1800	33	32.60 ± 0.41	30.8 ± 0.26	1.15 ± 0.28	0.64
1800-5000	9	7.23 ± 0.08	7.05 ± 0.05	0.05 ± 0.00	0.13
60-5000	8062669	8069466.56 ± 13616.33	7970740.72 ± 13614.34	89465.75 ± 219.88	9259.72

Table 6.14: Predicted SM background and observed data yields as a function of the dielectron invariant mass. The uncertainties of the total background predictions include only the statistical component.

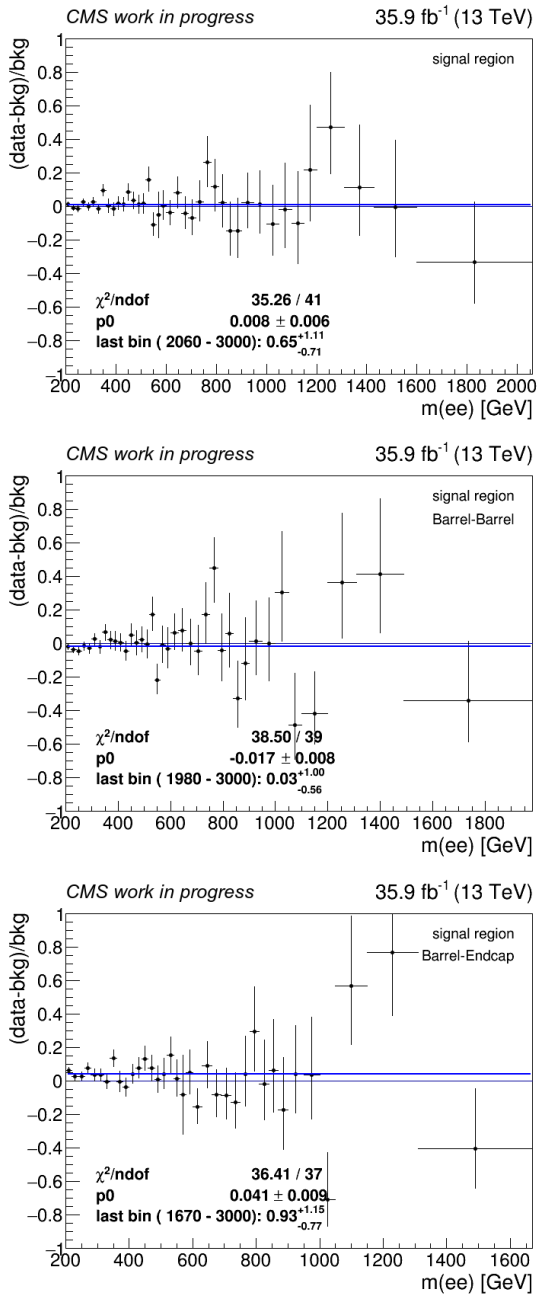

Figure 6.27: Data minus background over background distribution for all selected events (top), events in the barrel-barrel category (middle) and events in the barrel-endcap category (bottom) [118].

6.7 Statistical interpretation

Given that no excesses over the SM expectation are seen in the invariant mass spectra, upper limits on the ratio of cross sections of a new resonance to the Z resonance are computed.

Methodology

The statistical treatment of the results follows a Bayesian method with an unbinned extended likelihood function [125, 126, 127].

The probability density function (pdf) is modeled as the sum of a resonant signal pdf and a steeply falling background pdf as follows:

$$f(m|\boldsymbol{\theta},\boldsymbol{\nu}) = q_1 \cdot f_{\mathrm{S}}(m|\boldsymbol{\theta},\boldsymbol{\nu}) + (1-q_1)\cdot f_{\mathrm{B}}(m|\boldsymbol{\theta},\boldsymbol{\nu}) \tag{6.5}$$

where m is the observable under study (specifically the dilectron invariant mass), $\boldsymbol{\theta}$ is the vector of parameters of interest and $\boldsymbol{\nu}$ the vector of nuisance parameters. The probability of a signal event is given by q_1.

The signal pdf f_S is modeled as the convolution of a Breit-Wigner function, describing the natural shape of the resonance, with the resolution function $\mathrm{Res}(m|\sigma,\boldsymbol{\theta})$:

$$f_{\mathrm{S}}(m|\Gamma,\sigma,\boldsymbol{\theta}) = \mathrm{BW}(m|\Gamma) \otimes \mathrm{Res}(m|\sigma,\boldsymbol{\theta}) \tag{6.6}$$

where Γ is the intrinsic width of the signal and σ is the mass resolution. Possible other parameters of the resolution function are left as implicit to avoid a heavy notation. As described in Section 6.4 the resolution function $\mathrm{Res}(m|\sigma,\boldsymbol{\theta})$ is described by a RooCruijff function.

The background pdf has instead an ad-hoc shape derivation computed using simulated background events. An analytic function is used to describe the background shape in the search region above 200 GeV of dielectron invariant mass. The function that was found to fit the shape of the combined background coming from all SM processes summed together can be expressed by:

$$f_B(m \mid a,b,c,d,k) = Ae^{a+b\times m+c\times m^2+d\times m^3}m^k \tag{6.7}$$

where A is a normalization parameter. The rest of the parameters is determined by a fit to the simulated dielectron mass spectrum performed in the mass range 200-5500 GeV. The background spectra together with the fitted functional forms are shown in Figure 6.28 for the BB and BE categories.

The parameter of interest used for this analysis is R_σ, defined as the ratio between the cross section times branching ratio (BR) to electron pairs of a generic new resonance normalized to the same quantity for the Z resonance in the mass region 60-120 GeV:

$$R_\sigma = \frac{\sigma_{\mathrm{Z}'}\cdot \mathrm{BR}(\mathrm{Z}' \to \mathrm{ee})}{\sigma_{\mathrm{Z}}\cdot \mathrm{BR}(\mathrm{Z}\to \mathrm{ee})} \tag{6.8}$$

This choice has the important advantage that certain uncertainties, e.g. the uncertainty on the integrated luminosity or any other E_{T}-independent effect, cancel out or are at least greatly reduced. The R_σ quantity can be connected to the signal event yield via the following relation:

$$\mu_{\mathrm{S}} = R_\sigma \frac{(\mathrm{Acc}\times\epsilon)_{\mathrm{Z}'}}{(\mathrm{Acc}\times\epsilon)_{\mathrm{Z}}} N_{\mathrm{Z}} \tag{6.9}$$

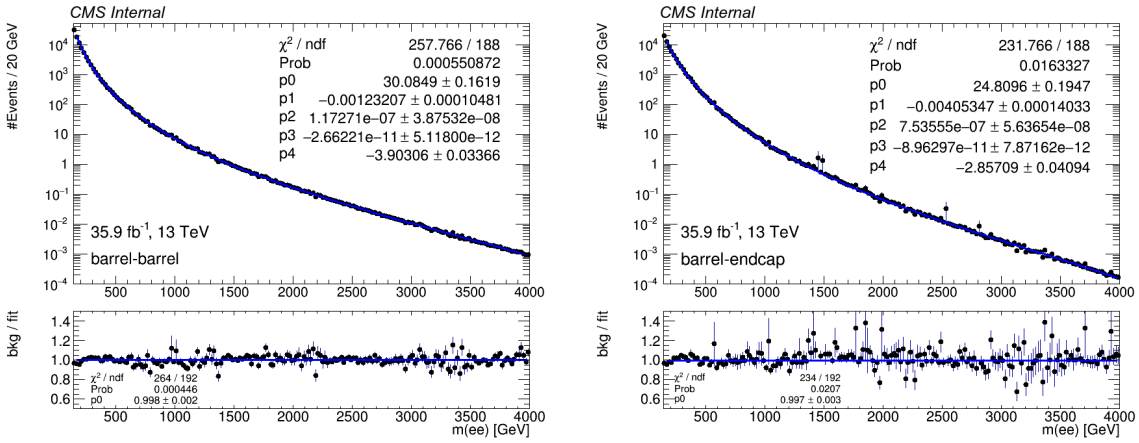

Figure 6.28: The total SM background together with the fitted functional forms used for the limit setting tool for the barrel-barrel (left) and barrel-endcap (right) channels.

where $(\mathrm{Acc}\times\epsilon)_{\mathrm{Z}'}$ and $(\mathrm{Acc}\times\epsilon)_{\mathrm{Z}}$ are the acceptance times efficiency of the Z$'$ and the Z boson respectively and N_{Z} is the number of selected Z events, defined in the mass region 60-120 GeV.

The unbinned likelihood is defined as:

$$\mathcal{L}\left(\boldsymbol{m}|R_\sigma,\boldsymbol{\nu}\right) \;=\; \prod_{i=1}^{N} f\left(m_i|R_\sigma,\boldsymbol{\nu}\right) \tag{6.10}$$

where the product is over the events in the dataset and $\boldsymbol{m}$ is the vector of corresponding dielectron masses. Proceeding in extending the likelihood with a poissonian normalization component in front of Equation (6.10) and inserting the equation for the signal and background pdfs detailed above, one obtains:

$$\mathcal{L}(\boldsymbol{m}|R_\sigma,\boldsymbol{\nu}) = \frac{\mu^N e^{-\mu}}{N!}\cdot\prod_{i=1}^{N}\left(\frac{\mu_{\mathrm{S}}(R_\sigma,\boldsymbol{\nu})}{\mu}f_{\mathrm{S}}(m_i|R_\sigma,\boldsymbol{\nu}) + \frac{\mu_{\mathrm{B}}(R_\sigma,\boldsymbol{\nu})}{\mu}f_{\mathrm{B}}(m_i|R_\sigma,\boldsymbol{\nu})\right) \tag{6.11}$$

with μ_{S} and μ_{B} being the signal and background yields and μ the sum of the two yields.

The uncertainties on the nuisance parameters in the vector $\boldsymbol{\nu}$ are taken into account by modeling the nuisance parameter as

$$\nu = \hat{\nu}\cdot(1+\delta\nu)^{\beta} \tag{6.12}$$

where $\hat{\nu}$ is the estimate of ν, $\delta\nu$ is the corresponding systematic uncertainty and β is a random number drawn from a gaussian distribution with mean value at zero and second order moment equal to 1 (denoted as $Gauss(\beta|0,1)$). The likelihood is then weighted by $Gauss(\beta|0,1)$ for each nuisance parameter giving

$$\mathcal{L}(\boldsymbol{m}|R_\sigma,\boldsymbol{\nu}) = \prod_j(\mathcal{L}(\boldsymbol{m}|R_\sigma,\hat{\nu_j}\cdot(1+\delta\nu_j)^{\beta_j})\cdot Gauss(\beta_j|0,1) \tag{6.13}$$

where the product is done over the nuisance parameters. The two categories of the analysis are independent categories, hence the total likelihood can be obtained by multiplying the two separated likelihoods.

With this definition of the likelihood function, 95% confidence level (CL) upper limits can be computed using the Bayes theorem, which states:

$$f(R_\sigma, \boldsymbol{\nu}|\boldsymbol{m}) \cdot p(\boldsymbol{m}) = \mathcal{L}(\boldsymbol{m}|R_\sigma, \boldsymbol{\nu}) \cdot p(R_\sigma, \boldsymbol{\nu}) \tag{6.14}$$

where $p(R_\sigma, \boldsymbol{\nu})$ is the prior pdf for the parameter of interest of the model. In this analysis, the pior is taken as a log-normal distribution for the uncertainties, and a uniform (positive) prior for the parameter of interest.
After integrating over the nuisance parameters $\boldsymbol{\nu}$, one obtains

$$p(R_\sigma|\boldsymbol{m}) \cdot p(\boldsymbol{m}) = \mathcal{L}(\boldsymbol{m}|R_\sigma) \cdot p(R_\sigma) \tag{6.15}$$

The expression for the posterior pdf immediately follows as

$$p(R_\sigma|\boldsymbol{m}) = \frac{\mathcal{L}(\boldsymbol{m}|R_\sigma) \cdot p(R_\sigma)}{p(\boldsymbol{m})} = \frac{\mathcal{L}(\boldsymbol{m}|R_\sigma) \cdot p(R_\sigma)}{\int \mathcal{L}(\boldsymbol{m}|R_\sigma) \cdot p(R_\sigma)\, \mathrm{d}R_\sigma} \tag{6.16}$$

Given the posterior pdf for the parameter of interest R_σ, the 95% C.L. upper limit R_σ^{95} is defined by the following constraint:

$$\int_0^{R_\sigma^{95}} p(R_\sigma|\boldsymbol{m})\, \mathrm{d}R_\sigma = 0.95 \tag{6.17}$$

where the integration is done using the Metropolis-Hasting algorithm [128, 129].

Finally, it is worth noting (see Equation (6.9)) that the limit setting procedure described above requires also the parametrization of the acceptance times efficiency (Acc$\times\epsilon$). This parametrization is shown in Figure 6.29. For convenience, all the inputs required for the limit setting tools are listed in Table 6.15, as well as the number of data events and acceptance at the Z peak region.

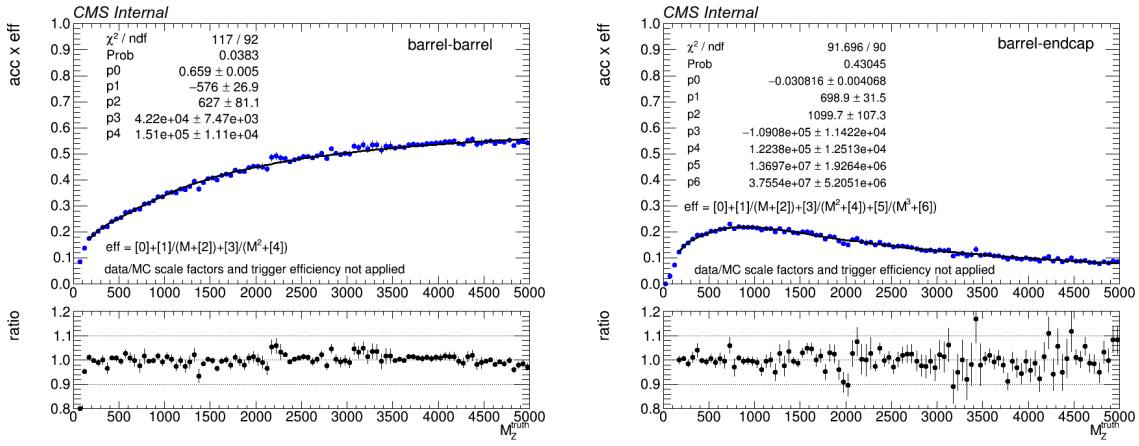

Figure 6.29: The acceptance times efficiency for a spin-1 particle to be selected by the analysis in the barrel-barrel region (left) and barrel-endcap region (right) together with the fitted functional forms used in the limit setting tools [118].

Expected limits

Expected upper limits on R_σ under the background-only hypothesis are obtained computing the median of a set of limits derived using an ensemble of randomly drawn pseudo-data. The limits for the pseudo-data are estimated using the same procedure as described for the

Variable	BB	BE
$(\text{Acc} \times \epsilon)_{\text{Z}}$	0.0895	0.0318
Error on $(\text{Acc} \times \epsilon)_{\text{Z}'}/(\text{Acc} \times \epsilon)_{\text{Z}}$	6%	8%
N_{data} (Mass region>200 GeV)	28154	25054
N_{Z}	5739714	2036781

Table 6.15: Input parameters for the upper limit computation.

observed limits. The pseudo-data are generated by drawing the event yield as a random number from a Poissonian distribution whose mean is

$$\mu_B = \hat{\mu}_B \cdot (1 + \delta\mu_B)^{\beta_B} \tag{6.18}$$

where β_B is again a random number extracted from a normal distribution as the case of Equation (6.12). The value of $\hat{\mu}_B$ is estimated by integrating the background shape over the observable range, where the shape is normalized over a sideband in the data below 200 GeV of dielectron invariant mass.

Repeating this procedure many times, the distribution of the expected limits under the background-only hypothesis is built and so the median and the $\pm1\sigma$ and $\pm2\sigma$ bands can be computed.

6.7.1 Systematic uncertainties

As already mentioned above, the results are expressed in terms of upper limits on the parameter R_σ, defined as a ratio of cross sections, hence all systematic uncertainties which are E_{T}-independent cancel out and can be ignored. The main uncertainties are mainly due to the MC modeling of the electron efficiency at high E_{T} as well as NLO and PDF effects on the Drell-Yan background. The PDF uncertainties on the DY background, already described in Section 6.5.1, are summarised in Table 6.11 and range from 1.54% at 400 GeV to 19.9% at 6 TeV.

In addition, a 3% (5%) uncertainty is assigned in barrel (endcaps) for the data/MC HEEP selection scale factor at high E_{T}. The jet background uncertainty is evaluated to be 50% and the uncertainty on the non DY background is taken to be 7% based on cross section uncertainties. A 2% (1%) energy scale uncertainty is assigned in barrel (endcaps)[17].

The total background uncertainty ranges from approximately 3% at 200 GeV to 12% at 5 TeV. The background uncertainty is mainly driven, at low mass, by normalization uncertainties while, at high mass, the PDF and NLO corrections are the dominating ones.

6.7.2 Upper limits

Using the methodology described above, 95% CL upper limits are computed for the R_σ parameter. The limits are calculated in a mass window of ±6 times the signal width, where the window is symmetrically enlarged until it contains at least 100 events.

The expected and observed upper limits for a resonance width of 0.6% is shown in Figure 6.30. The signal R_σ curves are shown on the plot in order to obtain a mass limit on two specific Z$'$ signal model. The limits are obtained by dividing the LO cross section of the specific model calculated using the PYTHIA 8.2 program with the NNPDF2.3 PDFs for the Z boson cross section. As the limits presented here are set on the on-shell cross section and the PYTHIA event generator includes off-shell effects, the cross section

[17] See also the discussion about the "slew rate issue" of the ECAL electronics in Section 5.1.2.

is calculated in a mass window of $\pm 5\% \cdot \sqrt{s}$ centered on the resonance mass [130]. The validity of this procedure for the Z'_{SSM} and Z'_{ψ} models was explicitly checked [130] and is found to be accurate up to approximately 5-7% level. Finally to account for NLO effects, the obtained signal LO cross sections are multiplied by a k-factor of 1.3.

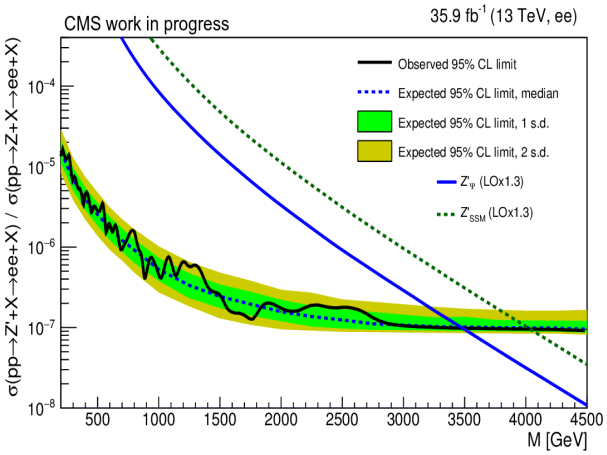

Figure 6.30: The 95% CL upper limits on R_σ for a spin 1 resonance with a width equal to 0.6% of the resonance mass. The shaded bands correspond to the 68% and 95% quantiles for the expected limits. Theoretical predictions for the spin 1 Z'_{SSM} and Z'_{ψ} resonances are also shown [118].

In parallel to the search for new resonances in the dielectron final state, a similar search was performed using the dimuon final state [123]. The expected and observed upper limits for a Z' resonance with natural width of 0.6% and decaying in the dimuon final state is shown in Figure 6.31.

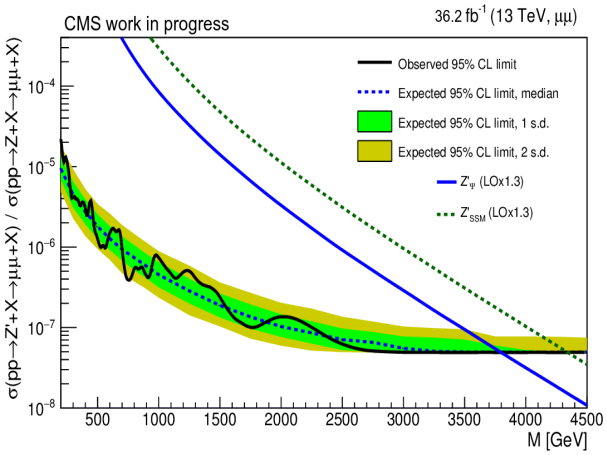

Figure 6.31: The 95% CL upper limits on R_σ for a spin 1 resonance with a width equal to 0.6% of the resonance mass for the dimuon final state. The shaded bands correspond to the 68% and 95% quantiles for the expected limits. Theoretical predictions for the spin 1 Z'_{SSM} and Z'_{ψ} resonances are also shown [118].

Given that the sensitivity of the two searches is comparable and, under the assumption that the BR to dielectron and dimuon final state is the same, the upper limits coming from the two separate searches can be combined.
Figure 6.32 shows the upper limits for the combination of the two analysis with dielectron and dimuon final states.
Moreover, results for widths equal to 0.6%, 3% and 5% of the resonance mass are shown in Figure 6.33.

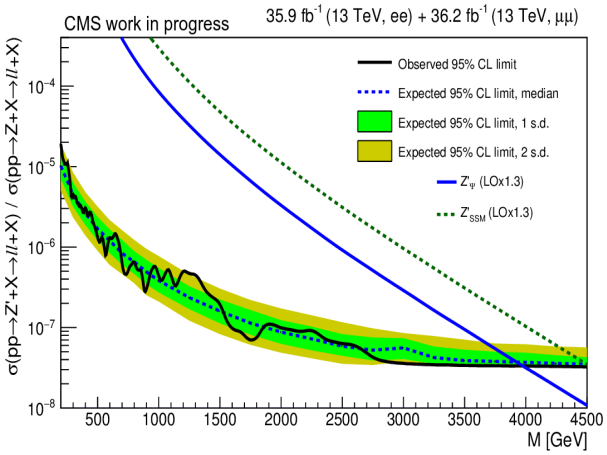

Figure 6.32: The 95% CL upper limits on R_σ for a spin 1 resonance with a width equal to 0.6% of the resonance mass for the dielectron and dimuon final states combined. The shaded bands correspond to the 68% and 95% quantiles for the expected limits. Theoretical predictions for the spin 1 Z'_{SSM} and Z'_{ψ} resonances are also shown [118].

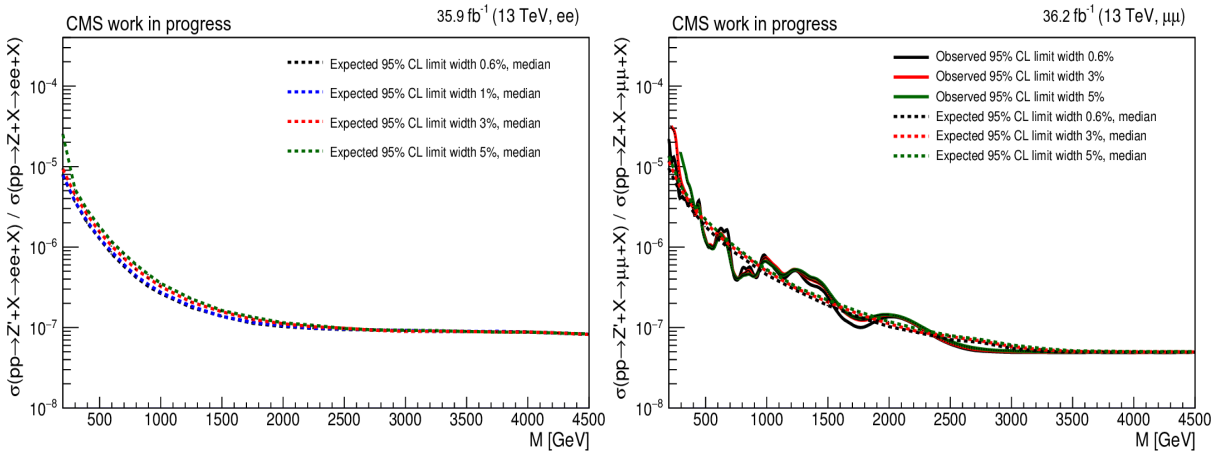

Figure 6.33: The 95% CL upper limits (only median) on R_σ for a spin 1 resonance with a width equal to 0.6%, 3.0% and 5% of the resonance mass for the dielectron (left) and dimuon (right) final states [118].

Finally, Table 6.16 lists the lower limit on the resonance mass for the Z'_{SSM} and Z'_{ψ} models, for the dielectron and dimuon final states separately and for the combination of the two channels.

	Z'_{SSM}		Z'_{ψ}	
Channel	Observed (TeV)	Expected (TeV)	Observed (TeV)	Expected (TeV)
ee	4.00	4.00	3.50	3.50
$\mu\mu$	4.35	4.35	3.80	3.80
ee + $\mu\mu$	4.60	4.60	4.00	4.00

Table 6.16: The observed and expected 95% CL lower limits on the masses of spin 1 Z'_{SSM} and Z'_{ψ} bosons for the combination of 13 TeV data, assuming a signal width of 0.6% of the resonance mass for Z'_{ψ} and 3% for Z'_{SSM}.

6.8 Summary

A search for high mass narrow resonances decaying in the dielectron final state has been performed using data from proton-proton collisions at $\sqrt{s} = 13$ TeV taken by the CMS experiment. The results of the analysis have been also combined with those of the analogous search in the dimuon final state. The integrated luminosity for the dielectron sample is 35.9 fb^{-1} and for the dimuon sample 36.2 fb^{-1}. Since no evidence for non-SM physics is found, upper limits at 95% confidence level on the parameter of interest R_σ have been derived and limits are set on the masses of hypothetical particles that could appear in two BSM scenarios.

My personal contributions to the analysis include the definition of the HEEP event selection (see Section 6.3) for the RunII data-taking, the estimation of the irreducible background coming from $t\bar{t}$ and $t\bar{t}$-like processes using the eμ invariant mass spectrum (see Section 6.5.2) and the weekly monitoring of the dielectron invariant mass spectrum during the data-taking.
In particular, given that the HEEP selection is used in other CMS analyses, since January 2016 I have been the HEEP contact for the exotica (EXO) group. In addition, I was responsible for the studies on the mass scale and resolution of the ECAL detector, which are key inputs for the computation of the limits, given that they are used to define the signal model. Two major steps are needed: the first one considers only the Z peak region via comparing the dielectron invariant mass shapes in data and MC for the two analysis categories, while the second step inspects the higher invariant mass region considering the difference between the generated invariant mass at simulation level and the reconstructed invariant mass. Finally a parametric model for the mass resolution function is provided as a function of the dielectron invariant mass itself, with the aim of allow flexibility in the limits computation. The procedure has been described in details in Section 6.4.

The techniques described in this chapter have been applied to different datasets over the years, collected according to the LHC schedule. This results in several publications by the CMS collaboration: two PAS documents, one corresponding to the analysis of 2.6 fb^{-1} of integrated luminosity collected in 2015 [131] and the other corresponding to the analysis of 12.9 fb^{-1} of integrated luminosity collected in the first half of 2016 [132] and a paper [133] where the results obtained analyzing the 2015 dataset are combined with the ones obtained at RunI, at a center-of-mass energy of 8 TeV. Moreover, another PAS document, where the full 2016 dataset is analyzed, is going to be published in the immediate future. The latest available results from CMS (showed in this chapter) put 95% confidence level lower mass limits on two particular BSM scenarios. For the Z'_{SSM} particle, which arises in the sequential standard model, and for the superstring inspired Z'_{ψ} particle, 95% confidence level lower mass limits are found to be 4.6 TeV and 4.0 TeV respectively, for the combination of both dielectron and dimuon final state. These results significantly extend the limits based on the 8 TeV datasets collected during RunI (see Section 2.2.4).

The ATLAS collaboration has also provided several publications on the same topic: three conference notes [134, 135, 136], corresponding respectively to the analysis of 3.2 fb^{-1} collected in 2015, 13.3 fb^{-1} collected in the first half of 2016 and 36.1 fb^{-1} collected during the full 2016 period, and a paper [137] with the results on the 2015 dataset. The latest available results from ATLAS [136] put a 95% confidence level lower mass limit of 4.5 TeV for the Z'_{SSM} model and 3.8 TeV for the Z'_{ψ} one, once the combination of both dielectron and dimuon final states is considered. These results are well in agreement with the ones obtained by CMS and showed in this chapter.

Chapter 7

Search for new physics in the diphoton final state at CMS

The production of high mass resonances decaying in the diphoton final state is a generic prediction of several extensions of the SM. From general assumptions [138, 139], the spin of a resonance decaying to two photons can be restricted to be either 0 or an integer greater or equal to 2. The production of heavy spin-0 resonances decaying to two photons is predicted by SM extensions with non-minimal Higgs sectors[1]*, while the production of spin-2 resonances is predicted by models postulating the existence of additional space-like dimensions (see Chapter 2).*
Previous results from both ATLAS [141] and CMS [142] collaborations based on the LHC 2015 data at center-of-mass energy $\sqrt{s}$ = 13 TeV, reported the observation of modest deviations from the background-only expectations, compatible with the resonant production of two photons in the diphoton invariant mass region $\sim$ 750 GeV.

In this chapter the latest CMS results on the search for new physics in the diphoton invariant mass spectrum at 13 TeV collected in 2016 will be described.
The analyzed dataset corresponds to an integrated luminosity of 12.9 fb^{-1}.

7.1 Data and MC samples

The analysis employs the data collected in pp collisions at $\sqrt{s}$ = 13 TeV by the CMS experiment up to the first half of the 2016 data taking, for an integrated luminosity of 12.9 fb^{-1}. Prompt calibration and alignment conditions were used to reconstruct the used data. The analyzed sample fulfilled standard data quality criteria for all components of the CMS detector ("golden json").

The list of datasets used in the analysis is reported in Table 7.1, together with the corresponding integrated luminosity. The DoubleEG dataset is used for the nominal analysis, while SingleElectron and SinglePhoton datasets are used for calibrations and background control studies. As explained in Section 6.1, the DoubleEG dataset requires at trigger level the presence of two electromagnetic objects compatible with two electrons/photons in the final state, the SingleElectron dataset requires only one electromagnetic object compatible with an electron, while SinglePhoton dataset requires the presence at trigger level of one electromagnetic object compatible with a photon. The run ranges are adjusted to only include certified runs.

The different Monte Carlo (MC) simulated samples used in this analysis were centrally produced during the "RunIISpring16" campaign. Table 7.2 reports the samples related to

[1] The simplest extension of the SM Higgs sector consists in the addition of a second doublet of scalar fields to the theory. Those models are known as two-Higgs-doublet models (2HDM) [140].

Dataset	Integrated luminosity
/DoubleEG/Run2016B-PromptReco-v2/MINIAOD	5.9 fb^{-1}
/DoubleEG/Run2016C-PromptReco-v2/MINIAOD	2.6 fb^{-1}
/DoubleEG/Run2016D-PromptReco-v2/MINIAOD	4.4 fb^{-1}
/SingleElectron/Run2016B-PromptReco-v2/MINIAOD	5.9 fb^{-1}
/SingleElectron/Run2016C-PromptReco-v2/MINIAOD	2.6 fb^{-1}
/SingleElectron/Run2016D-PromptReco-v2/MINIAOD	4.4 fb^{-1}
/SinglePhoton/Run2016B-PromptReco-v2/MINIAOD	5.9 fb^{-1}
/SinglePhoton/Run2016C-PromptReco-v2/MINIAOD	2.6 fb^{-1}
/SinglePhoton/Run2016D-PromptReco-v2/MINIAOD	4.4 fb^{-1}

Table 7.1: Datasets used in the analysis.

the different background sources for the analysis[2]. The main background processes are the irreducible background sources from the direct production of two photons (and jets) in the final state (first block of Table 7.2), as well as the reducible components due to $\gamma +$ jets (second block of Table 7.2) and multijet final states coming from QCD processes (third block of Table 7.2), where the jets are misidentified as photons. Finally, the last row of Table 7.2 reports the sample corresponding to the Drell-Yan (DY) production of two electrons in the final state. The samples were simulated respectively with the SHERPA 2.1 [143], , MADGRAPH5_AMC@NLO 2.2 [144] (interfaced with PYTHIA 8.2 [106] for parton showering and hadronization), and PYTHIA 8.2 generators. The DY background was generated with POWHEG v2 [100] for the next-to-leading order (NLO) matrix elements using the NNPDF3.0 parton distribution functions, interfaced with PYTHIA 8.2 for parton showering and hadronization.

Dataset	Cross-section [pb]
/GGJets_M-60To200_Pt-50_13TeV-sherpa/	5.971
/GGJets_M-500To1000_Pt-50_13TeV-sherpa/	1.510e-01
/GGJets_M-1000To2000_Pt-50_13TeV-sherpa/	1.181e-02
/GGJets_M-2000To4000_Pt-50_13TeV-sherpa/	3.690e-04
/GGJets_M-4000To6000_Pt-50_13TeV-sherpa/	2.451e-06
/GGJets_M-6000To8000_Pt-50_13TeV-sherpa/	1.753e-08
/GGJets_M-8000To13000_Pt-50_13TeV-sherpa/	7.053e-11
/GJets_DR-0p4_HT-40To100_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	18420.0
/GJets_DR-0p4_HT-100To200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	4881.0
/GJets_DR-0p4_HT-200To400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	1092.0
/GJets_DR-0p4_HT-400To600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	126.3
/GJets_DR-0p4_HT-600ToInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	44.75
/QCD_Pt-20to30_EMEnriched_TuneCUETP8M1_13TeV_pythia8/	557600000
/QCD_Pt-30to50_EMEnriched_TuneCUETP8M1_13TeV_pythia8/	6.876e+06
/QCD_Pt-50to80_EMEnriched_TuneCUETP8M1_13TeV_pythia8/	2.180e+06
/QCD_Pt-80to120_EMEnriched_TuneCUETP8M1_13TeV_pythia8/	4.140e+05
/QCD_Pt-120to170_EMEnriched_TuneCUETP8M1_13TeV_pythia8/	7.517e+04
/QCD_Pt-170to300_EMEnriched_TuneCUETP8M1_13TeV_pythia8/	1.899e+04
/QCD_Pt-300toInf_EMEnriched_TuneCUETP8M1_13TeV_pythia8/	1.249e+03
/DYToEE_NNPDF30_13TeV-powheg-pythia8/	2008

Table 7.2: Background MC samples with their cross-sections.

[2]For the final results, the background estimation follows a data-driven approach described in Section 7.5.1, hence the MC samples are not used for this purpose. However, they are used for other studies, specified in the main text.

Simulated signal samples of spin-0 and spin-2 resonances decaying to two photons were generated at leading order (LO) with the PYTHIA 8.2 event generator, using the NNPDF2.3 [111] parton distribution functions, with values of the resonance mass in the range 0.5−4.5 TeV. They are summarized in Table 7.3. The first block of Table 7.3 reports the simulated samples of heavy spin-0 resonances. The spin-0 case is treated like a heavier SM-like Higgs boson production, via gluon-gluon fusion only. Three width scenarios are considered where the natural width of the resonance Γ divided by the mass M of the resonance itself $\Gamma/M = 0.014\%$, 1.4% and 5.6%. The second block reports the simulated samples of the RS gravitons, chosen as a reference for the spin-2 resonance search. In this case, the relative width of the resonance is parametrized as $\Gamma/M = 1.4k^2$ and thus for the same Γ/M values as for the spin-0 scenario, k=0.01,0.1 and 0.2.

Dataset
/GluGluSpin0ToGG_W-0p014_M-XXX_TuneCUEP8M1_13TeV-pythia8/
/GluGluSpin0ToGG_W-1p4_M-XXX_TuneCUEP8M1_13TeV-pythia8/
/GluGluSpin0ToGG_W-5p6_M-XXX_TuneCUEP8M1_13TeV-pythia8/
/RSGravToGG_kMpl-001_M-XXX_TuneCUEP8M1_13TeV-pythia8/
/RSGravToGG_kMpl-01_M-XXX_TuneCUEP8M1_13TeV-pythia8/
/RSGravToGG_kMpl-02_M-XXX_TuneCUEP8M1_13TeV-pythia8/

Table 7.3: MC samples for different signal hypotheses. Masses between 500 GeV and 7000 GeV were considered.

Different mass points were considered and left implicit as "M-XXX" in Table 7.3. For the narrow width hyphotesis (i.e. $\Gamma/M = 0.014\%$) the samples were produced with a fine scan in mass:

- Every 5 GeV from 740 GeV to 770 GeV.
- Every 250 GeV from 1000 GeV to 4000 GeV.
- Every 500 GeV elsewhere from 500 GeV to 7000 GeV.

while for the larger width hypotheses, a coarser sampling was produced.
The simulation of the CMS detector response in both background and signal samples were performed using the GEANT4 package [117].

Finally, all the MC samples include the simulation of pile-up interactions (see Section 3.2.2) comparable with the pile-up distribution in the dataset under study. Moreover, the simulated pile-up distribution in MC was reweighted to match the corresponding distribution in the analyzed data. In Figure 7.1, the data and MC distributions of the number of vertices are compared after applying the diphoton event selection (see Section 7.3) for the two analysis categories where both photons are in the ECAL barrel (EBEB) or one photon is in the barrel and the other in one of the endcaps (EBEE). It can be seen that the absolute data/MC agreement is not particularly good: in this figure the MC distribution is normalized to the luminosity in data and no additional k-factor is considered, given that its value was not available. In later studies (see Section 7.5.1) the LO cross-section is scaled by an "effective" k-factor of 1.4 to take into account higher order corrections.

7.2 Trigger

A dedicated trigger path, requiring two photon candidates with a transverse momentum larger than 60 GeV was developed for this analysis. This trigger is hence labeled as

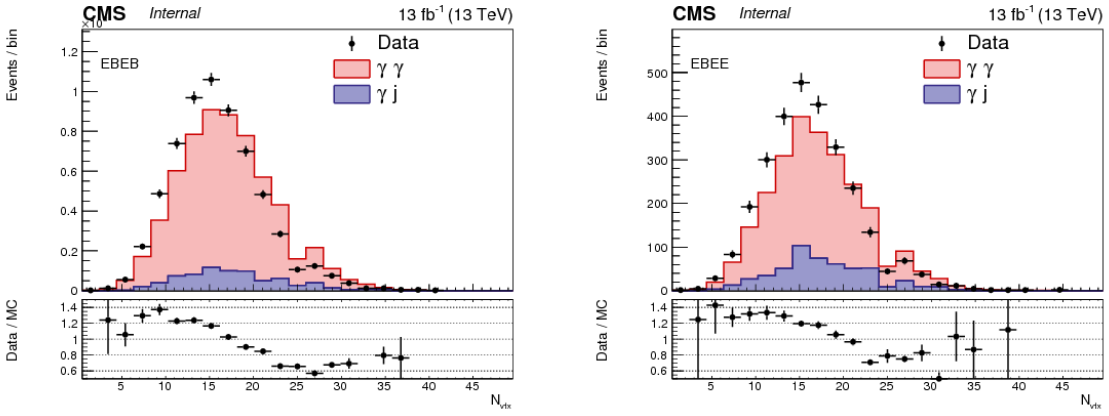

Figure 7.1: Distribution of the number of vertices for EBEB (left) and EBEE (right) for data (black points with error bars) and the sum of the two main sources of background, after pile-up reweighting [145].

"HLT_DoublePhoton_60" and referred to as "main analysis trigger". In addition, the ratio of the hadronic and electromagnetic energy deposit (the H/E ratio defined in Section 6.3 and reported in Section 7.3 for completeness) is required to be below 0.15 and no other identification requirements are applied. The E_{T} requirements of the trigger, together with the requirement on the H/E ratio are referred to in the text as "HLT selection".

Preliminarily, in order to study the efficiency of the main analysis trigger in data, a control sample has been obtained by requiring the events to be triggered by the loose working point (WPLoose) of the SingleEle_35 trigger[3] path (whose efficiency has been measured and reported later in this section). In addition, the diphoton selection used in the analysis (see Section 7.3.2) was required, loosening the E_{T} thresholds to 40 and 25 GeV for the leading and subleading photon candidate respectively. The leading photon candidate was also required to match an HLT candidate associated with the above mentioned SingleEle_35_WPLoose trigger path. The electron veto requirement of the selection (see again Section 7.3.2) has been removed. In this way, the control sample obtained is almost completely composed by events where a Z boson decays to a dielectron pair.

The efficiency of the main analysis trigger has been studied in two steps: first the efficiency of the level-1 (L1) trigger has been checked, then the efficiency of the HLT selection has been considered.
The HLT_DoublePhoton_60 trigger is seeded by the OR of three L1 triggers: SingleEG40, SingleJet200, and DoubleEG_22_12, respectively requiring one localised deposit in ECAL with $E_{\mathrm{T}} > 40$ GeV, an object compatible at L1 with a jet of transvere momentum above 200 GeV and two localised objects in ECAL with $E_{\mathrm{T}} > 22(12)$ GeV. The efficiency of the first L1 trigger has been evaluated simply by counting the fraction of subleading photon candidates that match the SingleEG40 trigger. Figure 7.2 shows such a fraction as a function of the photon candidate p_{T} for the barrel and endcap regions.

For photon candidates in the barrel with $75 < p_{\mathrm{T}} < 300$ GeV the efficiency of the L1 trigger is found to be above 97% and above 90% for p_{T} in the range between 300 and 400 GeV. Assuming no correlated source of inefficiency between the two legs, one can therefore estimate a L1 trigger efficiency higher than 99% for events with two photons in the barrel region and $p_{\mathrm{T}} < 400$ GeV. A similar conclusion can be taken for events with at least one photon in the endcap region with $p_{\mathrm{T}} < 300$ GeV.
While more data are required to characterize the phase space with higher transverse

[3]The "35" in the trigger name specifies the E_{T} treshold applied.

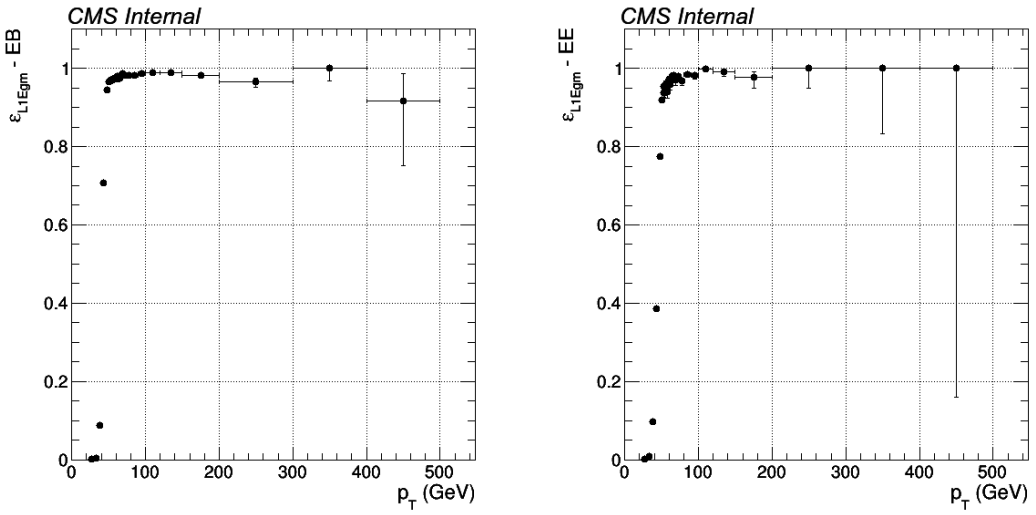

Figure 7.2: Efficiency for the subleading photon candidate to match an e/γ L1 trigger object of transverse energy above 40 GeV for the barrel (left) and endcap (right) regions [145].

momenta, the H/E requirement applied by the L1 trigger (0.03 and 0.06 for barrel and endcaps respectively) is expected to reduce the efficiency of the SingleEG40 starting at $p_{\mathrm{T}} \sim 500$ GeV. In this case, the SingleJet200 seed is expected to recover most of the inefficiency.

The efficiency of the HLT selection is studied by counting the fraction of subleading photon candidates that match an HLT candidate associated with the main analysis trigger path. The resulting curves are shown in Figure 7.3, where it can be seen that the requirement is fully efficient for events where the subleading photon candidate p_{T} is above 65 GeV.

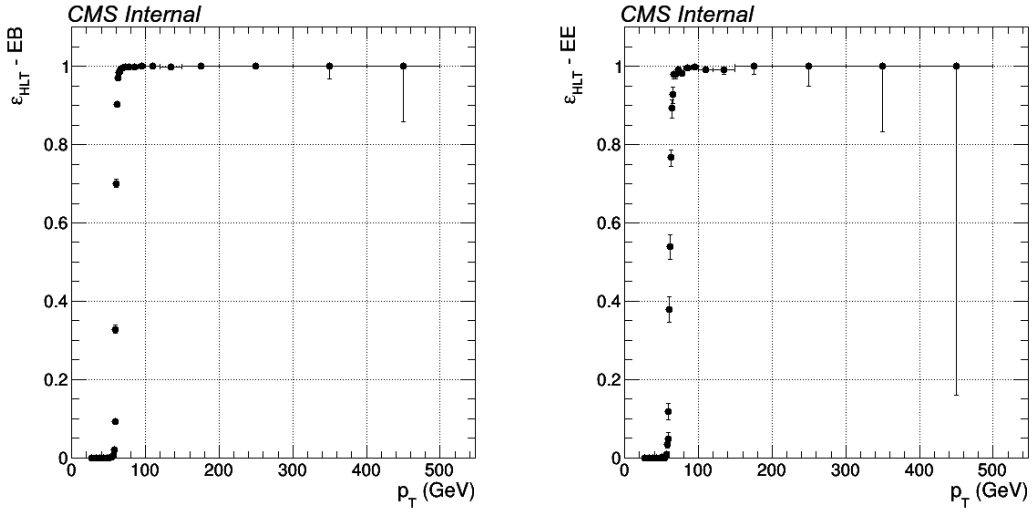

Figure 7.3: Efficiency of HLT_DoublePhoton_60 trigger as a function of p_{T} of the subleading photon. The left and right plots refer to photon candidates in the barrel and endcaps regions respectively [145].

Assuming no correlation between the two legs, the HLT selection can be estimated to be fully efficient for all the events entering the final analysis selection (where the photon candidates are required to have a p_{T} of at least 75 GeV; see Section 7.3.2). Since the trigger efficiency measured in data is consistent with unity no additional corrections are applied to MC.

In addition to the main analysis trigger efficiency, also the trigger efficiency of the above mentioned SingleEle_35_WPLoose has been considered given that, as explained above, this is the trigger used in order to derive the Z→ ee control sample. The efficiency of the SingleEle_35 HLT has been measured with the tag-and-probe technique [119] already described in Section 6.2. For this measurement:

- The events have to pass the loose working point of the HLT_Ele27_eta2p1 trigger where $|\eta|$ requirements are restricted up to 2.1.
- The tag is requested to have $p_{\mathrm{T}}>$ 40 GeV and pass the tight working point of the electron selection developed by the electron/photon POG in 2015. The tag must also fire the HLT_Ele27_eta2p1_WPLoose trigger.
- The probe must have $p_{\mathrm{T}}>$ 20 GeV and pass the full photon selection detailed in Section 7.3.2.
- The invariant mass of the tag-and-probe system is required to be within 70 GeV and 110 GeV.

The efficiency of the SingleEle_35_WPLoose trigger is evaluated by counting the fraction of probes that match an HLT candidate associated with the concerned path. The resulting efficiency curves are shown in Figure 7.4 and used to correct the simulation predictions, since its deviation from unity is not negligible.

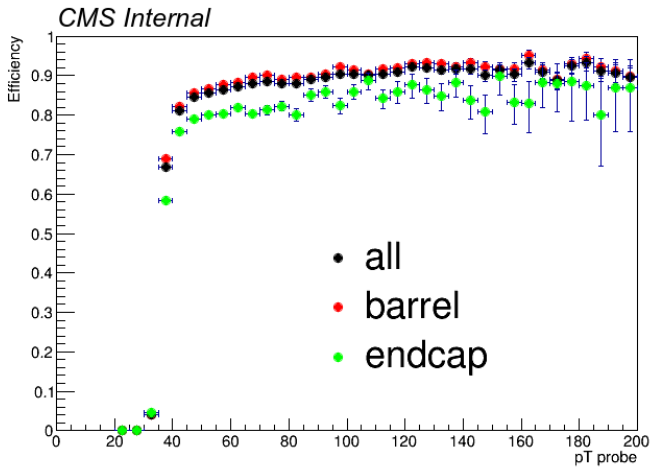

Figure 7.4: Efficiency of the SingleEle_35_WPLoose trigger as a function of the p_{T} of the probe photon [145].

7.3 Event selection

The analysis searches for a localised excess of events in the diphoton invariant mass spectrum over the SM background. The event selection is meant to be as simple and robust as possible. In particular, events are required to satisfy the trigger selection described in Section 6.2. The resulting events are then combined to form diphoton candidates which in turn have to first pass the preselection detailed in Section 7.3.1 and then the photon identification and isolation criteria described in Section 7.3.2.

All the criteria were studied in 2015 and applied for the 2016 data as well after having verified the changes between the two datasets.

7.3.1 Events preselection

The diphoton candidates are required to satisfy the following kinematic criteria:

- The p_{T} of both candidates has to be above 75 GeV.
- The absolute value of the pseudo-rapidity of both candidates, computed with respect to the centre of the CMS detector and denoted η_{SC}in the following, is required to be below 2.5 and not between 1.4442 and 1.566.
- At least one of the photon candidates has to have $|\eta_{SC}|$ below 1.4442 (i.e. events with both photon candidates in the endcap regions are rejected).

  The events are hence split in two categories: one in which both photons are in the barrel (EBEB) and the other one where one photon is in the barrel and the other in one of the endcaps (EBEE). Further categorizations based on the shower shape of the photons, like the R_9 variable, were also tested but the improvement was found to be negligible (1-3% on expected limit and significance) hence, for sake of simplicity, they are discarded.
- The diphoton invariant mass is required to be above 230 GeV. For events where one of the photon candidates is in the endcap region, the minimal invariant mass is asked to be 320 GeV.

Both photon candidates of the pair are then required to satisfy the identification criteria described in Section 7.3.2. If more than one diphoton candidate satisfies the set of selection criteria described, only the pair with the largest scalar sum of photon momenta is retained.

A primary vertex is assigned to each diphoton candidate and the photon candidates kinematic properties are computed under this assumption. The standard algorithm used in CMS ranks the primary vertices according to the sum of the squares of the tracks transverse momentum. In the case of events with neutral particles this algorithm may lead to suboptimal performances, while the performances can be improved through the analysis of the correlation between the tracks recoil and the diphoton system. Such algorithms have been successfully used in the search for the diphoton decay of the Higgs boson [146], hence in this analysis the same algorithm used for the $H \to \gamma\gamma$ search has been employed.

7.3.2 Photon identification and isolation requirements

A dedicated set of photon identification and isolation criteria (simply referred to as "photon identification")[4], targeting high p_{T} objects, has been developed for the analysis. The reason for developing a custom set is that the criteria developed by the electron/photon POG are typically tuned for objects of much lower transverse momenta than those used in this analysis. Therefore optimal analysis performance requires dedicated tuning.

The variables used are defined as follows:

- Iso_{ch}: the scalar sum of the transverse momenta of the charged PF hadron candidates assigned to the chosen primary vertex for which the distance in terms of ΔR between the photon candidate (labeled with a γ as subscript) and the generic PF candidate (labeled with a *cand* as subscript) must be:

$$\Delta\mathrm{R} = \sqrt{(\eta_\gamma - \eta_{cand})^2 + (\phi_\gamma - \phi_{cand})^2} < 0.3.$$

[4]The photon identification requirements together with the preselection requirements described earlier define the full "photon selection".

- Iso_γ: the scalar sum of the transverse momenta of the PF photon candidates for which $\Delta\mathrm{R} < 0.3$.

- H/E: the ratio of the energy deposited in the HCAL tower behind the photon candidate and the energy of the candidate itself.

- $\sigma_{i\eta i\eta}$: the (weighted) spatial second order moment of the photon candidate in the η direction, computed as

$$\sigma_{i\eta i\eta} = \sqrt{\frac{\sum_{i\in 5\times 5}\left(\eta_i - \bar{\eta}\right)^2 w_i}{\sum_{i\in 5\times 5} w_i}}, \qquad w_i = \max\left(0, 4.7 + \log(E_i/E_{5\times 5})\right).$$

The photon identification and isolation criteria used in the analysis are shown in Table 7.4. In addition to the criteria described earlier it can be noted that the photon candidates are required to satisfy the so-called "conversion-safe electron veto" criterion, whose purpose is to reduce the contamination coming from electrons, rejecting events with no missing hits in the tracker.

Photon category	Iso_{ch}(GeV)	Iso_γ(GeV)	H/E	$\sigma_{i\eta i\eta}$
$\lvert\eta_{SC}\rvert < 1.4442$	< 5	< 2.75	$< 5\times 10^{-2}$	< 0.0105
$\lvert\eta_{SC}\rvert > 1.566$	< 5	< 2.0	$< 5\times 10^{-2}$	< 0.028
conversion-safe electron veto applied for all categories				

Table 7.4: Photon identification and isolation criteria used in the analysis. The full photon selection is defined by these criteria, together with the preselection criteria described earlier.

The dependence of the chosen identification variables as a function of pile-up and transverse momentum of the photon was studied. Only in the case of Iso_γ the observed effect was found to be non-negligible and specific corrections were applied in order to correct the dependence of the 95% quantile of the variable. The general form of the correction is:

$$\bar{Iso}_\gamma = \alpha + Iso_\gamma - \rho\cdot A - \kappa\cdot p_{\mathrm{T}}$$

where ρ is defined as the transverse energy density due to pile-up interactions calculated using the Fastjet framework [147] and α, A and κ are factors to be found. Obviously, the α term in the definition of the corrected photon isolation variable could be dropped and its value subtracted from the corresponding thresholds. The reason to introduce it is to maintain the intuitive meaning of the Iso_γ variable, such that the median of the corrected distribution $\bar{Iso}_\gamma$ would be roughly 0, as intuitively expected for isolated photon candidates. The values of α, A and κ were derived using a MultiVariate approach and separately for different rapidity regions. The coefficients were found to be consistent with the results (obtained with the 2015 MC campaign) of previous analyses [148, 149] within about 5-10%. In order to have a uniform set of identification criteria for the 2015 and the 2016 datasets (given that the results coming from the two datasets will be combined at the end of this chapter), the numerical values from [148, 149] have been used. They are listed in Table 7.5.

To conclude the section, the diphoton identification efficiency as a function of the generated diphoton invariant mass is reported in Figure 7.5. The overall diphoton identification efficiency is over 85% for diphoton candidates in the barrel-barrel (EBEB) category while it is about 80% in the barrel-endcap (EBEE) one.

Region	α (GeV)	A	κ
$\lvert\eta_{SC}\rvert < 0.9$	2.5	0.17	4.5×10^{-3}
$0.9 < \lvert\eta_{SC}\rvert < 1.4442$	2.5	0.14	4.5×10^{-3}
$1.566 < \lvert\eta_{SC}\rvert < 2.0$	2.5	0.11	3×10^{-3}
$2.0 < \lvert\eta_{SC}\rvert < 2.2$	2.5	0.14	3×10^{-3}
$2.2 < \lvert\eta_{SC}\rvert < 2.5$	2.5	0.22	3×10^{-3}

Table 7.5: Corrections factors for the Iso_γ variable.

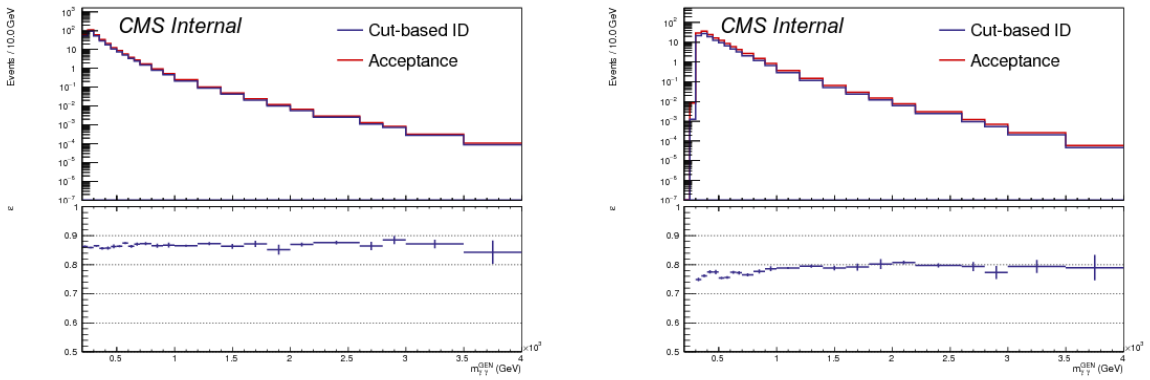

Figure 7.5: Diphoton identification efficiency as a function of the diphoton invariant mass at generation level, evaluated on the irreducible $\gamma\gamma$ background. The efficiency is computed with respect to the diphoton candidates passing the preselection detailed in Section 7.3.1. Events in the EBEB (EBEE) category are shown on the left (right) [145].

7.3.3 Selection efficiency

The tag-and-probe technique (see Section 6.2) is used to determine the efficiency of the photon selection[5] described in the previous sections (preselection plus identification and isolation requirements) using $Z \rightarrow ee$ events in data and MC.

Data events are selected in the SingleElectron dataset, requiring the loose working point of the HLT_Ele27_eta2p1 ($|\eta|$-restricted up to 2.1 and with E_T treshold at 27 GeV). In the MC samples, the trigger information is missing, so no trigger is required in this case. For the efficiency measurement:

- The tag electron must have $p_T > 30$ GeV and pass the tight working point of the official electron selection developed by the electron/photon POG in 2015.
- The probe photon must have $p_T > 20$ GeVand pass the inverted electron veto.
- The invariant mass of the tag-and-probe system is required to be within 70 GeV and 110 GeV.

The p_T and η spectra of the probe photons entering the selection efficiency measurement are shown in Figure 7.6 for the probes in data and in MC. In Figures 7.7 and 7.8 the distributions of the variables used in the photon selection in data and MC are compared. After rescaling the MC event yields to match those in data, the agreement is good in the majority of the cases. The background contamination coming from fake photons, which is visible in the H/E or Iso_γ distributions, cannot anyway affect the efficiency results since a fit to the invariant mass spectrum is performed to determine the signal and background composition.

[5]For this study the electron veto requirement has been inverted in order to not reject the electrons.

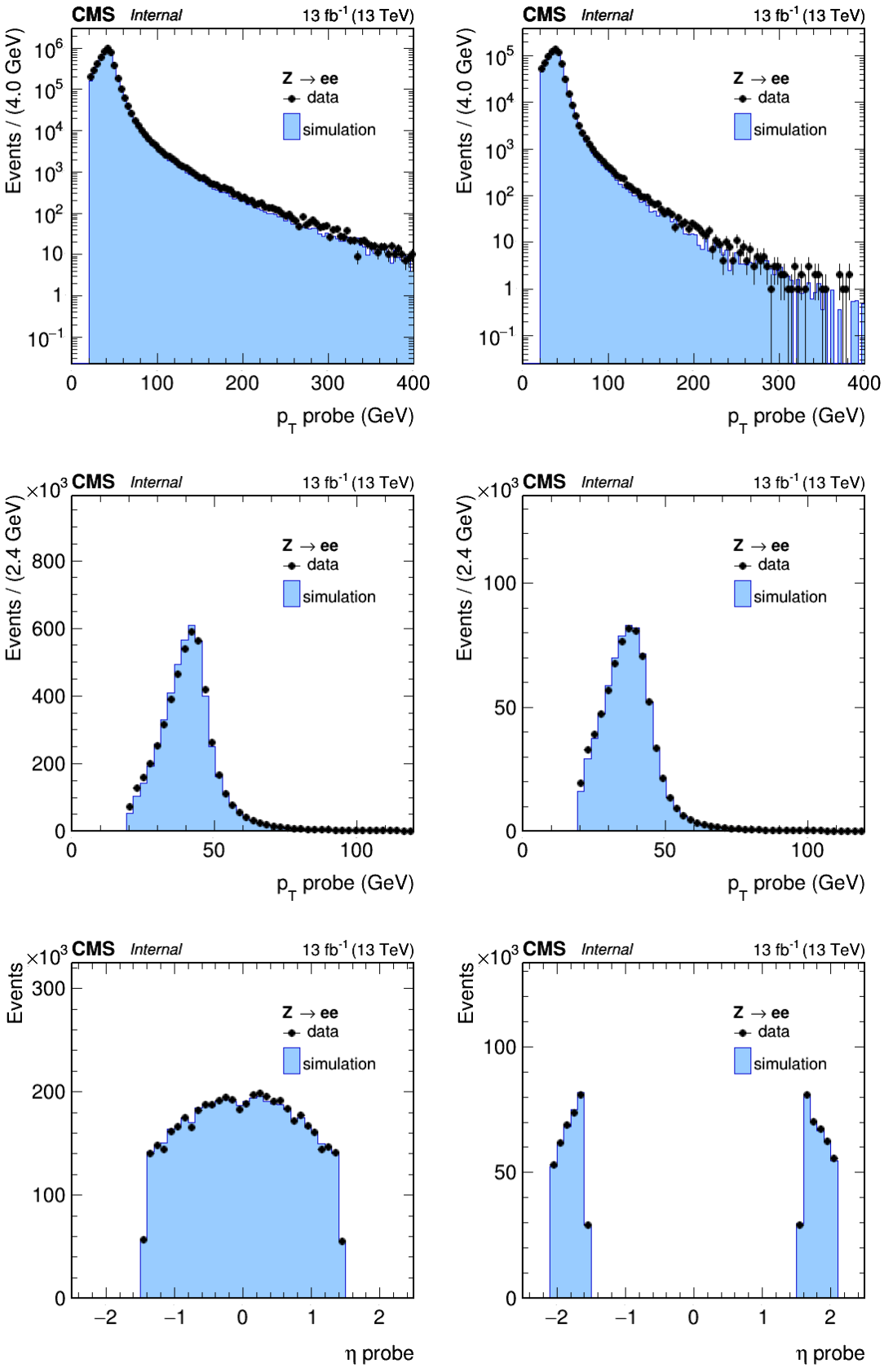

Figure 7.6: Transverse momentum (top and middle, in logarithmic and linear scale) and η (bottom) spectra of the probes. Probes in the barrel (endcaps) are shown on the left (right) [145].

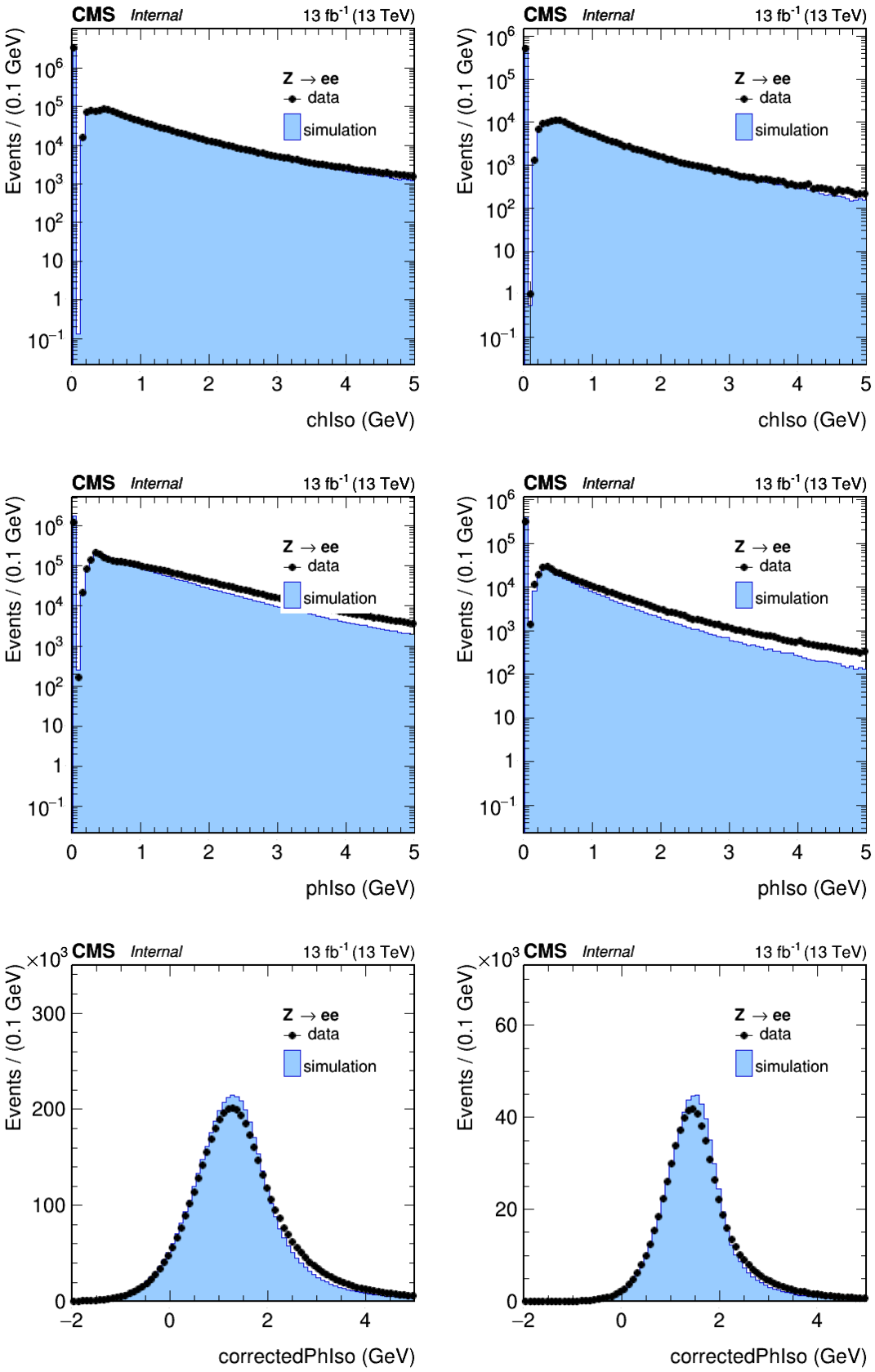

Figure 7.7: Charged component of the isolation for probe photons Iso_{ch} (top) and photon component of the isolation Iso_{γ} for probe photons before (middle) and after (bottom) the corrections for pileup and p_{T} dependances described in the previous section. Probes in the barrel (endcaps) are shown on the left (right) [145].

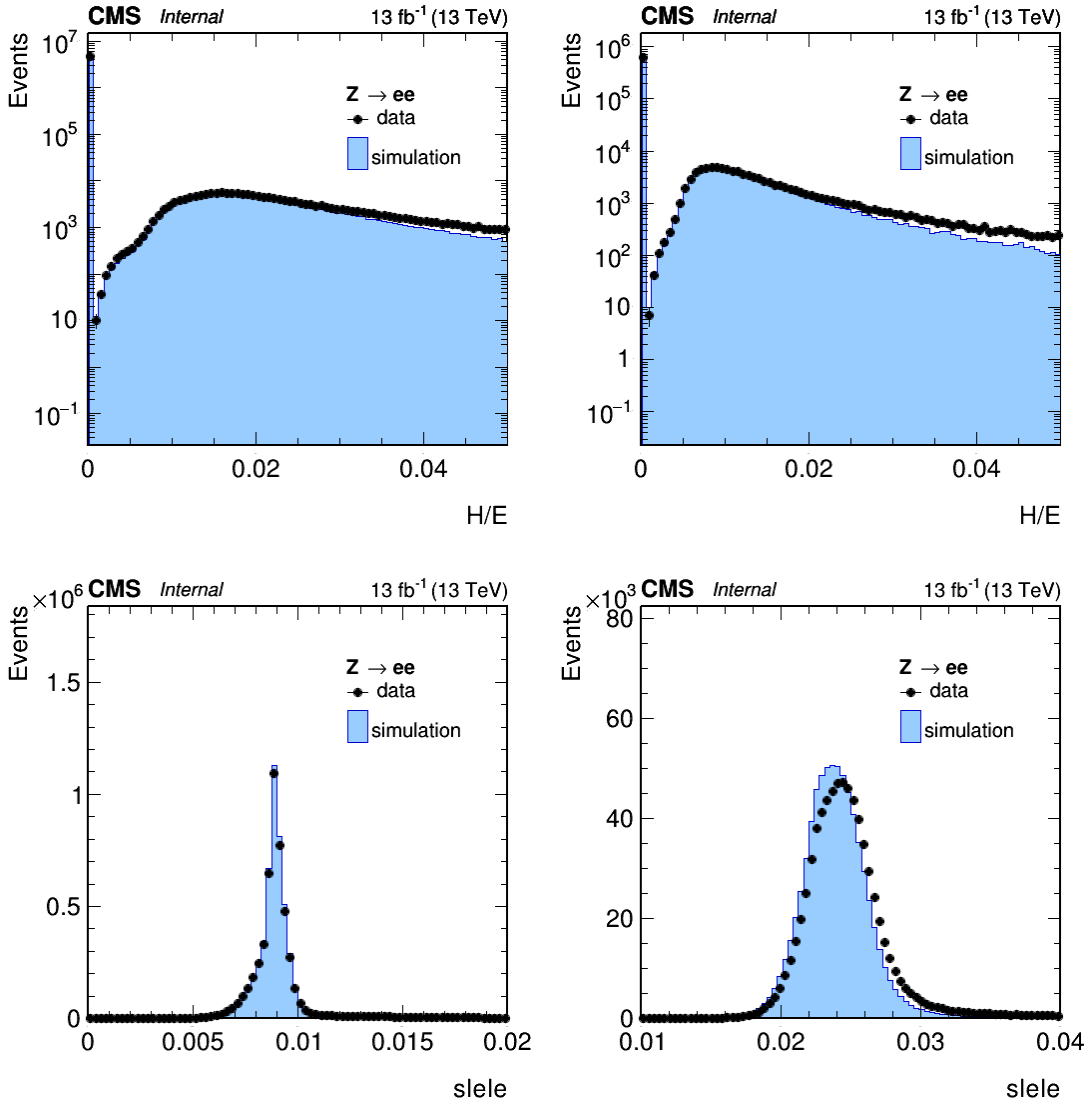

Figure 7.8: H/E (top) and $\sigma_{i\eta i\eta}$ (bottom) distribution for probe photons. Probes in the barrel (endcaps) are shown on the left (right) [145].

The variable which shows the largest disagreement between data and MC is the Iso_γ (bottom plots in Figure 7.7), even after applying the corrections for p_T and pile-up dependances discussed in the previous section.

The selection efficiency has been computed separately for photons in the barrel and in the endcaps as a function of p_T. The p_T-binning has been chosen to keep the statistical error within a few percents. To estimate the efficiency in MC, a simple cut-and-count method is performed, following the recommendations of the electron/photon POG while, in data, the selected events are fitted with a "signal plus background" model, simultaneously for probes passing and failing the selection. In particular, for the nominal efficiency values, the signal probability density function is defined starting from a MC template generated by the DY invariant mass shape in MC (referred to as "Z line-shape") convoluted with a simple Gaussian function. The background probability density function (pdf) is instead defined as a falling exponential function. The impact on the efficiency measurement coming from the choice of the signal and background models has been evaluated by changing the chosen pdf and repeating the procedure. The alternative signal pdf is the Z line-shape convoluted with a Crystal Ball instead of a simple Gaussian[6] while the alternative background pdf is a RooCMSShape[7]. The maximal difference from the nominal fit, bin-per-bin in p_T, is taken as systematic error on the measurement.
As an example, the fit results for the dielectron invariant mass spectra built requiring that the probe passes the photon selection (Passing probes), fails the photon selection (Failing probes) or with no requirements on the probe (All probes) are shown in Figure 7.9 for the p_T bin defined by the range: $50\,\text{GeV} < p_T < 60\,\text{GeV}$.

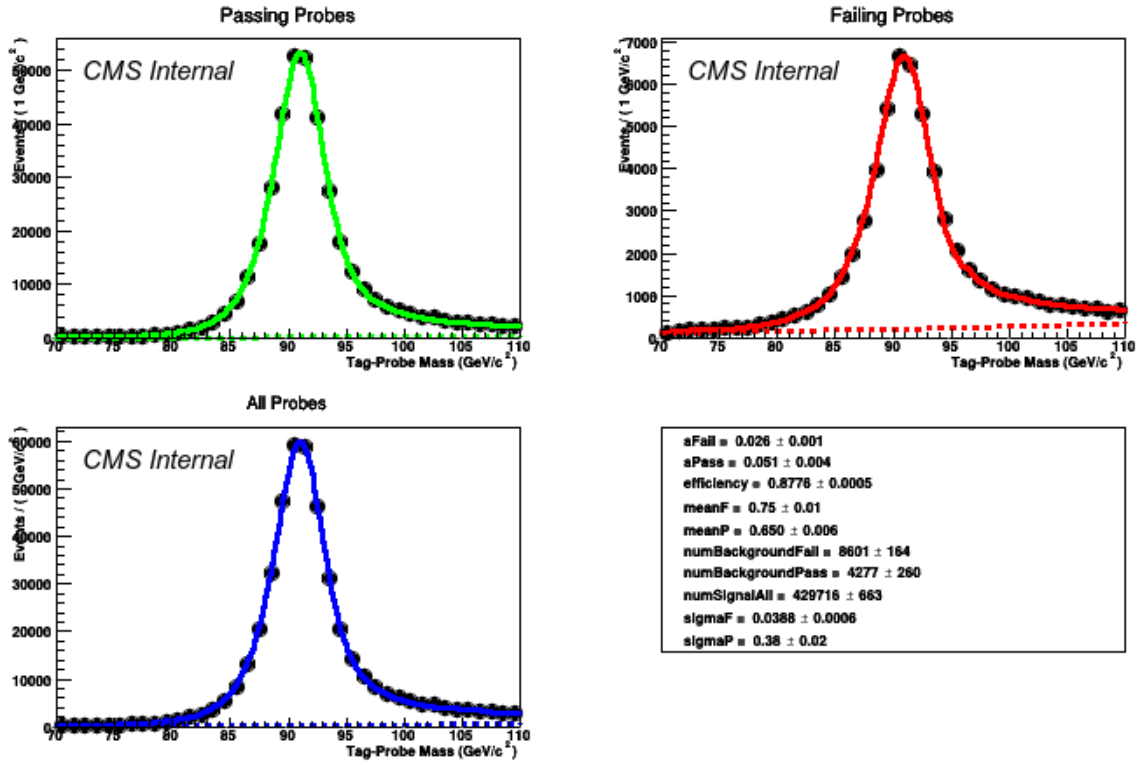

Figure 7.9: Fit results (listed in the bottom right side of the figure) for the dielectron invariant mass spectra built requiring that the probe passes the photon selection (top left), fails the photon selection (top right) and with no requirements on the probes (bottom left). The p_T bin considered is defined by: $50\,\text{GeV} < p_T < 60\,\text{GeV}$ [145].

[6]The Crystal Ball function is defined as a Gaussian core connected with a power-low tail.
[7]The RooCMSShape is an error function connected with an exponential decrease.

Table 7.6 reports the efficiencies for data and MC together with the statistical and systematic errors.

Barrel	Data			MC
p_T[GeV]	Efficiency	Stat. Err.	Syst. Err	Efficiency
20-30	0.7842	0.0011	0.0297	0.8331
30-40	0.8466	0.0004	0.0018	0.8820
40-50	0.8775	0.0002	0.0014	0.9092
50-60	0.8776	0.0005	0.0034	0.9115
60-80	0.8733	0.0012	0.0114	0.9109
80-110	0.8774	0.0024	0.0117	0.9107
110-150	0.8862	0.0037	0.0092	0.9158
150-200	0.8892	0.0059	0.0049	0.9160
200-270	0.8752	0.0090	0.0094	0.9256
270-350	0.8982	0.0150	0.0270	0.9238
350-500	0.8735	0.0260	0.0068	0.9255
Endcap	Data			MC
p_T[GeV]	Efficiency	Stat. Err.	Syst. Err	Efficiency
20-30	0.6202	0.0017	0.0175	0.6870
30-40	0.7059	0.0009	0.0017	0.7586
40-50	0.7547	0.0008	0.0032	0.8012
50-60	0.7702	0.0019	0.0016	0.8182
60-80	0.7816	0.0033	0.0093	0.8319
80-110	0.8042	0.0062	0.0063	0.8549
110-150	0.8131	0.0097	0.0115	0.8672
150-200	0.8189	0.0168	0.0159	0.8958
200-350	0.8961	0.0175	0.0162	0.9167

Table 7.6: Photon selection efficiency measured in barrel and endcaps up to $|\eta|$ <2.1 using the tag-and-probe method.

The efficiencies and the ratio between data and MC (scale factors) are also shown in Figure 7.10. In both cases, for the endcaps results the $|\eta|$ region is restricted up to $|\eta|$ <2.1. This choice is motivated by the fact that the signal for this analysis is more central (i.e. its decay products tend to have smaller $|\eta|$ values) than the SM Z→ee process.

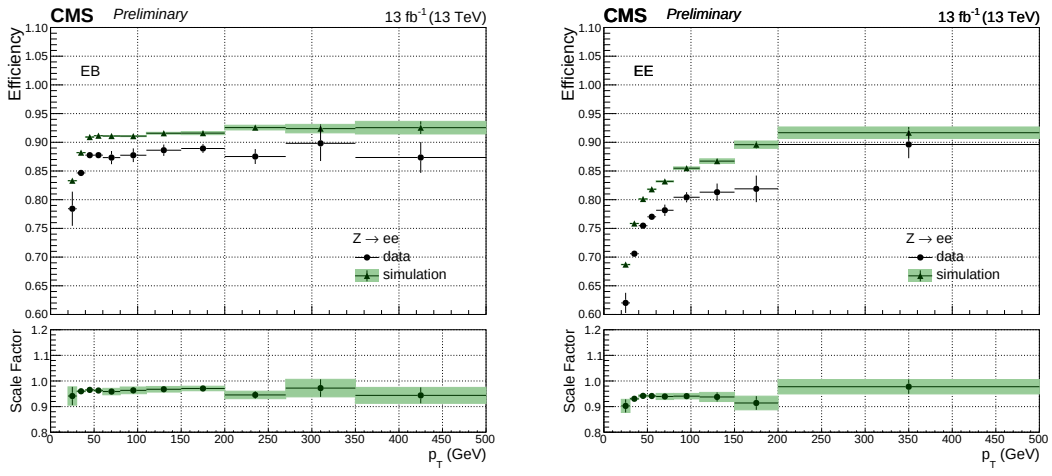

Figure 7.10: Photon selection efficiency measured in barrel (left) and endcaps up to $|\eta|$ <2.1 (right) using the tag-and-probe method. Statistical and systematic errors are shown [145].

Figure 7.11 shows the efficiency and the scale factors as a function of η of the probe

photon. Differently with respect to 2015 results, the scale factor is not compatible with 1

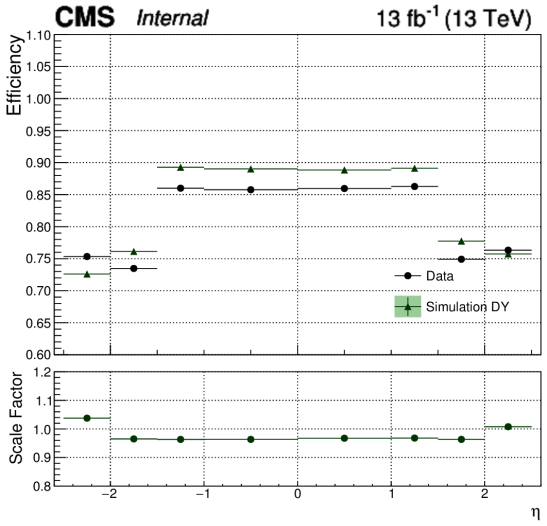

Figure 7.11: Photon selection efficiency measured using the tag and probe method as a function of the pseudorapidity. Statistical errors only are shown [145].

in all the p_{T} bins and as a result an additional correction has been applied in the analysis. Also, a corresponding uncertainty of 6% is propagated to the signal model (see Section 7.7). This additional correction has been derived by fitting with a constant function the scale factor distributions shown in the bottom side of Figure 7.10. The corresponding numbers are listed in Table 7.7.

Region	Photon selection scale factor
$\|\eta\|$ <1.5	0.9627 ± 0.0002
1.5< $\|\eta\|$ <2.1	0.9352 ± 0.0007

Table 7.7: Results of the fits to constant function for the data/MC scale factors related to the photon selection efficiency.

As said, in order to use the Z peak in the tag-and-probe method, starting from the full photon selection criteria, the electron veto has been inverted. The efficiency of the electron veto itself has not been measured here, but one can refer to the electron/photon POG measurement, done by using Z $\rightarrow\mu^{+}\mu^{-}\gamma$ events. The results of this measurement have been extracted from [150] and reported in Table 7.8. Given that the data/MC scale factors for the electron veto are close to 1 for all categories they are not propagated at the analysis level.

Region	Data		Simulation		Data/MC ratio	
	Eff.	Stat. Err.	Eff.	Stat. Err.	Eff.	Stat. Err.
Barrel; R_9 >0.85	0.9932	0.0005	0.9972	0.0008	0.9960	0.0010
Barrel; R_9 <0.85	0.9750	0.0018	0.9824	0.0041	0.9924	0.0045
Endcap; R_9 >0.90	0.9851	0.0013	0.9852	0.0025	0.9999	0.0029
Endcap; R_9 <0.90	0.9525	0.0071	0.9658	0.0113	0.9862	0.0137

Table 7.8: Efficiency of the conversion-safe electron veto [150] in different categories (the R_9 variable is defined in Section 5.1.3). The data/MC ratio is also shown with its uncertainty.

7.4 Energy scale and resolution corrections

Data-driven energy scale and resolution corrections are derived for the photon energy using the Z peak dielectron invariant mass[8] coming from $Z \to ee$ decays, through a workflow which is described below and referred to as "calibration procedure". The calibration procedure is organized in several steps, where the results at step $n-1$ are used as inputs for the n-th step. The dielectron invariant mass spectra are built requiring the loose working point of the electron identification criteria developed by the electron/photon POG and a cut on the electron transverse energy $E_T > 20$ GeV. Moreover the electrons are required to be ECAL driven (see Section 5.1.6 about the track seeding). In addition, some discrepancies have been observed in the R_9 distributions in data and MC and, consequently, the R_9 distribution in MC has been reweighted in order to match the distribution in data.

In the first step, run dependent energy scale corrections are derived and applied to data, in order to correct for possible time dependent energy scale variations during the data taking. In particular, both data and MC dielectron invariant mass distributions are built in the range 80-100 GeV for several run range periods. An unbinned maximum-likelihood fit is then performed using a Breit-Wigner (BW) function convoluted with a Crystal Ball (CB) function. The BW parameters (mean and width) are fixed to the ones taken from the Particle Data Group (PDG) for the mass and natural width of the Z boson, while the CB parameters (Δm and $\Delta\sigma_{CB}$) can be interpreted as the shift in the Z peak position with respect to the PDG value (Δm) and a detector resolution term ($\Delta\sigma_{CB}$) which broaden the invariant mass distribution; anyway the $\Delta\sigma_{CB}$ term is nowhere used in the procedure. The fitted Z peak position is hence defined as:

$$\text{Z peak position} = \text{PDG value} + \Delta m \tag{7.1}$$

The energy scale in data is then corrected so that the Z peak position fitted in data matches the Z peak position fitted in MC, for the different run range periods considered. The behaviour of the Δm shift in data and MC is summarized in Figure 7.12 for barrel (EB) and endcaps (EE) divided in different $|\eta|$ categories: while the overall behavior of the scale in EB is rather stable in the first part of the dataset, a significant drop at the level of 0.5 GeV is observed in the last part of the dataset; moreover, some significant variations are observed in EE as can be seen in Figure 7.12. The effect is particularly evident for the high-η region of EE, where a variation of about 1-1.5 GeV is clearly visible. In more details, in order to align the Z peak position in data to be the one predicted by the MC, data/MC run dependent energy scale factors (SF) are applied to the data in the simple form, at the end of the first step:

$$\text{SF}^{\text{step1}}(\text{run number}, \eta) = \frac{\text{peak position fitted in MC}}{\text{peak position fitted in data}}$$

After applying the SF^{step1} corrections to data, the invariant mass spectra are fitted again, with the exact same procedure exposed before and in the same run range periods, as a closure test of the step. As shown in Figure 7.13, a much more stable behaviour is obtained as a function of run number and the Δm values in data are closer to the corresponding MC values.

In the second step of the procedure, on top of the SF^{step1} corrections, a set of energy corrections is obtained in 8 categories ($4\eta \times 2R_9$), using what is called "smearing method". This method, instead of using an analytical function to fit the mass distributions in data and MC, as done in the first step, employs the MC invariant mass distribution itself as a binned pdf. In general terms (full details are given later), the MC invariant mass

[8] As explained in Section 5.1.3, two energy regressions are available: the photon-tuned one and the electron-tuned one. Given that this analysis uses photons, the photon-tuned regression has been used for the calibration procedure.

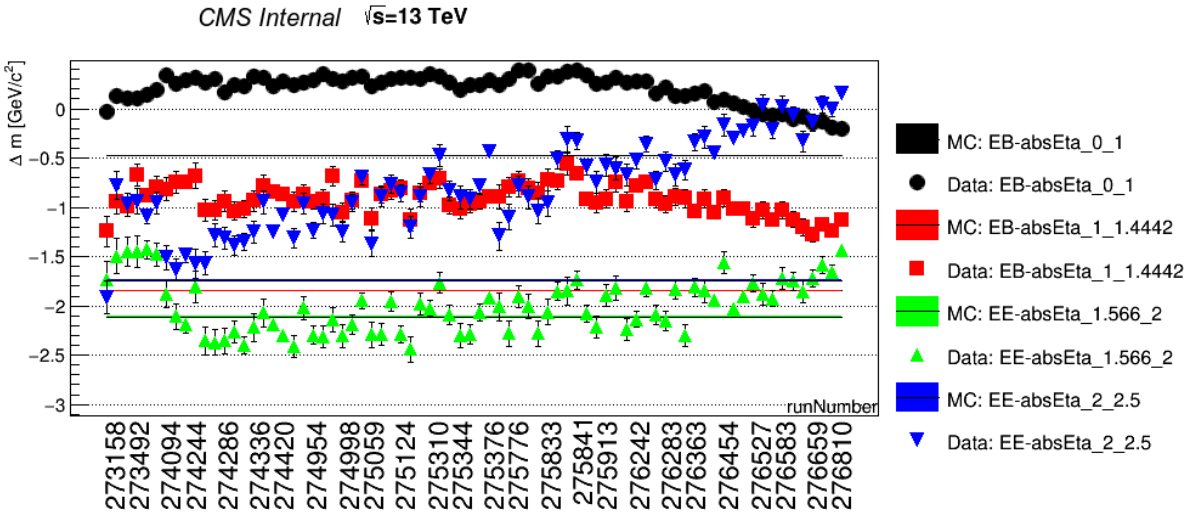

Figure 7.12: $Z \to ee$ mass shift Δm with respect to the PDG value for the Z boson mass ($m_Z = 91.188$ GeV) in several run ranges for 4 η categories: 2 categories in EB, 2 categories in EE. MC value is displayed as a continuous line.

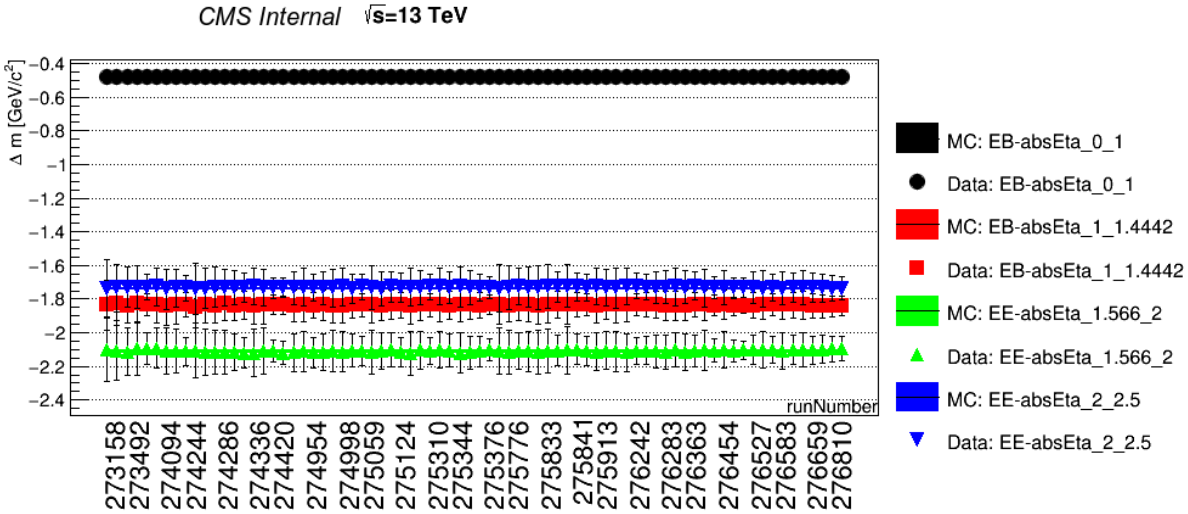

Figure 7.13: $Z \to ee$ mass shift Δm with respect to the PDG value for the Z boson mass ($m_Z = 91.188$ GeV), after applying the scale corrections derived at the end of the first step of the calibration procedure, in several run ranges for 4 η categories: 2 categories in EB, 2 categories in EE. MC value is displayed as a continuous line.

distribution is modified via a certain rule in order to better fit the data distribution, guiding the procedure through the binned likelihood function computed between data and MC.

In particular, the smearing method procedure consists in:

- Obtain the histogram of the invariant mass distribution in data.
- Obtain the histogram of the invariant mass distribution in MC.
- Fit the data histogram using the MC as a pdf, adapting the MC shape to the one in data. Energy scale (ΔP) and energy resolution corrections ($\Delta\sigma$, also referred to as

"smearing" corrections) are defined, as functions of $\eta \times \mathrm{R}_9$, to improve the agreement between the MC and the data distributions. These factors allow to redefine the energy per electron in the MC with the rule:

$$E_{MC} \to E_{MC} \times Gauss(1 + \Delta P, \Delta\sigma)$$

where $Gauss(1 + \Delta P, \Delta\sigma)$ is a value extracted from a Gaussian distribution with $(1 + \Delta P)$ as mean and standard deviation $\Delta\sigma$. As a marginal note, the use of a Gaussian factor assumes that no extra behaviour exists in the tails of the invariant mass distribution in data[9].

- Electrons are defined in categories $(4\eta \times 2R_9)$ and the values of ΔP and $\Delta\sigma$ which optimise the agreement between the data and the modified MC (labeled as "MC smear" in the figures) distribution are found maximising the likelihood between the two distributions, assuming the MC distribution as the pdf. In particular a binned likelihood (L) is defined, using the multinomial expression:

  $$L = \prod_i p_i^{N_i}$$

  where probabilities p_i are taken directly from the normalized MC distribution in the i-th bin range, and N_i identifies the number of events in data in that particular bin.

- At the end of the procedure, the found ΔP (as the one that maximizes the logarithm of the likelihood, i.e. minimizes the negative logarithm of the likelihood function) is used to define the scale factor for this second step as $\mathrm{SF}^{\mathrm{step2}} = (1 - \Delta P)$. The minus sign in the definition of $\mathrm{SF}^{\mathrm{step2}}$ comes from the fact that the smearing method works adapting the MC distribution to match the data, while at analysis level the opposite is needed: the energy scale in data must be shifted to match the one in MC.

Technically also the $\Delta\sigma$ parameters are found at step 2, but only ΔP corrections are considered[10], in order to derive the resolution parameters only when the scale corrections are already applied. Hence the final set of energy scale corrections is defined as:

$$\mathrm{SF} = \mathrm{SF}^{\mathrm{step1}}(\text{run number}, |\eta|) * (1 - \Delta P(|\eta|, \mathrm{R}_9)) \tag{7.2}$$

where it can be seen that the final SF are run-,$|\eta|$- and R_9-dependent.

After applying the SF defined in Equation (7.2) to data, the third step of the calibration procedure is performed, where a set of smearing corrections, in the same 8 categories $(4\eta \times 2R_9)$ of the second step, is found with the above defined smearing method, again minimizing the likelihood between the smeared MC and the data. At the same time, the third step serves also as closure test for eventual residual energy shifts ΔP. Hence, while the set of scale corrections is run-,$|\eta|$- and R_9-dependent, the smearing corrections are only $|\eta|$- and R_9-dependent. The latter can be found in Table 7.9, with their statistical errors are labeled as $\Delta_{stat}\sigma$. The 8 categories in $\eta \times R_9$ are specified too: it can be seen that the barrel is split in a "low-η" region ($|\eta| < 1$) and a "high-η" region ($|\eta| > 1$), both splitted according the R_9 value. The same is done for the endcaps, splitted in $|\eta| < 2$ and $|\eta| > 2$ regions and then according to R_9. The data energy scale corrections are not reported in the table simply due to the large amount of run-,$|\eta|$- and R_9-bins which have been defined (about 100 run periods times the 8 $\eta \times \mathrm{R}_9$ categories).

[9] This assumption proved to be effective from past experiences but in order to get even better data/MC agreeements this could be re-thought for future analyses.

[10] A third step is proposed for $\Delta\sigma$ resolution corrections.

Category		$\Delta\sigma$[%]	$\Delta_{stat}\sigma$[%]
EB: $\|\eta\| < 1$	$R_9 > 0.94$	0.84	0.03
EB: $\|\eta\| < 1$	$R_9 < 0.94$	0.97	0.03
EB: $\|\eta\| > 1$	$R_9 > 0.94$	1.08	0.11
EB: $\|\eta\| > 1$	$R_9 < 0.94$	1.62	0.04
EE: $\|\eta\| < 2$	$R_9 > 0.94$	1.82	0.08
EE: $\|\eta\| < 2$	$R_9 < 0.94$	2.20	0.05
EE: $\|\eta\| > 2$	$R_9 > 0.94$	2.29	0.05
EE: $\|\eta\| > 2$	$R_9 < 0.94$	2.51	0.06

Table 7.9: MC energy smearing corrections for the first half of 2016 dataset obtained using $Z \to ee$ events in 8 categories: $(4\eta \times 2R_9)$.

Figure 7.14 shows the invariant mass distribution in data after applying the full set of energy scale corrections, the initial MC distribution and the one at the end of the calibration procedure, where the energy resolution corrections have been applied. As an example, two categories are shown: both electrons in EB with $|\eta| < 1$ and $R_9 > 0.94$, and both electrons in EE with $|\eta| < 2$ and $R_9 < 0.94$.

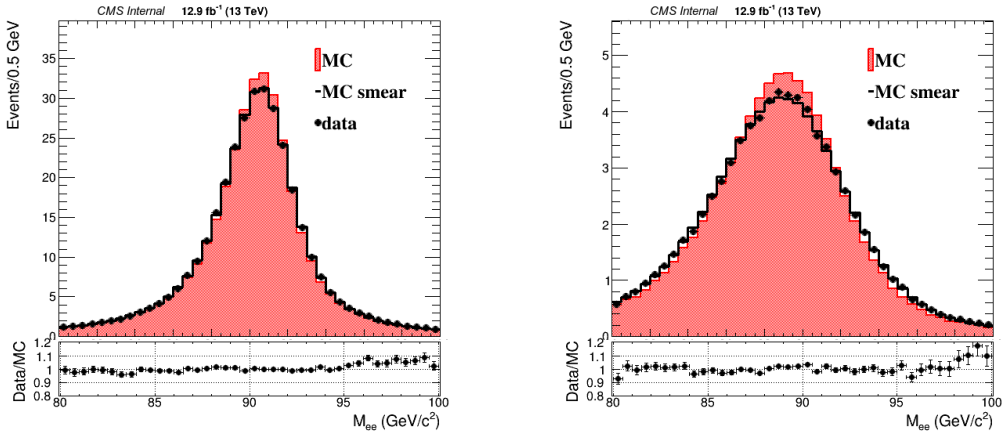

Figure 7.14: Invariant mass distribution in data after applying the full set of energy scale corrections, compared with the initial MC distribution (red histogram) and the one at the end of the calibration procedure (MC smear), where the energy resolution corrections have been applied. Two categories are shown: both electrons in EB with $|\eta| < 1$ $R_9 > 0.94$ (left), and both electrons in EE with $|\eta| < 2$ $R_9 < 0.94$ (right).

The final validation of both the energy corrections (scale and resolution) and the selection efficiency scale factors presented in Section 7.3.3 is shown in Figure 7.15 where the dielectron invariant mass distribution in data is compared with the MC one, after requiring the photon identification criteria described in Section 7.3.2 (inverting the electron veto) and applying the selection efficiency scale factors to MC together with the efficiency of the HLT selection (see Section 6.2). The SingleEle_35_WPLoose trigger has been required in data. The agreement between data and simulation is good for both the analysis categories: EBEB (left plot of Figure 7.15) and EBEE (right plot of Figure 7.15) As further check, Figure 7.16 shows the same regions of Figure 7.15 (EBEB and EBEE) buth further splitted in R_9 categories. Again, the data/MC agreement is rather good.

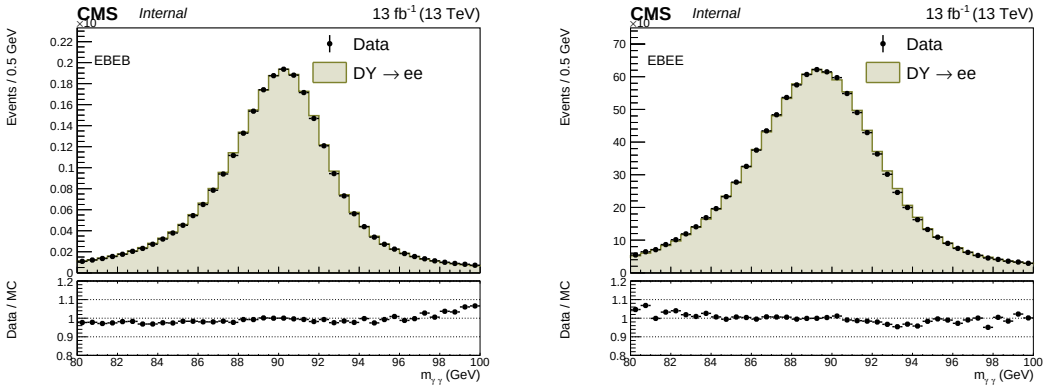

Figure 7.15: Comparison between the predicted and observed invariant mass distribution of electron pairs obtained after the application of the energy corrections (scale to data and resolution to MC) and the selection efficiency scale factors to MC for events with both electrons in the barrel region (left) and events with one electron in the barrel region and one in the endcaps (right).

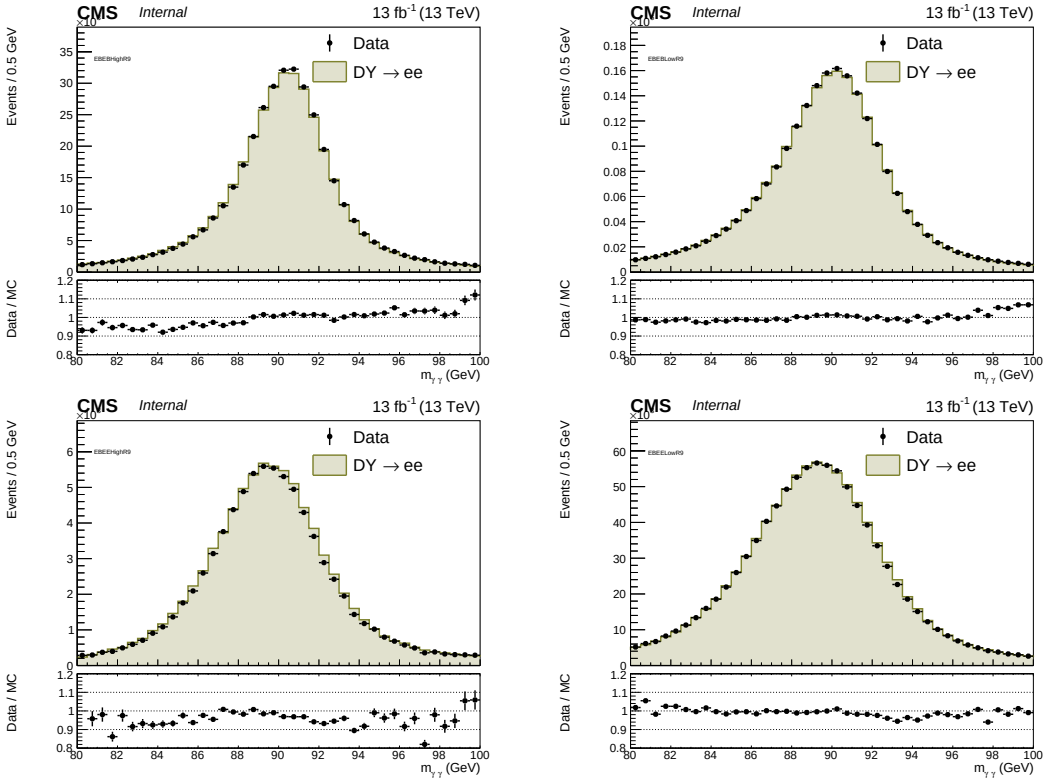

Figure 7.16: Comparison between the predicted and observed invariant mass distribution of electron pairs obtained after the application of the energy corrections (scale to data and resolution to MC) and the selection efficiency scale factors to MC. Distributions are shown for events where both electrons are reconstructed in the barrel, with at least one having $R_9 > 0.94$ (top left) and at least one having $R_9 < 0.94$ (top right), as well as where one electron is reconstructed in barrel and the other in the endcap with at least one having $R_9 > 0.94$ (bottom left) and at least one having $R_9 < 0.94$ (bottom right).

7.4.1 Systematic uncertainties on energy corrections

The total systematic uncertainty on the energy corrections is broken into several components which are then added in quadrature assuming that they are uncorrelated. In particular, an important source of systematic uncertainty in the determination of the energy corrections to be applied in the search for high mass resonances decaying in the

diphoton final state is the intrinsic difference between electrons and photons. This systematic source can be estimated by recomputing the energy corrections (only with the smearing method) using the electron-tuned energy regression (instead of the photon-tuned one used in the derivation of the nominal corrections) and taking the difference with respect to the nominal values. This systematic uncertainty is defined as Δ_{ele} in the text. The event selection has also been varied, using the medium and tight working points for the electron identification criteria of the electron/photon POG and the cut on the electron transverse energy E_{T}, from 20 GeV to 25 GeV. This systematic uncertainties are called respectively Δ_{sel} and $\Delta_{E_{\mathrm{T}}}$. The total systematic uncertainty is then defined via the following relation:

$$\Delta^{tot}_{syst} = \Delta^2_{ele} + \Delta^2_{sel} + \Delta^2_{E_{\mathrm{T}}} \tag{7.3}$$

Table 7.10 shows the different systematic uncertainties on the energy scale corrections in the 8 $\eta \times R_9$ categories, as well as the total systematic uncertainty and the statistical one. Table 7.11 has the same structure but for the energy resolution corrections.

Category		Δ_{ele}[%]	Δ_{E_T}[%]	Δ_{sel}[%]	Δ^{tot}_{syst}[%]	Δ_{stat}[%]
EB: $\lvert\eta\rvert < 1$	$R_9 > 0.94$	0.02	0.00	0.01	0.02	0.01
EB: $\lvert\eta\rvert < 1$	$R_9 < 0.94$	0.00	0.00	0.01	0.02	0.01
EB: $\lvert\eta\rvert > 1$	$R_9 > 0.94$	0.02	0.01	0.03	0.05	0.04
EB: $\lvert\eta\rvert > 1$	$R_9 < 0.94$	0.02	0.02	0.05	0.06	0.01
EE: $\lvert\eta\rvert < 2$	$R_9 > 0.94$	0.05	0.04	0.04	0.09	0.04
EE: $\lvert\eta\rvert < 2$	$R_9 < 0.94$	0.05	0.00	0.04	0.07	0.02
EE: $\lvert\eta\rvert > 2$	$R_9 > 0.94$	0.02	0.03	0.01	0.05	0.02
EE: $\lvert\eta\rvert > 2$	$R_9 < 0.94$	0.05	0.09	0.02	0.13	0.03

Table 7.10: Systematic uncertainties on the energy scale corrections in the 8 $\eta \times R_9$ categories.

Category		Δ_{ele}[%]	Δ_{sel}[%]	$\Delta_{E_{\mathrm{T}}}$[%]	Δ^{tot}_{syst}[%]	Δ_{stat}[%]
EB: $\lvert\eta\rvert < 1$	$R_9 > 0.94$	0.04	0.01	0.01	0.04	0.03
EB: $\lvert\eta\rvert < 1$	$R_9 < 0.94$	0.07	0.05	0.01	0.09	0.03
EB: $\lvert\eta\rvert > 1$	$R_9 > 0.94$	0.09	0.16	0.06	0.19	0.11
EB: $\lvert\eta\rvert > 1$	$R_9 < 0.94$	0.01	0.05	0.01	0.05	0.04
EE: $\lvert\eta\rvert < 2$	$R_9 > 0.94$	0.01	0.15	0.03	0.15	0.08
EE: $\lvert\eta\rvert < 2$	$R_9 < 0.94$	0.03	0.02	0.03	0.05	0.05
EE: $\lvert\eta\rvert > 2$	$R_9 > 0.94$	0.01	0.08	0.02	0.08	0.05
EE: $\lvert\eta\rvert > 2$	$R_9 < 0.94$	0.07	0.07	0.02	0.10	0.06

Table 7.11: Systematic uncertainties on the energy resolution correction in the 8 $\eta \times R_9$ categories.

7.4.2 Energy scale behaviour at high-E_{T}

Given the importance in this analysis, a further check is performed in order to ensure that the energy scale corrections are still valid at high-E_{T} regions. This study exploits again the dielectron invariant mass distribution, but after having applied the nominal set of energy scale corrections (described in the previous sections) to data, and using boosted $\mathrm{Z} \to \mathrm{ee}$ events, where at least one the decay products carries large E_{T}, with the advantage of preserving the Z peak which can still be used as a handle.

In particular, the smearing method is used to verify if there is a need for extra scale corrections in different E_{T} categories, especially for $E_{\mathrm{T}} > 150$ GeV. Given the reduced

amount of statistics in the high-E_T region, the 8 $\eta \times R_9$ categories defined for the calibration procedure have been merged in a single EB category and a single EE category. The data/MC comparison for EB and EE are shown in Figures 7.17 and 7.18 respectively, as well as the negative logarithm of the likelihood (NLL) used in the derivation of the residual ΔP corrections. Even if the amount of statistics at high-E_T is such that the likelihood profiles cannot be considered to be "smooth", a minimum in the NLL is clearly visible in both EB and EE cases. Morever the purpose of this study is not the exact derivation of additional corrections at high-E_T but rather to proof that the energy scale in data is under control. It can be seen that the residual ΔP corrections do not exceed the systematic uncertainty of 1% assigned to the energy scale in the final analysis results (see Section 7.8.1), being at the level of 0.3 (0.8)% for EB (EE).

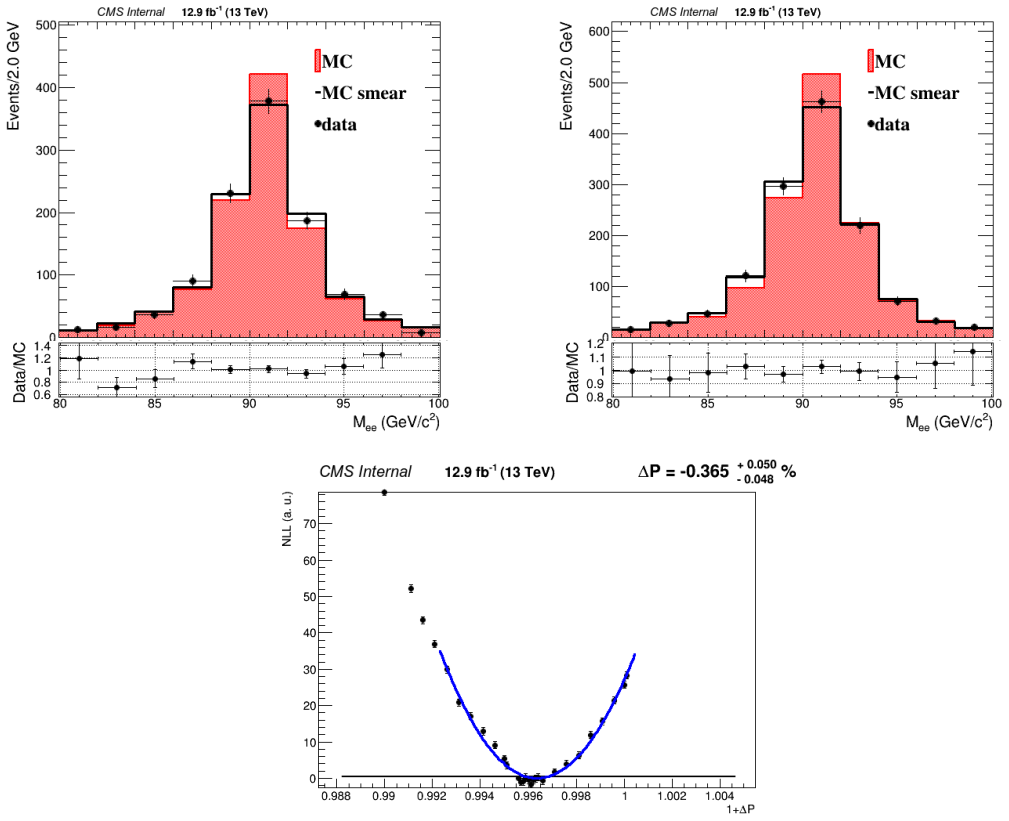

Figure 7.17: Comparison between the invariant mass distribution in data and MC for two event categories with both electrons in EB: both electrons with $100 < E_T < 150$ GeV (top left), one electron with $100 < E_T < 150$ GeV and the other one with $E_T > 150$ GeV (top right). The profile of the negative logarithm of the likelihood (NLL) used in order to derive the residual scale correction ΔP is also shown (bottom) for the single electron category with $E_T > 150$ GeV.

7.5 Background model

The strategy of the analysis is meant to be as simple as possible: it looks for a localised excess of events in the diphoton invariant mass spectrum $m_{\gamma\gamma}$ over the SM background, which needs to be characterized.
There are three background sources for the search of high mass resonances decaying in the diphoton final state:

- An irreducible source of background comes from the SM production of two prompt photons in the final state either via the qq$\rightarrow \gamma\gamma$ process or via the gg$\rightarrow \gamma\gamma$ process

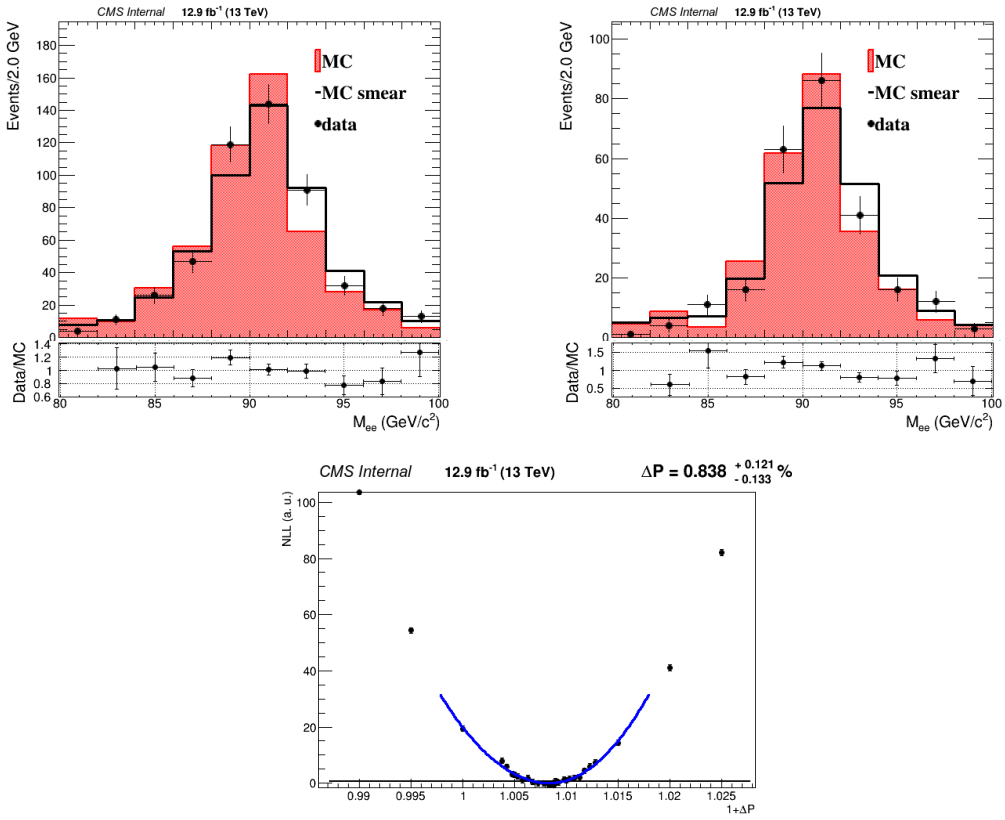

Figure 7.18: Comparison between the invariant mass distribution in data and MC for two event categories with both electrons in EE: both electrons with $100 < E_{\mathrm{T}} < 150$ GeV (top left), one electron with $100 < E_{\mathrm{T}} < 150$ GeV and the other one with $E_{\mathrm{T}} > 150$ GeV (top right). The profile of the negative logarithm of the likelihood (NLL) used in order to derive the residual scale correction ΔP is also shown (bottom) for the single electron category with $E_{\mathrm{T}} > 150$ GeV.

(which proceeds through quark loop). In the following the irreducible background will be referred to as $\gamma\gamma$.

- A second (reducible) source of background comes from the qg$\rightarrow \gamma j$, where the jet j in the final state is misidentified as a photon. This second type of background will be referred to as γj.

- The third (reducible) source of background comes from QCD processes producing at least two jets in the final state, misidentified as photons. This third type of background will be referred to as jj.

Instead of relying on the MC description of these sources of background, a data-driven approach has been chosen where the shape of background in the $m_{\gamma\gamma}$ spectrum is described as a parametric function, whose parameters are derived by performing a fit on the $m_{\gamma\gamma}$ distribution in data (see next section).

7.5.1 Background parametrization and bias estimation

The choice of an appropriate parametric function able to describe the $m_{\gamma\gamma}$ spectrum is crucial in order to ensure the accuracy of the background prediction and, at the same time, guarantee a good analysis sensitivity.
First, the background shape has been parametrized, for both analysis categories, via the

ansatz functional form $g(m_{\gamma\gamma})$ defined as:

$$g(m_{\gamma\gamma}) = m_{\gamma\gamma}^{a\ +\ b\log(m_{\gamma\gamma})} \tag{7.4}$$

As an example, Figure 7.19 shows the results of the fit to the $m_{\gamma\gamma}$ spectra obtained by using the ansatz funcional form defined above on the Asimov pseudo-dataset[11].

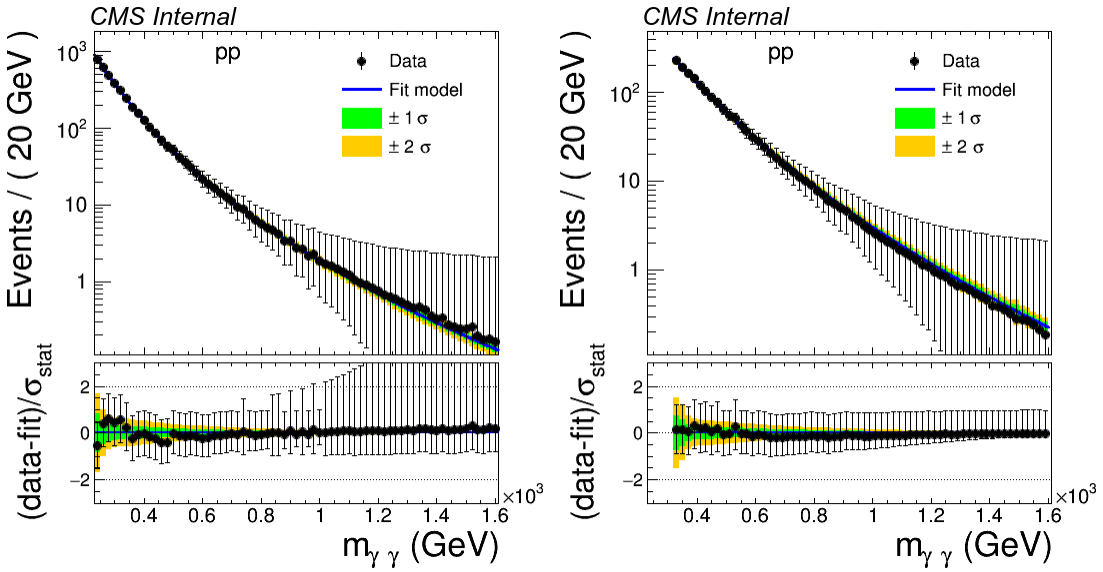

Figure 7.19: The Asimov pseudo-dataset, showing the results of the fit to the $m_{\gamma\gamma}$ spectrum, for the EBEB (left) and for the EBEE category (right) obtained by using the background parametrization $g(m_{\gamma\gamma}) = m_{\gamma\gamma}^{a\ +\ b\log(m_{\gamma\gamma})}$. The statistical uncertainty on the pseudo-data is reflecting the expected uncertainty in data for 10 fb^{-1}. In order to avoid introducing local distortions in the $m_{\gamma\gamma}$ spectrum, the events are not weighted to match the data pile-up distribution [145].

The accuracy of the background prediction obtained by fitting the $m_{\gamma\gamma}$ spectrum with an unbinned maximum-likelihood fit and so deriving the a and b coefficients of the $g(m_{\gamma\gamma})$ function, is evaluated through the following procedure:

- A "true" background distribution $h(m_{\gamma\gamma})$ is derived by making histograms of the $m_{\gamma\gamma}$ distribution of the $\gamma\gamma$ MC samples (the SHERPA samples detailed in Table 7.2). The available MC samples have been generated in several mass bins covering the $m_{\gamma\gamma}$ interval $60 < m_{\gamma\gamma} < 13000$ GeV. Furthermore, the LO cross-section is scaled by a k-factor of 1.4 to take into account higher order corrections (this has to be considered as an effective estimation of the non-available k-factor from the full computation of the cross-sections at higher orders in the perturbative theory).

- Pseudo-data are extracted from the true distribution $h(m_{\gamma\gamma})$ making toy experiments labeled as t_i. The number of events in the t_i corresponds to the expected number of events for an integrated luminosity of 10 fb^{-1}.

- An unbinned maximum likelihood fit is performed for each toy experiment with the functional form $g(m_{\gamma\gamma})$, obtaining the a and b coefficients for the experiment i. Hence, to each experiment t_i correspond a fitted g function, labeled as $\hat{g}_i(m_{\gamma\gamma})$.

[11]The Asimov pseudo-dataset is defined as the particular binned dataset, where the number of events for each bin is taken directly from the corresponding MC simulation, while the error in the same bin is taken from to the expectations in data. The name *Asimov* is inspired by the short story *Franchise*, by Isaac Asimov, where elections are held by selecting the single most representative voter to replace the entire electorate.

- The predicted number of events, obtained by integrating the fitted function $\hat{g}_i(m_{\gamma\gamma})$ over a certain mass window, is compared with "true" number of events, obtained by integrating the "true" background distribution $h(m_{\gamma\gamma})$ in the same mass window. This comparison is done in several mass windows j and the pull test statistics is constructed as:

$$p_i^j = \frac{N_{\hat{g}_i}^j - N_h^j}{\sigma(N_{\hat{g}_i}^j)}$$

  where $\sigma(N_{\hat{g}_i}^j)$ is the error on the predicted number of events coming from the errors on the fitted parameters of $\hat{g}_i$. The j index runs over the mass windows, while the i index runs over the pseudo-experiments.

- Given a set of mass window intervals j (three sets have been considered, see later), the background parametrization $g(m_{\gamma\gamma})$ is considered to be accurate if, for the median value of the pull tests over the toy experiments, holds the following relation:

$$b^j = \mid median_i \left(p_i^j\right) \mid < 0.5 \tag{7.5}$$

  for all the mass windows j.
  This request can be interpreted as the request that the deviation between the "true" number of background events and the fitted number of background events, coming from the chosen functional form does not exceed half of the statistical uncertainty coming the fit. The systematic uncertainty coming from the choice of the fitting function is then neglected as it is "reasonably" smaller[12] than the statistical accuracy [151].

- If the criterion from Equation (7.5) is not accomplished, the pull test statistics is modified as follows:

$$\tilde{p}_j^i = \frac{N_{\hat{g}_i}^j - N_h^j}{\sqrt{\sigma^2(N_{\hat{g}_i}^j) + \beta_j^2}} \tag{7.6}$$

  where β_j represents an additional uncertainty ("bias term") that is additionally assigned to the model. The β_j term can be written as the integral in the mass window j of a certain β function (which is determined for convenience, since it allows a useful parametrization of the bias term): $\beta_j = \int_j \beta(m_{\gamma\gamma}) dm_{\gamma\gamma}$. The bias criterion can then be modified exchanging p with $\tilde{p}$.
  Consequently, the criterion that allows to validate the accuracy of the background parametrization writes as:

$$\tilde{b}^j = \mid median_i \left(\tilde{p}_i^j\right) \mid < 0.5. \tag{7.7}$$

The three sets of mass window intervals considered in the study were:

- The "default" set was built with variable bin width from 500 GeVup-to 5500 GeV: these mass windows correspond to the set used for the analysis of the 2015 dataset only, reported in a previous version of this analysis [149].
- A second set, listed in Table 7.13, was built to have a 5% width of the $m_{\gamma\gamma}$ value.
- A third set, listed in Table 7.14, was built with a 10% width.

[12] In the searches for exotic new resonances the threshold is set to 0.5, while in the search for the SM Higgs boson decaying to photons $H \to \gamma\gamma$ this threshold was tightened to be 0.2 [151].

$m_{\gamma\gamma}$ min (GeV)	$m_{\gamma\gamma}$ max (GeV)	$m_{\gamma\gamma}$ min (GeV)	$m_{\gamma\gamma}$ max (GeV)
500	550	900	1000
550	600	1000	1200
600	650	1200	1800
650	700	1800	2500
700	750	2500	3500
750	800	3500	4500
800	900	4500	5500

Table 7.12: List of mass window regions considered for the bias determination with a variable bin width from 500 GeVto 5500 GeV. This was the list used for the analysis of the 2015 dataset [149].

$m_{\gamma\gamma}$ min (GeV)	$m_{\gamma\gamma}$ max (GeV)	$m_{\gamma\gamma}$ min (GeV)	$m_{\gamma\gamma}$ max (GeV)	$m_{\gamma\gamma}$ min (GeV)	$m_{\gamma\gamma}$ max (GeV)
500	525	1040	1090	2160	2270
525	550	1090	1145	2270	2380
550	580	1145	1205	2380	2500
580	610	1205	1265	2500	2625
610	640	1265	1325	2625	2760
640	670	1325	1395	2760	2895
670	705	1395	1465	2895	3040
705	740	1465	1535	3040	3195
740	775	1535	1615	3195	3350
775	815	1615	1695	3350	3520
815	855	1695	1780	3520	3695
855	900	1780	1865	3695	3880
900	945	1865	1960	3880	4000
945	990	1960	2060		
990	1040	2060	2160		

Table 7.13: List of mass window regions considered for the bias determination with a bin width of 5%.

$m_{\gamma\gamma}$ min (GeV)	$m_{\gamma\gamma}$ max (GeV)	$m_{\gamma\gamma}$ min (GeV)	$m_{\gamma\gamma}$ max (GeV)
500	550	1425	1570
550	605	1570	1725
605	665	1725	1900
665	730	1900	2090
730	805	2090	2295
805	885	2295	2525
885	975	2525	2780
975	1070	2780	3060
1070	1180	3060	3365
1180	1295	3365	3700
1295	1425	3700	4000

Table 7.14: List of mass window regions considered for the bias determination with a bin width of 10%.

Moreover, several variations of the $m_{\gamma\gamma}$ spectrum have been taken into account when constructing the "true" background distribution $h(m_{\gamma\gamma})$. For each considered variation of the $m_{\gamma\gamma}$ spectrum, the study detailed above has been repeated. The following variations have been considered:

- The $m_{\gamma\gamma}$ spectrum predicted by the Sherpa MC samples is used as "default". A fit of the chosen background parametrization $g(m_{\gamma\gamma})$ in the default case has already been shown in Figure 7.19.
- A $m_{\gamma\gamma}$-dependent k-factor [148] is applied to take into account higher order correc-

tions up to NNLO ("pNNLO").

- An artificial spectrum is generated, inverting (i.e. instead of applying the k-factor itself, applying a "(2-k)-factor") the k-factor ("mNNLO").
- The contribution from reducible background is taken into account, parametrizing the background composition measured in the 2015 data [149] ("pFakes").

The modified bias criterion in Equation (7.7) was met for all the three sets of mass window regions considered and for all the $m_{\gamma\gamma}$ spectrum variations using as β functions the ones reported in Table 7.15, which allow a convenient parametrization of the bias term.

category	$m_{\gamma\gamma}$ range	$\beta(m_{\gamma\gamma})$/L(fb/GeV)
EBEB	> 230 GeV	$0.1 \times m_{\gamma\gamma}^{(2.2-0.4\times\log(m_{\gamma\gamma}))}$
EBEE	> 320 GeV	$0.01 \times \left(\frac{m_{\gamma\gamma}}{600\ \text{GeV}}\right)^{-5} + 2 \times 10^{-6}$

Table 7.15: β functions, divided by the integrated luminosity L, for the two analysis categories.

The median of the pull tests (b) and the median of the pull tests modified with the inclusion of the bias term ($\tilde{b}$) are shown for the default test region of Table 7.12 in Figure 7.20 for the EBEB and EBEE category. It can be seen that while the b values are sometimes outside the gray band corresponding to the 0.5 threshold, the $\tilde{b}$ are always inside this band. The results for the set of mass window regions with a 5% and 10% width are shown in Figures 7.21 and 7.22. Given that the chosen bias term parametrization is able to meet the requirement of Equation (7.7) for all the different sets of mass window regions considered and for all the MC shape variations, the background parametrization procedure is considered to be validated.

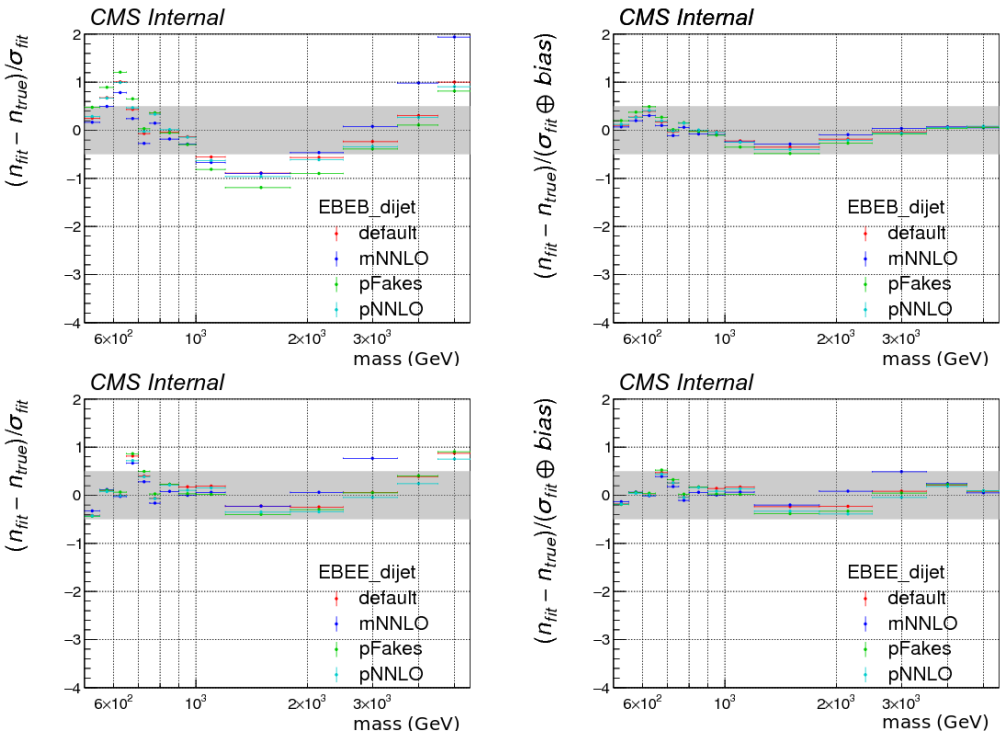

Figure 7.20: Median of the pull (left) and modified pull (right) for all considered mass window regions according to Table 7.12 for EBEB (top row) and EBEE (bottom row). Different datasets correspond to different MC shape variations as specified in the main text [145].

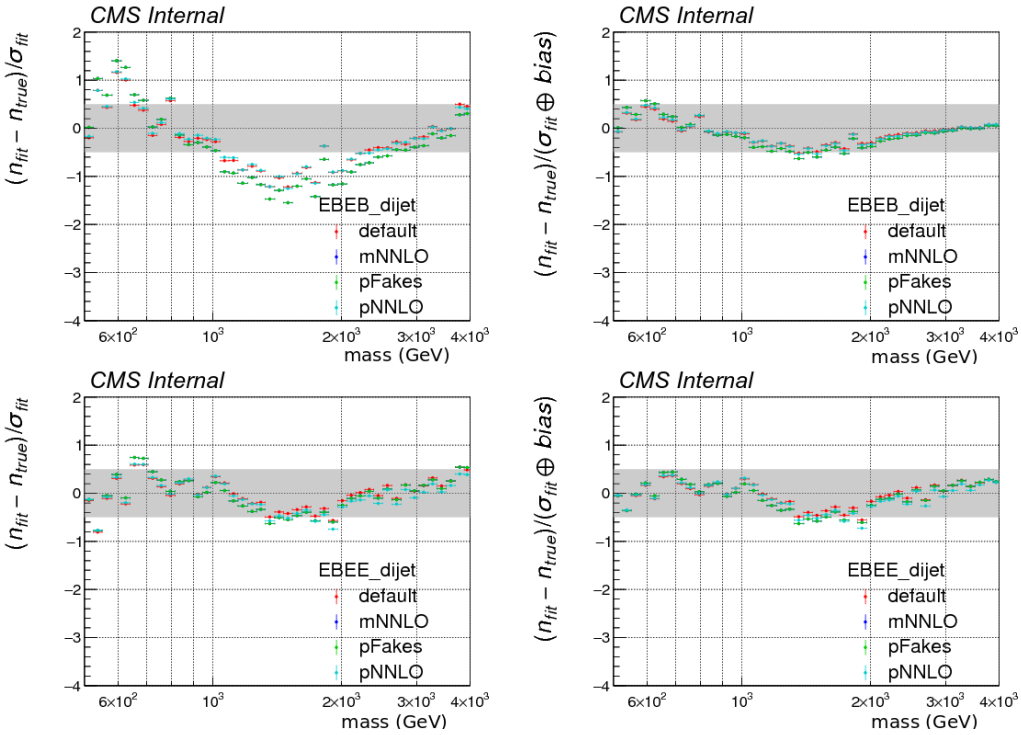

Figure 7.21: Median of the pull (left) and modified pull (right) for all considered mass window regions according to Table 7.13 for EBEB (top row) and EBEE (bottom row). Different datasets correspond to different MC shape variations as specified in the main text [145].

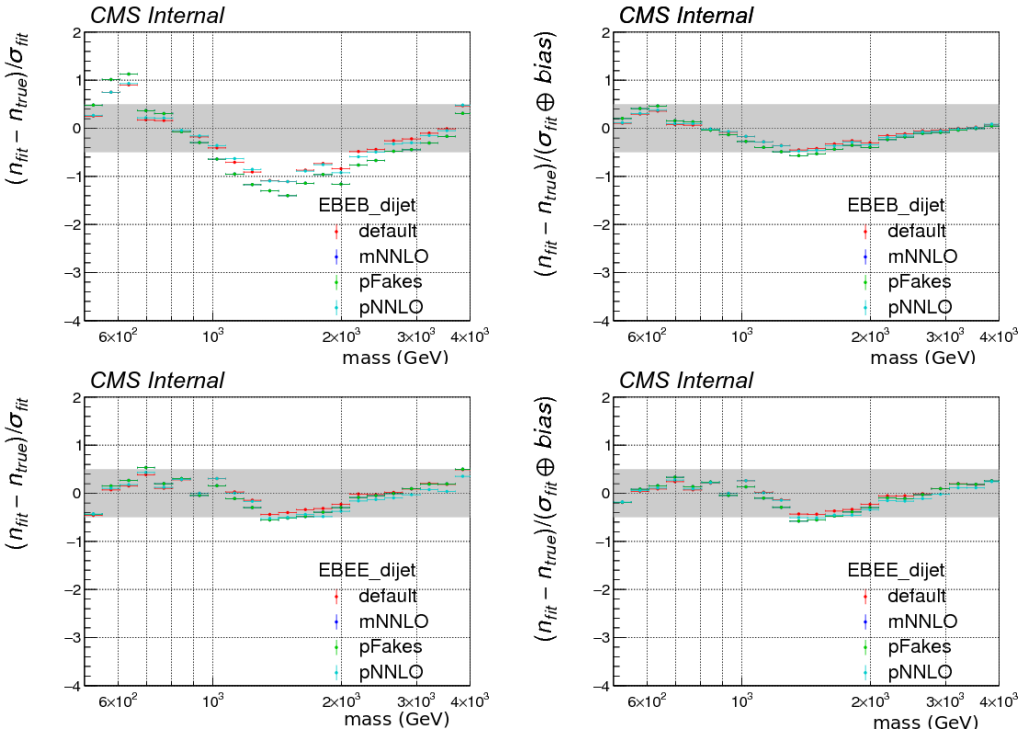

Figure 7.22: Median of the pull (left) and modified pull (right) for all considered mass window regions according to Table 7.14 for EBEB (top row) and EBEE (bottom row). Different datasets correspond to different MC shape variations as specified in the main text [145].

7.6 Invariant mass spectra

The observed diphoton invariant mass spectra are shown in Figure 7.23 for the EBEB and EBEE categories. Events are selected requiring the main analysis trigger (see Section 6.2) and the full photon selection detailed in Section 7.3. The first half of the 2016 dataset has been considered for the results reported here, for an integrated luminosity of 12.9 fb^{-1}. For both the analysis categories, the mass spectra are parametrized using the functional form

$$g(m_{\gamma\gamma}) = m_{\gamma\gamma}^{a\ +\ b\ \log(m_{\gamma\gamma})}$$

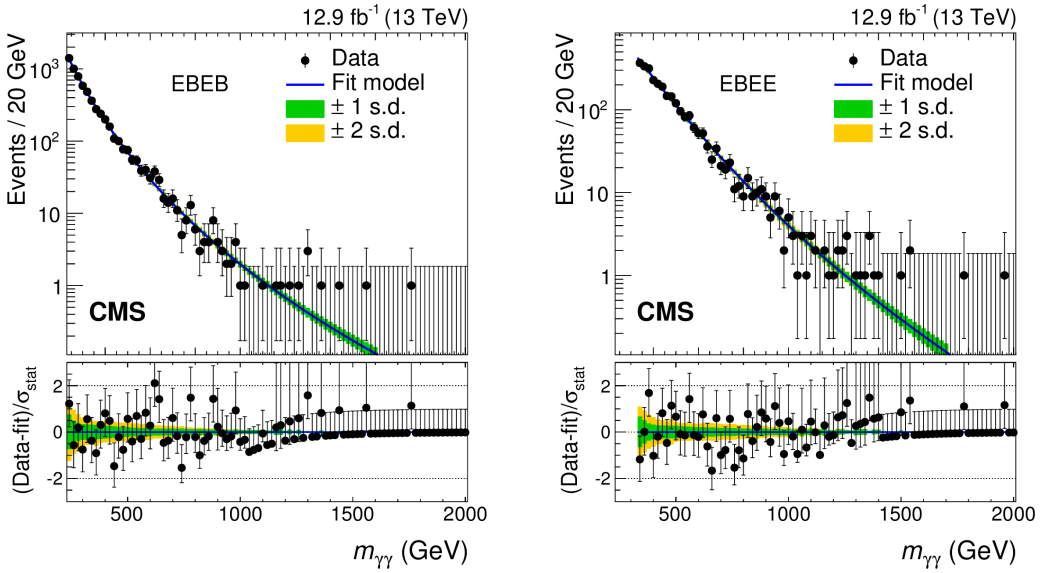

Figure 7.23: Observed diphoton invariant mass spectrum in the EBEB (left) and EBEE (right) category. The result of the parametric fit is superimposed to the data points, together with bands representing the statistical uncertainties on the knowledge of the background shape [133].

In Figure 7.24 the diphoton invariant mass spectrum in data is compared to the MC simulation predictions (the two main sources of background $\gamma\gamma$ and γj are considered) for the two categories used in the analysis. The MC predictions are rescaled by a k-factor equal to 1.4, in order to account for higher order effects.

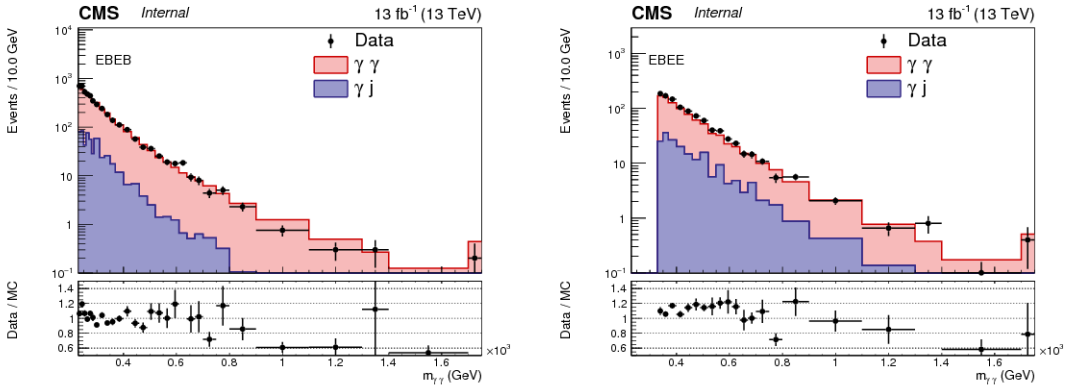

Figure 7.24: Data/MC comparison of the diphoton invariant mass distribution for EBEB (left) and EBEE (right) [145].

The total number of selected events N_{ev} is also reported in Table 7.16 divided per analysis category. The full mass region is defined by the diphoton invariant mass threshold

$m_{\gamma\gamma} > 230$ GeV for the EBEB case, and $m_{\gamma\gamma} > 320$ GeV for the EBEE case. The number of events in the search region $m_{\gamma\gamma} > 500$ is also quoted for both analysis categories.

Mass cut	$m_{\gamma\gamma} > 230$ GeV	$m_{\gamma\gamma} > 320$ GeV	$m_{\gamma\gamma} > 500$ GeV	$m_{\gamma\gamma} > 500$ GeV
Category	EBEB	EBEE	EBEB	EBEE
N_{ev} in data	6284 ± 79	2791 ± 53	461 ± 21	800 ± 28

Table 7.16: Observed number of events in the two analysis categories.

As can be seen from the bottom part of Figure 7.23 which shows the difference between the number of events in data and the number of background events coming from the background fit using the functional form g, divided by the statistical uncertainty of the fit, no striking sign of new physics is visible in the diphoton mass spectra with a maximum deviation of $\sim 2\sigma$ for $m_{\gamma\gamma} \sim 600$ GeV in the EBEB category, while a deficit of $\sim 1.5\sigma$ is observed for $m_{\gamma\gamma} \sim 750$ GeV[13].

Comparison with 2015 dataset

The number of selected events in the first half of the 2016 dataset has been also compared, in data and MC, with the one obtained for the 2015 analysis [149]. In Table 7.17 the number of selected events N_{ev} is compared in the full analysis region and in the search region for the two datasets normalized to an integrated luminosity of 1 fb^{-1} for the two analysis categories. Again, a k-factor of 1.4 is applied to the MC predictions for both datasets.

Mass cut	$m_{\gamma\gamma} > 230$ GeV		$m_{\gamma\gamma} > 500$ GeV	
EBEB				
Dataset	2015	2016	2015	2016
N_{ev} in $\gamma\gamma$	389	460	34	35
N_{ev} in γj	124	57	6	2.8
Total N_{ev} in MC	513	517	40	38
N_{ev} in data	458	487	36	36
EBEE				
Dataset	2015	2016	2015	2016
N_{ev} in $\gamma\gamma$	152	168	47	47
N_{ev} in γj	82	44	23	11
Total N_{ev} in MC	234	212	70	58
N_{ev} in data	230	216	68	62

Table 7.17: Number of events selected in data and MC in the EBEB category (first block) and in the EBEE category (second block). A k-factor of 1.4 is applied to the MC predictions. Numbers refer to a luminosity of 1 fb^{-1}.

The agreement in the (normalized) number of events between the 2015 and 2016 datasets has been considered in order to ensure that no major inefficiencies have been introduced between the two analyzed datasets. Hence, given that no clear excesses are seen over the SM background expectations, upper limits on the production cross-section times branching ratio for a heavy resonance decaying in the diphoton final state are computed. The statistical treatment is explained in details in Section 7.8. Preliminarly, the signal model description is explained in the next section, as it is an input for the upper limit computation.

[13] A more accurate statistical interpretation of the results is reported in Section 7.8.2. The statements reported here about the deviation in term of σ have to be considered purely heuristic.

7.7 Signal model

In order to statistically interpret the results, it is necessary to have a description of the signal shape distribution (f_S in the Equation (7.10)). Two possible approaches have been considered: a parametric model and a non-parametric model based on the MC samples. The former is the one used in the analysis, while the latter was the one used for the 2015 analysis and no longer used. The old method is briefly described for reference and also to allow a comparison between the two techniques. The main advantage when using a parametric model is the possibility to perform a mass scan that could be fine at will, starting from a simulated sample with only generator level quantities.

7.7.1 Parametric signal model

The building strategy for the parametric signal model starts by obtaining the diphoton invariant mass distribution for the signal MC samples, after the application of the energy corrections described in Section 7.4 and the efficiency scale factors described in Section 7.3.3. The aim is to describe the simulated diphoton invariant mass shape with an analytic function in which the hypothetical signal mass (m_X) represents a parameter which can vary continuously for any value in the range of interest of the search. This strategy can be extended in order to include as additional parameter the natural width of the new resonance (Γ_X). The procedure is defined below.

The theoretical signal shape of the heavy resonance is described by the functional form of a relativistic Breit-Wigner centered at m_X and with the expected natural width of Γ_X. The relativistic Breit-Wigner distribution can be fitted with a double-sided crystal ball (dCB) function[14]. To prove this statement the generated diphoton invariant mass for a resonance mass of 750 GeV is obtained and a dCB fit superimposed. The results of the fit are shown in Figure 7.25, where it can be seen that the dCB function describes well the shape at generator level.

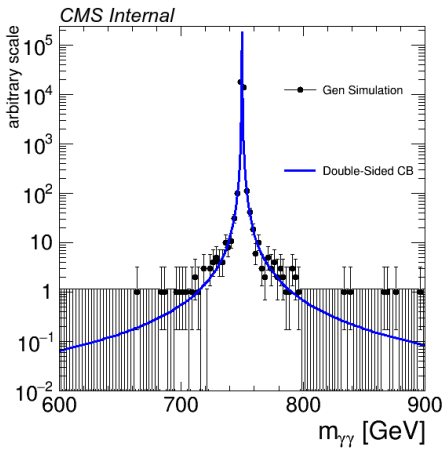

Figure 7.25: Double-sided crystal ball fit (blue line) to the generated diphoton invariant mass distribution for $m_X = 750$ GeV, under the narrow width hypothesis ($\Gamma_X/m_X = 0.01$) [145].

The reduced mass Δm (response distribution), defined as the difference between the reconstructed mass (m_{reco}) and the generated one (m_{true}):

$$\Delta m = m_{reco} - m_{true} \tag{7.8}$$

[14]The dCB function has already been defined in Section 6.4. As a reminder, the dCB is defined as a gaussian core with mean m_0 and sigma σ and two asymmetric power-law tails defined by two different different n which are connected to the gaussian core at $\alpha \cdot \sigma$ distance from the mean.

is computed for each available reconstructed mass point (m_x= 500, 750, 1000, 1250, 1500, 1750, 2000, 2250, 2500, 2750, 3000, 4000 GeV). In order to construct the parametric model the Δm distribution is fitted again using a dCB function.
In Figure 7.26 the fit to the reduced mass is shown for the signal mass point $m_X = 750$ GeV in linear and logarithmic scale, for both analysis categories.

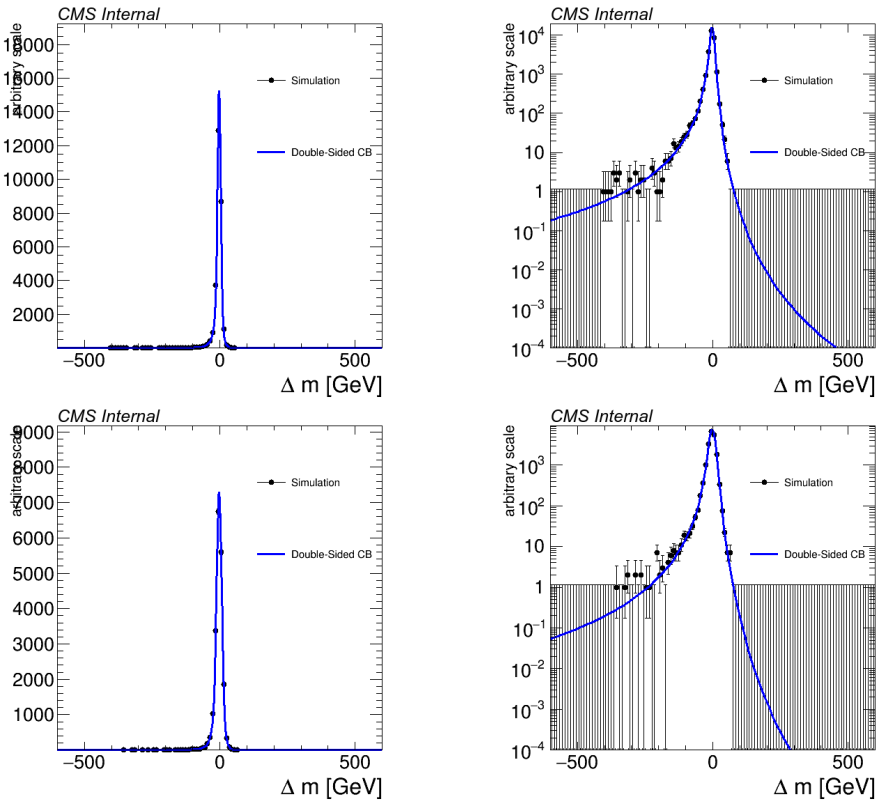

Figure 7.26: Double-sided crystal ball fit (blue line) to the Δm distributions for $m_X = 750$ GeV in linear scale (left) and logarithmic scale (right). Barrel-barrel category is shown on the top and barrel-endcap category on the bottom [145].

The fit function of the theoretical diphoton invariant mass shape is then convoluted with the fit function coming from the Δm distribution to account for the experimental resolution of the detector. The convoluted shape is compared with the reconstructed mass shape in MC, as a closure test of the fitting model. The closure test is shown in Figure 7.27 for $m_X = 750$ GeV, where a close-to-optimal level of agreement is found.

As final closure test, the convoluted model obtained in the previous step is used to throw a pseudo-dataset (Asimov), which is again fitted with a dCB function. The closure test fit to the pseudo-dataset is shown in Figure 7.28 for $m_X = 750$ GeV and for the narrow width hypothesis ($\Gamma_X/m_X = 0.01$). This last fitting model represents the final parametric description of the diphoton mass distribution for a resonance of mass m_X and for a given width Γ_X, as used in the limit setting tools (see Section 7.8).

The signal model derived in the previous steps depends continuously on m_X through the parameters of the dCB function which defines the model itself: mean, σ, $\alpha_{L/R}$, $n_{L/R}$. For each category and for each width hypothesis the behaviour of these parameters as a function of m_X is studied and modelled with a polynomial function, using m_X as the only independent variable.

The parametric signal model obtained is shown in Figure 7.29 as a function of m_X for the different width hypotheses.

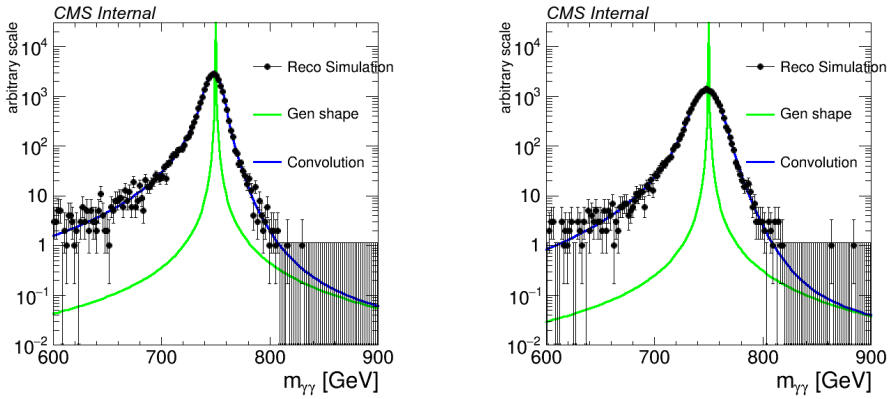

Figure 7.27: Convoluted model (blue line) of the generator-level function (green line) and the response function compared with the reconstructed diphoton invariant mass distribution for a resonance of mass $m_X = 750$ GeV for the barrel-barrel (left) and barrel-endcap category (right), both in logarithmic scale [145].

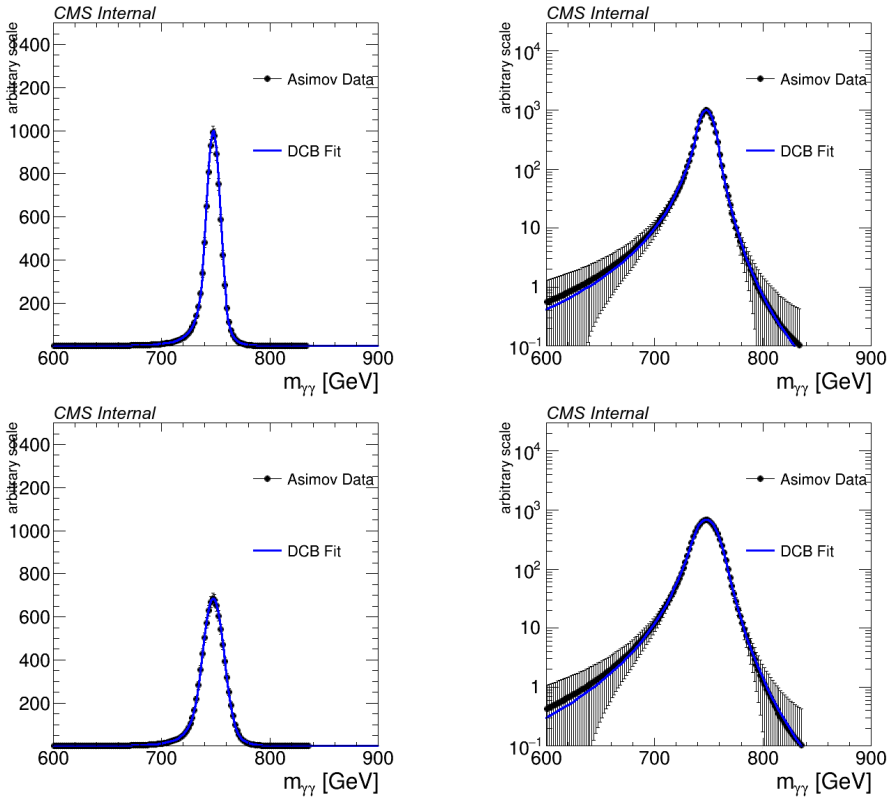

Figure 7.28: Final dCB fit (blue line) of the diphoton mass distribution from a pseudo-experiment generated from the convoluted model described in the text, for a resonance mass of $m_X = 750$ GeV in linear scale (left) and logarithmic scale (right). Barrel-barrel category is shown on the top and barrel-endcap category on the bottom [145].

7.7.2 Non-parametric signal model

For the 2015 analysis, a non-parametric signal model was employed. A description of the method is given here for reference, and so that the performance of the parametric signal model can be compared with it.
First, a binned signal model for all available simulated signal samples is obtained. The

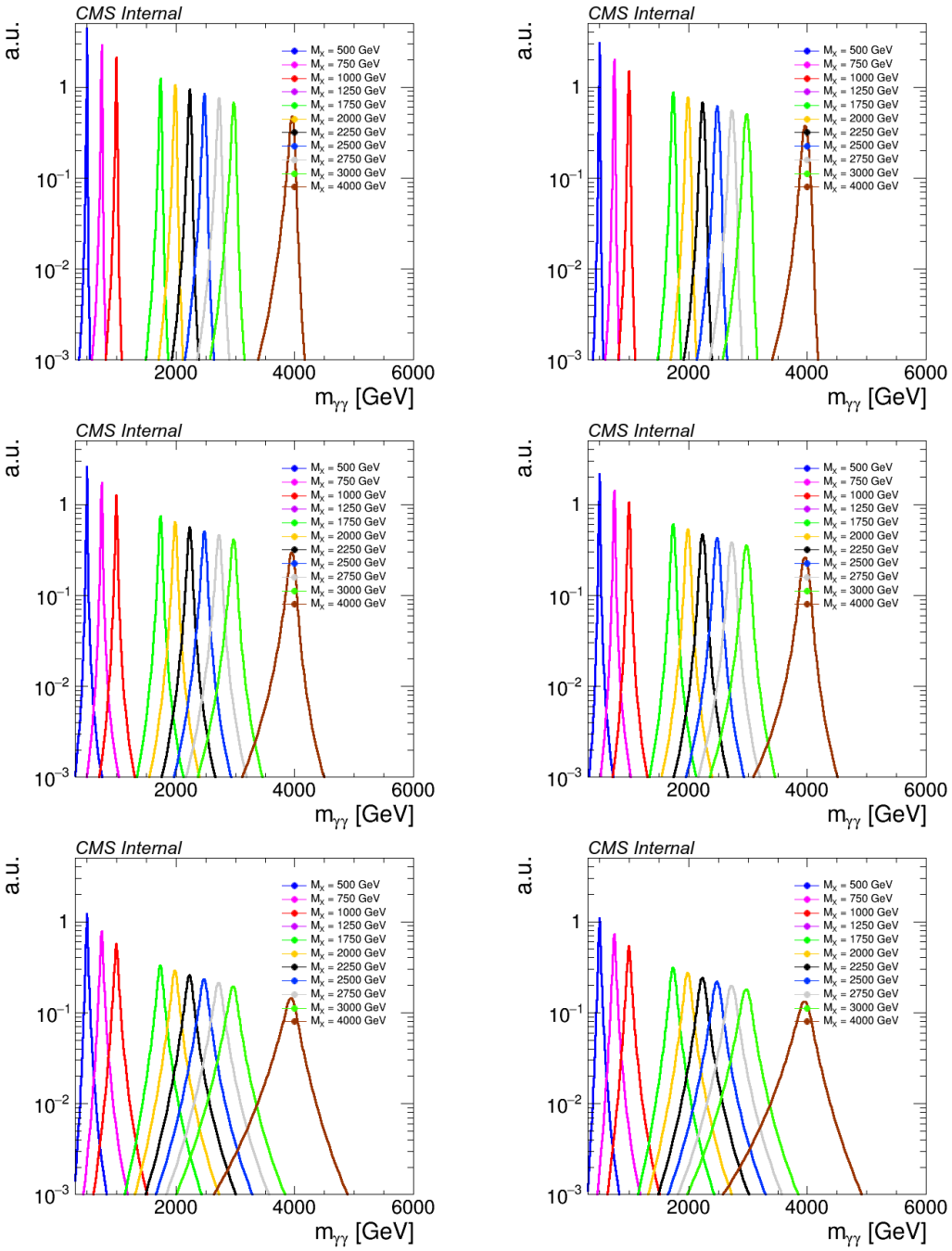

Figure 7.29: Parametric signal model as a function of the resonance mass m_X for different width hypotheses in the barrel-barrel category (left) and barrel-endcap category (right). From top to bottom the width increases as follows: $\Gamma/m = 0.14, 1.4\%, 5.6\%$ [145].

diphoton invariant mass distribution is obtained using the MC samples as input after the application of the energy corrections and efficiency scale factors.
Then, each histogram is fitted with an analytic function for all the available mass points. Finally, the signal model for a generic mass point is defined by an interpolation of the fitted histograms obtained in the previous step using a morphing technique documented in [153].

In order to evaluate the difference between the two signal models, expected limits on the cross-section times branching ratio for a new resonance decaying in diphoton final

state have been computed as described in Section 7.8.2 using both the parametric signal model and the non-parametric signal model described above. The comparisons for both spin 0 and spin 2 hypotheses and for the different width hypotheses are shown in Figure 7.30. The expected limits are found compatible within at most 3.5% for all spin and width hypotheses. With this result, the parametric signal model is considered to be validated. As a marginal information, it can be noticed from the luminosity labels on the plots in Figure 7.30 that this closure test has been performed assuming an integrated luminosity of 7.6 fb^{-1}.

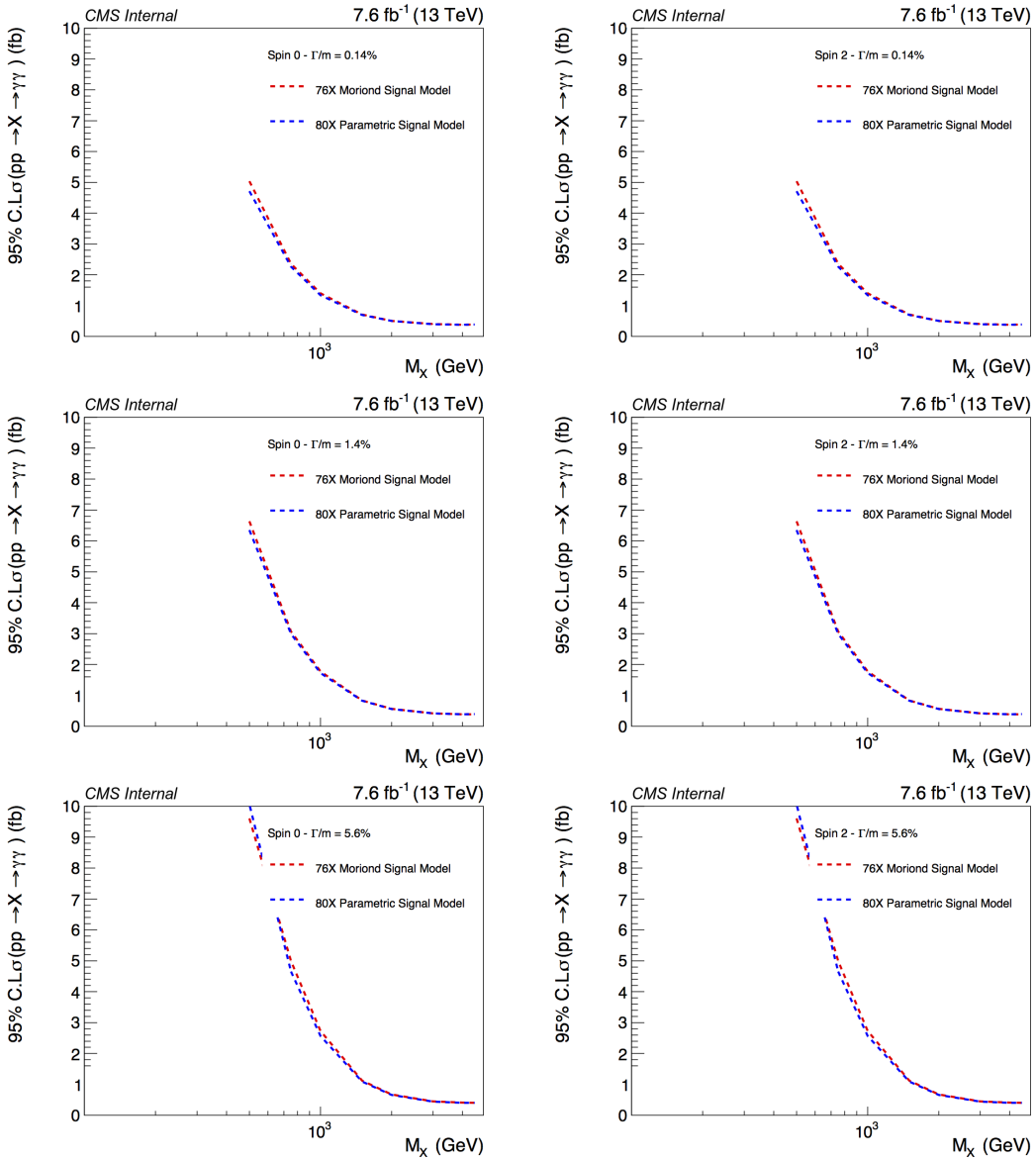

Figure 7.30: Expected sensitivity obtained with the statistical procedure described in Section 7.8 using the non parametric signal model (labeld "76X Moriond Signal Model" in the legend) and the parametric signal model (labeled "80X Parametric Signal Model") for the spin 0 hypothesis on the left and spin 2 hypothesis on the right. From top to bottom the width increases as follows: $\Gamma/m = 0.14, 1.4\%, 5.6\%$ [145].

7.8 Statistical interpretation

Given that no clear excesses over the SM expectation are seen in the diphoton invariant mass spectra, upper limits on the production cross-section of a new resonance decaying in the diphoton final state are computed following the methodology described below.

Methodolody

The statistical treatment of the results follows the modified frequentist method, commonly known as CL_s [154], in conjunction with the use of asymptotic formulas [155].
The first step is the construction of the likelihood function given the parameter of interest μ (which will be defined later)[15], and the vector of systematic uncertainties (nuisance parameters)[16] $\boldsymbol{\theta}$ (see Section 7.8.1).
The likelihood function $\mathcal{L}(data|\mu, \boldsymbol{\theta})$ is written as:

$$\mathcal{L}(data|\mu, \boldsymbol{\theta}) = Poisson(data|\mu, \boldsymbol{\theta}) \cdot p(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta}) \tag{7.9}$$

where *data* represents either the actual experimental observation or the pseudo-data used for the computation of the expected limits (see the dedicated section) and $p(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta})$ is the pdf associated to the vector of systematic errors and $\tilde{\boldsymbol{\theta}}$ is the default value of the vector of nuisance parameters. All source of uncertainties are taken to be either 100% correlated or uncorrelated: this allows one to include all the constraints in the likelihood function in a clean factorised form. Typically, for the different sources of uncertainty normal and log-normal distributions are used depending on the case.
It is also common to define the parameter of interest μ as the signal strength modifier, which can be understood by explicitly writing the poissonian unbinned likelihood of observing k events in data (or pseudo-data):

$$Poisson(data|\mu, \boldsymbol{\theta}) = (k!)^{-1} \prod_i (\mu \cdot S(\boldsymbol{\theta}) \cdot f_S(x_i) + B(\boldsymbol{\theta}) \cdot f_B(x_i)) e^{-(\mu \cdot S(\boldsymbol{\theta}) + B(\boldsymbol{\theta}))} \tag{7.10}$$

where $f_S(x)$ and $f_B(x)$ are the signal and background probability density functions respectively (described in Sections 7.7 and 7.5), x_i is the i-th value taken by the observable under study (in the case of this analysis the diphoton invariant mass of the i-th selected event), while S and B are the total event rates expected for signal and background.
As it can be seen from Equation (7.10), the variable of interest μ is a multiplicative factor of the reference signal yield S. Once the result on μ is obtained (see next section) it is easy to interpret the result in terms of the equivalent cross-section (of production of a heavy resonance) times branching ratio (of the decay in the diphoton final state). In order to do this, the last input which is still needed is the characterization of the acceptance times efficiency curves for the analysis categories.
The acceptance times efficiency ($\epsilon \times$ A) of the analysis, for both EBEB and EBEE categories, as well as the sum of the two is shown in Figure 7.31 for different spin hypotheses as a function of the resonance mass m_X as obtained in MC simulations, dividing the fraction of events passing the full photon selection (see Section 7.3) by the total number of generated events.

In order to compare the compatibility of the data (or pseudo-data) with the "background-only" (also referred to as "null" hypothesis) and "signal plus background" hypotheses,

[15] The parameter μ sometimes is also called r in literature.
[16] A nuisance parameter is any parameter that is not under investigation but still has an impact on the predictions.

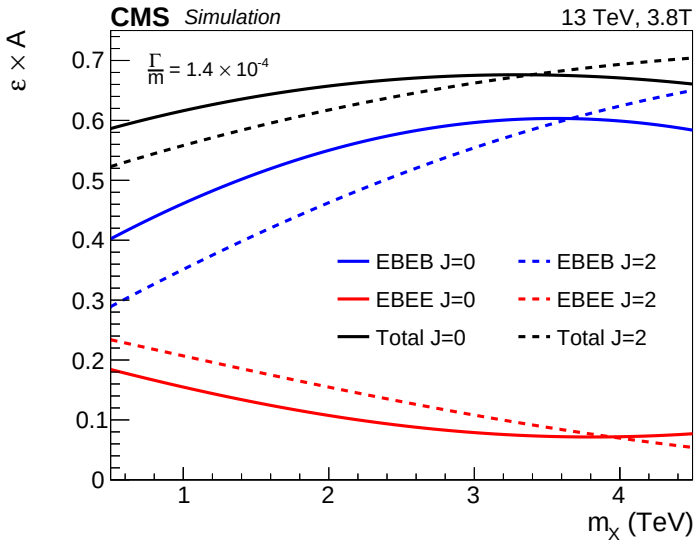

Figure 7.31: Acceptance times efficiency ($\epsilon \times$ A) curves for EBEB (blue), EBEE (red) categories and the total sum of the two (black) for the spin 0 (solid lines) and spin 2 (dashed lines) hypotheses [152].

where the signal is allowed to be scaled by the factor μ, the test statistic q_μ [156], based on the profile likelihood ratio, is defined:

$$q(\mu) = -2 \log \frac{\mathcal{L}(\mu \cdot S + B|\hat{\boldsymbol{\theta}}_{\boldsymbol{\mu}})}{\mathcal{L}(\hat{\mu} \cdot S + B|\hat{\boldsymbol{\theta}})} \tag{7.11}$$

where $\hat{\theta}_\mu$ refers to the conditional maximum-likelihood estimators of θ, given the signal strength μ and data (or pseudo-data). The estimators $\hat{\mu}$ and $\hat{\theta}$ correspond to the global maximum of the likelihood. Note also, in the definition of q_μ, the factor 2 in the second member of Equation (7.11), which asymptotically brings its distribution to a χ^2 one.

Computation of the observed Limit

In order to quote the 95% confidence level (CL) observed upper limit on μ the strategy is the following:

- Compute the observed value of the test statistic $\tilde{q}_\mu^{obs}$ for a given signal strength modifier μ under test.
- Find values of the nuisance parameters $\hat{\boldsymbol{\theta}}_{\mathbf{0}}$ and $\hat{\boldsymbol{\theta}}_{\boldsymbol{\mu}}$ best describing the experimentally observed data (i.e. maximizing the likelihood as given in Equation (7.9)), for the "background-only" and "signal plus background" hypotheses, respectively.
- Generate toy pseudo-data to construct the probability density functions $f(\tilde{q}_\mu; \mu, \hat{\boldsymbol{\theta}}_{\boldsymbol{\mu}})$ and $f(\tilde{q}_\mu; 0, \hat{\boldsymbol{\theta}}_{\mathbf{0}})$ assuming a signal with strength μ in the "signal plus background" hypothesis and $\mu = 0$ for the background-only hypothesis.
- Having constructed the $f(\tilde{q}_\mu; \mu, \hat{\boldsymbol{\theta}}_{\boldsymbol{\mu}})$ and $f(\tilde{q}_\mu; 0, \hat{\boldsymbol{\theta}}_{\mathbf{0}})$ distributions, two p-values are defined, associated with the actual observation for the "signal plus background" and "background-only" hypotheses, p_μ and p_b:

$$p_\mu = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{obs}|signal + background) = \int_{\tilde{q}_\mu^{obs}}^{\infty} f(\tilde{q}_\mu; \mu, \hat{\boldsymbol{\theta}}_{\boldsymbol{\mu}}) d\tilde{q}_\mu$$

$$1 - p_b = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{obs} | background\text{-}only) = \int_{\tilde{q}_0^{obs}}^{\infty} f(\tilde{q}_\mu; 0, \hat{\boldsymbol{\theta}}_{\mathbf{0}}) d\tilde{q}_\mu$$

and calculate $CL_s(\mu)$ as a ratio of these two probabilities:

$$CL_s(\mu) = \frac{p_\mu}{1 - p_b}. \tag{7.12}$$

- Finally to quote 95% CL upper limits on μ, denoted as $\mu^{95\%}$, the value of μ is adjusted until the condition $CL_s = 0.05$ is reached.

Computation of the expected limit

Under the "background-only" hypothesis, a large number of pseudo-data is generated and for each of them, the $\mu^{95\%}$ is computed as explained in the previous section. At that point, by integrating the distribution of $\mu^{95\%}$ obtained for all the pseudo-data, one can build the cumulative probability distribution of the results. The point at which the cumulative probability distribution crosses the quantile at 50% defines what is referred to as "median expected limit". The $\pm 1\sigma$ (68%) band is defined by crossing the 16% and 84% quantiles. Crossing at 2.5% and 97.5% quantiles define the $\pm 2\sigma$ (95%) band.

The results corresponding to the observed and expected upper limits set in this analysis are shown in details in Section 7.8.2. However, before coming to the final results, the different sources of systematic uncertainty are considered in the next section.

7.8.1 Systematic uncertainties

The impact of the systematic uncertainties is sub-dominant in this analysis, which is in fact statistically dominated.

The uncertainties associated to the signal modelling are summarized below. All normalization uncertainties are assigned to the overall signal yield (i.e. they refer to pair of photons where applicable).

- Luminosity uncertainty: 6.2% on the signal normalization was assigned to reflect the uncertainty on the knowledge of the total integrated luminosity as prescribed by the luminosity POG.
- Selection efficiency uncertainties: a 6% uncertainty on the signal normalization was included to reflect the uncertainty on the knowledge of the data/MC scale factors (see Section 6.2).
- Parton distribution functions: a 6% uncertainty on the signal normalization was assigned in order to account for the variation in the kinematic acceptance of the analysis coming from the use of alternative PDF sets.
- Photon energy scale uncertainty: a 1% energy scale uncertainty was assumed. This number was derived to take into account the knowledge of the energy scale uncertainty at the Z peak as well as the knowledge on the extrapolation to high mass, following the studies presented in Section 7.4.
- Photon energy resolution uncertainty: this uncertainty was evaluated by summing and subtracting in quadrature a 0.5% constant term from the estimated resolution correction term measured at the Z peak. The 0.5% value for the constant term was chosen to match the statistical uncertainty on the energy resolution corrections.

The parametric background model (see Section 7.5) has no associated systematic uncertainties, except for the treatment of the bias term in the hypothesis test described below.

Inclusion of the bias term in the hypothesis test

In order to account for the bias term (see Section 7.5) in the test statistics used in the limit setting procedure, the background model $g(m_{\gamma\gamma})$ is modified to include an additional term which has the same distribution as the signal hypothesis under test $s(m_{\gamma\gamma})$, but rescaled by a factor proportional to $\theta_\beta \cdot N_\beta$ as defined in Equation (7.13):

$$\tilde{g}(m_{\gamma\gamma}|\theta_\beta) \cdot p(\theta_\beta) = \left(\frac{N_{bkg} - \theta_\beta N_\beta}{N_{bkg}} g(m_{\gamma\gamma}) + \frac{\theta_\beta N_\beta}{N_{bkg}} s(m_{\gamma\gamma}) \right) \cdot Norm(\theta_\beta) \tag{7.13}$$

In particular, N_β in Equation (7.13) is the integral of the bias function defined at the end of Section 7.5 multiplied by the signal shape function, hence $N_\beta = \int \beta(m_{\gamma\gamma}) \cdot s(m_{\gamma\gamma})$[17], and $Norm(\theta_\beta)$ is the normal distribution of the nuisance parameter θ_β, introduced to take into account to the bias term.

The size of the N_β term for the widest mass hypotheses tested in the analysis is quantified in terms of equivalent cross-section and shown in Table 7.18.

Signal hypothesis		$N_\beta/(\varepsilon \cdot A \cdot L)$(pb)		Signal hypothesis		$N_\beta/(\varepsilon \cdot A \cdot L)$(pb)	
m_G (GeV)	Γ/M (%)	EBEB	EBEE	m_G (GeV)	Γ/M (%)	EBEB	EBEE
500	5.6	2.54	6.75	2000	5.6	0.04	0.04
750	5.6	0.94	1.71	3000	5.6	0.01	0.04
1000	5.6	0.43	0.59	4000	5.6	0.004	0.07
1500	5.6	0.12	0.15				

Table 7.18: Equivalent cross section of the uncertainty associated to the bias term for the widest signal hypotheses tested in the analysis for spin $J = 2$.

Effect of the bias term on the analysis sensitivity

In order to evaluate the impact of the N_β term introduced in the previous paragraphs on the sensitivity of the analysis, the effect on the median expected upper limits, obtained with the additional term included or not has been considered. The expected limits in both cases were computed using the same technique exposed in the previous sections (which is also of course the same one used for the final results) assuming an integrated luminosity of 10 fb^{-1}. The results for the median expected limits computed in both scenarios (with and without the N_β term) are reported in Table 7.19 and in Table 7.20 for the spin 0 and spin 2 hypotheses, respectively. A comparison of the median expected limit for the largest width RS graviton hypothesis is shown in Figure 7.32. As expected, the largest effect is seen for the mass hypotheses with the widest width ($\Gamma/M = 5.6\,\%$) and at the low end of the search region ($m_G \sim 500 - 800$ GeV). In this range, a $5 - 15\%$ worsening of the analysis sensitivity is estimated. The expected significance for a mass of $m_{\gamma\gamma} = 750$ GeV is reduced by up to 6% when the largest signal width hypothesis is considered. For narrower and heavier mass hypotheses, the size of the effect is smaller. For the narrow width (large width) hypothesis the effect is at most 1% for masses above 1 TeV(1.6 TeV).

[17] Given that the signal shape is localised around the resonance mass m_X, the result of the integration can be aproximated by $N_\beta \sim \beta(m_X) \cdot FWHM$ where FWHM stands for the full width at half maximum of the signal shape distribution.

$m_{\gamma\gamma}$(GeV)	Γ/M (%)	Expected limit (fb) with bias	Expected limit (fb) without bias	Difference (%)
500	0.014	2.61	2.43	7
	5.6	5.37	4.80	12
750	0.014	1.28	1.25	2
	5.6	2.72	2.57	6
1000	0.014	0.82	0.81	1
	5.6	1.58	1.49	6
3000	0.014	0.31	0.31	< 1
	5.6	0.34	0.34	< 1

Table 7.19: Comparison of the median expected upper limit on the signal strength for the narrow and large width hypotheses for the spin $J = 0$ hypothesis using the bias parametrization from Table 7.15. An integrated luminosity of 10 fb^{-1} is assumed.

$m_{\gamma\gamma}$(GeV)	Γ/M (%)	Expected limit (fb) with bias	Expected limit (fb) without bias	Difference (%)
500	0.014	3.34	3.06	9
	5.6	6.63	5.88	13
750	0.014	1.58	1.54	3
	5.6	3.25	3.07	6
1000	0.014	0.97	0.96	1
	5.6	1.83	1.75	5
3000	0.014	0.31	0.31	< 1
	5.6	0.35	0.35	< 1

Table 7.20: Comparison of the median expected upper limit on the signal strength for the narrow and large width hypotheses for the spin $J = 2$ hypothesis using the bias parametrization from Table 7.15. An integrated luminosity of 10 fb^{-1} is assumed.

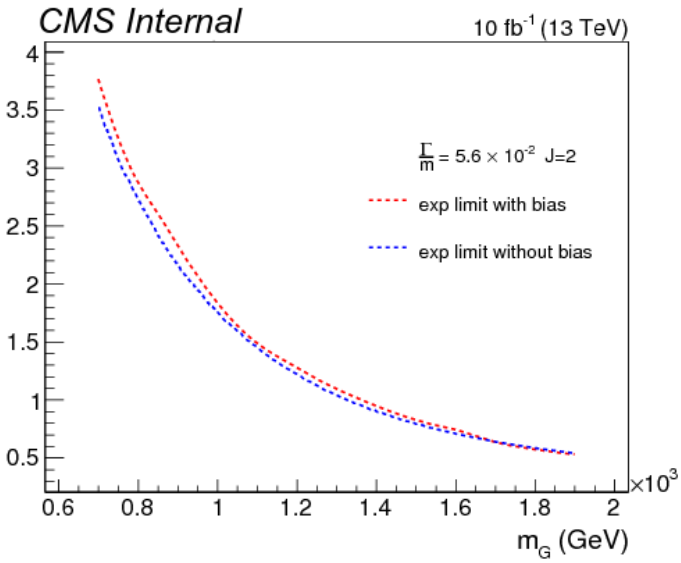

Figure 7.32: Comparison of the median expected limit for the largest width and spin 2 hypothesis in the mass range 700 GeV $< m_{\gamma\gamma} <$ 2 TeV computed with and without the inclusion of the bias term in the hyphotesis test [145].

7.8.2 Upper limits

To set upper limits on the signal strength, two families of models are considered:

- Generic scalar resonances produced via gluon-gluon fusion (having spin $J = 0$).

- RS gravitons (having spin $J = 2$).

The ADD graviton model is discarded here, given that the characteristic mass separation of the model would be so small that individual resonances cannot be resolved from each other [157, 158, 159, 160, 161].

For each of the two spin hypotheses, three kind of width scenarios are considered Γ/M: 0.014%, 1.4% and 5.6%. In the case of the RS graviton resonances, the width is parametrized as $\Gamma/M = 1.4k^2$ and thus $k = 0.01, 0.1, 0.2$.
Figure 7.33 shows the median expected (dashed lines) and the observed exclusion limits (solid lines) for the different signal hypotheses under study. No relevant excesses over the SM-only expectations are seen for any of the spin-width hypotheses considered.

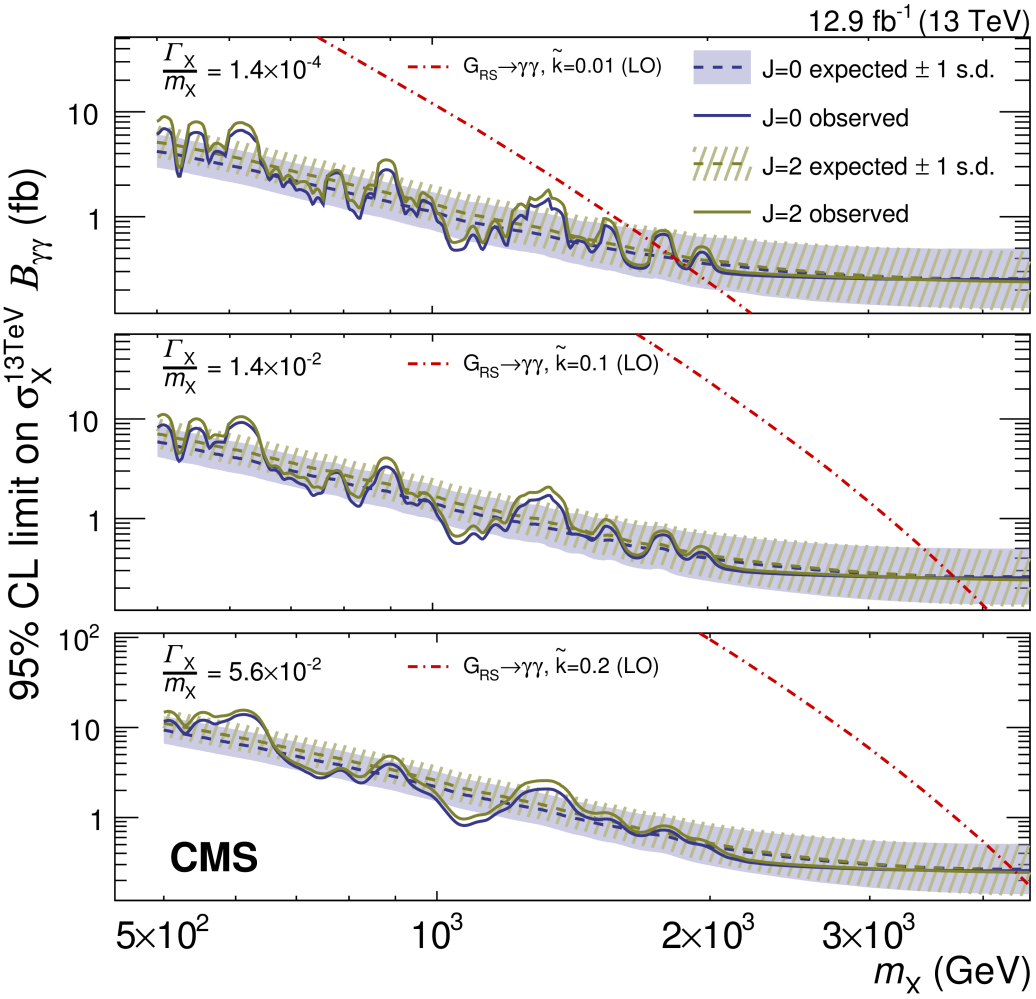

Figure 7.33: Expected and observed upper limit for a spin 0 (blue) resonance and for a RS graviton (grey) under different width hypotheses. From top to bottom the width under consideration increases [133].

Moreover, the compatibility of the data with the background-only hypothesis was evaluated computing the background-only p-value p_0, i.e. the probability to obtain a $q(0)$ above the observed one in the background-only hypothesis. The results are shown in Figure 7.34. As it can be seen, no significant deviation is observed from the SM-only predictions, and in particular in the region around 750 GeV for what concerns the 2016 dataset.

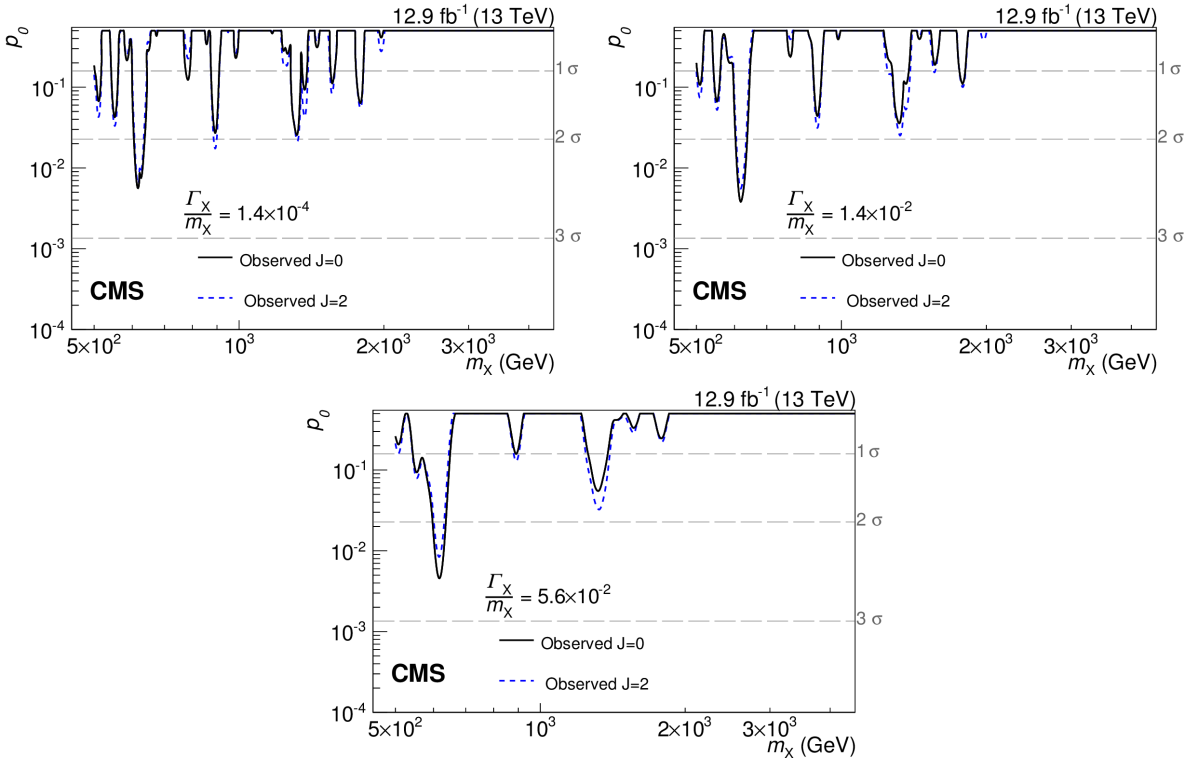

Figure 7.34: Observed background-only p-value (p_0) as a function of the resonance mass for a spin 0 (solid black line) resonance and for a RS graviton (dashed blue line) under the different width hypotheses: small width (upper left), medium width (upper right) and large width (bottom) hypotheses [133].

7.8.3 Combination with previous results

The results obtained analyzing the 2016 dataset are then combined with those obtained by previous analyses with different datasets [149, 162]. In particular, three searches are combined:

- The results obtained analyzing the first half of the 2016 dataset, for an integrated luminosity of 12.9 fb^{-1}at $\sqrt{s} = 13$ TeV (discussed in this chapter).
- The results obtained analyzing the full 2015 dataset, for an integrated luminosity of 3.3 fb^{-1} at $\sqrt{s} = 13$ TeV (including the dataset taken when the CMS magnet was off, corresponding to an integrated luminosity of 0.66 pb^{-1}).
- The results obtained analyzing the full 2012 dataset, for an integrated luminosity of 19.7 fb^{-1} at $\sqrt{s} = 8$ TeV.

In order to combine the 2015 and 2016 results, the same background coefficients (a and b in the definition of the background shape $g(m_{\gamma\gamma})$; see Section 7.5) are used for the 2015 dataset with the CMS magnet at nominal magnet field of 3.8 T and the 2016 dataset. The normalisation coefficients are assumed to be independent, while the bias uncertainty is taken to be fully correlated. The background shapes obtained with independent fits to the two datasets are shown in Figure 7.35. Fully correlated signal uncertainties were considered.

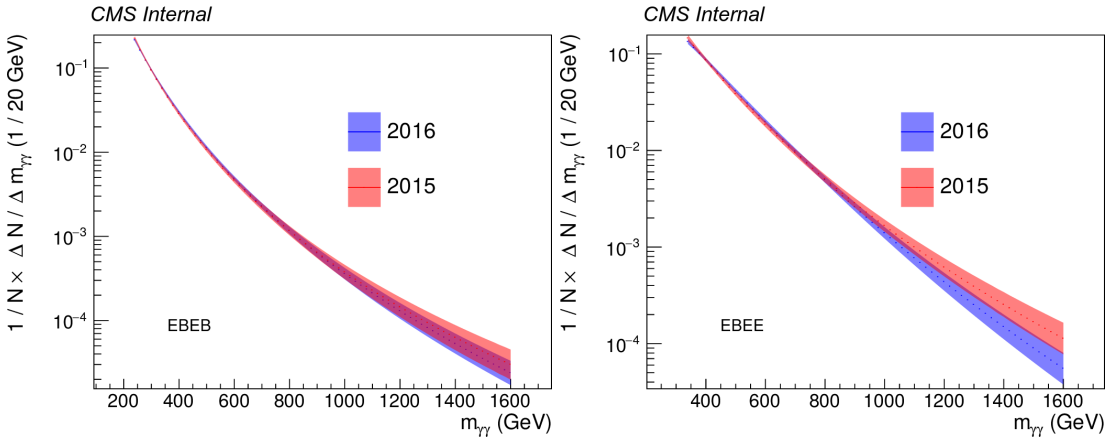

Figure 7.35: Background shapes obtained using independent fits to the 2015 3.8T dataset and the 2016 dataset. The EBEB and EBEE categories are shown on the left and the right respectively [145].

Figure 7.36 shows the combined (observed) background-only p-values p_0 obtained using the 2015 and 2016 datasets. As it can be seen from the figure (which hosts dedicated windows where the diphoton mass region is zoomed), the $\sim$ 2.9 standard deviations[18] excess observed in the 2015 dataset around 750 GeV under the narrow signal hypothesis is reduced to less than 1 standard deviation when the two datasets are combined.

Finally, the 13 TeV results (coming from both 2015 and 2016 datasets) are further combined with those obtained at 8 TeV [162]. The combined background-only p-values p_0 are shown in Figure 7.37.

The compatibility of the three results is also estimated by studying the likelihood function defined in Equation (7.9) as a function of the signal strength μ. In this way, three likelihood scans are obtained, depending on the dataset considered, and the corresponding μ values that maximize the likelihood (i.e. minimize the negative logarithm of the

[18]The look-elsewhere effect is not considerate in this estimate.

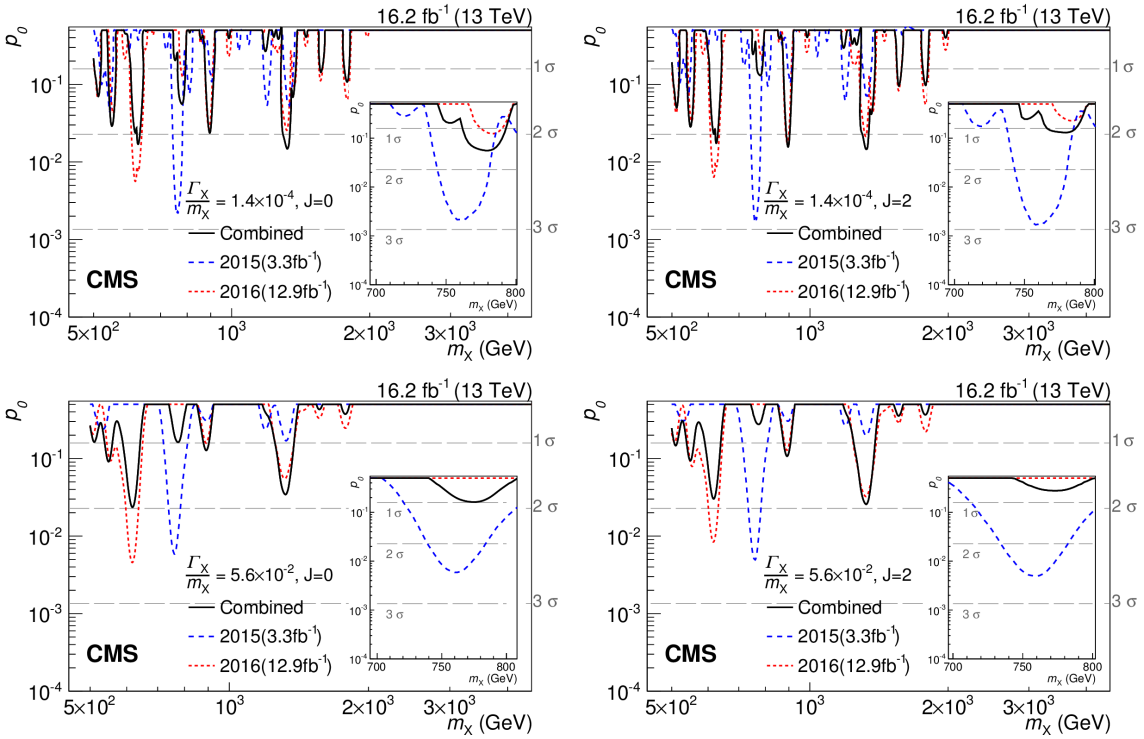

Figure 7.36: Observed p_0 as a function of the resonance mass obtained from the combination of the 2015 and 2016 results for a spin 0 (left) resonance and for a RS graviton (right) under the small width (top) and large width (bottom) hypotheses [133].

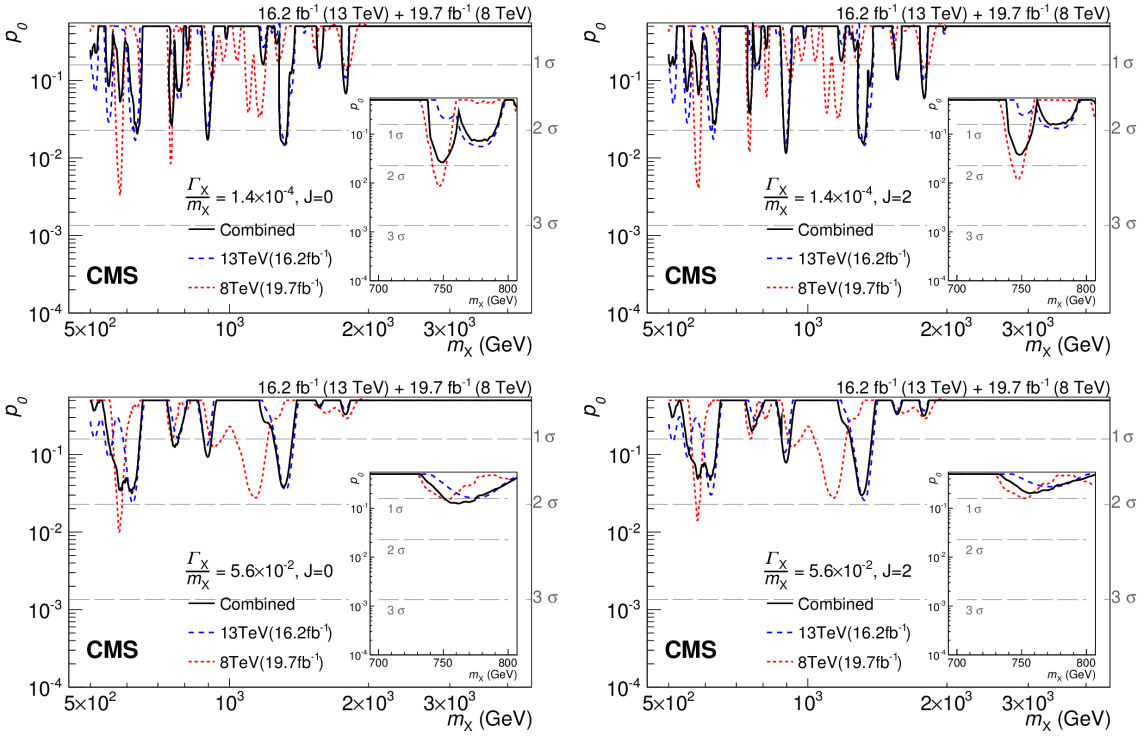

Figure 7.37: Observed p_0 as a function of the resonance mass obtained from the combination of the 8 and 13 TeV results for a spin 0 (left) resonance and for a RS graviton (right) under the small width (top) and large width (bottom) hypotheses [133].

likelihood) are obtained. It is easy then to interpret the results obtained for μ in terms of a signal cross-section. Figure 7.38 shows the likelihood scan for the cross-section (times branching ratio) of a spin 0 resonance under the narrow width hyphothesis having a mass of 750 GeV. The plot shows results obtained with the individual 13 TeV analyses (blue line for the 2015 dataset, red for the 2016 dataset) and 8 TeV analysis (in light blue), as well as the combination of the three results (in black). It is a well known result in statistics that the error on the maximum-likelihood estimator can be inferred "graphically" simply by throwing a constant line at 1 and computing the intercepts with this constant line and the likelihood function itself[19] [163]. As it can be seen from the figure, the compatibility between the different datasets is at the level of ~ 2.5 standard deviations.

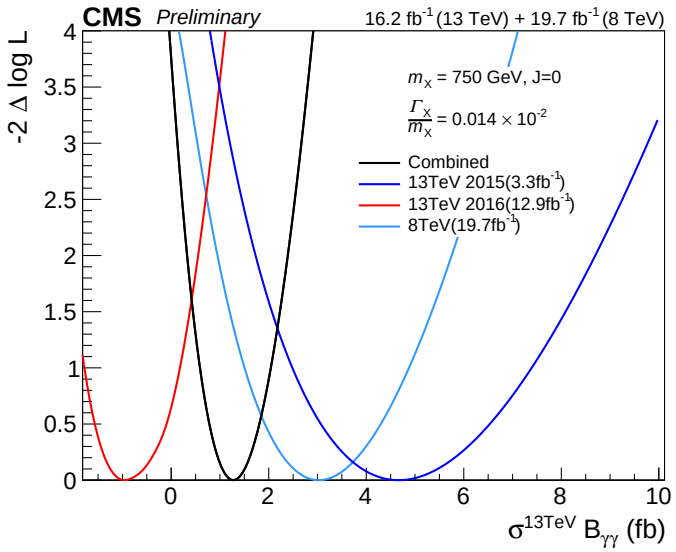

Figure 7.38: Likelihood scan for the cross-section (times branching ratio) of a spin 0 resonance under the narrow width hyphothesis having a mass of 750 GeV. The plot shows results obtained with the individual 13 TeV analyses (blue for the 2015 dataset, red for the 2016 dataset) and 8 TeV analysis (in light blue), as well as the combination of the three results (in black) [152].

[19] This statement assumes that the minimum of the NLL is conveniently put at 0 and that a factor 2 is placed in front of the NLL, as it is the case of Figure 7.38.

7.8.4 Alternative analysis

In parallel to the main analysis, a major cross-check has been developed, performing an alternative analysis that uses a completely independent data analysis framework, but analyzes the same datasets. The event selection, object reconstruction, and energy corrections used are identical to the ones used for the standard analysis. The two independent softwares have been carefully synchronized and cross-checked on the 2015 dataset [164], including a near-perfect agreement in selected events and diphoton mass values, and excellent agreement, within 4% at most, in the relevant parameters for signal and background models, and expected and observed limits and p-values.

Having fully understood the performances of the alternative analysis on the 2015 dataset, it has been applied on the 2016 dataset and its results compared with those obtained by the standard analysis. There are 9083 events selected by the alternative analysis, compared to 9068 for the standard analysis, of which 9032 events are in common. The remaining events not in common are attributed either to differences related to the numerical precision near the boundaries of the various cut values which define the photon selection or to known small difference in the MVA primary vertex selection used by the two analyses. For the 9032 events in common, the diphoton mass values are compared, and are in agreement at the level of 10^{-6}, with differences in the core of the distribution consistent with numerical precision. There is a small number of outliers, with 100 events with a relative difference of more than 1%. These events are mainly cases where a different primary vertex is selected.

The diphoton mass distributions of the selected events are compared between the cross-check and standard analyses in Figure 7.39, with differences consistent with the level of synchronization in the selected events and mass values.

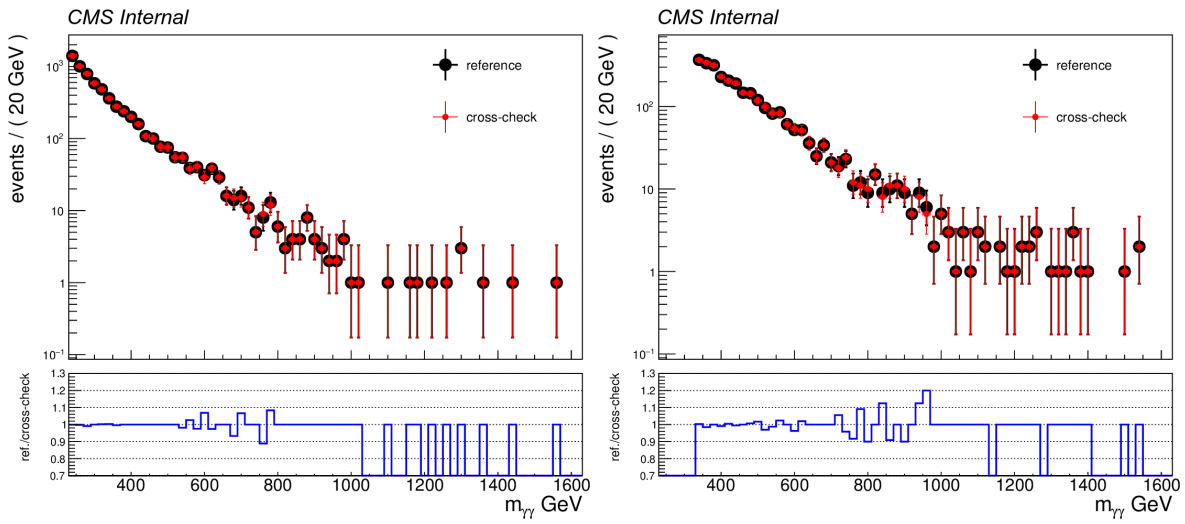

Figure 7.39: Comparison between the diphoton mass distributions obtained for the selected events from the alternative analysis ("cross-check") compared to the standard analysis ("reference") in the EBEB category (left) and the EBEE category (right) [145].

Exclusion limits and background-only p-values have been also computed, but due to lack of time, the additional photon efficiency scale factors derived for the 2016 dataset (see Section 7.3) have been neglected, so implying an approximation at the level of 7%. The expected and observed exclusions limits for a spin 0 resonance, under the narrow width hyphotesis, are shown in Figure 7.40, as well as the observed p-value in Figure 7.41.

The exclusion limits and p-value obtained with the alternative analysis are compared

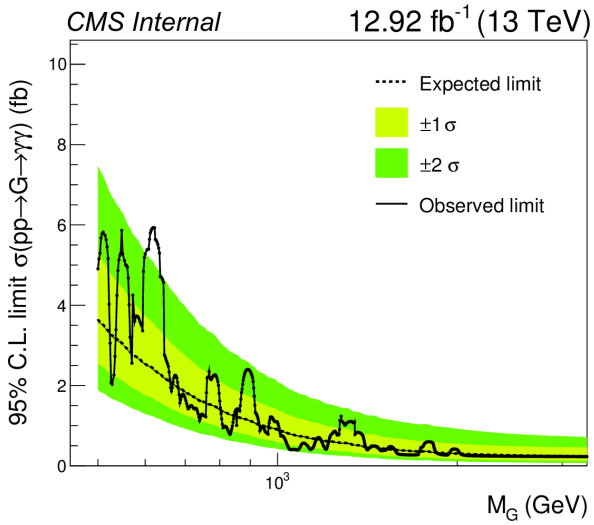

Figure 7.40: The expected and observed 95% exclusion limits for a narrow spin 0 signal hypothesis in the alternative analysis [145].

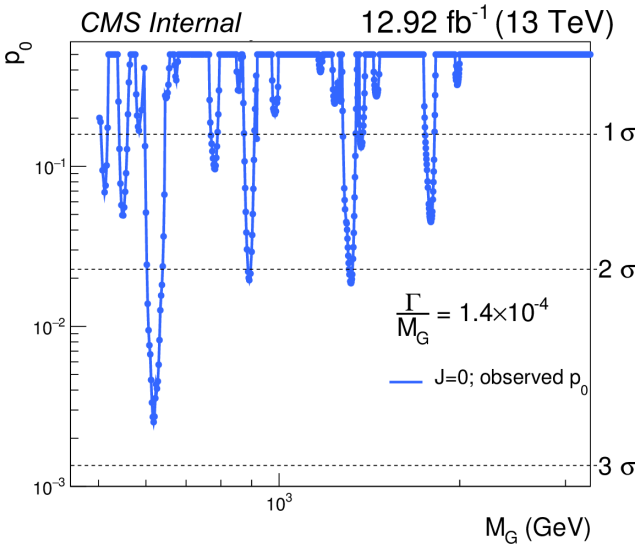

Figure 7.41: The observed p-value with respect to the background-only hypothesis for a narrow spin 0 signal hypothesis in the alternative analysis [145].

to those obtained for the standard analysis in Figure 7.42.
The shape of both the limit and p-value curves as a function of the diphoton invariant mass are in excellent agreement. The normalizations, on the other hand, agree at the 10% level, of which an effect of approximately 7% is known to be attributed to the lack of efficiency scale factors in the alternative analysis.

7.9 Summary

This chapter described the latest CMS search for high mass resonances decaying in the diphoton final state. The analysis employs 12.9 fb^{-1} of pp collisions at center-of-mass energy $\sqrt{s} = 13$ TeV, collected by CMS in the first half of 2016. The analysis strategy has been described in details, in particular for what concerns the selection and photon

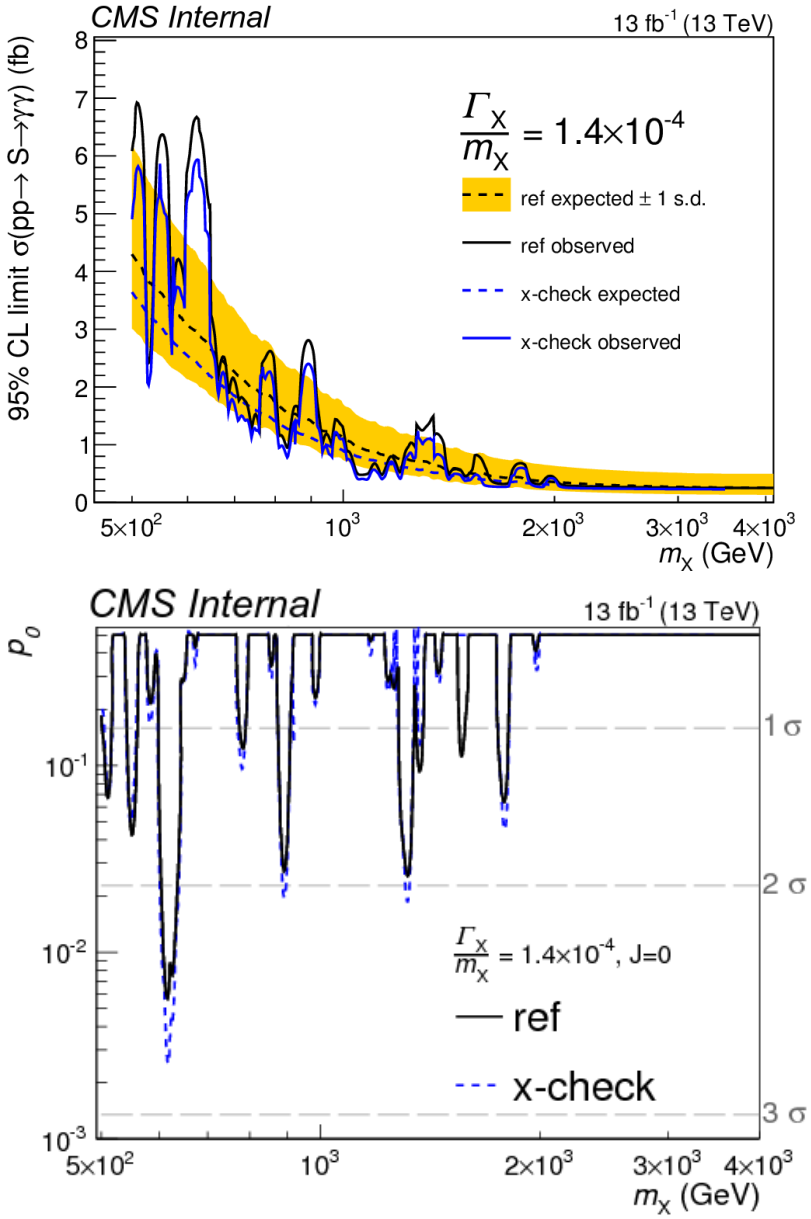

Figure 7.42: Comparison between the expected and observed 95% exclusion limits (top) for a narrow spin 0 signal hypothesis obtained in the alternative analysis and those obtained in the standard analysis. The comparison of the p-values obtained by the two analyses is shown on the bottom [145].

identification criteria, the energy scale and resolution corrections, the signal and background models. The results of the analysis are shown in Section 7.8.2 where it can be seen that, overall, the observed data are consistent with the expectations coming from the background-only hyphothesis.

My personal contribution to the analysis, reported in details in Section 7.4, accounts for the studies on the energy scale and resolution of the ECAL detector. Their determination is based on a multi-step procedure which uses events where the SM Z boson decays in the dielectron final state. In the first step a maximum-likelihood fit is performed on the invariant mass spectra, in data and MC, as a function of data-taking periods. The following steps use the MC simulation as a probability density function in order to fit the data invariant mass distribution. Once the complete set of scale and resolution corrections have been obtained at the Z peak region, boosted Z events are considered, in order to

explore the high-p_{T} region, while still exploiting the Z peak as a handle. In addition, I was involved in the computation of the upper limits and p-values (see Section 7.8.2), that were frequently updated during the data-taking.

The techniques described in this chapter have been applied to different datasets provided by the CMS experiment according by the LHC data-taking schedule, resulting in several publications over the years. The analysis of the 2015 dataset, corresponding to 2.6 fb^{-1} of integrated luminosity [165] and the following ones where also the dataset collected when the magnetic field of the detector was at 0 T was considered [166, 142], showed an excess of events (quantified in $\sim$2.9 standard deviations in terms of local p-value) in the diphoton invariant mass region around 750 GeV. This result lead to a certain excitement in the particle physics community, expecially because the ATLAS collaboration also reported a similar excess in the same mass region [167, 141] and caused big expectations from the analyses, by both ATLAS and CMS, of the factor $\sim$4 bigger dataset collected in the first half of 2016.

The results obtained on the 2016 dataset collected by the CMS experiment up to July 2016 (12.9 fb^{-1}) have been shown in this chapter, in combination with those obtained analyzing the full 2015 dataset, corresponding to an integrated luminosity of 3.3 fb^{-1} and those obtained on the full 2012 dataset, corresponding to an integrated luminosity of 19.7 fb^{-1} collected by the CMS experiment at a center-of-mass energy $\sqrt{s} = 8$ TeV. When the three datasets are combined, the $\sim$ 2.9 standard deviations excess observed in the 2015 dataset alone, around a mass value of $\sim$ 750 GeV under the narrow width signal hypothesis, is reduced to less than 1 standard deviation, hence the existence of a new particle decaying in photon pairs is not confirmed. An alternative analysis performed with an independent analysis framework confirms the results obtained by the main analysis. The results have been published by the Physics Letter B journal [168].

A similar non-confirmation statement was obtained by the ATLAS collaboration by analyzing the 2016 dataset corresponding to 12.2 fb^{-1} of integrated luminosity, in combination with the 3.2 fb^{-1} of integrated luminosity collected in 2015 [169].

Chapter 8

Conclusions and perspectives

This thesis presented the latest available results from two separate searches for new physics beyond the standard model, targeting the dielectron and the diphoton final state respectively, with the CMS detector at the Large Hadron Collider. The two searches presented here are independent, but share many similarities expecially for what concerns their strategies, which are meant to be as simple and robust as possible. Different datasets have been analyzed: the former analysis considers the full 2016 dataset recorded by the CMS experiment during 2016 at a center-of-mass energy $\sqrt{s} = 13$ TeV, for a total integrated luminosity of 35.9 fb^{-1}, while the latter considers the first half of the 2016 dataset, for a total integrated luminosity of 12.9 fb^{-1}, which is then further combined with the 2015 dataset, also at a center-of-mass energy $\sqrt{s} = 13$ TeV, and with the results coming from the 2012 dataset at a center-of-mass energy $\sqrt{s} = 8$ TeV. The results obtained by applying the same analysis techniques described here on the 2015 dataset alone are not described in details, but only cited in Sections 6.8 and 7.9. In particular, the analysis of the 2015 dataset, resulting in an integrated luminosity of 2.6 fb^{-1} (3.3 fb^{-1} when considering also the dataset collected when the magnetic field of the CMS detector was at 0 T), showed a mild excess of events, quantified in $\sim$2.9 standard deviations in terms of local p-value in the diphoton invariant mass region $\sim$750 GeV, with an interesting coincidence with the ATLAS case in the same invariant mass region. This fact drew a certain amount of attention on the channel by the community and the eagerness to analyze a bigger amount of data, which is in fact what has been done with the 2016 dataset as presented here.

The search for new heavy resonances decaying in the dielectron final state was described in details in Chapter 6. The analysis exploits an electron selection that was specifically optimized in order to have high efficiency for high-energy electrons. The main background for the search, which comes from the standard model Drell–Yan process, was estimated from simulations. Other sources of background with prompt leptons in the final state, which include $\mathrm{t\bar{t}}$ and WW production, were also estimated from simulations. The third source of background, which comes from processes of quantum chromo dynamics producing jets misidentified as electrons, was instead estimated using a data-driven approach. The events were categorized according to the position of the electron candidates inside the detector. Two categories have been defined: one with events with both the electron candidates in the detector barrel, the other with events where one electron candidate was reconstructed in the barrel and the other in one of the endcaps. Events with both electron candidates reconstructed in the detector endcaps were discarded. The two channels were combined for the final limit calculation. The observed dielectron invariant mass spectra are in agreement with those predicted by the standard model. As a consequence, 95% confidence level upper limits have been set on the ratio of cross section times branching ratio of a new resonance to the one from the Z boson resonance, using a Bayesian ap-

proach. With the measured upper limits on the cross-section ratio, lower limits on the resonance masses have been set for particles predicted by various models. In particular, for spin 1 resonances, masses below 4.0 TeV, for the Z'_{SSM} particle from the sequential standard model, and below 3.5 TeV for the superstring inspired Z'_{ψ} particle could be excluded with the dielectron channel alone. The combination with the dimuon channel, whose analysis is not described in detail in this thesis, further improve the 95% confidence level lower mass limits to 4.6 TeV for the Z'_{SSM} particle and 4.0 TeV for the Z'_{ψ} particle. The ATLAS collaboration has also provided results targeting the same final state. The latest available results from ATLAS put a 95% confidence level lower mass limit of 4.5 TeV for the Z'_{SSM} model and 3.8 TeV for the Z'_{ψ} one. These results are obtained by combining the dielectron and dimuon final states and are compatible with the ones obtained by the CMS collaboration. They are the most stringent limits to-date on the topic and they superseed the previous results obtained in the RunI era by the CMS and ATLAS experiments themselves, extending the excluded mass region by roughly 1.5 TeV (for example, in the RunI case, a Z'_{SSM} decaying in the dilepton final state was excluded at 95% confidence level for masses below ∼2.9 TeV, by both ATLAS and CMS).

The second analysis presented in this thesis is the search for new resonances decaying in the diphoton final state. As for the dielectron case, the photon selection of this analysis was specifically optimized in order to be highly efficient in selecting photons with high transverse momentum. The background estimation followed a completely data-driven approach, via a parametrization of the observed diphoton invariant mass spectrum. The event categorization is the same one as for the dielectron case: two categories have been defined where at least one the photons in the final state has to be reconstructed in the barrel side of the electromagnetic calorimeter, while the other can be either in the barrel or in one of the endcaps. Events with both photon candidates reconstructed in the endcaps were discarded. The two channels were then combined for the final results. No significant excesses are seen in the measured diphoton invariant mass spectra over the expectations coming from the standard model. As a consequence, 95% confidence level upper limits have been set on the production cross-section times branching ratio of such a new resonance, using the modified-frequentist approach known as CLs. Moreover, the compatibility of the data with the background-only hypothesis was evaluated computing the background-only p-values which showed no significant excess over the standard model in the full spectrum and, in particular, in the diphoton invariant mass region ~ 750 GeV. The results obtained on the 2016 dataset have been combined with those obtained analyzing the full 2015 dataset, corresponding to an integrated luminosity of 3.3 fb^{-1} and those obtained on the full 2012 dataset, corresponding to an integrated luminosity of 19.7 fb^{-1}, at a different center-of-mass energy ($\sqrt{s} = 8$ TeV). With the combination of the three datasets, the ~ 2.9 standard deviations excess observed in the 2015 dataset alone, around a mass value of ~ 750 GeV reduced to less than 1 standard deviation. A similar conclusion is reported also by the ATLAS collaboration, where the 3.4 standard deviations (in terms of local p-value) observed in the 2015 dataset only, under the relative width hypothesis $\Gamma/M = 10\%$, is reduced to roughly 2 standard deviations when the dataset is combined with the 2016 one. Moreover, 95% confidence level lower mass limits are set, excluding the mass region below 1.95 - 4.45 TeV depending of the relative width scenario considered for the new resonance. The result significantly improves the one obtained at RunI by CMS [51], where the excluded mass region was between 1.4 - 2.8 TeV, depending on the width scenario considered. The ATLAS results at RunI inspected only the "low" mass region up to 650 GeV.

The RunII era, with its dramatic increase in the center-of-mass energy with respect to the RunI case, opened a completely new phase-space for the discovery of new phenomena

beyond the standard model, and the analysis of the first datasets taken in this new era has certainly been a goal of major importance. Unfortunately, no evidence of new physics has been seen in the first part of the RunII data-taking. From now on, the next important step in order to increase the sensitivity of the analyses is to cumulate the biggest possible amount of statistics and explore even higher invariant mass regions. In order to achieve this task, and according to the LHC schedule, the RunII data-taking will continue until 2018, when a long shutdown period is planned during 2019 and 2020. After that, the RunIII period will start in 2021 and last until 2023. At the end of Phase-1 (which includes RunI, RunII and RunIII) the collected integrated luminosity is expected to reach 300 fb^{-1}.

Following another long shutdown period, the Phase-2, also referred to as "high luminosity LHC" (HL-LHC), will start. The foreseen operating scenario is to reach the instantaneous luminosity of $5 \cdot 10^{34}$ $\text{cm}^{-2}\text{s}^{-1}$ and a potential peak value of $2 \cdot 10^{35}$ $\text{cm}^{-2}s^{-1}$ at the beginning of the fills. Due to radiation damage, aging, and the challenges of higher instantaneous luminosities at the HL-LHC, a number of major upgrades to the CMS detector will be required in order to preserve the ability to carry out the diverse physics program of the CMS experiment.
In particular, the upgrade must ensure that the precision electromagnetic calorimetry and robust jet and missing transverse energy reconstruction capability are maintained at the HL-LHC. Issues are particularly severe in the endcap region, where the present calorimeters would suffer from radiation damage, hence both the electromagnetic and hadronic endcap calorimeters must be replaced. For what concers the ECAL case, due to the cumulated radiation in the endcaps, the crystals would develop very low light output. Therefore, CMS plans to replace the endcap calorimeter of the ECAL subdetector with a high granularity calorimeter made of silicon sensors and tungsten absorber. In the barrel case, the crystal light output will be sufficient to ensure good performance, but hadron fluence will cause a growth in the APD dark current, and consequently in the electronic noise. A reduction of the APD dark current of a factor about 2 could be obtained by lowering the operating temperature of the calorimeter from the present 18° C to about 8° C. Also, the ECAL electronics must be replaced to cope with the increase in trigger latency and in readout rate. The new design will transmit all the ECAL data off-detector with fast optical links. The upgrade of the electronics also includes a new preamplifier, faster and better matched to the APD dark current increase.
Moreover, the new tracker subdetector will have substantially less material and in addition to providing improved tracking capability in a dense charged particle environment, will also provide tracks to the L1 trigger, allowing a substantial increase in trigger functionality. In fact, the CMS trigger algorithm will be more complex in order to select the interesting events on top of the high number of concurrent pile-up interactions per bunch crossing and the track information in the L1 trigger algorithms is one of the key improvements in this direction.
Another major change at the HL-LHC will be the unprecedented average number of pile-up interactions of about 140. The event vertices will be spread in position along the beam axis with a displacement of 5 cm, and in time with a spread of hundreds of picoseconds. This characteristic can be used to mitigate the pile-up using the time of flight information, provided that the calorimeter hits have a precise timing measurement associated to them. Finally, concerning the data-taking schedule, the HL-LHC foresees to deliver $\sim$250 fb^{-1} of integrated luminosity per year, during 10 years of operation, thus collecting, at the end of operation, a factor roughly 100 more of integrated luminosity with respect to the one used for the results showed in this document. Thanks to the massive increase in cumulated statistics, a completely new era of precision measurements will be opened and

the strategies of the analyses presented in this thesis will need to be re-defined and optimized in order to face the unprecedented experimental conditions, while the amount of available data will allow the possibility to target more complex theoretical scenarios, where, for example, broad resonances appear instead of a resonant peak over a continuous background.

Bibliography

[1] Particle Data Group. "Review of Particle Physics". In: *Chin. Phys.* C40.10 (2016), p. 100001. DOI: 10.1088/1674-1137/40/10/100001.

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

[3] P. W. Higgs. "Broken symmetries, massless particles and gauge fields". In: *Phys. Rev. Lett.* (1964), pp. 132–133. DOI: 10.1016/0031-9163(64)91136-9.

[4] 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: *Phys. Lett.* B716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

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

[6] Halzen and Martin. "Quarks and Leptons". In: John Wiley and Sons (1984).

[7] B. Cabrera and M. E. Peskin. "Cooper-pair mass". In: *Phys. Rev. B* 39 (10 1989), pp. 6425–6430. DOI: 10.1103/PhysRevB.39.6425. URL: https://link.aps.org/doi/10.1103/PhysRevB.39.6425.

[8] Gerald S. Guralnik. "Gauge Invariance and the Goldstone Theorem". In: *Mod. Phys. Lett.* A26 (2011), pp. 1381–1392. DOI: 10.1142/S0217732311036188. arXiv: 1107.4592 [hep-th].

[9] Gerard 't Hooft and M.J.G. Veltman. "Regularization and Renormalization of Gauge Fields". In: *Nucl.Phys* (1972), B44:189–213. DOI: 10.1016/0550-3213(72)90279-9.

[10] M. Peskin. "Beyond the Standard Model". In: *Lectures notes of the 1996 European School of High-Energy Physics* (Carry-le-Rouet, France, 1996).

[11] David J. Gross and Frank Wilczek. "Ultraviolet Behavior of Nonabelian Gauge Theories". In: *Phys. Rev. Lett.* (1973), 30:1343–1346. DOI: 10.1103/PhysRevLett.30.1343.

[12] H. David Politzer. "Reliable Perturbative Results for Strong Interactions?" In: *Phys. Rev. Lett.* (1973), 30:1346–1349. DOI: 10.1103/PhysRevLett.30.1346.

[13] Jonathan L. Rosner. "Forward - backward asymmetries in hadronically produced lepton pairs". In: *Phys. Rev.* D54 (1996), pp. 1078–1082. DOI: 10.1103/PhysRevD.54.1078. arXiv: hep-ph/9512299 [hep-ph].

[14] Vera C. Rubin and W. Kent Ford Jr. "Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions". In: *Astrophys. J.* 159 (1970), pp. 379–403. DOI: 10.1086/150317.

[15] Katherine Freese. "Review of Observational Evidence for Dark Matter in the Universe and in upcoming searches for Dark Stars". In: *EAS Publ. Ser.* 36 (2009), pp. 113–126. DOI: 10.1051/eas/0936016. arXiv: 0812.4005 [astro-ph].

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

[17] Y. Abe et al. "Indication for the disappearance of reactor electron antineutrinos in the Double Chooz experiment". In: *Phys. Rev. Lett.* 108 (2012), p. 131801. DOI: 10.1103/PhysRevLett.108.131801. arXiv: 1112.6353 [hep-ex].

[18] Stephen P. Martin. "A Supersymmetry primer". In: *Adv. Ser. Direct. High Energy Phys.18,1* (1997). DOI: 10.1142/9789812839657_0001, 10.1142/9789814307505_0001. arXiv: hep-ph/9709356 [hep-ph].

[19] Adam G. Riess et al. "Observational evidence from supernovae for an accelerating universe and a cosmological constant". In: *Astron. J.* 116 (1998), pp. 1009–1038. DOI: 10.1086/300499. arXiv: astro-ph/9805201 [astro-ph].

[20] S. Perlmutter et al. "Measurements of Omega and Lambda from 42 high redshift supernovae". In: *Astrophys. J.* 517 (1999), pp. 565–586. DOI: 10.1086/307221. arXiv: astro-ph/9812133 [astro-ph].

[21] The CMS and ATLAS Collobarations. "Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV". In: *JHEP* 08 (2016), p. 045. DOI: 10.1007/JHEP08(2016)045. arXiv: 1606.02266 [hep-ex].

[22] R. Barbier et al. "R-parity violating supersymmetry". In: *Phys.Rept.* 420 (2005). DOI: 10.1016/j.physrep.2005.08.006. arXiv: 0406039 [hep-ex].

[23] G. Altarelli, B. Mele, and M. Ruiz-Altaba. "Searching for new heavy vector bosons in $p\bar{p}$ colliders". In: *Z. Phys. C* 45 (1989), p. 109. DOI: 10.1007/BF01556677.

[24] A. Leike. "The Phenomenology of extra neutral gauge bosons". In: *Phys.Rept.* 317 (1999), p. 143. DOI: 10.1016/S0370-1573(98)00133-1.

[25] J. L. Hewett and T. G. Rizzo. "Low-Energy Phenomenology of Superstring Inspired E(6) Models". In: *Phys.Rept.* 183 (1989), p. 193. DOI: 10.1016/0370-1573(89)90071-9.

[26] L. Randall and R. Sundrum. "An Alternative to compactification". In: *Phys. Rev. Lett.* 83 (1999), p. 4690. DOI: 10.1103/PhysRevLett.83.4690.

[27] L. Randall and R. Sundrum. "A Large mass hierarchy from a small extra dimension". In: *Phys. Rev. Lett.* 83 (1999), p. 3370. DOI: 10.1103/PhysRevLett.83.3370.

[28] H. Georgi and S. Glashow. "Unity of All Elementary Particle Forces". In: *Phys. Rev. Lett.* 32 (1974), p. 438. DOI: 10.1103/PhysRevLett.32.438.

[29] R. C. King, L. Dehuai, and B. G. Wybourne. "Symmetrized Powers of Rotation Group Representations". In: *J. Phys.* A14 (1981), pp. 2509–2538. DOI: 10.1088/0305-4470/14/10/009.

[30] Edward Witten. "Symmetry Breaking Patterns in Superstring Models". In: *Nucl. Phys.* B258 (1985), p. 75. DOI: 10.1016/0550-3213(85)90603-0.

[31] N. Arkani-Hamed, S. Dimopoulos, and G. Dvali. "The hierarchy problem and new dimensions at a millimeter". In: *Phys. Lett. B* 429 (1998), p. 263. DOI: 10.1016/S0370-2693(98)00466-3. arXiv: 9803315 [hep-ph].

[32] John H. Schwarz. "Introduction to superstring theory". In: *NATO Sci. Ser. C* 566 (2001), pp. 143–187. DOI: 10.1007/978-94-010-0522-7_4. arXiv: 0008017 [hep-ex].

[33] H. Davoudiasl, J. L. Hewett, and T. G. Rizzo. "Phenomenology of the Randall-Sundrum Gauge Hierarchy Model". In: *Phys. Rev. Lett.* 84 (2000), p. 2080. DOI: 10.1103/PhysRevLett.84.2080.

[34] The CDF Collaboration. "Search for High-Mass e^+e^- Resonances in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV". In: *Phys. Rev. Lett.* 102 (2009), p. 031801. DOI: 10.1103/PhysRevLett.102.031801. arXiv: 0810.2059 [hep-ex].

[35] The D0 Collaboration. "Search for a heavy neutral gauge boson in the dielectron channel with 5.4 fb^{-1} of $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV". In: *Phys. Lett.* B695 (2011), pp. 88–94. DOI: 10.1016/j.physletb.2010.10.059. arXiv: 1008.2023 [hep-ex].

[36] The CDF Collaboration. "Search for Randall-Sundrum gravitons in the diphoton channel at CDF". In: *Phys. Rev. D* 83 (2011), p. 011102. DOI: 10.1103/PhysRevD.83.011102. arXiv: 1012.2795 [hep-ex].

[37] The CDF Collaboration. "Search for New Dielectron Resonances and Randall-Sundrum Gravitons at the Collider Detector at Fermilab". In: *Phys. Rev. Lett.* 107 (2011), p. 051801. DOI: 10.1103/PhysRevLett.107.051801. arXiv: 1103.4650 [hep-ex].

[38] The CDF Collaboration. "Search for New Dielectron Resonances and Randall–Sundrum Gravitons at the Collider Detector at Fermilab". In: *Phys. Rev. Lett.* 107 (2011), p. 051801. DOI: 10.1103/PhysRevLett.107.051801. arXiv: 1103.4650 [hep-ex].

[39] The ATLAS Collaboration. "Search for high-mass resonances decaying to dilepton final states in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector". In: *JHEP 1211* B429 (2012), p. 138. DOI: 10.1007/JHEP11(2012)138.

[40] The CMS Collaboration. "Search for narrow resonances in dilepton mass spectra in pp collisions at $\sqrt{s} = 7$ TeV". In: *Phys.Lett.* B714 (2012), p. 158. DOI: 10.1016/j.physletb.2012.06.051.

[41] The CMS Collaboration. "Search for heavy narrow dilepton resonances in pp collisions at $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV". In: *Phys.Lett.* B720 (2013), p. 63. DOI: 10.1016/j.physletb.2013.02.003.

[42] The CMS Collaboration. "Search for physics beyond the standard model in dilepton mass spectra in proton-proton collisions at $\sqrt{s} = 8$ TeV". In: *JHEP* 04 (2015), p. 025. DOI: 10.1007/JHEP04(2015)025. arXiv: 1412.6302 [hep-ex].

[43] The ATLAS Collaboration. "Search for high-mass dilepton resonances in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector". In: *Phys. Rev. D* 90 (2014), p. 052005. DOI: 10.1103/PhysRevD.90.052005. arXiv: 1405.4123 [hep-ex].

[44] The ATLAS Collaboration. "Search for contact interactions and large extra dimensions in the dilepton channel using proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector". In: *Eur.Phys.J.* C74 (2014), p. 3134. DOI: 10.1140/epjc/s10052-014-3134-6.

[45] The CMS Collaboration. "Search for Large Extra Spatial Dimensions in Dielectron Production with the CMS Detector". In: *CMS Physics Analysis Summaries* CMS-PAS-EXO-12-031 (2012).

[46] The CMS Collaboration. "Search for Extra Dimensions in Dimuon Events in pp Collisions at $\sqrt{s} = 8$ TeV". In: *CMS Physics Analysis Summaries* CMS-PAS-EXO-12-027 (2012).

[47] The ATLAS Collaboration. "Search for Extra Dimensions in diphoton events using proton-proton collisions recorded at $\sqrt{s} = 7$ TeV with the ATLAS detector at the LHC". In: *New J. Phys.* 15 (2013), p. 043007. DOI: 10.1088/1367-2630/15/4/043007. arXiv: 1210.8389 [hep-ex].

[48] The CMS Collaboration. "Search for signatures of extra dimensions in the diphoton mass spectrum at the Large Hadron Collider". In: *Phys. Rev. Lett.* 108 (2012), p. 111801. DOI: 10.1103/PhysRevLett.108.111801. arXiv: 1112.0688 [hep-ex].

[49] The ATLAS Collaboration. "Search for scalar diphoton resonances in the mass range 65-600 GeV with the ATLAS detector in pp collision data at $\sqrt{s} = 8$ TeV". In: *ATLAS Conference Notes* ATLAS-CONF-2014-031 (2014).

[50] The CMS Collaboration. "Search for an Higgs-like resonance in the diphoton mass spectra above 150 GeV with 8 TeV data". In: *CMS Physics Analysis Summaries* CMS-PAS-HIG-14-006 (2014).

[51] The CMS Collaboration. "Search for High-Mass Diphoton Resonances in pp collisions at $\sqrt{s}$ =8 TeV with the CMS detector". In: *CMS Physics Analysis Summaries* CMS-PAS-EXO-12-045 (2012).

[52] L. Evans and P. Bryant. "LHC Machine". In: *JINST* 3 (2008). DOI: 10.1088/1748-0221/3/08/S08001.

[53] The CMS Collaboration. "The CMS experiment at the CERN LHC". In: *JINST* 3 (2008). DOI: 10.1088/1748-0221/3/08/S08004.

[54] The ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *JINST* 3 (2008). DOI: 10.1088/1748-0221/3/08/S08003.

[55] The LHCb Collaboration. "The LHCb Detector at the LHC". In: *JINST* 3 (2008). DOI: 10.1088/1748-0221/3/08/S08005.

[56] The ALICE Collaboration. "The ALICE experiment at the CERN LHC". In: *JINST* 3 (2008). DOI: 10.1088/1748-0221/3/08/S08002.

[57] The CMS Collaboration. *CMS Luminosity - Public Results*. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults.

[58] W.J. Stirling. *Public plots*. http://www.hep.ph.ic.ac.uk/wstirlin/plots/crosssections2012_v5.pdf.

[59] Bo Andersson et al. "Parton Fragmentation and String Dynamics". In: *Phys. Rept.* 97 (1983), pp. 31–145. DOI: 10.1016/0370-1573(83)90080-7.

[60] The H1 Collaboration. "Inclusive Deep Inelastic Scattering at High Q2 with Longitudinally Polarised Lepton Beams at HERA". In: *PoS* EPS-HEP2013 (2013), p. 474. arXiv: 1310.0968 [hep-ex].

[61] The H1 and ZEUS Collaborations. "Deep inelastic scattering at HERA". In: *Acta Phys. Polon. Supp.* 6 (2013), pp. 709–714. DOI: 10.5506/APhysPolBSupp.6.709.

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

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

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

[65] H.-L. Lai et al. "New parton distributions for collider physics". In: *Phys.Rev.* D82 (2010), 074024. DOI: 10.1103/PhysRevD.82.074024.

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

[67] A. Martin, W. Stirling, R. Thorne, and G. Watt. "Parton distributions for the LHC". In: *Eur.Phys.J.* C63 (2009), p. 189. DOI: 10.1140/epjc/s10052-009-1072-5.

[68] R. D. Ball et al. "Parton distributions with LHC data". In: *Nucl.Phys.* B867 (2013), p. 244. DOI: 10.1016/j.nuclphysb.2012.10.003.

[69] *The Durham Hep-Data Project - PDF plotter*. URL: http://hepdata.cedar.ac.uk/pdf/pdf3.html.

[70] John C. Collins, Davison E. Soper, and George F. Sterman. "Factorization of Hard Processes in QCD". In: *Adv. Ser. Direct. High Energy Phys.* 5 (1989), pp. 1–91. DOI: 10.1142/9789814503266_0001. arXiv: 0409313 [hep-ph].

[71] W.J. Stirling. *Public plots*. URL: http://www.hep.ph.ic.ac.uk/wstirlin/plots/lumiRatio.pdf.

[72] The CMS Collaboration. "CMS Physics Technical Design Report, Volume II: Physics Performance". In: *Journal of Physics G: Nuclear and Particle Physics* 34.6 (2007), p. 995. URL: http://stacks.iop.org/0954-3899/34/i=6/a=S01.

[73] Christian Piasecki. "Development of the CMS tracker and reconstruction of secondary vertices of b and c hadrons". In: *IEKP-KA-2006-7* (2006).

[74] The CMS Collaboration. *The CMS electromagnetic calorimeter project: Technical Design Report*. Geneva: CERN, 1997.

[75] A.A Annenkov, M.V Korzhik, and P Lecoq. "Lead tungstate scintillation material". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 490.1–2 (2002), pp. 30 –50. ISSN: 0168-9002. DOI: 10.1016/S0168-9002(02)00916-6.

[76] The CMS Collaboration. *The CMS hadron calorimeter project: Technical Design Report*. Geneva: CERN, 1997.

[77] The CMS Collaboration. *The CMS muon project: Technical Design Report*. Geneva: CERN, 1997.

[78] M Raymond et al. "The MGPA electromagnetic calorimeter readout chip for CMS". In: *9th Workshop on Electronics for LHC Experiments* (2003). DOI: 10.5170/CERN-2003-006.83.

[79] Emanuele Di Marco. *CMS electromagnetic calorimeter calibration and timing performance during LHC Run I and future prospects*. Tech. rep. CMS-CR-2014-410. Geneva: CERN, 2014. URL: https://cds.cern.ch/record/1975982.

[80] P Adzic and E Auffray. "Reconstruction of the signal amplitude of the CMS electromagnetic calorimeter". In: *Eur. Phys. J. C* 46 S1 (2006), pp. 23–35. URL: https://cds.cern.ch/record/1062402.

[81] S Gadomski et al. "The deconvolution method of fast pulse shaping at hadron colliders". In: *Nucl. Instrum. Methods Phys. Res., A* 320.CERN-PPE-92-024 (1992), 217–227. 28 p. URL: https://cds.cern.ch/record/232960.

[82] M T Heath. *Numerical methods for large sparse linear least squares problems*. Tech. rep. ORNL-CSD-114. Oak Ridge, TN: Oak Ridge Nat. Lab., 1983. URL: https://cds.cern.ch/record/144406.

[83] The CMS Collaboration. *Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in pp Collisions at $\sqrt{s}$ = 7 TeV*. Tech. rep. P09009. JINST 8: Oak Ridge Nat. Lab., 2013.

[84] Zhang Liang Ying et al. "Monitoring light source for CMS lead tungstate crystal calorimeter at LHC". In: *IEEE Trans. Nucl. Sci.* 48.3 pt.1 (2001), pp. 372–8. URL: https://cds.cern.ch/record/521707.

[85] The CMS Collaboration. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". In: *J. Instrum.* 9 (2014), P10009. DOI: 10.1088/1748-0221/9/10/P10009.

[86] Franz Hamilton, Tyrus Berry, and Timothy Sauer. "Kalman-Takens filtering in the presence of dynamical noise". In: Eur. Phys. J. C (2016). arXiv: 1611.05414 [physics.data-an].

[87] Afiq A. Anuar. "Electrons and photons at High Level Trigger in CMS for Run II". In: *J. Phys. Conf. Ser.* 664.8 (2015), p. 082001. DOI: 10.1088/1742-6596/664/8/082001.

[88] K. Krajewska, C. Müller, and J. Z. Kami ński. "Bethe-Heitler pair production in ultrastrong short laser pulses". In: *Phys. Rev. A* 87 (6 2013), p. 062107. DOI: 10.1103/PhysRevA.87.062107. URL: https://link.aps.org/doi/10.1103/PhysRevA.87.062107.

[89] Livia Soffi. "Search for new exotic particles decaying to photons with the CMS experiment at the LHC". In: *The European Physical Journal Plus* 131.8 (2016), p. 283. ISSN: 2190-5444. DOI: 10.1140/epjp/i2016-16283-3. URL: http://dx.doi.org/10.1140/epjp/i2016-16283-3.

[90] The CMS Collaboration. *The CMS Physics Technical Design Report*. CERN/LHCC 2006-001, CMS TDR 8.1. 2006.

[91] The CMS Collaboration. "Particle-Flow event reconstruction in CMS and performance for jets, taus and MET". In: *CMS Physics Analysis Summaries* CMS-PAS-PFT-09-001 (2009).

[92] The CMS Collaboration. "Commissioning of the particle-flow event reconstruction with leptons from J/Ψ and W decays at 7 TeV ". In: *CMS Physics Analysis Summaries* CMS-PAS-PFT-10-003 (2010).

[93] The CMS Collaboration. "Performance of τ-lepton reconstruction and identification in CMS". In: *J. Instrum.* 7.arXiv:1109.6034. CMS-TAU-11-001. CERN-PH-EP-2011-137 (2011), P01001. 33 p. URL: https://cds.cern.ch/record/1385560.

[94] E K Friis. "Tau reconstruction and identification at the compact muon solenoid". In: *Nucl. Phys. B, Proc. Suppl.* 218 (2011), pp. 256–261. URL: https://cds.cern.ch/record/2002999.

[95] The CMS Collaboration. "Search for new physics with high-mass tau lepton pairs in pp collisions at $\sqrt{s} = 13$ TeV with the CMS detector". In: *CMS Physics Analysis Summaries* CMS-PAS-EXO-16-008 (2016). URL: https://cds.cern.ch/record/2160363.

[96] Sanmay Ganguly. *Jet Measurements In CMS*. Tech. rep. arXiv:1303.6038. 2013. URL: https://cds.cern.ch/record/1532824.

[97] The CMS Collaboration. "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS". In: *J. Instrum.* arXiv:1107.4277. CMS-JME-10-011. CERN-PH-EP-2011-102 (2011), P11002. 67 p. URL: https://cds.cern.ch/record/1369486.

[98] Ivan Marchesini. *Identification of b-quark Jets in the CMS experiment*. Tech. rep. CMS-CR-2013-412. Geneva: CERN, 2013. URL: https://cds.cern.ch/record/1629516.

[99] Saranya Ghosh. "Performance of MET reconstruction and pileup mitigation techniques in CMS". In: *Nucl. Part. Phys. Proc.* 273-275.arXiv:1502.05207 (2015). Proceedings of the 37th International Conference on High Energy Physics, ICHEP 2014, 2512–2514. 3 p. URL: https://cds.cern.ch/record/1992977.

[100] Paolo Nason. "A new method for combining NLO QCD with shower Monte Carlo algorithms". In: *JHEP* 11 (2004), p. 040. DOI: 10.1088/1126-6708/2004/11/040. arXiv: 0409146.

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

[102] Simone Alioli et al. "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX". In: *JHEP* 06 (2010), p. 043. DOI: 10.1007/JHEP06(2010)043. arXiv: 1002.2581 [hep-ph].

[103] Simone Alioli et al. "NLO vector-boson production matched with shower in POWHEG". In: *JHEP* 07 (2008), p. 060. DOI: 10.1088/1126-6708/2008/07/060. arXiv: 0805.4802 [hep-ph].

[104] Stefano Frixione, Paolo Nason, and Giovanni Ridolfi. "A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction". In: *JHEP* 09 (2007), p. 126. DOI: 10.1088/1126-6708/2007/09/126. arXiv: 0707.3088 [hep-ph].

[105] Emanuele Re. "Single-top Wt-channel production matched with parton showers using the POWHEG method". In: *Eur. Phys. J. C* 71 (2011), p. 1547. DOI: 10.1140/epjc/s10052-011-1547-z. arXiv: 1009.2450 [hep-ph].

[106] Torbjörn Sjöstrand et al. "An Introduction to PYTHIA 8.2". In: *Comput. Phys. Commun.* 191 (2015), p. 159. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].

[107] Piotr Golonka and Zbigniew Was. "PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays". In: *Eur. Phys. J.* C45 (2006), pp. 97–107. DOI: 10.1140/epjc/s2005-02396-4. arXiv: 0506026 [hep-ph].

[108] Ye Li and Frank Petriello. "Combining QCD and electroweak corrections to dilepton production in FEWZ". In: *Phys. Rev. D* 86 (2012), p. 094034. DOI: 10.1103/PhysRevD.86.094034. arXiv: 1208.5967 [hep-ph].

[109] Aneesh Manohar et al. "How bright is the proton? A precise determination of the photon parton distribution function". In: *Phys. Rev. Lett.* 117.24 (2016), p. 242002. DOI: 10.1103/PhysRevLett.117.242002. arXiv: 1607.04266 [hep-ph].

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

[111] Richard D. Ball et al. "Parton distributions with LHC data". In: *Nucl. Phys. B* 867 (2013), p. 244. DOI: 10.1016/j.nuclphysb.2012.10.003. arXiv: 1207.1303 [hep-ph].

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

[113] Richard D. Ball et al. "Parton distributions for the LHC Run II". In: *JHEP* 04 (2015), p. 040. DOI: 10.1007/JHEP04(2015)040. arXiv: 1410.8849 [hep-ph].

[114] M. R. Whalley, D. Bourilkov, and R. C. Group. "The Les Houches accord PDFs (LHAPDF) and LHAGLUE". In: *HERA and the LHC: A Workshop on the implications of HERA for LHC physics. Proceedings, Part B.* 2005. arXiv: 0508110 [hep-ph].

[115] D Bourilkov, R C Group, and M R Whalley. "LHAPDF: PDF use from the Tevatron to the LHC". arXiv:0605240 (2006).

[116] Andy Buckley et al. "LHAPDF6: parton density access in the LHC precision era". In: *Eur. Phys. J. C* 75 (2015), p. 132. DOI: 10.1140/epjc/s10052-015-3318-8. arXiv: 1412.7420 [hep-ph].

[117] S. Agostinelli et al. "GEANT4: a simulation toolkit". In: *Nucl. Instrum. Meth. A* 506 (2003), p. 250. DOI: 10.1016/S0168-9002(03)01368-8.

[118] B. Clerbaux at al. "Search for high mass dielectron resonances with the full 2016 data". In: *CMS Notes* CMS AN-2016-404 (2017).

[119] The CMS Collaboration. "Generic Tag and Probe Tool for Measuring Efficiency at CMS with Early Data". In: *CMS Notes* CMS AN-2009-111 (2009).

[120] B. Clerbaux et al. "Dielectron resonance search in Run 2 at $\sqrt{s} = 13$ TeV pp collisions". In: *CMS Notes* CMS AN-2015-222 (2015).

[121] CMS cross-sections twiki page. https://twiki.cern.ch/twiki/bin/view/CMS/StandardModelCrossSectionsat13TeV?rev=24.

[122] D. Bourilkov and G. Daskalakis. *PDF Uncertainties for Z' searches at 13 TeV with Electron Pair or Muon Pair Final States*. CMS Note 2016/053. CERN, 2016.

[123] Abbiendi et al. "Search for High-Mass Resonances Decaying to Muon Pairs in pp Collisions at $\sqrt{s} = 13$ TeV with the full 2016 data set of 37 fb^{-1}and combination with 2015 result". In: *CMS Notes* CMS-AN-2016-391 (2016).

[124] B. Clerbaux et al. "Search for High Mass Di-Electron Resonances with 2016 Data". In: *CMS Notes* CMS AN-2016-190 (2016).

[125] S. Folgueras, N. Neumeister, and J.-F. Schulte. *Statistical Analysis for a Search for a Narrow Resonance*. CMS Note 2016/307. CERN, 2016.

[126] S. Schmitz et al. "Statistical Inference in a Search for a Narrow Resonance". In: *CMS Notes* CMS-AN-2012-185 (2012).

[127] V. Giakoumopoulou et al. "Z′ stats meeting". In: *CMS Indico Pages* (2015). URL: https://indico.cern.ch/event/387187/.

[128] Nicholas Metropolis et al. "Equation of State Calculations by Fast Computing Machines". In: *The Journal of Chemical Physics* 21.6 (1953), pp. 1087–1092. DOI: http://dx.doi.org/10.1063/1.1699114. URL: http://scitation.aip.org/content/aip/journal/jcp/21/6/10.1063/1.1699114.

[129] W. K. Hastings. "Monte Carlo sampling methods using Markov chains and their applications". In: *Biometrika* 57.1 (1970), pp. 97–109. DOI: 10.1093/biomet/57.1.97. eprint: http://biomet.oxfordjournals.org/content/57/1/97.full.pdf+html. URL: http://biomet.oxfordjournals.org/content/57/1/97.abstract.

[130] Elena Accomando et al. "Z′ at the LHC: Interference and Finite Width Effects in Drell-Yan". In: *JHEP* 10 (2013), p. 153. DOI: 10.1007/JHEP10(2013)153. arXiv: 1304.6700 [hep-ph].

[131] The CMS Collaboration. "Dielectron resonance search in Run 2 at $\sqrt{s} = 13$ TeV with 2.6 fb^{-1}of pp collisions". In: *CMS Physics Analysis Summaries* CMS-PAS-EXO-15-005 (2015).

[132] The CMS Collaboration. "Search for high-mass resonances in dilepton final state with 2016 data". In: *CMS Physics Analysis Summaries* CMS-PAS-EXO-16-031 (2016).

[133] The CMS Collaboration. "Search for narrow resonances in dilepton mass spectra in proton-proton collisions at $\sqrt{s} = 13$ TeV and combination with 8 TeV data". In: *Phys. Lett.* B768 (2017), pp. 57–80. DOI: 10.1016/j.physletb.2017.02.010. arXiv: 1609.05391 [hep-ex].

[134] The ATLAS Collaboration. "Search for new phenomena in the dilepton final state using proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *ATLAS Conference Notes* ATLAS-CONF-2015-070 (2015).

[135] The ATLAS Collaboration. "Search for new high-mass resonances in the dilepton final state using proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *ATLAS Conference Notes* ATLAS-CONF-2016-045 (2016).

[136] The ATLAS Collaboration. "Search for new high-mass phenomena in the dilepton final state using proton-proton collisions $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *ATLAS Conference Notes* ATLAS-CONF-2017-027 (2017).

[137] The ATLAS Collaboration. "Search for high-mass new phenomena in the dilepton final state using proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Phys. Lett.* B761 (2016), pp. 372–392. DOI: 10.1016/j.physletb.2016.08.055. arXiv: 1607.03669 [hep-ex].

[138] L. D. Landau. "The moment of a 2-photon system". In: *Dokl. Akad. Nauk SSSR* 60 (1948), p. 207.

[139] C. N. Yang. "Selection Rules for the Dematerialization of a Particle into Two Photons". In: *Phys. Rev.* 77 (1950), p. 242. DOI: 10.1103/PhysRev.77.242.

[140] Nathaniel Craig, Jamison Galloway, and Scott Thomas. "Searching for signs of the second Higgs doublet". arXiv:1305.2424 (2013).

[141] The ATLAS Collaboration. "Search for resonances decaying to photon pairs in 3.2 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *ATLAS Conference Notes* ATLAS-CONF-2015-081 (2015). URL: http://cds.cern.ch/record/2114853.

[142] The CMS Collaboration. "Search for new physics in high mass diphoton events in 3.3 fb^{-1}of proton-proton collisions at $\sqrt{s} = 13$ TeV and combined interpretation of searches at 8 TeV and 13 TeV". In: *CMS Physics Analysis Summaries* CMS-PAS-EXO-16-018 (2016). URL: https://cds.cern.ch/record/2139899.

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

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

[145] The CMS Diphoton working group. "Search for new physics in the diphoton spectrum at high mass at 13 TeV with 2016 data". In: *CMS Notes* CMS-AN-2016-167 (2016).

[146] The CMS Hgg working group. "Measurements of the Higgs boson at 125 GeV in the two photon decay channel in the Full LHC Run 1 Data at CMS". In: *CMS Notes* CMS-AN-2013-253 (2013).

[147] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "FastJet user manual". In: *The European Physical Journal C* 72.3 (2012), p. 1896. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-012-1896-2. URL: http://dx.doi.org/10.1140/epjc/s10052-012-1896-2.

[148] The CMS Diphoton working group. "Early search for new physics in the diphoton spectrum at high mass at 13 TeV". In: *CMS Notes* CMS-AN-2015-241 (2015).

[149] The CMS Diphoton working group. "Update on the search for new physics in the diphoton spectrum at high mass at 13 TeV". In: *CMS Notes* CMS-AN-2016-079 (2016).

[150] The CMS Hgg working group. "H$\rightarrow \gamma\gamma$ measurements at $\sqrt{s}$=13 TeV using 2016 data". In: *CMS Notes* CMS-AN-2016-209 (2016).

[151] The CMS Collaboration. "Observation of the diphoton decay of the Higgs boson and measurement of its properties". In: *Eur. Phys. J.* C74.10 (2014), p. 3076. DOI: 10.1140/epjc/s10052-014-3076-z. arXiv: 1407.0558 [hep-ex].

[152] The CMS Collaboration. "Search for resonant production of high mass photon pairs using 12.9 fb^{-1} of proton-proton collisions at $\sqrt{s} = 13$ TeV and combined interpretation of searches at 8 and 13 TeV". In: *CMS Physics Analysis Summaries* CMS-PAS-EXO-16-027 (2016). URL: http://cds.cern.ch/record/2205245.

[153] M Baak et al. "Interpolation between multi-dimensional histograms using a new non-linear moment morphing method". In: *Nucl. Instrum. Meth. A* 771 (2015), p. 39. DOI: 10.1016/j.nima.2014.10.033. arXiv: 1410.7388.

[154] The ATLAS and CMS Collaborations, LHC Higgs Combination Group. *Procedure for the LHC Higgs boson search combination in Summer 2011.* ATL-PHYS-PUB/CMS NOTE 2011-11, 2011/005. 2011. URL: http://cdsweb.cern.ch/record/1379837.

[155] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *Eur.Phys.J.* C71 (2011), p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0,10.1140/epjc/s10052-013-2501-z. arXiv: 1007.1727 [physics.data-an].

[156] A L Read. "Presentation of search results: the CLs technique". In: *Journal of Physics G: Nuclear and Particle Physics* 28.10 (2002), p. 2693. URL: http://stacks.iop.org/0954-3899/28/i=10/a=313.

[157] Gian F. Giudice, Riccardo Rattazzi, and James D. Wells. "Quantum gravity and extra dimensions at high-energy colliders". In: *Nucl. Phys.* B544 (1999), pp. 3–38. DOI: 10.1016/S0550-3213(99)00044-9. arXiv: 9811291 [hep-ph].

[158] Eugene A. Mirabelli, Maxim Perelstein, and Michael E. Peskin. "Collider signatures of new large space dimensions". In: *Phys. Rev. Lett.* 82 (1999), pp. 2236–2239. DOI: 10.1103/PhysRevLett.82.2236. arXiv: 9811337 [hep-ph].

[159] Tao Han, Joseph D. Lykken, and Ren-Jie Zhang. "On Kaluza-Klein states from large extra dimensions". In: *Phys. Rev.* D59 (1999), p. 105006. DOI: 10.1103/PhysRevD.59.105006. arXiv: 9811350 [hep-ph].

[160] JoAnne L. Hewett. "Indirect collider signals for extra dimensions". In: *Phys. Rev. Lett.* 82 (1999), pp. 4765–4768. DOI: 10.1103/PhysRevLett.82.4765. arXiv: 9811356 [hep-ph].

[161] H. Davoudiasl, J. L. Hewett, and T. G. Rizzo. "Phenomenology of the Randall-Sundrum Gauge Hierarchy Model". In: *Phys. Rev. Lett.* 84 (2000), p. 2080. DOI: 10.1103/PhysRevLett.84.2080. arXiv: 9909255 [hep-ph].

[162] The CMS Diphoton working group. "Updated combination of the results at $\sqrt{s} = 8$ TeV and 13 TeV in the search for exotic production of di-photon pairs". In: *CMS Notes* CMS-AN-2016-094 (2016).

[163] Glen Cowan. "Statistical Data Analysis". In: Oxford Science Publications (1998).

[164] D. Anderson et al. "Search for High Mass Diphoton Resonances at 13 TeV using the 2015 dataset". In: *CMS Notes* CMS AN-2016-191 (2016).

[165] The CMS Collaboration. "Search for new physics in high mass diphoton events in 2.6 fb^{-1}of proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *CMS Physics Analysis Summaries* CMS-PAS-EXO-15-004 (2016).

[166] The CMS Collaboration. "Search for Resonant Production of High-Mass Photon Pairs in Proton-Proton Collisions at $\sqrt{s} = 8$ and 13 TeV". In: *Phys. Rev. Lett.* 117.5 (2016), p. 051802. DOI: 10.1103/PhysRevLett.117.051802. arXiv: 1606.04093 [hep-ex].

[167] The ATLAS Collaboration. "Search for resonances in diphoton events at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *JHEP* 09 (2016), p. 001. DOI: 10.1007/JHEP09(2016)001. arXiv: 1606.03833 [hep-ex].

[168] The CMS Collaboration. "Search for high-mass diphoton resonances in proton–proton collisions at 13 TeV and combination with 8 TeV search". In: *Phys. Lett.* B767 (2017), pp. 147–170. DOI: 10.1016/j.physletb.2017.01.027. arXiv: 1609.02507 [hep-ex].

[169] The ATLAS Collaboration. "Search for scalar diphoton resonances with 15.4 fb^{-1} of data collected at $\sqrt{s} = 13$ TeV in 2015 and 2016 with the ATLAS detector". In: *ATLAS Conference Notes* ATLAS-CONF-2016-059 (2016).