Beyond the Standard Model Higgs Physics with Photons with the CMS Detector

Rafael Teixeira de Lima

Northeastern University

Boston - Massachusetts

CERN-THESIS-2017-189
11/04/2017

Supervised by

Toyoko Orimoto,

Professor of Physics

A thesis submitted for the degree of

Doctor of Philosophy

April 11, 2017

Acknowledgements

First and foremost, I would like to thank my Ph.D. advisor, Prof. Toyoko Orimoto, for her encouragement and support, particularly at tough times. After spending almost four years based at CERN hearing horror stories about life-draining and mind-destroying Ph.D. advisors, I consider myself extremely lucky to have worked with someone who would push me into new projects and adventures while still considering my well-being.

I would like to thank the second in command, one of the most upbeat people I have ever met, Italian kitchen master and ECAL connoisseur Andrea Massironi. A good part of what I have learned during my years at CERN were from white-board conversations with him and/or at pizza parties.

I would like to thank the ECAL DAQ team, particularly Giacomo Cucciati. Learning the ropes of the ECAL online software while cheering in the office whenever the code compiled was surely a great way to work.

I would like to thank, in general, the ECAL team, and, specifically, David Barney, André David, Evgueni Vlassov, David Petyt and Pedro Parracho. Their expertise is only surpassed by their willingness to teach newcomers about the wonders of ECAL.

I would like to thank the $b\bar{b}\gamma\gamma$ team, specially Amina Zghiche and Alexandra Carvalho, for the support and encouragement even when the energies were gone but there was still a lot to do.

I would also like to thank the Northeastern HEP team, specially Profs. Darien Wood and Emanuela Barberis, for their continuous useful input and discussions during our group meetings. Bruno, Dina, Mark and Mitchell, who made my one year in Boston an unforgettable experience. And finally, Nancy Wong, who repeatedly made sure I wouldn't lose my RA by forgetting to sign up for classes.

I would like to thank Álvaro and Ana, for your CERN visits have always been a great source of fun and brigadeiro-topped carrot cake.

A special warm thanks to Rafael Coelho Lopes de Sá, who followed me from D0 to CMS so he could continue teaching me about particle physics and the wonders of the academic life. His help during the always-so-difficult job-searching period probably kept me from throwing my laptop against a wall a couple of times.

Brenda, Brisa, Rebeca and Yasmin, you know the drill.

Finalmente, agradeço a minha mãe, Roseane Teixeira de Lima, que me apoia mesmo se eu deixar de ligar para ela na hora em que sempre combinamos - ela sabe que não faço por mal. Mas agradeço especialmente a ela pela minha formação como uma pessoa de sonhos e objetivos concretos, e que trabalha por eles. A minha mãe foi o meu primeiro exemplo de que as coisas melhoram com trabalho duro e dedicação. As coisas melhoraram tanto que ela já conheceu vários países da Europa e vai voltar para mais.

Abstract of Dissertation

The experimental discovery of the Higgs boson is one of the latest successes of the Standard Model of particle physics. Although all measurements have confirmed that this newly discovered particle is the Higgs boson predicted by the Standard Model, with no deviations to suggest otherwise, the Higgs boson can guide us to new models which modify the electroweak symmetry breaking mechanism or predict new states that couple to the Higgs. Therefore, it's paramount to directly look for modifications of our current model with the help of the recently discovered particle. In this thesis, two analyses involving beyond the Standard Model physics tied to the Higgs sector will be explored. First, looking at exotic Higgs decays, an analysis searching for the final state with photons and missing transverse energy will be presented. Then, the search for Higgs pair production, both resonantly and non-resonantly (a process predicted by the Standard Model, albeit at very low rates), in the final state with two bottom quark jets and two photons will be detailed. An important part of these results is the performance of the Compact Muon Solenoid's Electromagnetic Calorimeter, which will also be presented.

Table of Contents

Chapter 1

Introduction

1.1 The Standard Model of Particle Physics

The Standard Model of particles and fields (SM) describes the most basic constituents of matter and their dynamics, as currently known. It is built upon the hypothesis that fundamental, point-like particles exist (fermions) and that their interactions, via bosonic fields, can be described in a relativistic quantum field theoretical framework. Fundamental interactions are described in the SM by the gauge invariance principle, meaning that interactions will appear in the framework as gauge fields after imposing local and continuous (gauge) symmetries. An early example of this principle is the emergence of electromagnetism in the Schrödinger equation by imposing a local and continuous $U(1)$ symmetry.

As of today, the SM describes three types of interactions:

The Strong Interaction Also known as quantum chromodynamics or QCD. It appears in the SM via an $SU(3)$ gauge symmetry. This gauge group implies the existence of 8 gauge bosons, called gluons. Fermions particles that interact via QCD are called **quarks**, and are organized in triplets of a "color" charge. Fermions that do not interact via QCD are called **leptons**.

The Electromagnetic Interaction It appears in the SM via an $U(1)$ gauge symmetry, with one gauge boson associated to it (photon). The particles that interact via the electromagnetic interaction have electric charge.

The Weak Interaction Appears in the SM via an $SU(2)$ gauge symmetry. It has three different gauge bosons associated with it; two have electric charge (W^+ and W^-), while one is neutral (Z^0). The weak interaction charge is commonly called weak hypercharge.

The two last interactions have been further combined in a single theoretical model, called the electroweak interaction [1, 2, 3]. It is based on the fact that, before a spontaneous symmetry breaking occurs, both interactions can be described by a single gauge group, namely $SU(2) \otimes U(1)$. A main ingredient of this unification process is the spontaneous symmetry breaking mechanism based on the Higgs field [4, 5, 6].

The organization of the field content of the SM was completed with the observation that both quarks and leptons are organized into three generations. The first generation of leptons

contains the electron (e) and the electron neutrino (ν_e), while the first generation of quarks contains the up quark (u) and the down quark (d). The second generation of leptons contains the muon (μ) and the mu neutrino (ν_μ), while the second generation of quarks contains the charm quark (c) and the strange quark (s). The third generation of leptons contains the tau (τ) and the tau neutrino (ν_τ), while the third generation of quarks contains the top quark (t) and the bottom quark (b). The generation structure is particularly important by noting that particles of the same family, with left-handed chirality, form a weak isospin doublet, while the right-handed fermions are weak singlets:

$$\begin{aligned} L_l &= \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \quad \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, \quad \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L; \\ L_q &= \begin{pmatrix} u \\ d \end{pmatrix}_L, \quad \begin{pmatrix} c \\ s \end{pmatrix}_L, \quad \begin{pmatrix} t \\ b \end{pmatrix}_L; \end{aligned} \tag{1.1}$$

$$R_l = e_R,\, \mu_R,\, \tau_R;\; R_q = u_R,\, d_R,\, c_R,\, s_R,\, t_R,\, b_R. \tag{1.2}$$

1.2 Quantum Chromodynamics

QCD is the fundamental interaction in the SM that arises from the gauge symmetry under the $SU(3)$ gauge group [7]. Fermions that carry the color charge are called quarks and are represented as a triplet under $SU(3)$, a hypothesis that explains, for example, the baryonic structure discovered in the early 60's (the Eightfold Way). The QCD lagrangian can be written as:

$$\mathcal{L}_{\text{QCD}} = \bar{\psi} i\gamma^\mu \left(\partial_\mu + i g_s T^a G^a_\mu\right) \psi - \frac{1}{4} G^a_{\mu\nu} G^{a,\mu\nu} \tag{1.3}$$

where ψ are the $SU(3)$ triplet quark wave functions, G^a_μ are the gauge fields introduced by the symmetry (gluons), $G^{a,\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g_s f^{abc} G^b_\mu G^c_\nu$ is the field strength tensor of QCD, T^a are the 8 $SU(3)$ generators, and g_s is the QCD coupling constant.

The fact that the $G^a_{\mu\nu} G^{a,\mu\nu}$ term in the QCD lagrangian has terms in which the gluon fields interact with themselves, meaning that gluons also have color charge, leads to interesting behaviors of the theory. One of these interesting features is *confinement* [8], which forbids colored states to be detected in isolation. Therefore, only color neutral states, such as baryons (bound states with three quarks) and mesons (bound states with a quark and an anti-quark), are seen in nature. Another noteworthy feature is the *asymptotic freedom* [9, 10], which is the property that dictates that the QCD interaction between colored particles becomes asymptotically weaker at higher energies (smaller distances). This means that quarks bounded in protons and neutrons, for example, behave asymptotically as free particles.

1.3 The Electroweak Theory of Glashow-Weinberg-Salam

Through observations that date back to the early twentieth century, the electroweak interactions were built upon a set of experimental fundaments:

Universality The weak interactions are blind with respect to quarks and leptons generations. This means that the coupling strength is the same to any of the three families described previously.

Massless and Left-Handed Neutrinos For the energy regime to which the SM addresses, the neutrinos are idealized as massless (even though experiments have proved that they do have mass, even if below eV scale), and they only exist (or only interact with the SM) in their left-handed chirality.

Chirality The electroweak interactions are a chiral: the structure of the theory should treat right-handed and left-handed particles differently, thus breaking parity symmetry.

Mixing The electroweak eigenstates of quarks is not the same as their mass eigenstates. Therefore, the electroweak interaction acts on mixed states of quarks mass eigenstates. This quark mixing is described by the CKM matrix.

The unified electroweak theory is based on the $SU(2)_L \otimes U(1)_Y$ gauge group. The indices L and Y address the fact that the $SU(2)$ gauge bosons ($\vec{B}_\mu$) will only act upon left-handed particles, while the $U(1)$ gauge boson (A_μ) will couple to particles that contain weak hypercharge. After the spontaneous symmetry breaking, these four gauge bosons will become the three weak bosons and the photon. We can write a lagrangian, symmetric under this gauge group, as:

$$\begin{aligned}\mathcal{L}_{\text{GWS}} &= \mathcal{L}_{Fermions} + \mathcal{L}_{Gauge} \\ &= \left\{ \bar{R} i\gamma^\mu \left(\partial_\mu + i\frac{a}{2} A_\mu Y \right) R + \bar{L} i\gamma^\mu \left(\partial_\mu + i\frac{a}{2} A_\mu Y + i\frac{b}{2} \vec{\sigma} \cdot \vec{B}_\mu \right) L \right\} \\ &\quad -\frac{1}{4} \left\{ F^M_{\mu\nu} F^{\mu\nu}_M + f_{\mu\nu} f^{\mu\nu} \right\},\end{aligned} \tag{1.4}$$

where R and L represents the right-handed and left-handed fermions, respectively; a and b represents the coupling constants of the model; γ and σ represent the Dirac and Pauli matrices, respectively; and F and f are the field strengths tensors of the electroweak bosons.

1.3.1 Spontaneous Symmetry Breaking and Mass Generation

One important detail of the lagrangian in Equation 1.4 is the fact that mass terms for the electroweak bosons are forbidden by gauge invariance. Particularly, classical mass terms such as $m_B^2 \vec{B}_\mu \cdot \vec{B}^\mu$ violate the model's gauge symmetry given that $\vec{B}_\mu$ transforms as:

$$B^M_\mu \quad \rightarrow \quad B^M_\mu + \frac{1}{b} \partial_\mu \alpha^M(x) + \epsilon^M_{NO} B^N_\mu B^O_\mu. \tag{1.5}$$

The Higgs mechanism is the way the SM generates these mass terms while keeping the gauge invariance intact. In its non-Abelian form, it adds an $SU(2)_L$ scalar doublet field Φ and a quartic potential $V(\Phi)$ to the SM lagrangian described previously:

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \tag{1.6}$$

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda \left(\Phi^\dagger \Phi\right)^2, \tag{1.7}$$

with ϕ^+ e ϕ^0 as complex scalar fields.

As a doublet, the covariant derivative that acts upon Φ is similar to the one that acts on left-handed leptons:

$$D_\mu = \partial_\mu + i\frac{a}{2}A_\mu Y + i\frac{b}{2}\vec{\sigma}\cdot\vec{B}_\mu. \tag{1.8}$$

Therefore, we can write the final SM lagrangean, including the Higgs mechanism, as:

$$\mathcal{L}_{Higgs+GSW} = (D\Phi)^\dagger (D\Phi) + \mu^2 \Phi^\dagger \Phi + \lambda \left(\Phi^\dagger \Phi\right)^2 + \mathcal{L}_{GSW}. \tag{1.9}$$

The shape of $V(\Phi)$ depends on the values of μ and λ. Therefore, we can place bounds on these parameters based on the role $V(\Phi)$ must perform. First, $V(\Phi)$ must not have a global minimum at $|\Phi| \to \infty$, which means $\lambda > 0$. Second, in order to properly generate the gauge boson masses, the $V(\Phi)$ minimum must not be at zero. This is accomplished by setting $\mu^2 < 0$, which creates a $V(\Phi)$ minimum at $|\Phi| = \sqrt{-\frac{\mu^2}{\lambda}} \equiv v$, also known as the electroweak vacuum expectation value. The found vacuum is spherically symmetric, meaning it remains unchanged by rotations such as $\Phi_{min} \to \exp\left(i\frac{\vec{\sigma}}{2}\cdot\vec{\xi}\right)\Phi_{min}$. With this fact, we can expand the Higgs field around its vacuum expectation value as:

$$\Phi \to \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}, \tag{1.10}$$

Writing this expression explicitly in the covariant derivative definition, calculating $(D\Phi)^\dagger (D\Phi)$ and gathering the terms involving the Higgs vacuum perturbation, we have:

$$\left[\frac{1}{4}\left(aA_\mu - bB_\mu^3\right)^2 + \frac{b^2}{4}\left(B_\mu^1 - iB_\mu^2\right)\left(B_\mu^1 + iB_\mu^2\right)\right]\left(\frac{1}{\sqrt{2}}\left(v + H(x)\right)\right)^2 \tag{1.11}$$

We can simplify this equation with the following redefinitions:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}\left(B_\mu^1 \pm iB_\mu^2\right), \tag{1.12}$$

$$Z_\mu^0 = \frac{bB_\mu^3 - aA_\mu}{\sqrt{b^2+a^2}}, \tag{1.13}$$

$$\mathcal{A}_\mu = \frac{bB_\mu^3 + aA_\mu}{\sqrt{b^2+a^2}}. \tag{1.14}$$

Using these new field definitions, we can expand $(v + H(x))^2$ in equation 1.11 to obtain their couplings to the Higgs vacuum and the Higgs perturbation:

$$\frac{v^2}{8}\left(b^2+a^2\right)Z_\mu^0 Z^{0\mu}+\frac{v^2b^2}{8}W_\mu^+W^{-\mu}+ \tag{1.15}$$
$$\frac{v}{8}\left(b^2+a^2\right)Z_\mu^0 Z^{0\mu}H\left(x\right)+\frac{vb^2}{8}W_\mu^+W^{-\mu}H\left(x\right)+$$
$$\frac{1}{8}\left(b^2+a^2\right)Z_\mu^0 Z^{0\mu}H\left(x\right)H\left(x\right)+\frac{b^2}{8}W_\mu^+W^{-\mu}H\left(x\right)H\left(x\right).$$

The first two terms in the equation above are canonical mass terms to the Z^0 and $W^\pm$ fields, but the field $\mathcal{A}$ remains massless. This takes the model back exactly to standard views of the weak interactions, with one massive neutral boson and two massive charged bosons, and on the electromagnetic interaction, with one massless neutral boson. The mass of the Z^0 and $W^\pm$ bosons can then be expressed based on the electroweak vacuum expectation value and on the electroweak couplings, the latter being well known experimentally, as: $m_W^2=\frac{v^2b^2}{4}$ and $m_Z^2=\frac{v^2}{4}\left(b^2+a^2\right)$. We can express v^2 as a function of m_W and relate it to the Fermi coupling constant (G_F), such that $v=\sqrt{\frac{-\mu^2}{\lambda}}=\left(\sqrt{2}G_F\right)^{1/2}\approx 246\,\text{GeV}$ [11], where G_F is precisely obtained through muon lifetime measurements.

We can also investigate the shape of the Higgs potential after the vacuum has been fixed. Using the previous definition of v, we have:

$$V\left(\Phi\right) = \frac{-v\lambda}{2}\left(v+H\left(x\right)\right)^2+\frac{\lambda}{4}\left(v+H\left(x\right)\right)^4 \tag{1.16}$$
$$= -\frac{\lambda v^4}{2}+\lambda v^2H^2\left(x\right)+\lambda vH^3\left(x\right)+\frac{\lambda}{4}H^4\left(x\right). \tag{1.17}$$

The first important term in the equation above, $\lambda v^2H^2\left(x\right)$, can be interpreted as a mass term for a scalar field. This means that perturbations of the Higgs field around the vacuum will be realized as a scalar boson, the Higgs boson, with mass equal to $m_H^2=2\lambda v^2$. The following terms describe how the Higgs boson interacts with itself via triple and quartic couplings. It is interesting to note that the Higgs boson can be understood as an excitation in the transverse direction to the direction in which the vacuum of the Higgs potential is symmetric. The other degrees of freedom present in the theory, namely the longitudinal excitations along the direction in which the vacuum is symmetric, were "eaten" by the massless fields $\vec{B}$ and A, as new degrees of freedom related to their masses.

1.3.2 Fermion Masses

In the previous section, the Higgs mechanism was used to break electroweak symmetry and give mass to three of the four gauge bosons present in the theory. However, the SM fermion sector also presents problems when dealing with canonical mass terms. Namely, knowing that left-handed and right-handed fermions are, respectively, doublets and singlets under the electroweak interaction, their gauge transformations will be different, and terms such as $m_f\bar{f}f=m_f\left(\bar{f}_Lf_R+\bar{f}_Rf_L\right)$ will break the gauge invariance.

However, we can note that terms such as $\bar{f}_L\Phi$ are an electroweak singlet and $\bar{f}_R\Phi^\dagger$ are a doublet (meaning that the Higgs interaction with fermions changes their chirality). Therefore,

one can construct a new type of mass term based on fermionic Yukawa interactions with the Higgs field:

$$g_f \left[\bar{f}_L \Phi f_R + \bar{f}_R \Phi^\dagger f_L \right] \tag{1.18}$$

which preserves the $SU(2)$ gauge invariance.

Using the choice of vacuum from the previous section, we can write an example of this mass term applied to the bottom quark:

$$\begin{aligned} \mathcal{L}_{\text{Yukawa}} &= g_b \left[\gamma^0 \begin{pmatrix} t_L^\dagger & b_L^\dagger \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} b_R + \text{c.c} \right] && (1.19) \\ &= \frac{g_b}{\sqrt{2}} \left(\bar{b}_L b_R + \bar{b}_R b_L \right) \left(1 + \frac{H(x)}{v} \right) && (1.20) \\ &= \frac{g_b v}{\sqrt{2}} \bar{b} b + \frac{g_b}{\sqrt{2}} \bar{b} b H(x) \,. && (1.21) \end{aligned}$$

The first term in the last step of the equation above is a canonical term for the b quark mass, while the second term dictates the b quark coupling to the Higgs boson. In order to give mass also to the top quark, we must use the charge conjugate to the Higgs doublet:

$$\Phi^c = i\sigma^2 \Phi^* = \frac{1}{\sqrt{2}} \begin{pmatrix} v + H(x) \\ 0 \end{pmatrix}, \tag{1.22}$$

which rotates the original Higgs doublet preserving the vacuum symmetry.

The same mechanism can be applied to the leptons. However, due to the non-existence of right-handed neutrinos, it cannot be applied to give mass to the left-handed neutrinos. As of today, we know neutrinos indeed have mass, albeit small, and specific mechanisms have been developed to address this issue. But, given that the neutrino masses are several orders of magnitude smaller than the characteristic scale of the electroweak interactions (vacuum expectation value), the approximation that the neutrinos are massless is well justified in the SM.

1.4 Higgs Boson Phenomenology

An important aspect of the Higgs mechanism is the prediction of a new mass eigenstate, the Higgs boson, with couplings to fermions and electroweak bosons completely defined by the model. The only free parameter present is the Higgs boson mass, $m_H^2 = 2\lambda v^2$, that dictates both the branching fraction of the Higgs decays, and its cross section at colliders. These two quantities, as a function of the Higgs mass, can be seen in Figure 1.1 [12].

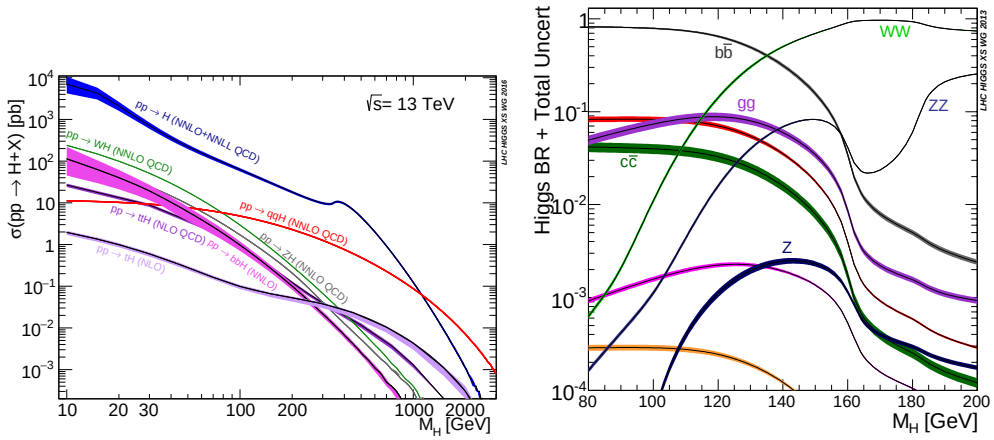

Figure 1.1: (Left) The Higgs production cross section, as a function of the Higgs boson mass, for proton-proton colliders at 13 TeV. (Right) The Higgs branching fractions to SM particles, as a function of the Higgs boson mass.

Even though the Higgs boson only directly couples to massive particles, processes via loops allow decays such as $H \to \gamma\gamma$ and production mechanisms such as $gg \to H$, where g are gluons. The latter example is particularly important, since it is the production mechanism with highest cross section at the Large Hadron Collider, a proton-proton collider. For other types of colliders, this picture might be different, such as for the Tevatron (proton-antiproton), in which the main production mechanism was through associated production with an electroweak boson.

1.5 Higgs Physics Current Status

While the Higgs boson, and the electroweak symmetry breaking mechanism applied to the SM itself, was predicted in the 60s, it wasn't until 2012 that it was experimentally verified. The CMS and ATLAS experiments, working at the Large Hadron Collider at CERN, detected statistically significant signals compatible with the SM predictions for the Higgs boson, using approximately 5 fb^{-1} of 7 TeV and 8 TeV data [13, 14]. The discovery of the SM-like Higgs boson in 2012 led to the Nobel prize to be awarded to some of the leading theorists behind the electroweak symmetry breaking mechanism: Peter Higgs and François Englert. Robert Brout, Englert's long-time collaborator and also an important contributor to the model, unfortunately passed in 2011, before the discovery and the award.

In 2012, it was not possible to say with certainty that this anomaly was, without a doubt, the SM Higgs. However, five years later, the data that has been analyzed by the LHC experiments has confirmed the SM predictions for this new particle, with a mass measured as about 125 GeV, including its spin (scalar) and CP eigenvalue (even) [15, 16].

The Higgs boson decays to electroweak bosons, $H \to WW$, $H \to ZZ$ and $H \to \gamma\gamma$, have been measured and these branching fractions, assuming SM-like cross section, match the expected values. The Higgs decay modes to pairs of fermions have yet to be detected with the statistical significance seen in its bosonic modes. However, decays involving the third genera-

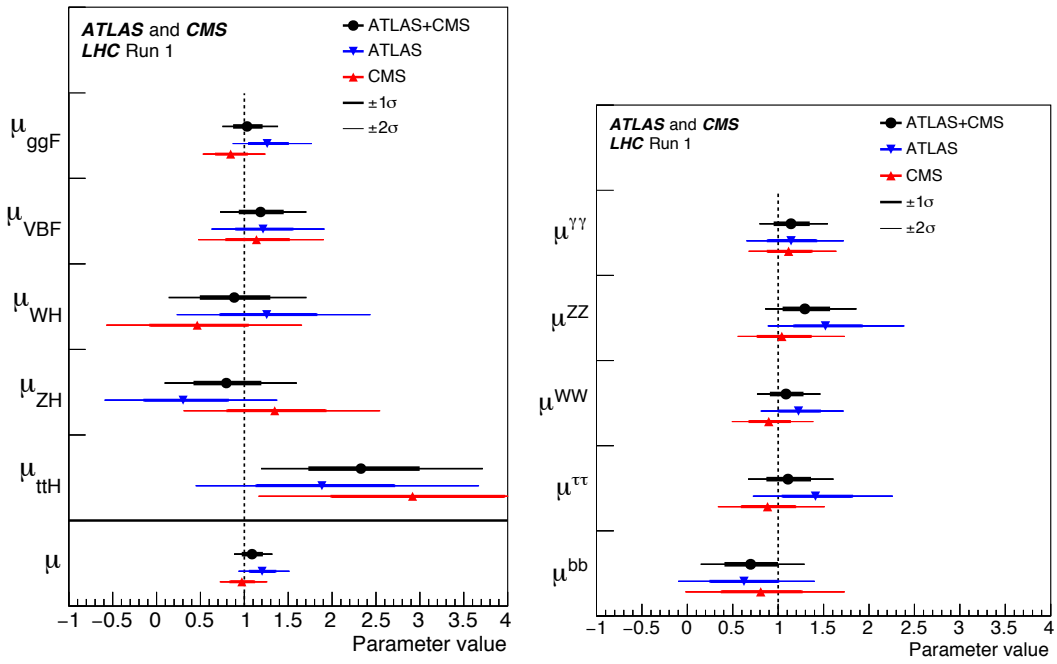

Figure 1.2: (Left) The signal strength measured for the five main Higgs boson production mechanisms, assuming SM branching fractions. (Right) The signal strength measured for the three bosonic Higgs decays and the two third family fermionic decays, assuming SM production.

tion of quarks and leptons, $H \to b\bar{b}$ and $H \to \tau\tau$, have seen evidence of such processes when combining searches in different production channels. The latest results summary published by CMS and ATLAS can bee seen in Figure 1.2 in terms of signal strengths (μ), defined as the ratio of the measured Higgs boson rate to its SM prediction [17].

Other than measuring signal strengths, an important part of studying the Higgs boson is understanding its differential properties. Measuring kinematic distributions related to the Higgs candidates, such as the transverse momentum and pseudo-rapidity profiles, can indicate possible corrections and beyond the SM contributions to the production mechanisms. For example, new particles appearing in the gluon fusion loop can have different coupling structures that induce kinematic correlations not present in the SM. In Figure 1.3, the CMS differential Higgs boson measurements using the $H \to \gamma\gamma$ channel in terms of the Higgs transverse momentum and rapidity using 8 TeV data [18].

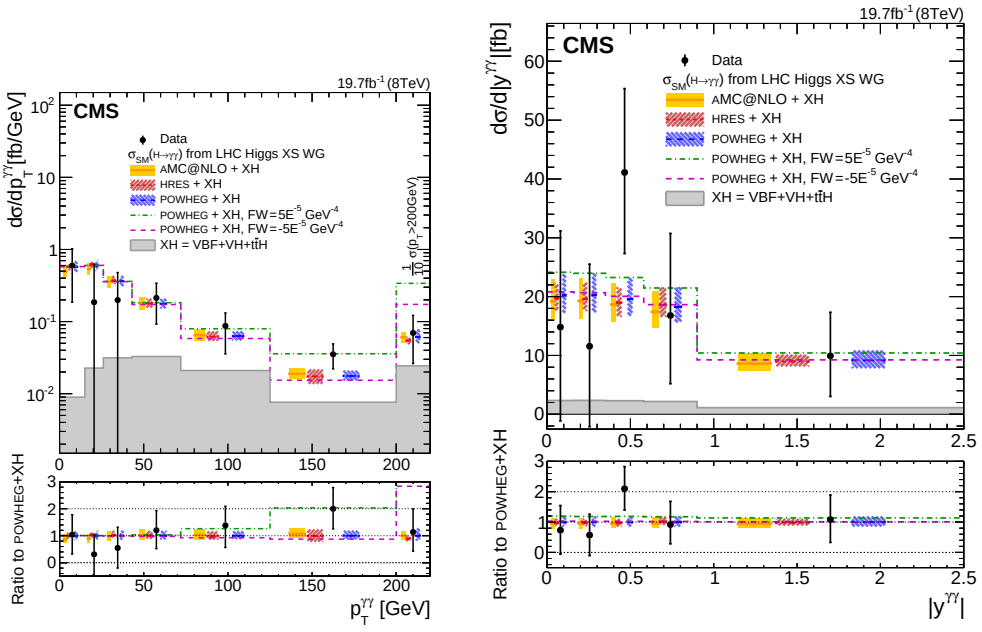

Figure 1.3: CMS differential Higgs boson measurements using the $H \to \gamma\gamma$ channel in terms of the Higgs transverse momentum (left) and rapidity (right) using 8 TeV data. The measurements are compared to different Monte Carlo generators.

1.6 Higgs as a Probe of New Physics

The experimental confirmation of the Higgs boson marks another success of the SM. However, the measured mass of the Higgs boson poses a naturalness conundrum: being a scalar, the Higgs mass is subject to higher order corrections from loops that contain spin 0, 1/2 and 1 particles, which diverge as a function of the energy scale. If the SM is valid for all energy scales, these corrections would take the Higgs mass to its highest possible energy scale, the Planck scale ($\approx 10^{19}$ GeV), which is in clear contrast with the Higgs mass value known today. In principle, it is possible that the higher order corrections to the Higgs mass cancel out in a way to leave $m_H = 125$ GeV, however, such an "accidental" fine tuning is not well motivated within the theory. The presence of these two distinct and far apart energy scales in the electroweak theory (Higgs mass and electroweak vacuum versus the Planck scale) is called the hierarchy problem.

Several models and extensions to the SM have been proposed to solve the Hierarchy problem. Among them, and perhaps the most famous, is supersymmetry (SUSY) [19, 20, 21, 22, 23, 24, 25]. SUSY explains the higher order cancelations by postulating a new symmetry of nature that correlates bosons and fermions. By symmetrizing the bosonic and fermionic contributions to the Higgs mass corrections, the higher order terms are canceled out naturally (since fermion loops contribute with an overall minus sign relative to boson loops).

The existence of SUSY implies that, for every elementary fermion in nature, there must exist a boson superpartner with the same mass (same applies to bosons and fermion superpartners). However, this has not been observed, which means that, if SUSY exists, it must be

a broken symmetry. In order to still address the Hierarchy problem, the SUSY breaking scale must be such that the superpartner masses should not exceed $\mathcal{O}(1$ TeV). This requirement also has the feature that it unifies the QCD and electroweak coupling constants at $\mathcal{O}(10^{16}$ GeV).

The simplest supersymmetric extension of the SM, known as the Minimal Supersymmetric SM (MSSM), adds the minimal amount of extra fields to the current model to realize SUSY - in the MSSM, each SM fermion/boson will have one boson/fermion superpartner. It also introduces changes to the Higgs sector: two Higgs doublets are required in order to perform the electroweak symmetry breaking, and give masses to both up and down type quarks and leptons. The existence of this extra doublet now implies that, instead of one Higgs boson mass state, five new bosons exist: two neutral CP-even bosons: h, H (one of which is usually identified as the particle discovered in 2012); one neutral CP-odd boson (pseudoscalar): A; and two charged charged bosons: $H^{\pm}$.

The MSSM is one example of model that adds to the Higgs sector of the SM an extra scalar doublet. Models with this characteristic are generally referred as Two-Higgs Doublet Models (2HDM) [26]. These models are also characterized by new phenomena related to the SM-like Higgs boson, either by exotic production or decay. For the first case to happen, at least one of the new resonances in the model, H for example, must be heavier than the SM-like Higgs (h), which then allows decays such as $H \to h + X$, where X can either be a Z, an A or even another h. Alternatively, if A, for example, is lighter than $m_h/2 \approx 62.5$ GeV, decays such as $h \to AA$ become possible.

These specific models serve as illustrations of how a modification to the SM can lead to exotic physics appearing in the Higgs sector. More generally, looking for new physics with the Higgs boson is a justifiable ansatz in both phenomenological terms - most extensions to the SM predict specific modifications of the Higgs sector - and in theoretical terms - the Higgs sector is intimately tied to some of the most undesirable features of the SM, such as the hierarchy problem. Another one of these undesirable features is the electroweak vacuum stability problem.

The electroweak vacuum in the SM is defined as the vacuum expectation value given by the Higgs potential minimum: $v \approx 246$ GeV. Even though the Higgs potential has one global minimum associated to this vacuum, higher order corrections change the shape of this potential and can turn this global minimum into a local minimum. The impact of these corrections are mostly dominated by the Higgs-top quark Yukawa coupling, and so, the stability of the electroweak vacuum depends on the Higgs mass and on the top quark mass, as seen in Figure 1.4 [27].

Given the current values for the Higgs boson and the top quark masses, our universe currently lies at a meta-stable point. These instabilities in the electroweak vacuum can have catastrophic consequences for the fate of the universe, since the vector boson masses, and therefore the strength of the weak interactions, are tied to this value. If it is indeed confirmed that we live in such a state, with more precise measurements for the mass parameters, new mechanisms must be in place to safeguard the universe against these features.

One important caveat of this stability analysis is the assumption that the tree-level Higgs potential has the quartic shape described in Equation 1.7. However, there are no strong theoretical indications that $V(|\Phi|)$ must have this exact shape. Theoretically, the SM $V(|\Phi|)$ is the simplest potential that correctly breaks the electroweak symmetry within theoretical

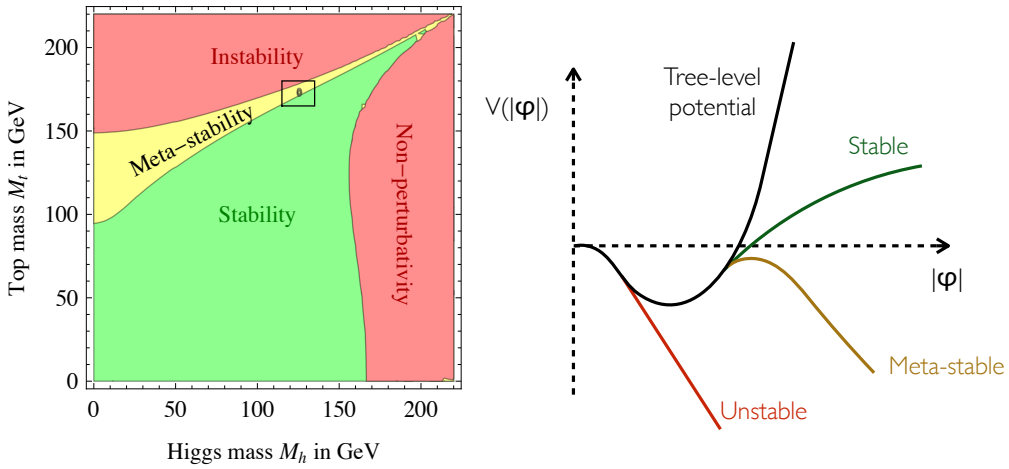

Figure 1.4: (Left) Stability of the Higgs potential after radiative corrections, as a function of the Higgs boson mass and the top quark mass. (Right) The effect of radiative corrections on the shape of the Higgs potential.

constraints, such as unitarity and renormalizability.

Experimentally, in order to probe its shape, the $V(|\Phi|)$ potential parameters must be measured directly. Through the detection of the Higgs boson and its mass measurement, the parameter μ in $V(|\Phi|)$ is determined, since $m_h = |\mu|$. As seen in Equation 1.17, the higher order terms of $V(|\Phi|)$ are related to Higgs self-interaction terms. It's important to note that, even if the *true* Higgs potential (the one that correctly breaks the electroweak symmetry and brings stability to the SM) is not $V(|\Phi|)$, an expansion of this general potential around the vacuum will generally create higher order terms with respect to H. Therefore, directly measuring the Higgs self-interaction terms, such as the Higgs triple coupling, is a crucial step towards understanding the SM.

Chapter 2

The Compact Muon Solenoid Experiment at the Large Hadron Collider

The Compact Muon Solenoid (CMS) experiment is one of the four experiments operating at the Large Hadron Collider (LHC) particle accelerator at the European Center for Nuclear Research (CERN), in the suburbs of Geneva, Switzerland. Along with ATLAS, LHCb and ALICE, they are the latest generation of high energy physics experiments to take place at high energy hadron colliders. CMS and ATLAS are multi-purpose experiments, looking for different types of processes arising from the proton-proton collisions at the LHC, while LHCb and ALICE are designed for specific purposes: rare processes involving flavor physics and QCD at high temperatures, respectively. In the following sections, the LHC will be briefly summarized, followed by a description of the CMS experiment.

2.1 The LHC Accelerator Complex

The Large Hadron Collider (LHC) is a circular proton-proton collider operated at European Organization for Nuclear Research (CERN) in Geneva, Switzerland [28]. The collider has a circumference of 27 km and is located in the underground tunnel that hosted the Large Electron Positron (LEP) collider, operated from 1989 until 2000. The LHC ring includes two adjacent beam pipes, each containing one of two colliding beams, which travel in opposite directions. The two beams are focused and bent in their circular trajectory by a system of more than 1600 superconducting magnets: 1232 dipole magnets operating at a temperature of less than 2 K generate a magnetic field of 8.3 T that maintains the circular motion of the two beams around the LHC; 392 quadrupole magnets are used to keep the beams focused as they travel inside the collider. The main part of the LHC physics program consists in operating the machine as a proton-proton (pp) collider, while part of the machine schedule is periodically dedicated to the delivery of heavy-ion collisions.

In the LHC, beams are accelerated by the electromagnetic field generated by radio-frequency cavities (eight per beam) located along the collider ring. Each of these cavities also operates in superconducting state, at a temperature of approximately 4.5 K, and can

deliver a voltage of 2 MV at a frequency of 400 MHz. Prior to their injection in the LHC, the colliding particles are grouped together in bunches and pre-accelerated in a chained system of smaller accelerators, which complete the CERN accelerator complex. A layout of the CERN accelerator complex can be seen in Figure 2.1.

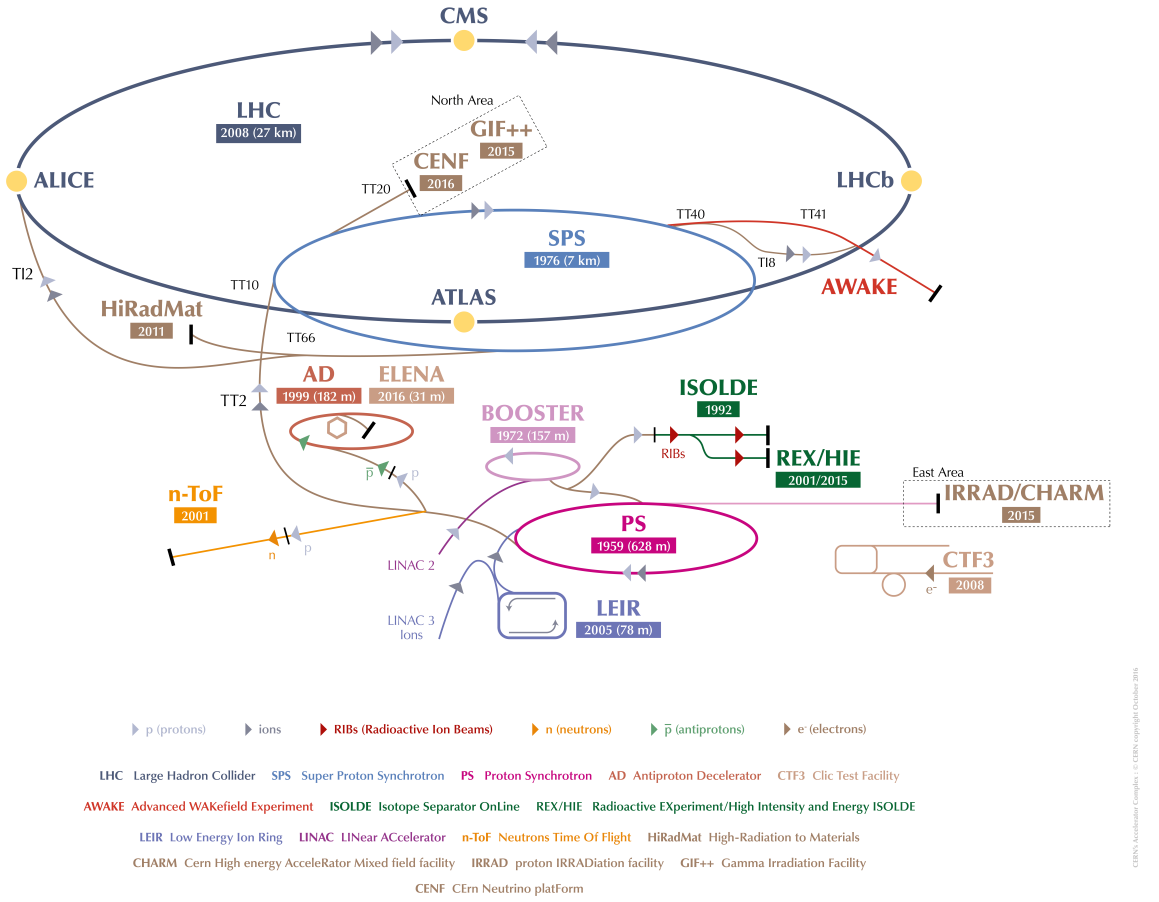

Figure 2.1: Sketch of the CERN accelerator complex, including the Large Hadron Collider.

The LHC was designed to deliver pp collisions at a center-of-mass energy of $\sqrt{s} = 14$ TeV with an instantaneous luminosity of 10^{34} cm^{-2}s^{-1}, using a bunch-spacing of 25 ns (time period between two consecutive bunch crossings). In the first years of data-taking, the LHC operated below the design values to reduce the time needed for commissioning the machine and to follow a safer strategy concerning the operation of the LHC magnets. The first run of the LHC physics program with pp collisions started in 2009 at $\sqrt{s} = 7$ TeV and continued through 2010 and 2011 reaching a maximum instantaneous luminosity of 3.5×10^{33} cm^{-2}s^{-1} with a bunch spacing of 50 ns, translated into 5 fb^{-1} collected by the CMS experiment. During 2012 and 2013, the center-of-mass energy was increased to $\sqrt{s} = 8$ TeV, and the machine reached a peak instantaneous luminosity of 7.7×10^{33} cm^{-2}s^{-1}. In this period the CMS experiment recorded a total integrated luminosity of 19.7 fb^{-1} with all subdetectors fully operational and nominal magnetic field.

From 2013 to 2015, the LHC went through upgrades aimed at operating the collider at its nominal configuration. The first stable collisions in 2015 were in May, at $\sqrt{s} = 13$ TeV, with a bunch spacing of 50 ns, then changed to 25 ns later in the year, reaching a peak luminosity of 5×10^{33} cm^{-2}s^{-1}. By the end of the first year of the LHC Run-2, the CMS experiment recorded a total integrated luminosity of 2.6 fb^{-1} with a magnetic field of 3.8 T and all subdetectors in full operation. The year of 2016 was a record breaking period for the LHC, with a peak luminosity of 1.4×10^{34} cm^{-2}s^{-1} (reaching its nominal value) and an integrated luminosity of about 40 fb^{-1} in ATLAS and CMS at $\sqrt{s} = 13$ TeV.

2.2 The CMS Experiment

The CMS experiment is one of the multi-purpose experiments at the LHC [29]. Its goal is to measure both Standard Model processes and search for physics beyond the Standard Model. In order to achieve this, it must precisely reconstruct objects that are typically present at high energy colliders (photons, electrons, muons and hadronic jets), including their positions and 4-momenta (3-dimensional momentum plus energy).

The overall CMS design is that of a cylindrical onion, meaning that every individual subdetector (specializing in measuring different types of event information) is stacked together concentrically, like onion layers. These different subdetectors are shown in Figure 2.2, and will be described in the following sections.

In order to measure charged particles momenta with the desired precision, a strong magnetic field must be present in the detector, so the momentum can be extracted from the particles' bending. The CMS magnet is a superconducting solenoid located between the CMS calorimeters and muon detectors. It is 13 meters long and 6 meters in inner diameter, and provides a 3.8T magnetic field. With this geometry, the direction of the magnetic field will be different at the CMS tracker system, in which charged particles trajectories are measured close to the interaction point, and at the CMS muon system, which is a tracking subdetector specialized in measuring muon trajectories. This optimizes the muon momentum measurement by having two magnetic bendings in the reconstructed muon path, instead of only one.

Muons are long-lived particles that interact only via the electromagnetic and weak interactions. Different from most particles produced in high energy collisions, muons are able to go through all layers of CMS. Electrons, on the other hand, are stopped by the CMS electromagnetic calorimeter (ECAL), where they produce a particle shower based mostly on bremsstrahlung radiation (at high energies), and lose their energies almost completely. Due to the muons higher mass, they do not emit bremsstrahlung radiation as much as an electron, and therefore, are able to go through the CMS ECAL. Photons are measured by the ECAL in a similar way to electrons, the difference being that a photon shower starts by an electron-positron pair production from the original high energy photon. Muons also go through the CMS hadronic calorimeter (HCAL), which measures the energy of hadronic particles in a similar way that the ECAL measures the energy of photons and electrons, but with showers caused by the strong interaction.

The CMS calorimeters are usually called destructive detectors, since, in order to measure the particle's momentum, the particle must not leave the detector. On the other hand, the tracking system aims to measure the particles' energy with little interaction to minimize the

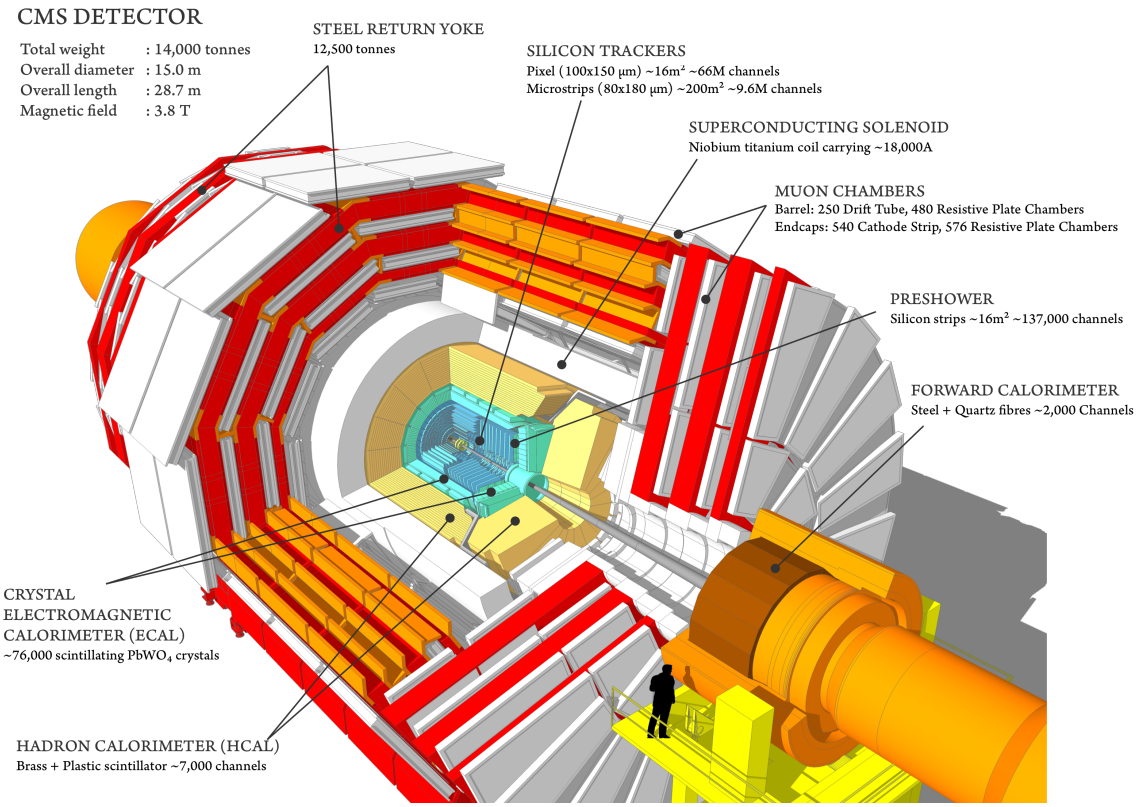

Figure 2.2: Sketch of the CMS experiment at the LHC, showing its different subdetectors stacked concentrically, like a cylindrical onion.

energy loss (so that it can be precisely measured by the calorimeters). This is performed, for example, through ionizing radiation in semi-conductor or gaseous devices.

2.2.1 The CMS Tracking Detectors

The CMS tracking system must provide precise reconstruction for charged particle trajectories with minimal interference in the particle energies. These trajectories are reconstructed in terms of tracks, which are built from the different interactions of the incoming particle with the tracking system (hits). It must also cope with the high charged particle flux at high luminosity, which requires a high detector segmentation to reconstruct large track multiplicities at the same time.

The CMS tracking system is entirely constructed of silicon, which detects hits by measuring electron-hole currents in the device produced by ionizing radiation. The geometry of these silicon sensors is different for three regions within the CMS detector, which are defined based on the charged particle flux from the LHC:

- Closest to the interaction vertex where the particle flux is the highest, pixel detectors are placed. The size of a pixel is $\approx 100 \times 150\ \mu\text{m}^2$, giving an occupancy of about 10^{-4} hits per pixel sensor per LHC crossing.

- In the intermediate region ($20 < r < 55$ cm), the particle flux is low enough to enable the usage of silicon microstrip detectors with a minimum cell size of 10 cm $\times$ 80 μm, leading to an occupancy of $\approx 2 - 3\%$ per LHC crossing.
- In the outermost region ($r > 55$ cm) of the inner tracker, the particle flux has dropped sufficiently to allow use of larger-pitch silicon microstrips with a maximum cell size of 25 cm $\times$ 180 μm, while keeping the occupancy to $\approx 1\%$.

The pixel layers of the CMS tracker consist of three cylindrical layers of pixel modules surrounding the interaction point at radii of 4.4, 7.3 and 10.2 cm, and are topped at each side with two pixel module disks at $|z| = 34.5$ cm and 46.5 cm. In total, it covers about 1 m^2, with 66 million pixels. The pixel detector delivers three high precision hits for each charged particle trajectory: the spatial resolution is measured to be about 10 μm for the $r-\phi$ measurement and about 20 μm for the z measurement

After the pixels, the strip tracker detector covers the radial range between 20 cm and 116 cm. It consists of three different subsystems: the tracker inner barrel, inner disks and outer barrel (TIB, TID and TOB). The inner subsystems (TIB and TID) extend in radius from 20 cm to 55 cm, are composed of 4 layers in the barrel and 3 layers at each end, and deliver up to 4 measurements in $r - \phi$ by track, with single hit resolution between 23 μm and 35 μm. The outer system (TOB) extends to a radius of 1.2m and to $|z| = 118$ cm, consists of 6 layers of micro-strip sensors, providing another 6 $r - \phi$ measurements with single point resolution between 35 μm and 53 μm.

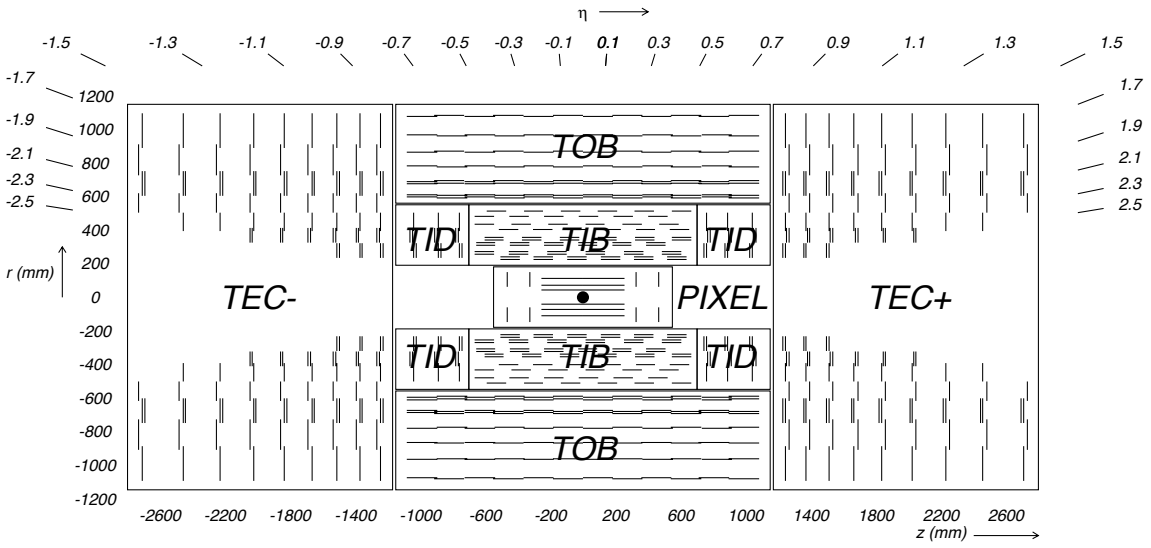

Figure 2.3: Sketch of the CMS tracker system.

Beyond the TOB, the tracker endcaps (TEC+ and TEC-, where the sign relates to the location along the z axis) cover the region 124 cm $< |z| <$282 cm and 22.5 cm $< |r| <$ 113.5 cm. Each endcap consists of 9 disks, each with up to 7 rings of silicon micro-strips, providing up to 9 $r - \phi$ measurements per track. Additionally, some of the modules on the TIB, TID, TOB and TEC also carry a second micro-strip sensor module, mounted with a stereo angle

of 100 mrad, to provide a measurement of the second coordinate (z in the barrel and r on the disks/endcaps). The overall CMS tracker module structure can be seen in Figure 2.3.

2.2.2 The CMS Calorimetry Detectors

The CMS calorimetry system consists of an electromagnetic calorimeter (ECAL), that measures the energies of electrons and photons, and a hadronic calorimeter (HCAL), that measures the energies of hadrons. The CMS HCAL will be described below, while the ECAL will be described in detail in Section 3.

2.2.2.1 The Hadronic Calorimeter

The strong interaction forces quarks and gluons to hadronize, a process in which hadrons (particles consisting of quarks) are produced copiously from an initial colored particle. Some of these hadrons are unstable and can further decay into stable particles (such as leptons and other stable hadrons). At the end of the hadronization process, a quark/gluon is seen by CMS as a cluster of mostly electrons (with photons from bremsstrahlung), muons, pions, kaons, protons and neutrons, which are combined to form an object called a jet. Therefore, for precise jet measurements, the CMS experiments must precisely measure the energy of hadrons within a jet shower.

The HCAL is limited by the ECAL in its inner radius, and by the CMS solenoid in its outer radius. In order to provide hermeticity and full coverage of the hadronic shower development, it must maximize the amount of absorber material in its region. This absorber material is a dense material with high atomic number that causes the hadrons within a jet to cascade decay into less energetic particles. The energy of these cascade decays are then measured by an active medium.

The CMS HCAL is divided into four sections, the barrel (HB), the outer barrel region just outside the CMS solenoid (HO), the endcaps (HE) and the forward region closer to the beam pipe (HF), which can be seen in Figure 2.4. The HB and HE consist of brass absorbers, as it has a reasonably short interaction length, is easy to machine and is non-magnetic, while the active medium consists of plastic scintillator tiles read out with embedded wavelength-shifting (WLS) fibres. The HO functions as a tail catcher and it consists of only a layer of active medium of plastic scintillator tiles. The HF uses steel as absorber and quartz fibers as active medium, a choice that is better suited for the more congested environment closer to the beam pipe. The photodetection readout is based on multi-channel hybrid photodiodes (HPDs) for HB, HE and HO, and on photomultiplier tubes (PMT) for HF.

The HB geometry consists of 32 towers covering the pseudorapidity region $-1.4 < \eta < 1.4$, resulting in 2304 towers with a segmentation $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. Each HCAL endcap consists of 14 η towers with 5^o ϕ segmentation, covering the pseudorapidity region $1.3 < |\eta| < 3.0$. For the 5 outermost towers (at smaller η) the ϕ segmentation is 5^o and the η segmentation is 0.087. For the 8 innermost towers the ϕ segmentation is 10^o, while the η segmentation varies from 0.09 to 0.35 at the highest η. The total number of HE towers is 2304. The HO detector contains scintillators with a thickness of 10 mm, which line the outside of the outer vacuum tank of the solenoid and cover the region $-1.26 < \eta < 1.26$. The tiles are grouped in 30^o-sectors, matching the ϕ segmentation of the muon chambers to which

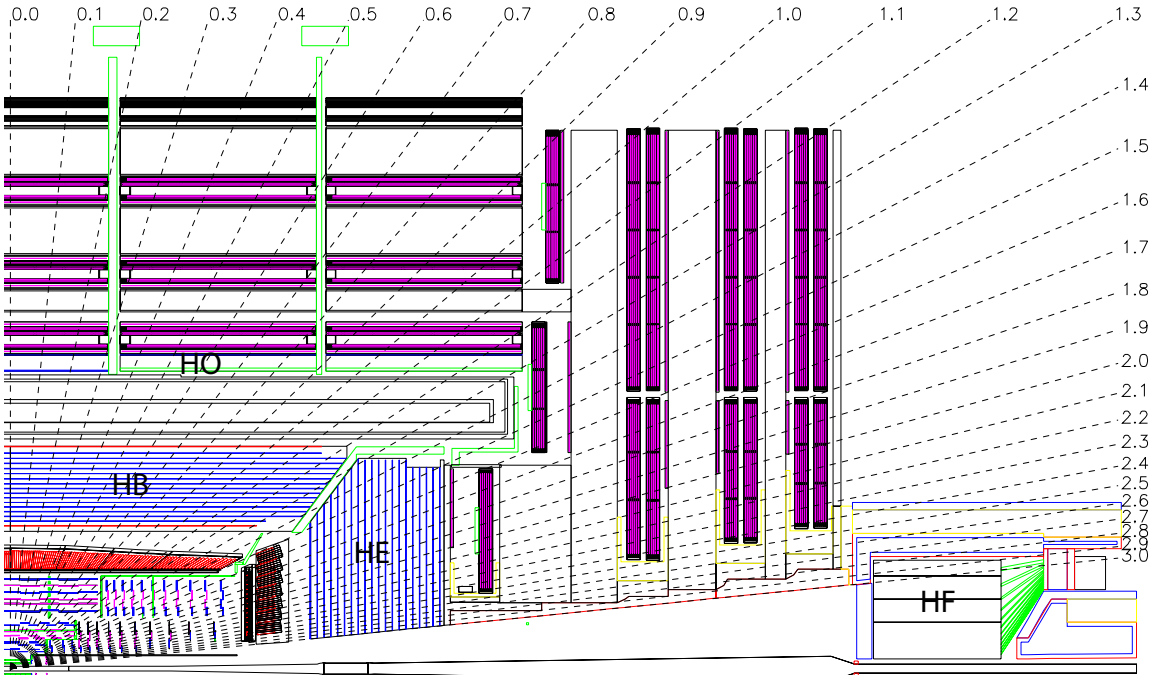

Figure 2.4: Sketch of the CMS hadronic calorimeter.

it is attached. Coverage between pseudorapidities of $3.0 < |\eta| < 5.0$ is provided by the HF calorimeter. There are 13 HF towers in η, all with a size given by $\Delta\eta \approx 0.175$, except for the lowest-η tower with $\Delta\eta \approx 0.1$ and the highest-η tower with $\Delta\eta \approx 0.3$. The ϕ segmentation of all towers is 10^o, except for the highest-η one which has $\Delta\phi = 20^o$. This leads to 900 towers and 1800 channels in the 2 HF modules.

2.2.3 The CMS Muon Detectors

Because of their mass, muons bend less in a magnetic field than electrons with the same momentum. Therefore, in order to precisely determine the momentum of muons, an extra tracking system must be added to CMS outside the solenoid. The CMS muon system is divided into two sections, with two different detector technologies in each: the drift tubes (DT) and resistive place chambers (RPC) located in the barrel, and the cathode strip chambers (CSC) and RPCs in the endcaps. These three technologies are based on gaseous detectors, differing on the geometry of the cell in which the gas is trapped and on the collection of charge from the gas ionization process due to the incoming muon. While the DTs and the CSCs are particularly suited for a precise determination of the muon's momentum, the RPCs provide a fast signal with good time resolution, but coarser spatial precision. The RPCs are used for bunch crossing determination in the muon system. Figure 2.5 shows a sketch of the muon system, highlighting its three different components.

In the barrel, the muon system is organized in chambers divided into 4 layers at radii of approximately 4.0, 4.9, 5.9 and 7.0 m from the beam axis. The layers are organized in 5 wheels along the beam axis, covering in total the region $|\eta| < 1.2$, and each wheel is divided

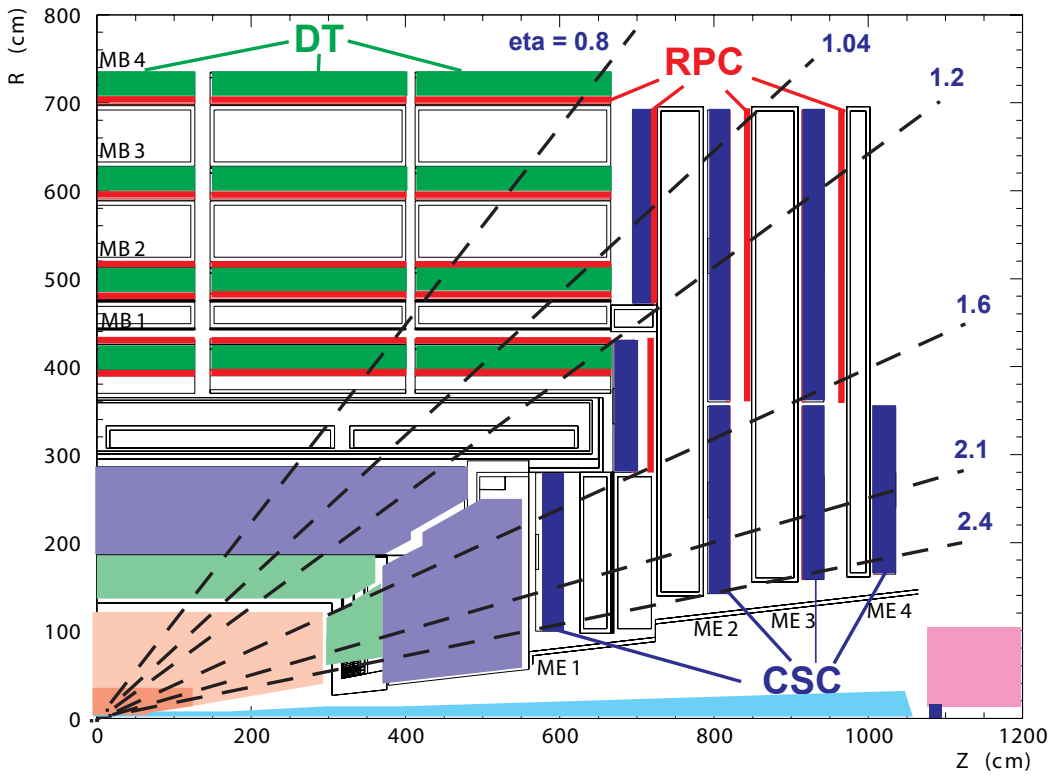

Figure 2.5: Sketch of the CMS muon detectors.

into 12 stations, with each covering a 30^o azimuthal angle. Chambers in different stations are organized so that a high-p_T muon produced near a sector boundary crosses at least 3 out of the 4 stations in each layer. In the two innermost layers, each DT chamber is coupled between two RPCs, while, in the two outermost layers, each chamber is coupled to one RPC.

In each of the endcaps, the CSCs and RPCs are arranged in 4 disks perpendicular to the beam, and in concentric rings, 3 rings in the innermost station, and 2 in the others, totaling in 486 CSC chambers for both endcaps. The chambers are trapezoidal and cover either 10^o or 20^o in ϕ. A muon in the pseudorapidity range $1.2 < |\eta| < 2.4$ crosses 3 or 4 CSCs. In the endcap-barrel overlap range, $0.9 < |\eta| < 1.2$, muons are detected by both the barrel DTs and endcap CSCs. The spatial resolution provided by each chamber from the CSCs is typically about 200 μm (100 μm for the innermost layer). The angular resolution in ϕ is of the order of 10 mrad.

2.2.4 Luminosity Detectors

The measurement of the instantaneous luminosity is a key element to perform physics analysis at colliders in order to estimate how many events a certain process will yield given its cross section. The CMS experiment uses five different detectors to monitor and measure the instantaneous luminosity, based on different occupancy measurements for each individual detector. The used subdetectors are: the barrel DTs, the HF, the Pixel Luminosity Telescope

(PLT) and the Fast Beam Condtions Monitor (BCM1f). The last two detectors use silicon pixel sensors and single-crystal diamonds, respectively, located close to the LHC beam pipe.

These luminosity measurements are calibrated with dedicated LHC runs called Van der Meer (VdM) scans. By scanning the two beams through one another in the transverse plane of the detector, VdM scans allow to measure the luminosity per colliding bunch pair from machine parameters.

2.2.5 The CMS Trigger and Data Acquisition Systems

The LHC provides bunch crossings with intervals of 25ns, which means that CMS produces collisions data at a rate of 40 MHz. The average event size produced by the CMS data acquisition system is of about 1 MB, which means that, if the experiment were to read and save every collision, it would need a data output bandwidth of approximately 40 TB/s. Such amount of data cannot be processed and saved by the current capabilities of the computing farms responsible for these tasks. However, most of these events are not of interest for the experiment: they consist of low energy QCD processes which would not be selected, in general, by any physics analysis. For this reason, CMS uses a trigger system that filters the collision data online. In general terms, the CMS trigger system must ensure a large acceptance for physics signals, while keeping the output rate and CPU time under control.

The first layer of the CMS trigger system is called the Level-1 (L1) trigger, and reduces the event rate from 40 MHz to 100 kHz. It works at hardware level, based on FPGA and custom ASIC integrated circuits technology, using information from the calorimeters and muon detectors to select, in about 4 μs, the most interesting events.

The L1 is divided into three parts: the calorimeter trigger, the muon trigger and the global trigger. The calorimeter trigger is itself divided into two steps: Layer-1 and Layer-2. In the first, it receives summarized data from the ECAL and HCAL in small packets called trigger primitives and computes basic information such as the sum of ECAL and HCAL energies and energy calibration as well as the computation of the H/E ratio (used to discriminate electrons and photon from hadronic jets). The information calculated by Layer-1 is parallelized: each Layer-1 processor calculates local information pertaining to a section of the detector. The following step, the Layer-2 calorimeter trigger, receives the information from the Layer-1 and calculates global variables (such as total energy sum and calorimeter based missing transverse energy) and reconstructs simple versions of jets, taus and electrons/photons.

The muon trigger receives information from the three muon subdetectors to ensure good coverage and redundancy. For the DTs and CSCs, the front-end trigger electronics identifies track segments from the hit information registered in multiple detector planes of a single measurement station, which are transmitted via optical fibers to regional track finder processors. These processors apply pattern recognition algorithms to identify muon candidates and measure their momenta from their bending under the CMS magnet. The combined muon trigger provides to the global trigger the transverse momentum of the muon candidates, along with their positions and quality scores, so they can be accepted/discarded by the global trigger.

The global trigger, the final L1 step, is responsible for selecting interesting events based on the global information received from the Layer-2 calorimeter trigger and the last step of the muon trigger. These decisions are taken based on L1 selection algorithms stored in the electronics. This decision is sent to the central CMS data acquisition system, which transmits

the full event data to processors that will format the data for the next trigger step, the High Level Trigger (HLT).

The HLT is the final online selection step at CMS, and reduces the event rate from 100 kHz (L1 output) to a sustainable level of 1 kHz. It is implemented in software running on a farm of commercial computers which includes about 16,000 CPU cores. The HLT runs a simplified version of the full CMS object reconstruction, detailed below, based on the CMS software architecture CMSSW, optimized for the CPU timing requirements that exist at trigger level.

2.2.6 Object Reconstruction at CMS

The general object reconstruction at CMS can be summarized in Figure 2.6. It takes advantage of the CMS geometry, the cylindrical onion, and gathers information from different subdetectors to reconstruct a single type of particle: charged hadrons will leave hits in the tracker and will be stopped by the HCAL; neutral hadrons will not leave hits in the tracker but will deposit their energy in the HCAL; electrons and photons will deposit their energies in the ECAL, while the former will also leave hits in the tracker; muons will leave hits in the tracker and the muon system, with no energy deposits in the calorimeters. This global event description, which links information from the different subdetectors, is usually referenced as the particle flow algorithm (PF) [30]. The particles created with this event description are called PF candidates and are mutually exclusive. This means, for example, that a single track cannot be used to reconstruct two different PF candidates. It's important to notice that using particle flow algorithms at CMS is only possible due to the high granularity and position resolution of all its parts, particularly its calorimeters.

The basic ingredients of the particle flow algorithm are the energy clusters on the calorimeters and the reconstructed trajectories obtained from the trackers and muon systems, which will be discussed below. Then, some of the objects reconstructed by CMS will be described, particularly the ones that are important in the analyses to be described in the following sections.

2.2.6.1 Particle Flow Ingredients

Tracks are the reconstructed trajectories of charged particles that leave hits in the CMS tracker [31]. Other than providing position information, these tracks are used to measure the particle's momentum via its bending in the CMS magnetic field. Track reconstruction starts with clusters of hits close to the beam pipe, in the inner pixel detector. Possible trajectories are then extrapolated and, if hits in the following tracker layers are found, they are added to the track and the track parameters are updated. After the algorithm hits the final tracker layers, and a first version of the full trajectory is reconstructed, another track reconstruction is performed, this time outside-in. The final track parameters are set to the weighted average of inside-out and outside-in results. The tracks have to fulfill quality criteria or are otherwise discarded. Finally, the used tracker hits are removed from the hit collection, and the procedure is repeated.

Primary vertices, which represent the point where a hard scatter event occurred during a bunch crossing, are reconstructed based on track clustering at the interaction region [31]. This region is at the center of CMS and has a typical length of $\approx$ 5-6 cm in the z direction,

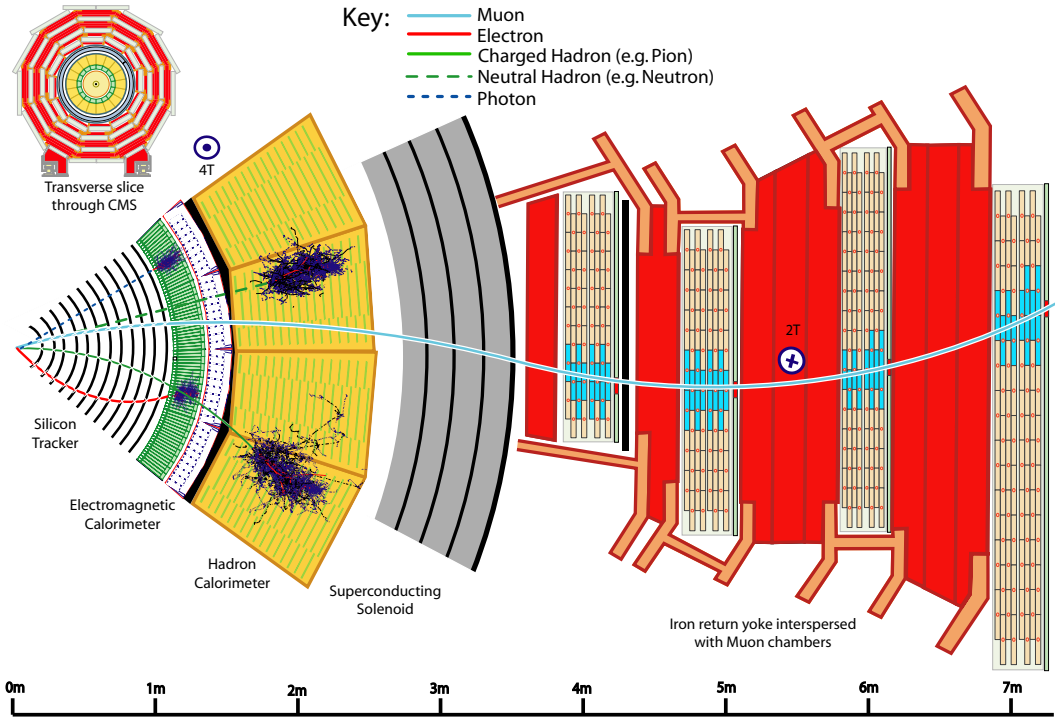

Figure 2.6: Sketch of different types of object reconstruction at CMS.

while its transverse dimension is $\approx$ 20-30 μm. Many primary vertices can be reconstructed for one bunch crossing. During the 2016 data taking period, events with over 50 primary vertices have been recorded by CMS. It is common to refer to the number of reconstructed primary vertices as "pile-up".

The fitting procedure used to extrapolate the tracks in the usual reconstruction does not perform well for electron tracks [31]. This happens because of the high amount of bremsstrahlung radiation electrons emit in the tracker, changing considerably the expected shape of their trajectory, which is not modeled by the standard algorithm. The electron-like tracks, selected based on their original fit quality and number of hits, are then refitted with a more appropriate method, the Gaussian-Sum Filter (GSF), which allows for substantial non-Gaussian energy loss along the trajectory.

Calorimeter clusters are reconstructed to detect and measure the energy of photons, electrons and hadrons, being the only source of measurement for photons and neutral hadrons [30]. The clustering algorithm, that gathers calorimeter deposits in groups, is applied separately for the different calorimeter partitions: ECAL barrel, ECAL endcaps, HCAL barrel, HCAL endcaps, and each of the two preshower layers.

The clustering procedure starts by searching for cluster seeds, which are local calorimeter-cell energy maxima, with respect to either its closest 8 or 4 cells, above a certain threshold. Topological clusters are grown from cluster seeds by adding together nearby seeds that share calorimeter cells. These topological clusters are used to estimate the energy sharing between clusters reconstructed from two different but nearby cluster seeds. Each individual cluster

position is calculated as $X = \sum_i (w_i X_i)/(\sum_i w_i)$, with $w_i = ln(E_i/E_{th})$, where E_{th} is an energy threshold and the i is summed over the cells within a cluster.

2.2.6.2 Photons and Electrons

Electrons are reconstructed based on seed electron tracks reconstructed with the GSF algorithm. After a good quality electron track is found, it must be linked to an energy cluster on the ECAL. Photons, on the other hand, rely solely on the ECAL clusters to be reconstructed, since they leave no signal on the trackers. The reconstruction algorithm for ECAL clusters, called supercluster algorithm, allows almost complete recovery of the energy of photons that convert due to the material in front of the ECAL [32]. The corrections and calibrations involved to ensure a precise energy reconstruction for these particles will be described in the chapter dedicated to the CMS ECAL.

Photons and electrons are identified based on a set of shower shape and isolation variables. The former measures the shape of the energy deposit on the ECAL, while the latter gives information about the event content surrounding the photon candidate. Isolation variables are particularly important when dealing with π^0 mesons, produced in hadronic showers, that decay to two almost collinear photons. These $\pi^0 \to \gamma\gamma$ objects from jets must be filtered out of the photons collection in a physics analysis looking for photons coming from the hard scatter interaction.

Shower shape variables help to distinguish between real photons and jets, but also give important information about non-collision backgrounds reconstructed as photons. One of these events are the ones in which hadrons interact with the ECAL photodetectors without showering in the crystals. This creates a signal with large amplitude that can be mis-identified as a signal from an energy deposit. These anomalous signals from direct interaction with the ECAL photodetectors are called spikes.

For the ECAL, electron and photon clusters look approximately the same, therefore, almost indistinguishable when using only calorimeter information. In order to separate electrons from the photon collection, one must veto on possible tracks appearing within the photon cone. This is done in two different ways in CMS: via a conversion-safe electron veto, which vetos on electron tracks that are not compatible with photon conversion tracks, and via the pixel seed veto, which vetos on hits in the innermost layers of the pixel detector assuming a photon cone coming from a certain primary vertex.

2.2.6.3 Jets

As mentioned above, colored particles cannot be observed in a color non-neutral state. Therefore, we can only access the information (such as kinematics) about a quark/gluon in the final state experimentally through its hadronization product, the jets. The process of obtaining a jet from a quark or a gluon, however, cannot be modeled by perturbative QCD calculations, making it susceptible to many theoretical uncertainties and causing a "jet" to be ill-defined theoretically. There are, however, methods of jet reconstruction that minimize those uncertainties.

The anti-k_t and the Cambridge-Aachen (CA) methods are currently used by CMS and provide a jet reconstruction that are both infrared safe (avoiding low energy non-perturbative sin-

gularities) and collinear safe (avoiding singularities due to radiation colinear to the hadronizing particle) [33]. Both algorithms work by combining PF candidates sequentially in a jet if a certain distance discriminant is smaller than a certain parameter. The analyses documented in this document use the anti-k_t algorithm with a distance parameter $R = 0.4$, as most analyses using the CMS Run 2 dataset.

The energy of the reconstructed jets must be corrected in order to match jets produced with particles generated from a full Monte Carlo simulation of the hadronization process. This mismatch is attributed to the non-uniform and non-linear response of the CMS calorimeters, electronics noise and pile-up. To correct for this, jet energy correction (JEC) factors are derived centrally by the CMS collaboration. These JEC factors correct the (raw) 4-momenta parameters of the reconstructed jet as a multiplicative factor.

An important contributor to performance degradation in jet energy measurement is the pile-up. Extra charged hadrons from softer primary vertices can be erroneously clustered within the jet and bias the energy reconstruction. Therefore, a pile-up subtraction method, called Charged Hadron Subtraction (CHS), is used in CMS. This is a particle-by-particle method that makes use of PF jets and exploits the excellent CMS tracking capabilities to identify and remove charged hadrons inside jets that are known to originate from pile-up vertices.

2.2.6.4 Identification of b-quark jets

Many interesting phenomena that are being searched for at the LHC involve bottom quarks. One important example is the $H \to b\bar{b}$ process, which has the highest branching fraction among the SM Higgs decays. Another example is the top quark decay, which almost always occurs in the channel with a bottom quark plus a W boson. These searches would be facing an enormous amount of background events if jets from bottom quarks were indistinguishable from lighter quark jets, copiously produced by the LHC. However, bottom quark jets (from now on simply referred to as b-jets) are strikingly different from their lighter counterparts due to the presence of a B meson (product of the bottom quark hadronization).

B mesons can only decay via the weak interaction through Cabbibo suppressed processes $b \to Wc$ and $b \to Wu$. This causes the average lifetime of a B meson to be of about 1.5 picoseconds. Assuming they are produced at the speed of light, these mesons can travel around 0.5 milimeters before decaying. This decay length is well within the reach of the current CMS tracking resolution, therefore, the vertex of the B meson decay (secondary vertex) can be measured. With this information, algorithms (b-taggers) can be constructed to distinguish b-jets from light jets.

The b-tagger algorithm utilized in the analyses presented in this thesis is the Combined Secondary Vertex (CSV) algorithm [34]. This method uses multiple reconstruction quantities related to the tracks associated to the jet constituents and the properties of secondary vertices reconstructed inside the jet, if present. The CSV algorithm is still applicable, however, when no secondary vertices are associated to the jet. The variables used by this method include the number of tracks in the jet, the significance of the tracks impact parameters, and, if available, the impact parameter, mass and number of tracks associated to the secondary vertex. These variables are used to determine a single likelihood discriminator associated with the jet, i.e. the CSV discriminator, proportional to the compatibility of the jet with a b-quark decay.

Chapter 3

The CMS Electromagnetic Calorimeter

The CMS electromagnetic calorimeter (ECAL) is a high-resolution, hermetic, and homogeneous electromagnetic calorimeter made of 75,848 scintillating lead tungstate crystals divided among a barrel region ($|\eta| < 1.48$) and two endcap regions ($1.48 < |\eta| < 3.0$) [29], as shown in Figure 3.1. The light emitted by these crystals is detected with avalanche photodiodes (APDs) in the barrel and vaccuum phototriodes (VPTs) in the endcaps. The signal readout is performed with two avalanche photodiodes (APDs) per crystal in the barrel, and one vacuum phototriode (VPT) in the endcaps. These characteristics, for providing precise energy and timing resolutions, are invaluable for the CMS physics program.

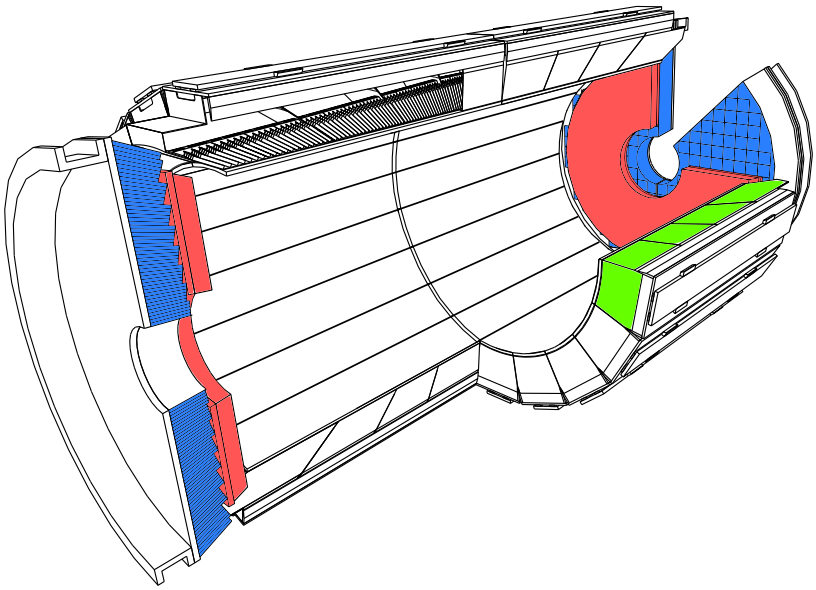

Figure 3.1: A schematic view of the CMS ECAL highlighting a barrel supermodule (green), the ECAL endcaps (blue) and the ECAL preshower (red).

Completing the CMS ECAL system is a preshower detector (ES), based on lead absorbers equipped with silicon strip sensors. It is installed in front of the ECAL endcaps, covering the region $1.65 < |\eta| < 2.6$. The fine granularity of the ES strips (2 mm wide) can resolve the signals of high-energy photons from the decays of neutral pions into two photons, when the separation angle between the photons is small, and can determine precisely the position of the electromagnetic deposits.

During the first year of Run 2 data taking at 13 TeV, the LHC provided a challenging environment, with one bunch crossing every 25 ns and an average of 10 interactions per crossing (pile up). The situation became even more challenging in 2016, with up to 40 pile up interactions. In the 2015 data taking period, the CMS ECAL operated with more than 98% of its channels active, and was responsible for less than 7% of CMS downtime during physics runs.

3.1 PbWO$_4$ Crystals and In-detector Electronics

The main component of the ECAL are the lead tungstate crystals, that serve as both active material and absorber, constituting a homogeneous calorimeter. The barrel and endcap crystals are shaped as trapezoidal prisms, with 23 and 22 cm of length, respectively, consisting of about $26X_0$. The crystal's front face (facing the interaction point) has an area of approximately 22×22 mm^2 (about $\Delta\eta \times \Delta\phi = 0.0175 \times 0.0175$) for the barrel and 25×25 mm^2 for the endcaps. Examples of such crystals are seen in Figure 3.2.

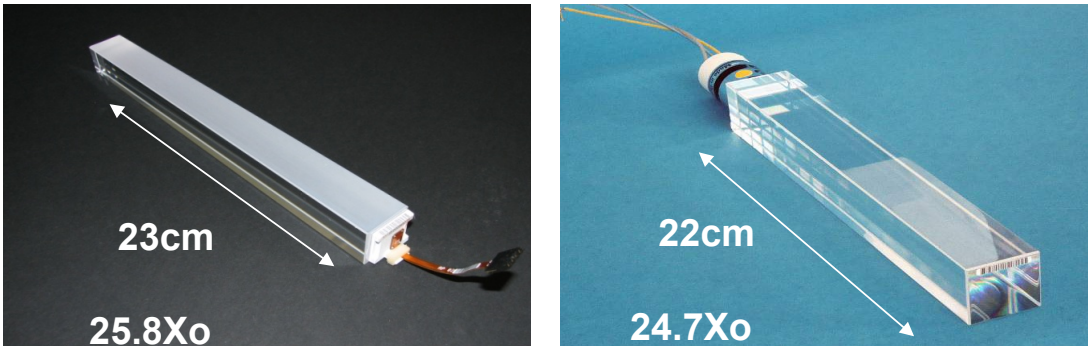

Figure 3.2: Examples of ECAL lead tungstate crystals for the barrel (left) and endcap (right).

The ECAL crystals are arranged to provide an off-pointing, pseudo-projective geometry (the crystal axes are tilted at 3^o with respect to the line from the nominal vertex position), shown in the left of Figure 3.3. In the barrel, the crystals are mounted in submodules (2×5 crystals in $\phi \times \eta$), which are then mounted into modules and grouped into supermodules. Each supermodule contains 4 modules, three with 10×4 submodules in $\phi \times \eta$ and one with 10×5. An ECAL supermodule is shown on the right of Figure 3.3.

An energy deposit is measured in the crystal through a scintillation process, in which the lattice atoms are excited and de-excite by the emission of photons. The emission scintillation peak for the ECAL crystals is at about 425 nm, with 80% of the light emission happening within 25 ns. The PbWO$_4$ crystals are also characterized by a small radiation length ($X_0 =$

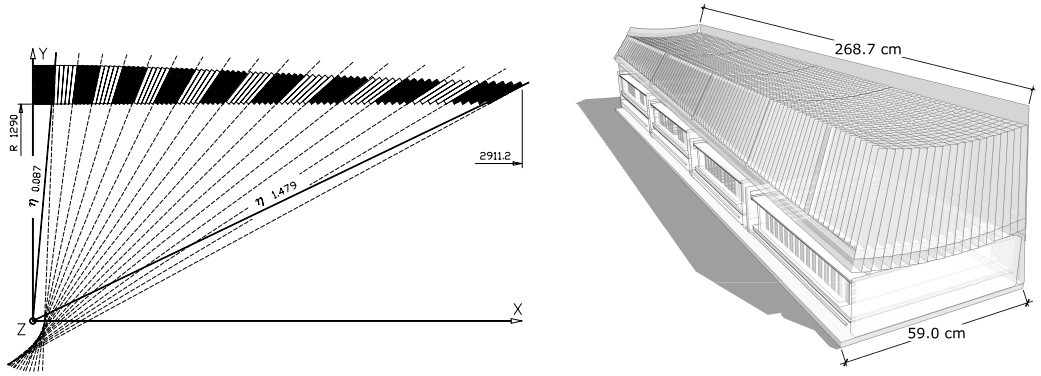

Figure 3.3: On the left, a schematic view of the off-pointing, pseudo-projective ECAL crystals geometry. On the right, the details of a barrel supermodule structure.

0.89 cm), the mean length of the path that takes a high energy electron to lose $(1 - 1/e)$ of its energy in that medium, and a small Molière radius ($R_M = 2.19$ cm), the radius of a cylinder that contains 90% of the shower energy.

On the other hand, the ECAL crystals have a low light output yield (about 30 photoelectrons per MeV), and therefore must be coupled to high gain light detectors. These detectors must also be radiation hard, to cope with the LHC environment, and be stable under a 4T magnetic field (the CMS solenoid magnet). For the ECAL barrel, avalanche photodiodes (APDs) were chosen due to their high gain (50). However, their temperature dependence requires ECAL to have a temperature control with a precision better than 0.1^oC (the $PbWO_4$ crystals light yield is also dependent on temperature, with lower light yields for lower temperatures).

The ECAL APDs are also characterized by their fast, 2 ns rise time, which allows time measurement with the crystal scintillation pulse shape. By the time of the ECAL construction, APDs were generally produced in small sizes in comparison to the crystal's rear side area, where they must be attached. For this, two APDs were used for each crystal. In the endcaps, the radiation dose is much higher than in the barrel, which does not allow the usage of APDs in the high η region. For this reason, vacuum VPTs are used instead. While more resistant to radiation, these light detectors generally have a smaller gain (8-10).

3.2 Trigger and Data Acquisition Systems

The ECAL trigger and data acquisition system (Trigger and DAQ) takes care of the the entire data flow from the crystals to the central CMS Trigger and DAQ systems. A schematic version of the system is shown in Figure 3.4.

3.2.1 Data Flow

The ECAL data flow starts right after the signal is detected by the photo detectors attached to the crystals. The first step in the chain is the amplification and digitization of the analog signal coming from the APDs and VPTs. This is performed with multi-gain pre-amplifiers

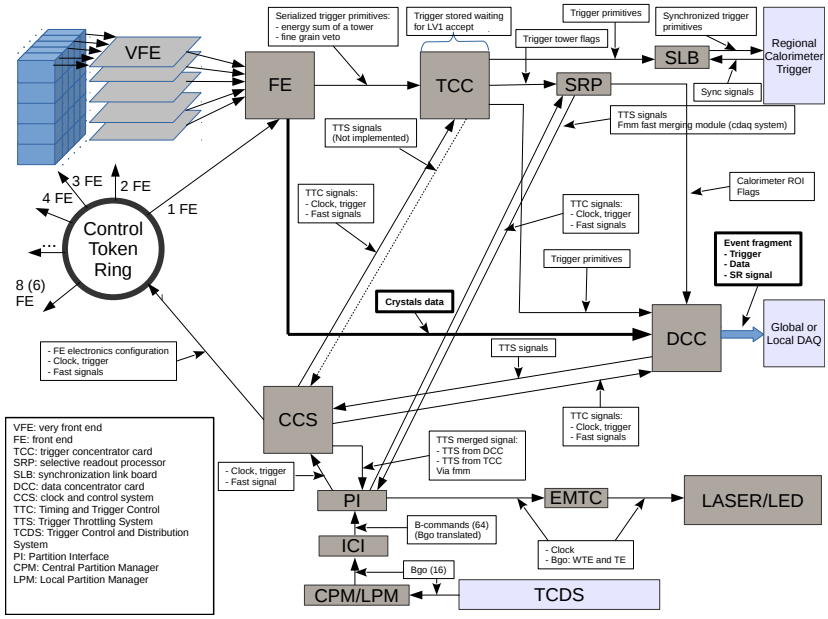

Figure 3.4: ECAL DAQ simplified scheme.

(MGPAs). The ECAL MGPAs work with three amplification gains: gain-12, gain-6 and gain-1. The shaping time associated with this amplification process is of about 40 ns. The signal is then sampled and digitized at a rate of 40 MHz. One pulse shape is then reconstructed using 10 of these samples. The electronics responsible for the amplification and digitization step is generally called the very-front-end electronics (VFE).

The VFE sends information to the front-end electronics (FE) at the sampling rate. The FE is responsible for creating trigger primitives, basic information such as energy and position, from arrays of 5×5 crystals (trigger towers). The trigger primitives are sent from the FE to the trigger concentrator cards (TCCs) at sampling rate, and passed to the first layer of the CMS Level-1 trigger. The VFE and the FE, known as the on-detector electronics, are placed right in the detector, close to the crystals, which is schematically shown in Figure 3.5. In Figure 3.5, the chips dedicated to generating and transmitting the serialized trigger primitives (FENIX chips) are shown, along with the gigabit optical hybrids (GOH) chips dedicated to data transmission and the chips dedicated to receiving and distributing timing and control data (CCU).

During the time it takes for a decision to be made by the L1 trigger (L1A signal) and to arrive back at the ECAL electronics, the trigger primitives are stored in the TCCs. In case the event is not accepted, these trigger primitives and the full crystal information stored at the FE are flushed from memory. If an event is accepted, the FE sends the serialized full individual crystals information to the data concentrator cards (DCCs), that will pack the event in a format readable by the CMS central DAQ.

In parallel, the TCC sends the trigger primitive information to the selective readout units (SRPs) after a L1A arrives. The SRP selects which portions of the detector should be unpacked by the DCCs based on an algorithm called Selective Readout (SR), reducing the load

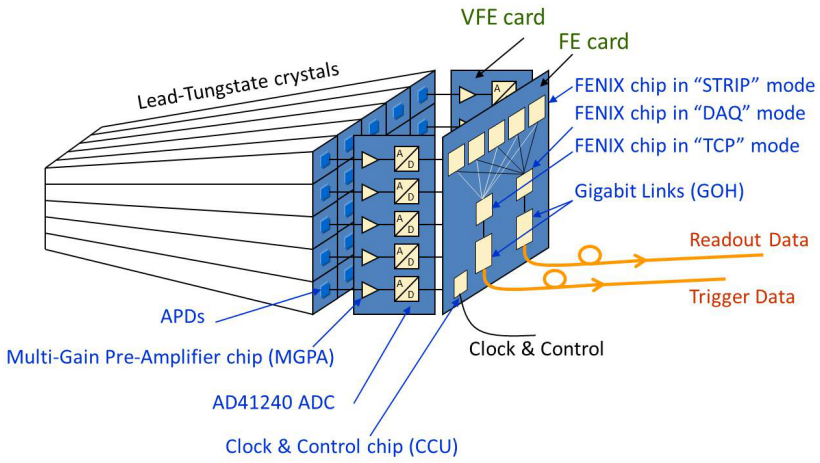

Figure 3.5: ECAL in-detector electronics simplified scheme.

of processing by the DCCs. The DCCs load is limited by the current electronics architecture and can be exceeded when the average activity per event is increased, such as in high pile up LHC runs. The SR algorithm classifies trigger towers as low interest ($E_{TT} < 1.5$ GeV), medium interest ($1.5 < E_{TT} < 2.5$ GeV) or high interest ($E_{TT} > 2.5$ GeV). If a trigger tower is classified as low interest, only the crystals with energy higher than 4.5 (6.5) ADC in the barrel (endcap) are unpacked. If a trigger tower is classified as medium interest, all crystals are unpacked, regardless of their energies. If a trigger tower is classified as high interest, all crystals in that trigger tower, and the 8 trigger towers around it are fully read.

3.2.2 Timing and Control Distribution

In order to ensure a synchronized working of all different parts of the ECAL DAQ system, a precise distribution of clock and control signals must be achieved. This is performed via a combination of different ECAL and central CMS electronics. For Run 2, CMS installed a new subsystem dedicated to distributing these signals among all CMS partitions, the Timing and Control Distribution System (TCDS). TCDS communicates timing/synchronization signals coming from the LHC and CMS (clock), L1As coming from the CMS Level-1 trigger, and configuration and control commands synchronously to all CMS components. These signals are then distributed among ECAL components via the clock and control system (CCS), including to the front-end electronics (through control token rings).

3.2.2.1 TCDS

The TCDS subsystem was installed on CMS during the Long Shutdown 1 in order to cope with the new running conditions for Run 2. Specifically, new detector subsystems were being installed in the experiment (such as the new pixel inner tracker layers). Instead of expanding the old system, a choice was made to construct a new infrastructure with updated hardware, firmware and software. The TCDS system consists on three types of boards: the Central Partition Manager (CPM), the CMS Interface boards (iCI) and the Partition Interface boards

(PI).

The TCDS CPM is fully controlled by the central TCDS system, and has information regarding the type of runs to be started by the central DAQ. This information defines, for example, at which point of the LHC orbit to send specific signals to the subdetectors. The iCIs must work as a library that translates the CPM signals into specific actions to be taken by the different subdetectors. This means that the iCIs must be configured properly and individually by each subdetector. In order to achieve that on ECAL, a new software interface to TCDS was written to configure and monitor the ECAL iCIs. The PIs act as fan-out boards to distribute the iCI messages among different subdetector components.

3.2.3 Software Architecture

The ECAL DAQ software architecture controls, configures and monitors the readout and trigger electronics described in the previous sections. The framework implementation is based on a CMS common framework developed by the central Trigger and DAQ group, mostly consisting of C^{++} xDAQ framework [35], while the top layer of the system is based on the JAVA RCMS [36] framework.

The overall system accomplishes several different tasks. It configures and monitors the different components of the FE and and off-detector electronics. It communicates with the central CMS TCDS, both sending the correct configuration parameters for the dedicated TCDS ECAL boards and receiving and translating these commands to the off-detector electronics. It is also able to perform different types of runs dedicated to calibration and detector developments, such as "MiniDAQ" and "local" runs.

With the xDAQ framework, the ECAL software communicates to the ECAL boards through SOAP (Simple Object Access Protocol) messages. SOAP messages are also used to communicate issues with the off-detector electronics directly to the central DAQ system.

There are five main components to the ECAL DAQ software architecture:

ECAL Function Manager Function managers (FM) work as the overall controller of actions happening to the ECAL software. It has the power to send commands to the ECAL supervisor to cause state transitions, triggering actions such as the beginning of a run. During global CMS runs, the central DAQ function manager communicates such actions to the ECAL function manager. In specific runs in which ECAL controls its own sequence, the ECAL function manager works by itself, and can be controlled via a web interface.

ECAL Supervisor The ECAL supervisor is at the top of software control chain and coordinates all other working pieces. It receives the commands from the ECAL function manager and, to any given command, proceeds with the specific steps to ensure the state transitions of the different ECAL components.

Services Supervisors For each different service described in the previous sections (TCC, DCC, DCS and SRP), dedicated software exists in order to perform the actions related to specific state transitions.

TCDS Supervisor A specific supervisor was developed to communicate with the central TCDS service. It sends the proper configuration parameters to the TCDS ECAL boards,

with the list of commands to be passed to the ECAL electronics when certain state transitions happen and during specific run conditions.

Monitoring Services In order to retrieve and display information about the status of the off-detector electronics, specific services based on xDAQ capabilities are used. These services also monitor the ECAL DAQ computing capabilities to make sure all machines are working properly. Depending on what type of error is detected by the monitoring services, SMS's and e-mails are sent to ECAL DAQ experts.

3.2.3.1 ECAL DAQ Finite State Machine

One of the main goals of the ECAL Supervisor is to ensure the proper state transitions of the different DAQ components through the allowed states of the ECAL DAQ Finite State Machine (FSM), summarized in Figure 3.6.

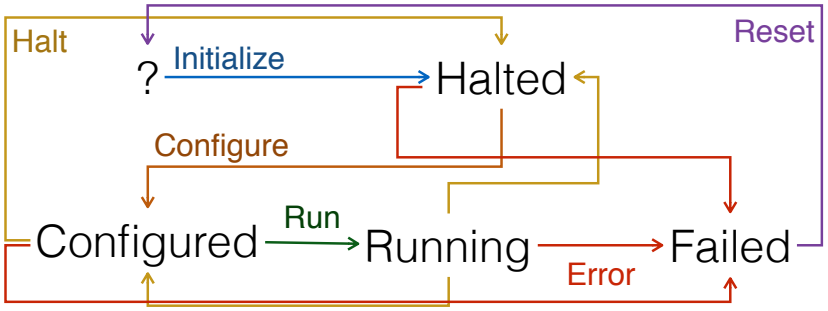

Figure 3.6: Simplified view of the ECAL Finite State Machine and its transitions.

Similarly to the central DAQ FSM, the transitions are:

Initialization The Initialization procedure tests the connection between the ECAL Supervisor and the different off-detector components and initializes all required drivers. The information about the hardware resources is then propagated back to the supervisor and the ECAL FM. The initialization transition takes the FSM to Halted state and makes it ready to be configured.

Configuration When the configuration signal is received, it reads the configuration parameters from the ECAL Configurations DataBase (ConfDB) and sends to the off-detector electronics, which are then loaded accordingly. Each different Service Supervisor independently ensures their components have been successfully configured and communicate it back to the ECAL Supervisor. This action takes the FSM to Configured state and is now ready to start producing data (running). It also enables the monitoring services.

Start The start command kicks-off the full data flow described at the beginning of this section, including trigger primitive production and delivery to the L1 trigger, and the delivery of the full data information upon a L1A. It also starts the delivery of monitoring data from the off-detector electronics to the monitoring services. This action takes the FSM to Running state.

Stop/Halt These actions take the system either from Running state to Configured state, or from Running or Configured states to Halted state, respectively. The Stop command interrupts the ECAL DAQ data flow, while the the Halt command interrupts the data flow and flushes the database parameters from the off-detector electronics.

Error Any error during the previous transitions will take the FSM to Failed state.

3.2.3.2 Types of Running Conditions with the ECAL

Global Runs A global run is the usual working condition of ECAL with the full CMS central DAQ infrastructure. In global runs, the ECAL FM receives commands directly from the central DAQ FM, including all state transition actions. These runs happen when the LHC is running, when CMS is taking cosmics data (with its magnetic field on or off) or when specific tests must be performed with the full central DAQ infrastructure.

MiniDAQ Runs These runs happen with a dedicated central DAQ infrastructure that is independent of global runs, and therefore can run in parallel. They are used in ECAL to test the state of the off-detector electronics and investigate possible errors seen in global runs. MiniDAQ runs have also been used in 2015 and 2016 to measure the response of the ECAL electronics to noise (pedestals) and to test pulses generated by the on-detector electronics.

Local Runs Local runs are completely independent from the central DAQ infrastructure. They rely on ECAL software to both directly configure the off-detector electronics and process the data that is produced by the DCCs. This freedom allows for more complex types of running conditions, such as the ones needed for more coherent measurements of ECAL pedestals (to be used in 2017).

3.3 Electron and Photon Energy Reconstruction

The photon and electron energy reconstruction based on ECAL energy deposits is based on the formula:

$$E_{e,\gamma} = \left[\sum_i \left(S_i(t) \times c_i \times A_i \right) \times G(\eta) + E_{ES} \right] \times F_{e,\gamma}, \tag{3.1}$$

with A_i, c_i and $S_i(t)$ as, respectively, the per individual channel amplitude, intercalibration constant and light monitoring constant, $G(\eta)$ is the ADC to GeV absolute scale, E_{ES} is the energy deposit in the preshower, and $F_{e,\gamma}$ are the cluster corrections (different for photons and electrons). The methods used to obtain the different terms in this equation will be detailed in the next sections.

3.3.1 Online Reconstruction

The time spacing between two consecutive samples from the ECAL readout electronics is 25 ns, which is the same time spacing between two colliding bunches in the LHC. This implies that, during the readout of one pulse, another scintillation process might start in the same crystal, compromising the in-time amplitude reconstruction. To mitigate this effect, also

called out-of-time (OOT) pile up, a new online pulse reconstruction method (multifit) was developed to replace the Run I method [37].

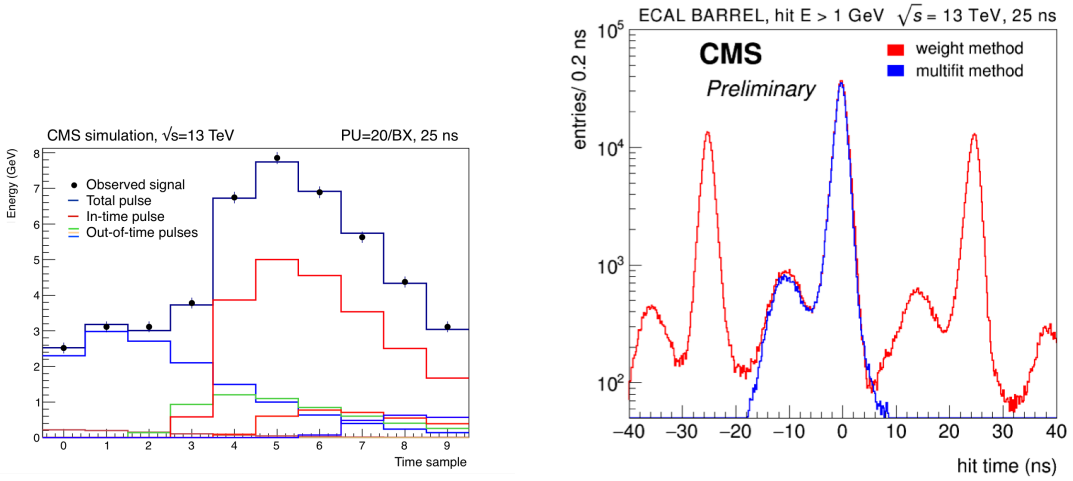

Figure 3.7: (Left) Fitted pulse shapes with the multifit method. (Right) In-time and out-of-time contributions to energy reconstruction with weights method (red) and multifit (blue).

In the multifit method, the pulse shape is reconstructed based on a fit to the time samples, minimizing $\chi^2 = \sum_{i=1}^{10} \left(\sum_j^M A_j p_{ij} - S_i\right)^2 / \sigma_{S_i}^2$. The samples ($S_i$) are fitted with one in-time pulse shape template, plus up to 9 out-of-time templates (p_{ij}) times their respective amplitudes (A_j). σ_{S_i} is noise generated by electronics associated with the crystal readout chain. The OOT templates have the same shape as the in-time one, but are shifted in time by multiples of 1 bunch crossing (1 BX = 25 ns), within a range of -5 to +4 BX around the in time signal (BX = 0). The pulse shapes have been measured in early 2015, in special runs in which the LHC delivered isolated bunches (no OOT pile up).

It has been observed in both data and simulation that, with the multifit method, OOT pile-up reconstruction is negligible. The energy resolution improvement, with respect to the Run I amplitude reconstruction method, is substantial especially for low E_T photons and electrons, given the larger contribution of deposits from pile-up to the total energy.

3.3.2 Response Monitoring

Time dependent corrections must be applied to the reconstructed amplitude due to changes in detector response with radiation exposure. These changes in response are due to decreases in crystal transparency and variations in VPT response in the endcaps.

The changes in the crystal transparency are due to ionizing radiation creating color centers in the lead tungstate. While the scintillation process remains intact, the amount of light detected by the photodetectors decreases. This effect is partially mitigated through thermal annealing, causing the transparency to increase in the absence of radiation.

A light monitoring system is used to monitor the overall changes in response in the ECAL [38]. It consists of a system of lasers (operating at 447 nm, close to the wavelength of peak

emission for lead tungstate) that injects light in each ECAL crystal, which is then read by the standard ECAL readout. The change in transparency per crystal (R/R_0) is then related to the ratio between reconstructed amplitude and the injected light amplitude (S/S_0) through the formula:

$$\frac{S}{S_0} = \left(\frac{R}{R_0}\right)^{\alpha}, \tag{3.2}$$

where α has been measured in beam tests and is ≈ 1.5. S/S_0 is then used as a correction factor to account for the response changes.

The light monitoring infrastructure is an integral part of the ECAL DAQ. It works alongside the main data flow, during specific periods in which the LHC bunches are empty (orbit gaps). During the orbit gaps, the laser system is shone in different parts of the ECAL barrel and endcaps. In about 40 minutes, one reading of the full ECAL is performed.

The history of response change measurements until the end of 2015 is summarized in Figure 3.8. The changes are up to 6% in the barrel and reach up to 30% at $|\eta| \approx 2.5$, the limit of the tracker acceptance. For high $|\eta|$ regions, changes are up to 70%. The recovery of the crystal response during the long shutdown period is visible. The response was not fully recovered, however, particularly in the region closest to the beam pipe. The monitoring corrections are validated by comparing isolated electron energy as measured by ECAL (E) and momentum as measured by the CMS Tracker (p), before and after light monitoring corrections. It is seen that the measured corrections bring stability to energy measurements with ECAL.

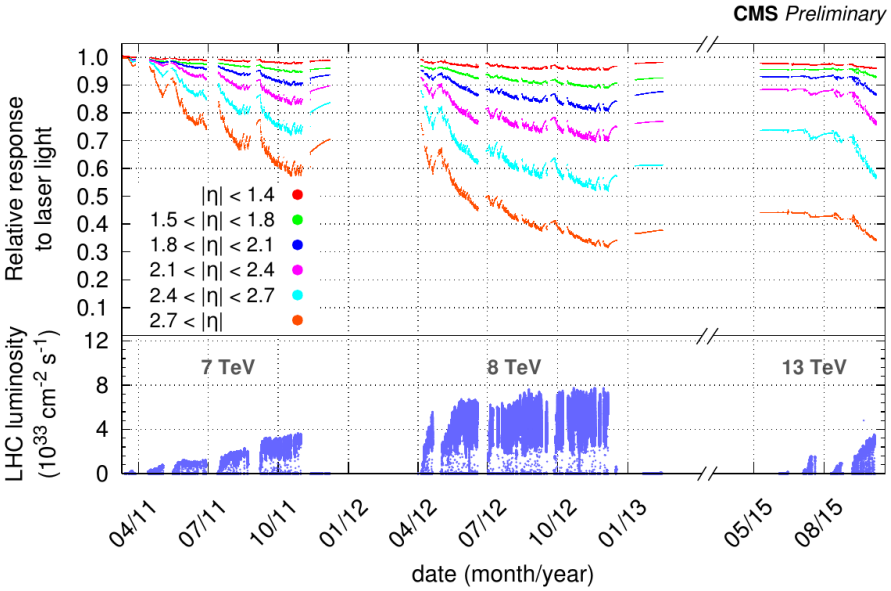

Figure 3.8: History of channel response changes as measured by the light monitoring system.

In Figure 3.9, it can be seen how the laser corrections impact the stability of the ECAL energy measurement. In this plot, the ratio between the energy of an isolated electron as measured by the ECAL (E) and by the CMS Tracker (p) is shown as a function of time. Due

to the response changes in the crystals and photon detectors, the ratio decreases when left uncorrected. After the laser correction is applied, ECAL achieves a stable energy measurement.

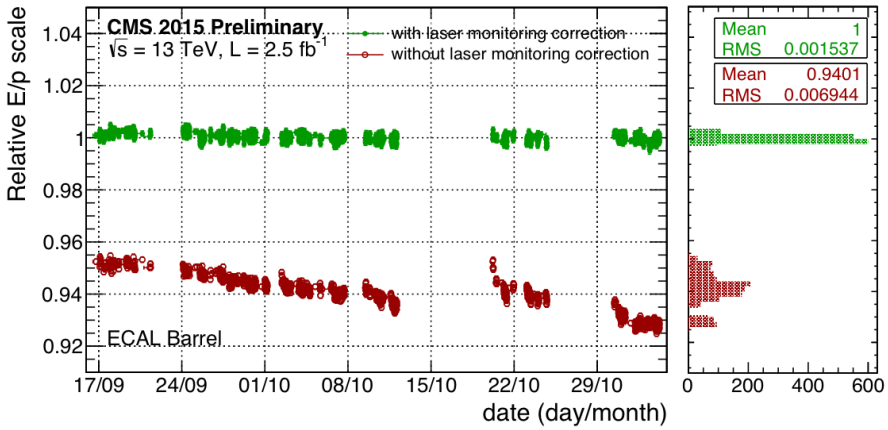

Figure 3.9: Stability of electron energy measurement with and without applying laser monitoring corrections.

3.3.3 Intercalibration

A relative calibration procedure in all ECAL channels is performed to ensure uniformity across the detector. Different and independent methods are used to calculate intercalibration constants (ICs), which are then combined to achieve the desired precision of $< 0.5\%$. The final 2015 version of the ICs have been calculated with the full 2.6 fb^{-1} dataset recorded by CMS with B=3.8 T. The following methods are the same as in Run I [39].

ϕ-Symmetry The ϕ-symmetry method is based on the expected uniformity of the energy flux along ϕ rings (region with fixed η). The ICs are calculated to correct non-uniformities in this flux. This method was used in 2015 to translate the latest ICs, calculated with the full 2012 dataset, to the 2015 detector conditions. This was done by scaling the 2012 ICs by the ratio between 2015 and 2012 ϕ-symmetry ICs.

π^0/η^0 The π^0/η^0 method consists of measuring the invariant mass of these resonances' decays to two photons and maximizing their resolutions by varying the ICs iteratively. This method does not utilize the nominal value of the π^0 and η^0 masses so as not to interfere with the absolute scale calibration.

E/p The E/p method employs the same logic as the light monitoring validation method, comparing isolated electron energy and momentum. An iterative method is used to minimize the spread of the E/p distribution.

The combined intercalibration was obtained from the mean of the individual ICs at a fixed value of η, weighted by their respective precisions. The residual miscalibration of an

intercalibration method, which is related to the final method precision, is calculated as the spread of the difference between the method's ICs and the other methods' ICs at a fixed value of η. This residual miscalibration can be seen in Figure 3.10 (left) for the ECAL barrel, where it is shown that the combination of ICs achieves the desired goal of less than 0.5% precision for the central barrel region. The overall impact of applying the intercalibration constants and the response monitoring corrections can be seen in Figure 3.10 (right), in $Z \to ee$ events.

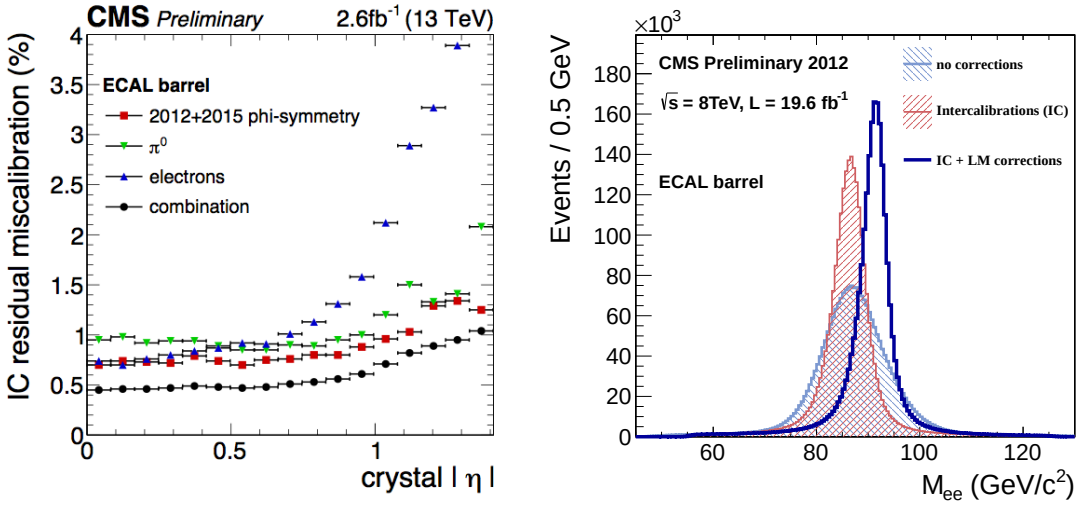

Figure 3.10: Residual mis-intercalibration with the different ECAL intercalibration methods and their combination.

3.3.4 Absolute Calibration

$Z \to ee$ events are used both to set the η scale and the absolute calibration [39]. The first is developed to ensure that different η regions have the same relative response, while the second (done separately for barrel and endcaps) sets the absolute energy scale.

A dedicated calibration was performed with 0 T data taken in 2015 to account for differences in shower shapes in the absence of magnetic field. For example, in 0 T there is no bremsstrahlung radiation outside the main electron cluster deposit, improving the reconstructed energy resolution.

In addition, the calibration was validated with high energy photons and electrons. The validation was performed by comparing data and Monte Carlo simulations for high energy electrons from $Z \to ee$. The calibration was found to be stable to 0.5% (0.7%) for electrons up to $p_T = 150$ GeV in the barrel (endcap). Possible saturation effects were corrected for with a multivariate technique, but those effects were found to be $< 2\%$ for photons arriving from resonance masses less than 1.4 TeV.

3.3.5 High Level Calibrations

The amount of material in front of ECAL, up to $2X_0$ in the barrel outer regions as seen in Figure 3.11 (left), produces a high rate of bremsstrahlung radiation from electrons and a high

probability of photon conversions. To mitigate this effect, a clustering algorithm is used to recombine energy deposits that come from those processes. The cluster energy is corrected via a multivariate technique, separately for photons [32] and electrons [40]. It also aims to correct other effects, such as in time pile up. The effect of these high level cluster corrections on the resolution of the $Z \to ee$ peak can be seen in Figure 3.11 (right).

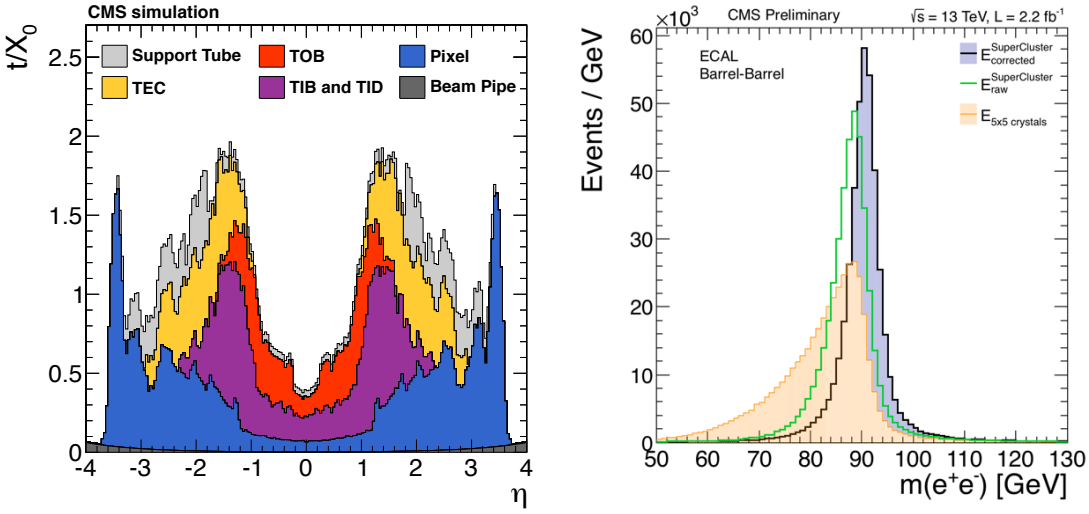

Figure 3.11: Material budget between the ECAL and the collision point on CMS (left). Effect of high level cluster corrections on the resolution of the $Z \to ee$ peak.

3.4 ECAL Energy Resolution

3.4.1 Energy Resolution of a Calorimeter

The intrinsic energy resolution of an electromagnetic calorimeter can be parametrized in terms of three sources, as in the following equation:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{A}{E}\right)^2 + \left(\frac{B}{\sqrt{E}}\right)^2 + C^2, \tag{3.3}$$

where the first term on the right-side of the equation is called the noise term, the second is called the stochastic term and the third, the constant term. A brief explanation of the physical nature of these uncertainties follows:

Noise term The noise term is the contribution to the energy resolution from the electronic noise in the readout chain. In the case of ECAL, this noise comes from the VFE and FE part of the detector, particularly in the light detection phase by the photo-detectors and on the amplification and sampling phase by the MGPA.

Stochastic term This contribution to the resolution comes from the physical development of the shower shape inside the detector. For homogeneous detectors, such as ECAL, this

term is usually small, since the full shower development is contained within the active medium.

Constant term Differently from the other two terms, this one is independent of the energy of the incoming particle in the detector. This uncertainty is related to geometrical and static features of the detector, such as non-uniformities and random instabilities in response that are independent on energy. In ECAL, this contribution is largely mitigated by the intercalibration and laser monitoring calibration.

During the ECAL test-beam commissioning period, these terms were measured directly as: A = 128 MeV, B = 0.028 $\sqrt{\text{GeV}}$ and C = 0.003 [41].

3.4.2 ECAL Energy Resolution with Run II data

The ECAL energy resolution is measured using $Z \to ee$ events, from an unbinned fit with a Breit-Wigner function convoluted with a Gaussian as signal model. Degradation effects come from the amount of material in front of ECAL and cracks between modules. The resolution, as a function of η, for low and high bremsstrahlung electrons in the barrel can be seen in Figure 3.12.

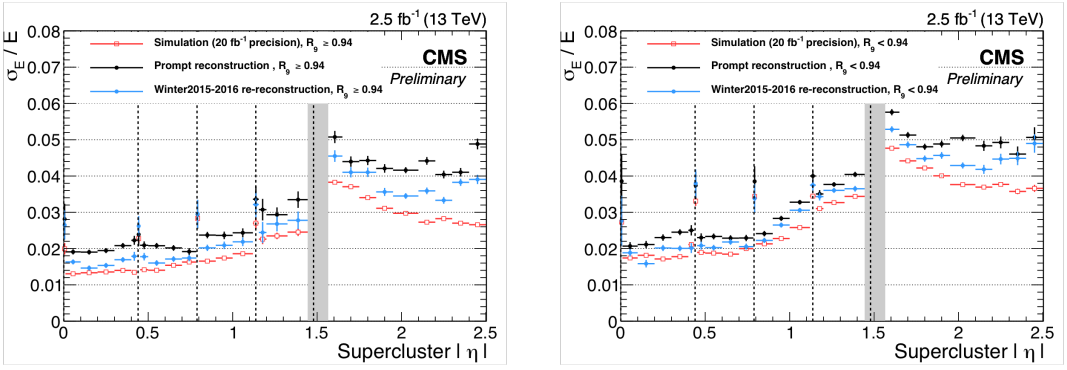

Figure 3.12: Energy resolution for low (left) and high (right) brehmsstrahlung $Z \to ee$ electrons as a function of η.

The energy resolution achieved with reprocessed data, which includes the latest intercalibration and calibration constants derived with 2015 data, achieves a resolution that is less than 2% for $Z \to ee$ low bremsstrahlung electrons in the central barrel region. The reprocessed data is especially better performing in the endcaps, when comparing to processed data with 2012 values for intercalibration and calibration. When applied to physics analysis, this energy resolution results in a $\sigma_{eff}/M_h \approx 1.5\%$ (from simulation) in the $H \to \gamma\gamma$ analysis, where σ_{eff} is the smallest interval in $M(\gamma\gamma)$ with 68.5% coverage [42].

3.5 The CMS ECAL Barrel Upgrade

3.5.1 The High Luminosity LHC

The performance of the Large Hadron Collider, and the physics program carried by ATLAS and CMS, have started remarkably. With the discovery of the Higgs boson in 2012, and the expected 40 fb^{-1} to be delivered in 2016 with record high instantaneous luminosity, the program has exceeded experimental expectations. The current LHC setup, with bunch spacing of 25 ns and peak instantaneous luminosity of the order of 10^{34} $cm^{-2}s^{-1}$, is expected to deliver around 300 fb^{-1} and operate until the end of 2022, when the Long Shutdown 3 (LS3) is planned to start.

During LS3, which is planned to last until 2025, the LHC is expected to replace the quadrupoles that focus the beam for ATLAS and CMS. Additionally, upgrades will be performed to optimize the bunch overlap at the interaction region. These improvements are expected to allow the peak luminosity to reach 2×10^{35} $cm^{-2}s^{-1}$ (High Luminosity LHC, HL-LHC). The full HL-LHC dataset of around 3000 fb^{-1} will open a brand new physics program at the LHC, involving precise measurements of the Higgs properties and rarer standard model phenomena, such as standard model Higgs pair production.

The new LHC configuration will also bring new experimental challenges. With the increased peak luminosity, a higher dose of radiation will be reaching the inner detectors of ATLAS and CMS, causing unrecoverable losses of efficiency in tracking reconstruction and calorimetry, for example. It will also challenge the online performance of the detectors, forcing the trigger systems to cope with higher event rates. To recover its performance to levels comparable or higher than during previous data taking periods, CMS has planned hardware upgrades to be deployed during LS3 [43].

3.5.2 ECAL Electronics Upgrade

In order to remain operational alongside the upgraded CMS L1 trigger and the CMS central DAQ, the ECAL front-end electronics will need to be replaced. As explained in the ECAL DAQ section, the full crystal information is currently read out of the ECAL FE at a rate of 100 kHz, which is the rate in which the L1 trigger sends accept signals. For the HL-LHC, this rate will be increased to 750 kHz, which is not allowed with the current ECAL FE bandwidth. The new ECAL FE will also have to deal with a decreased L1 latency of about 5 μs (from about 12 μs at the LHC).

Another issue for the ECAL EB upgrade will be to keep the physics performance at similar levels to the current ECAL. As mentioned in the previous section, the impact of incoming radiation on ECAL is to decrease the levels of transparency on the crystals. While some of this degradation is recovered through annealing, some of the transparency loss is permanent. With the incoming dose of radiation at the HL-LHC, the endcap crystals are expected to completely lose their transparencies. This prompted the CMS community to entirely replace the ECAL endcaps for its HL-LHC upgrade. The new ECAL endcap will be based on silicon technology, as a sampling calorimeter (HGCal).

In the ECAL barrel, the crystals are expected to robustly cope with the incoming dose of radiation (lower than the flux on the endcaps), and will therefore be kept in the detector. However, the new levels of radiation will also represent a challenge to the in-detector

electronics. For example, the dark current in the APDs (the electric current flowing through the photo-detector even in the absence of signal, created by random processes) is expected to increase, given the higher ionizing radiation dosage hitting the detector. This dark current is a large contributor to the noise term in the energy resolution formula, especially to homogeneous calorimeters, and interventions must be devised to keep it under control.

3.5.2.1 VFE Upgrade

For the HL-LHC upgrade, two ways to mitigate the increased detector noise have been investigated: replacing the VFEs and cooling the ECAL crystals and APDs from 18^{o}C to 8^{o}C. Cooling the APDs reduces the dark current.The downside of this upgrade is the decreased scintillation light yield from the lead tungstate crystals. The VFE upgrade intends to reduce the electronics noise by reducing the shaping time attached to the amplifying process, currently at 40 ns. This reduction limits the impact of electronics noise in the amplitude reconstruction, while also allowing for a more precise time resolution measurement with ECAL. These improvements can be seen in Figure 3.13 (left). Figure 3.13 (left) also shows that, in the scenario with no HL-LHC upgrades, the noise present on the ECAL barrel would reach levels almost 10 times higher than the current value (1.5 ADC counts). The upgrade reduces the noise at 3000 fb^{-1} by a factor of almost 2.

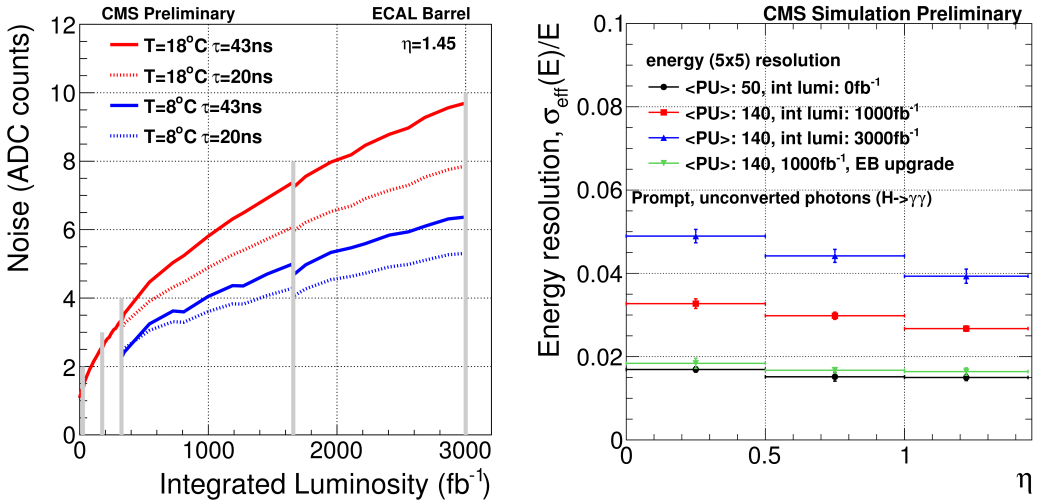

Figure 3.13: (Left) Noise levels on the upgraded ECAL barrel assuming different upgrade scenarios. (Right) Impact of full EB upgrade, including new VFE and cooling, on $H \to \gamma\gamma$ mass resolution.

The noise reduction is translated into physics performance, for example, in $H \to \gamma\gamma$ searches, as seen in Figure 3.13 (right). Figure 3.13 (right) shows the energy resolution of good quality $H \to \gamma\gamma$ photons under different integrated luminosity scenarios, with and without the ECAL barrel upgrade.

3.5.2.2 FE and Trigger Upgrade

With the upgraded FE, the full crystal information will be read out of the ECAL VFE at collision rate (40 MHz). New data links will be used to transmit all crystal data directly from the VFE to the back-end electronics, instead of being stored waiting for an L1A.

These spikes are created when a high energy hadron hits the APDs directly, without showering in the ECAL crystals. This creates an energy discharge in the APD that can fake a signal from a regular particle shower. Spikes are problematic when they happen with high frequency because they increase L1 trigger rates. These trigger must then have high energy thresholds in order for the rates to be acceptable by the CMS L1 trigger.

Another way currently implemented to suppress spikes at trigger level is through coarse geometric variables. These variables are computed at the trigger primitive generation step, in the FE electronics. Due to computational limits of the FE boards, these variables use limited information about the crystal array. This can be changed if the TP generation happened at the off-detector electronics instead. However, this would mean transmitting the full crystal information of the whole ECAL to the off-detector electronics at collision rate (40 MHz).

With extra computing power based on modern FPGAs at the back-end electronics, new strategies can be investigated to reduce the impact of anomalous signals (spikes) at L1 trigger level. One of these strategies is based on analyzing the signal pulse shape coming from the crystals. Given the smaller shaping time from the upgraded VFE, differences in pulse shapes from regular energy showers and spikes can become visible for ECAL.

Chapter 4

Higgs Exotic Decays with Photons and Missing Energy

4.1 Introduction

The final state consisting of a low transverse energy photon and low missing transverse energy ($\not{E}_T$) (also called the "monophoton" final state) can be used to constrain a variety of extensions of the standard model (SM). One such promising extension is supersymmetry (SUSY) [19, 20, 21, 22, 23, 24, 25], which has the attractive feature of stabilizing the radiative corrections to the Higgs boson mass (m_h), while also providing a natural dark matter (DM) particle candidate (χ) in the form of the lightest supersymmetric particle (LSP).

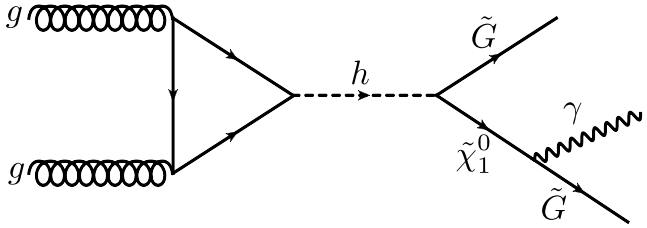

Figure 4.1: Feynman diagrams of a Higgs boson decay to gravitino LSP and a neutralino NLSP, which subsequently decays to a gravitino and photon.

In SUSY scenarios where the SUSY breaking scale is low ($\sqrt{f} \sim$ TeV) the newly discovered Higgs boson ($m_h = 125\,\text{GeV}$) [44, 45] may decay into a gravitino ($\widetilde{G}$) and neutralino ($\widetilde{\chi}_1^0$), with the neutralino subsequently decaying into a gravitino and a photon [46]. In this model, the gravitino is the LSP and the neutralino is the next-to-lightest supersymmetric particle (NLSP). Figure 4.1 shows the Feynman diagram for this process.

The fact that the gravitino is massive and stable (as the LSP) can give rise to cosmological problems (often quoted as the gravitino problem). This problem is related to the fact that the energy density of the gravitinos, which are produced in the early universe, may exceed the present critical density of the early universe if the gravitino mass is larger than 1 keV [47]. The benchmark model for this analysis, however, circumvents this issue with the fact that

$\sqrt{f} \sim$ TeV. This gives a gravitino mass that is $m_{3/2} \approx f/\sqrt{3}M_P \approx 10^{-3}$ eV, much smaller than the critical gravitino mass.

This decay mode produces a single isolated photon and $\not{E}_T$ from the undetected gravitinos. If $m_{\widetilde{\chi}_1^0} < m_h/2$, the decay process $h \to \widetilde{\chi}_1^0\widetilde{\chi}_1^0 \to \gamma\gamma + \not{E}_T$ would dominate. Therefore the kinematic region of interest for this search is $m_h/2 < m_{\widetilde{\chi}_1^0} < m_h$. Furthermore, since $m_h = 125\,\text{GeV}$, the photon transverse energy (E_T^γ) and $\not{E}_T$ will be relatively low.

In this document, we present a search for new physics in the low-E_T photon+$\not{E}_T$ final state, using an integrated luminosity at $7.3\,\text{fb}^{-1}$ of $\sqrt{s} = 8\,\text{TeV}$ LHC pp collision data collected with the CMS detector. This study is the first CMS search conducted in this low energy regime and it complements and expands upon previous high-energy monophoton searches for new physics conducted at the LHC [48, 49]. The results are interpreted in terms of the low-scale SUSY breaking model, as well as in a model independent manner.

4.2 Event selection

Events for this analysis are collected using a single HLT path, which is seeded by either a 20 GeV or a 22 GeV single e/γ L1 trigger depending on the running period. The HLT is used to select events with at least one photon with $E_\text{T} > 30\,\text{GeV}$ within the ECAL barrel region ($|\eta^\gamma| < 1.44$) and calorimetric $\not{E}_T$ >25 GeV with noise cleaning to suppress the anomalous noise in the HCAL barrel (HB) and endcap (HE) subdetectors due to characteristics of the hybrid photodiodes and the readout boxes [50]. The trigger further requires the photon to pass a loose calorimeter-based isolation selection and to exhibit shower shape characteristics consistent with unconverted photons. The main shower shape requirement is based on the R_9 variable, defined as the ratio of the energy deposited in a 3×3 crystal region centered around the crystal containing an energy deposit greater than all of its immediate neighbors (the "seed crystal") to the energy of the entire deposit of the photon ("supercluster"). The data recorded with this trigger corresponds to an integrated luminosity of $7.3\,\text{fb}^{-1}$ and was part of the CMS "data parking" program in 2012. With the data parking program, CMS recorded additional data with relaxed trigger requirements planning for a delayed offline reconstruction in 2013 after the completion of the LHC Run I.

The efficiencies of the trigger as a function of offline reconstructed E_T^γ and $\not{E}_T$ are measured using two prescaled control trigger paths. The first control trigger path accepts events with single photons with energy greater than 30 GeV with no further identification or isolation requirements on the photons. The second control trigger path has identical photon requirements to the signal path but without any selection on $\not{E}_T$. Figure 4.2 shows the efficiency turn-on curves as a function of E_T^γ and $\not{E}_T$, parameterized with an analytic function in the form of:

$$\varepsilon = \frac{p_2}{2} \cdot \left(1 + \text{Erf}\left(\frac{x - p_0}{p_1 \cdot \sqrt{2}}\right)\right). \tag{4.1}$$

In the offline selection, the events are required to have at least one well identified vertex with a distance less than 24 cm away from the nominal interaction point in z-direction and 2 cm away in the xy-plane. The vertex corresponding to the origin of the hard-scattering

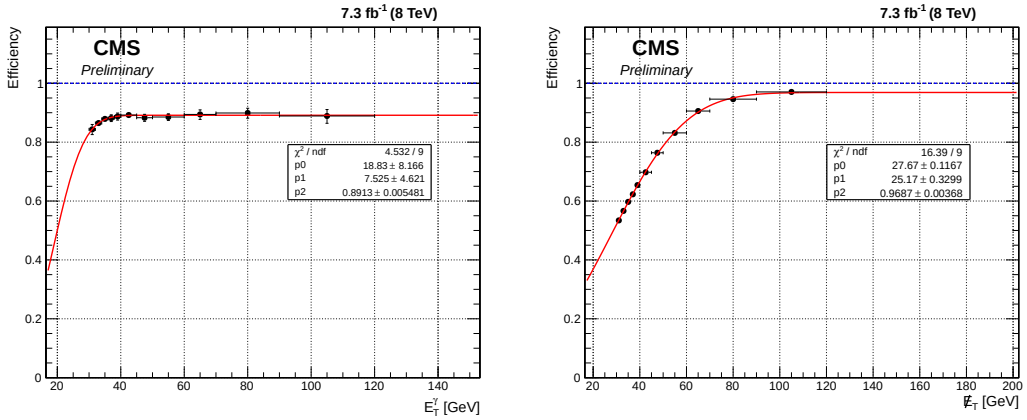

Figure 4.2: Trigger turn-on curves for $\mathrm{E}_{\mathrm{T}}^{\gamma}$ and $\not{E}_T$. The parameterization of efficiency as a function of offline $\mathrm{E}_{\mathrm{T}}^{\gamma}$ and $\not{E}_T$ are also shown in the form of an analytic function fitted to the turn-on distributions.

process with the largest value of $\sum p_{\mathrm{T}}^2$ of all associated tracks is identified as the primary vertex.

Each selected event is required to have at least one photon candidate with $\mathrm{E}_{\mathrm{T}}^{\gamma} > 45\,\mathrm{GeV}$ and $|\eta^{\gamma}| < 1.44$. The photon must also satisfy the following identification and isolation criteria: (a) to minimize the contribution from misidentified electrons, the shower is required to have no associated hits in the pixel detector, to be referred to as "pixel seed veto"; (b) the lateral extension of the shower, $\sigma_{i\eta i\eta}$, measured in terms of the energy-weighted spread within the 5×5 crystal should be consistent with that of a genuine photon; (c) the ratio between the energy collected by the HCAL cells behind the supercluster and the energy collected by the supercluster is required to be less than 0.05; (d) the sum of the E_{T} of all photons reconstructed with the particle flow (PF) algorithm within a cone of $\Delta R = 0.3$, excluding a strip in η of 0.015 around the supercluster, is required to be less than $0.7\,\mathrm{GeV} + 0.005\times\ \mathrm{E}_{\mathrm{T}}^{\gamma}$; (e) the sum of the E_{T} of all charged hadrons reconstructed with the PF algorithm within a hollow cone of $0.02 < \Delta R < 0.3$ around the supercluster is required to be less than $1.5\,\mathrm{GeV}$; (f) the sum of the E_{T} of all neutral hadrons reconstructed with the PF algorithm within a cone of $\Delta R = 0.3$ around the supercluster is required to be less than $1.0\,\mathrm{GeV} + 0.04\times\ \mathrm{E}_{\mathrm{T}}^{\gamma}$. To account for the effects of pile-up, the total energy density in the event is computed using the FastJet package [51] and is used to correct the isolation quantities. These pile-up-corrected isolation requirements correspond to a working point with a signal efficiency of approximately 85%. Furthermore, $R_9 > 0.9$ is also required to match the trigger requirements. The photon with the highest E_{T} in the event that satisfies all of the above requirements is selected as the photon candidate for the signal sample.

Spikes on the ECAL are rejected using additional shower shape requirements on the η and ϕ width of the shower. In addition, we reject showers that deposit more than 95% of their energy on the seed crystal [52].

To reduce the SM backgrounds arising from the leptonic decays of W and Z bosons, a lepton veto is applied. Events are rejected if they have at least one electron fulfilling loose

identification requirements [53] with $p_{\mathrm{T}}^{e} > 10\,\mathrm{GeV}$ and $|\eta^{e}| < 2.5$ (excluding the transition region of $1.44 < |\eta^{e}| \leq 1.55$) and are outside the cone defined by $\Delta R = 0.3$ around the photon candidate. Muons candidates which are identified using the PF algorithm using hits in the tracker and the muon systems are required to have $p_{\mathrm{T}}^{\mu} > 10\,\mathrm{GeV}$, $|\eta^{\mu}| < 2.1$, and $\Delta R(\gamma, \mu) > 0.3$ separation from the photon candidate. Events are rejected if any such muon is present in the event.

In addition to the selection requirements described above, the $\not\!\!E_T$ is required to be greater than 40 GeV. This level of selection is referred to as the preselection and is applied for both the model independent analysis and the analysis of the SUSY benchmark model. The additional applied selection requirements differ between the two analyses.

To define the jet candidates, identification criteria are used to separate pile-up jets from the jets originating from hard scattering. These identification criteria are based on the trajectory of tracks associated with the jets inside the tracker volume, the topology of the jet shape and multiplicity of the objects constituting these jets [54]. Only jets with $p_{\mathrm{T}}^{jet} > 30$ GeV and $|\eta^{jet}| < 2.4$ and that fulfill the non-pileup identification requirements are considered in the event. These jets must not overlap with the photon candidate within $\Delta R(\gamma,\mathrm{jet}) < 0.5$. In the model independent analysis, events with 2 or more jets are rejected and, if there is a jet in the event, we also require that $\Delta\phi(\gamma,\mathrm{jet}) < 2.5$.

In the analysis of the SUSY benchmark model, where no requirement is made on the jet multiplicity, more advanced selection is applied to reduce the background due to mismeasured $\not\!\!E_T$. Mismeasured $\not\!\!E_T$ can arise from many sources, including limited $\not\!\!E_T$ resolution, reconstruction and instrumental inefficiencies, and improper pattern recognition. Due to their large cross section, the γ+jets and multijet processes can contribute significantly to the background of this analysis, even though such events do not have genuine $\not\!\!E_T$. In order to minimize the contribution from these processes, we have used two different methods for identifying events with mismeasured $\not\!\!E_T$. The first one is the $\not\!\!E_T$ significance method [55] which takes into account the reconstructed objects in each event and their known measurement resolutions to compute an event-by-event estimation of the likelihood that the observed $\not\!\!E_T$ is consistent with zero. To complement this method we further developed the Missing H_{T} (MH_{T}) minimization method [56]. In the MH_{T} minimization method we first construct a χ^2 function with the form:

$$\chi^2 = \sum_{i=objects} \left(\frac{(p_{\mathrm{T}}^{reco})_i - (\widetilde{p}_{\mathrm{T}})_i}{(\sigma_{p_{\mathrm{T}}})_i} \right)^2 + \left(\frac{\widetilde{\not\!\!E}_x}{\sigma_{\not\!\!E_x}} \right)^2 + \left(\frac{\widetilde{\not\!\!E}_y}{\sigma_{\not\!\!E_y}} \right)^2 . \tag{4.2}$$

In the above equation, $(p_{\mathrm{T}}^{reco})_i$ are the transverse momenta of the reconstructed objects that pass the above mentioned identification criteria, the $(\sigma_{p_{\mathrm{T}}})_i$ are the expected resolutions of each object, the $\sigma_{\not\!\!E_{x,y}}$ are the resolution of the $\not\!\!E_T$ projection along the x-axis and the y-axis and the $(\widetilde{p}_{\mathrm{T}})_i$ are the free parameters allowed to vary in order to minimize the function. The first term of the equation is a scalar difference. The quantities $\widetilde{\not\!\!E}_{x,y}$ are functions of the free parameters;

$$\widetilde{\not\!\!E}_{x,y} = - \sum_{i=objects} (\widetilde{p}_{x,y})_i \tag{4.3}$$

In events with no genuine $\not\!\!E_T$, the mismeasured quantities can be re-distributed back into the particle momenta, resulting in a low χ^2 value. On the other hand, in events with genuine $\not\!\!E_T$ from undetected particles, the minimization generally will yield larger χ^2 values. For the analysis of the SUSY benchmark model, the re-calculated $\widetilde{\not\!\!E_T} = \sqrt{\widetilde{\not\!\!E_x}^2 + \widetilde{\not\!\!E_y}^2}$, i.e., in which the original object momenta are replaced with those obtained with the χ^2 minimization, is required to be > 45 GeV and the probability value obtained from the χ^2 minimization is required to be less than 10^{-3}.

To further suppress multijet backgrounds, events are rejected if the scalar sum of transverse momenta of the identified jets ($H_{\rm T}$) in the event is required to be greater than 100 GeV. An additional requirement is made on the angle (α) between the beam direction and the major axis of the supercluster in order to reject photons that have showers elongated along the beam line which is characteristic of non-prompt photons.

Finally, the transverse mass, $M_T = \sqrt{2p_{\rm T}^{\gamma}\not\!\!E_T(1-\cos\Delta\phi(\gamma,\not\!\!E_T))}$, formed by the photon candidate, $\not\!\!E_T$ and the angle between them, is required to be greater than 100 GeV. In order to easily interpret the results within the chosen benchmark model, we require the $\rm E_T^{\gamma}$ <60 GeV.

The final list of advanced selection used in both the model independent analysis and the analysis of the SUSY benchmark model with the relative cumulative efficiencies of the selection requirements relative to the preselection is given on table 4.1.

Selection requirements	Model independent		SUSY benchmark model		
Advanced selection	$Z\gamma \to \nu\bar{\nu}\gamma$	γ+jet	$Z\gamma \to \nu\bar{\nu}\gamma$	γ+jet	$M_{\widetilde{\chi}_1^0} = 120\,\text{GeV}$
Number of jets < 2	0.909	0.769	-	-	-
$\Delta\phi(\gamma,\text{jet}) < 2.5$	0.834	0.262	-	-	-
Transverse mass > 100 GeV	-	-	0.867	0.292	0.829
$H_{\rm T} < 100$ GeV	-	-	0.785	0.188	0.804
M$H_{\rm T}$ minimization: $\widetilde{\not\!\!E_T} > 45$ GeV	-	-	0.761	0.071	0.743
M$H_{\rm T}$ minimization: Prob(χ^2)$< 10^{-3}$	-	-	0.626	0.033	0.467
$\not\!\!E_T$ significance > 20	-	-	0.440	0.001	0.195
$\alpha > 1.2$	-	-	0.390	0.001	0.165
$\rm E_T^{\gamma} < 60$ GeV	-	-	0.074	0.0002	0.106

Table 4.1: Summary of selection for both model independent analysis and analysis with SUSY benchmark model with the cumulative efficiencies of the selection requirements relative to the preselection for $Z\gamma \to \nu\bar{\nu}\gamma$, γ+jet and $M_{\widetilde{\chi}_1^0} = 120\,\text{GeV}$.

4.3 Background estimation

The irreducible background for the γ+$\not\!\!E_T$ signal is the SM process $Z\gamma \to \nu\bar{\nu}\gamma$. Other SM backgrounds include Wγ, W $\to e\nu$, W $\to \mu\nu$, W $\to \tau\nu$, γ+jet, multijet (referred to as QCD background) and diphoton events. Background events that do not come from collision processes are also considered in the analysis. These backgrounds can be divided broadly into three categories:

- Backgrounds estimated with Monte Carlo (MC) simulation: $Z\gamma \to \nu\overline{\nu}\gamma$, $Z\gamma \to \ell\ell\gamma$, W$\gamma$, γ+jets, $\gamma\gamma$;
- Collision backgrounds estimated using data: jets misidentified as photons and electrons misidentified as photons;
- Non-collision backgrounds estimated using data.

4.3.1 Background estimates from MC simulation

The $Z\gamma \to \nu\overline{\nu}\gamma$ is the irreducible background in this study. It is estimated from a MC sample generated with MadGraph [57] and processed with pythia v6.426 [58] for showering and hadronization. To describe the topology of events in a realistic manner, up to two extra partons were included in the matrix elements during event generation. The final event yields are normalized using the next-to-leading-order (NLO) cross section calculated with the mcfm event generator [59]. The Wγ MC simulation was also produced with MadGraph, including up to two extra partons, and normalized using the NLO cross section calculated from mcfm.

The γ+jet is one of the most significant backgrounds in this analysis due to the presence of a genuine photon and the large production cross section. The events for this process are simulated using MadGraph, processed through pythia and normalized using the LO cross section. The normalization of this background is then corrected using control sample in data for two different event classes: no-jet and one or more jets. The control sample in data is obtained using events collected with the prescaled single-photon trigger and with the $\not\!\!E_T$ requirement reversed to ensure orthogonality to the signal phase space. Correction factors (C) are derived by normalizing the event yield in the simulation to match the data in the control region, separately for events with zero jets ($C = 1.7$) and one or more jets ($C = 1.1$). These correction factors are then used to normalize the γ+jet event yield from simulation in the signal region. An uncertainty of 16% is derived for these correction factors based on the difference between the corrected and uncorrected simulation and the relative fraction of zero jet ($\sim 10\%$ of the events in the control region) and one or more jet events.

The background estimate from the process $Z\gamma \to \ell\ell\gamma$ is obtained using MadGraph, and the background estimates from the processes W $\to \mu\nu$, W $\to \tau\nu$ are obtained using pythia. These processes are found to contribute only to a small fraction of the total background prediction.

All simulated events are processed through a Geant4-based simulation of the CMS detector [60] and the same reconstruction chain as the data. To take into account differences between data and simulation due to imperfect MC modeling, various scale factors (SF) are applied to correct the MC-based estimates. These SFs are defined as the ratio of the efficiency in data to the efficiency in simulation for a given selection. The SF for photon reconstruction and identification is estimated from $Z \to ee$ decays with $R_9 > 0.9$ using a standard "tag-and-probe" method [61] and are listed in Table 4.2. The trigger efficiency measured in Section 4.2 is applied to the simulated samples.

Table 4.2: Scale factors for the photon identification

Variable	E_T^γ $[30-40]$ GeV	E_T^γ $[40-50]$ GeV	$E_T^\gamma > 50$ GeV
Pixel Seed Veto	0.995 ± 0.004	0.970 ± 0.016	1.012 ± 0.014
Photon ID ($0 < \|\eta\| < 0.8$)	0.996 ± 0.010	0.996 ± 0.010	0.996 ± 0.010
Photon ID ($0.8 < \|\eta\| < 1.5$)	0.996 ± 0.010	0.997 ± 0.010	0.997 ± 0.010

4.3.2 Jets Identified as Photons Background

The contamination from jets misidentified as photons is estimated by using a data control sample, enriched in QCD multijet events, defined by $\not{E}_T < 40$ GeV. In this sample we measure the ratio of events that pass the signal photon identification and isolation criteria (numerator selection) relative to those that pass a looser photon identification and isolation criteria but also fail one of the nominal isolation requirements (denominator selection). The sample of events passing the numerator selection includes genuine photons whose contribution must be subtracted to obtain the true ratio of misidentified jets. This contribution is estimated by fitting templates of energy-weighted shower widths of genuine photons (determined from MC simulation of γ+jets events) and misidentified photons (determined from an isolation sideband in data) to the candidate distribution. The final corrected extrapolation factor, shown in Figure 4.3, is then used to scale a sample of events in data which pass the denominator selection in addition to the other non-orthogonal event selections.

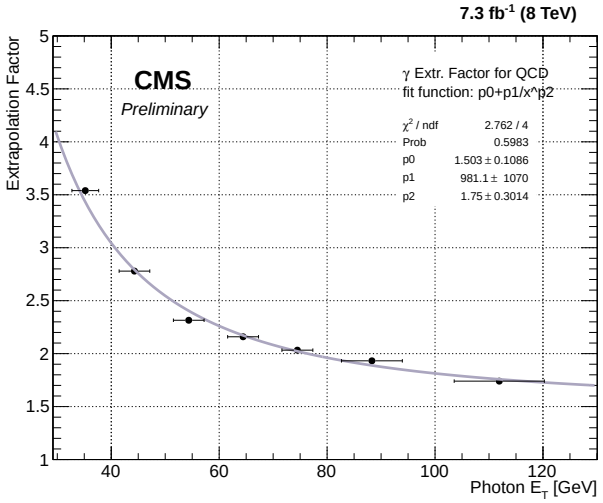

Figure 4.3: Ratio of number of photons passing signal-like requirements relative to those satisfying a very loose identification and isolation selection and at the same time failing at least one of the isolation criteria as a function of E_T^γ.

The systematic uncertainty in this method is dominated by the choice of the isolation sideband region and is conservatively estimated to be 35% by varying the charged-hadron isolation in the sideband region definition. The other sources of systematic uncertainty are determined by varying the bin size of the templates, the $\not{E}_T$ selection for the control region,

and the loose identification requirements on the photons, and are found to be comparatively small.

4.3.3 Electrons Identified as Photons Background

After the cuts for QCD-like events are applied, the background that arises from electrons-faking-photons becomes the dominant for the kinematic range of this analysis. We have developed a data-driven method to obtain an accurate estimation of the electron-faking-photon background. The main physical process behind this background is the $W \to e\nu$ production, which has a kinematic signature (fake photon E_{T}, transverse mass, etc) very similar to the signal sample and large production cross section.

The method for estimating the electron-faking-photon contamination in the signal sample was done by constructing a control sample, similar to the signal sample, but enriched by electrons and not signal photons. This was achieved by inverting the pixel seed veto (PSV) in the photon ID requirements. The pixel seed veto cut is not the standard electron rejection tool recommended by CMS; however, it has been shown to have a much smaller fake rate compared to the official electron rejection cut, the conversion safe electron veto. The PSV has been used in other photon-related analyses, such as the search for SUSY in final states involving photons and the high E_T monophoton analysis, in which the electron-faking-photon background is a dominant one.

After the control sample is created, it needs to be normalized by the number of expected electrons-faking-photons in the signal sample. This normalization can be obtained by the calculation of the fake rate of the PSV regarding electrons-faking-photons. It should be noted that this is not the fake rate of the full set of photon ID requirements, but only of the PSV cut.

Before describing the details of this method, the following definitions must be made:

- γ: Objects that pass the complete e/γ photon ID, including the PSV cut (photon object);
- γ_e: Objects that pass the photon ID with the PSV cut reversed (e-fake object);
- N_{γ_e}: Number of events in the control sample, which is made of γ_e;
- ϵ_{γ_e}: Efficiency of accepting γ_e objects (including acceptance);
- ϵ_{γ}: Efficiency of accepting γ objects (including acceptance);
- $F_{e\to\gamma}$: Fake rate of the PSV cut for electrons-faking-photons.

Therefore, the number of electrons-faking-photons in the signal sample is given by

$$N_{e\to\gamma} = \frac{N_{\gamma_e}}{\epsilon_{\gamma_e}} \times F_{e\to\gamma} = N_{\gamma_e} \times R \tag{4.4}$$

$$R = \frac{F_{e\to\gamma}}{\epsilon_{\gamma_e}}. \tag{4.5}$$

This ratio (R) can be related to the PSV fake rate because, for the PSV cut, the efficiency and fake rate obey the following relation:

$$\epsilon_{\gamma_e} + F_{e\to\gamma} = 1. \tag{4.6}$$

Therefore,

$$F_{e\to\gamma} = \frac{R}{1+R}. \tag{4.7}$$

The ratio and the fake rate can be measured using a tag-and-probe method. This method was applied to the signal dataset used in the analysis. Analyzing events in which the Z decays to electrons, we can reconstruct the number of events in four different cases:

- $Z_{e\gamma}$: one of the electrons of the Z decay is identified an electron and the other is reconstructed as a photon object;
- $Z_{e\gamma_e}$: one of the electrons of the Z decay is identified as an electron and the other is reconstructed as an e-fake;
- $Z_{\gamma_e\gamma}$: one of the electrons of the Z decay is identified an e-fake and the other as a photon;
- $Z_{\gamma_e\gamma_e}$: both electrons of the Z decay are identified as e-fakes;

Therefore, we can reconstruct the number of events in each Z peak case as:

$$\begin{aligned} N_{e\gamma} &= 2\times N'(Z\to ee)\times\epsilon_e\times F_{e\to\gamma} && (4.8)\\ N_{e\gamma_e} &= 2\times N'(Z\to ee)\times\epsilon_e\times\epsilon_{\gamma_e} && (4.9)\\ N_{\gamma\gamma_e} &= 2\times N'(Z\to ee)\times F_{e\to\gamma}\times\epsilon_{\gamma_e} && (4.10)\\ N_{\gamma_e\gamma_e} &= N'(Z\to ee)\times\epsilon_{\gamma_e}^2, && (4.11) \end{aligned}$$

where N_{xy} is the number of events of Z decaying to electrons when the electrons are reconstructed as x and y, and $N'(Z\to ee)$ overall number of expected $Z\to ee$ in the sample. The factors of two multiplying the first three equations are due to combinatorics.

With that system, we can infer that the calculation of the ratio can be performed in two ways:

$$R = \frac{F_{e\to\gamma}}{\epsilon_{\gamma_e}} = \frac{N_{e\gamma}}{N_{e\gamma_e}} = \frac{1}{2}\frac{N_{\gamma_e\gamma}}{N_{\gamma_e\gamma_e}} \tag{4.12}$$

The Z shape was obtained from the template of a DY$\to ee$ MC at generator level. This template was then convoluted with a Gaussian to simulate detector resolution effects. The parameters of the Gaussian were then fitted to the obtained invariant mass distribution from the different categories detailed above. For the calculation of N_{xy}, the signal function was integrated in the $\pm 2\sigma$ region to obtain the number of signal events. The fits for the $Z\to e\gamma$ and $Z\to e\gamma_e$ cases can be seen in Figures 4.4. The background shape for the fit was estimated by the convolution of an error function and a decaying exponential function (RooCMSShape).

The background in the Figures 4.4 originates from the combinatorics of the electron plus non-resonant photon-like objects, such as real photons, jets faking photons and electrons faking photons originated from other interactions. Because the electron is most likely coming from resonant Z production, the yield drops at lower values of $M_{e\gamma}$. The small disagreement between the fits and the data aroung 85 GeV comes from the trigger and selection acceptances, since our trigger photon object selection is 30 GeV and our probe selection p_{T} for this study is 35 GeV.

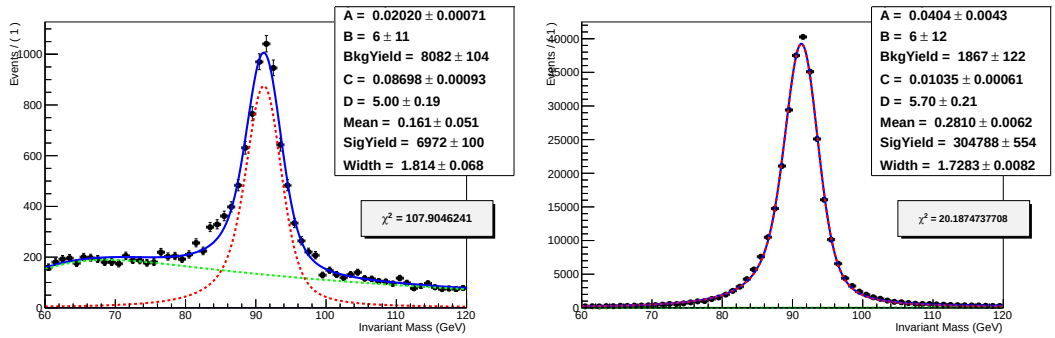

Figure 4.4: Fit of the Z invariant mass for the $Z \to e\gamma$ case (left) and the $Z \to e\gamma_e$ case (right). The fit parameters refer to the background shape, a convolution of an error function and an exponential function, and the gaussian smearing of the Z shape from MC.

Assuming a Poissonian error on the signal integral result, the error on the ratio and fake rate can be estimated as:

$$\sigma_R = \sqrt{\frac{1}{N^2_{e\gamma_e}}\sigma^2_{N_{e\gamma}} + \frac{N^2_{e\gamma}}{N^4_{e\gamma_e}}\sigma^2_{N_{e\gamma_e}}} \tag{4.13}$$

$$\sigma_{FR} = \frac{1}{(1+R)^2}\sigma_R. \tag{4.14}$$

With that, we obtain the following results:

$$R = (2.38 \pm 0.03)\,\% \tag{4.15}$$

$$F_{e\to\gamma} = (2.32 \pm 0.03)\,\%. \tag{4.16}$$

4.3.3.1 Electron → Photon Fake Rate Dependencies on p_{T}, N_{Vtx}, N_{Trk}

It has been shown before, for example in the SUSY photon analysis, that the PSV fake rate is dependent on variables such as the probe p_T, the number of tracks associated with the primary vertex and the number of reconstructed primary vertices in the event. The nature of the two last dependencies are rooted in the track reconstruction and matching algorithm.

To check the dependency of the fake rate in the three variables mentioned, the fake rate was calculated in exclusive bins. In each bin, the signal template used was the corresponding bin in the MC signal sample. Each fit was done individually for every bin.

In the plots in Figure 4.5, we see how the PSV fake rate depends on the probe p_T, number of tracks associated with the primary vertex and the number of reconstructed primary vertices, respectively. The red line in each plot represents the fake rate obtained previously, assuming that there are no dependencies, with the entire invariant mass spectrum. For now on, this first result will be referenced as the flat fake rate.

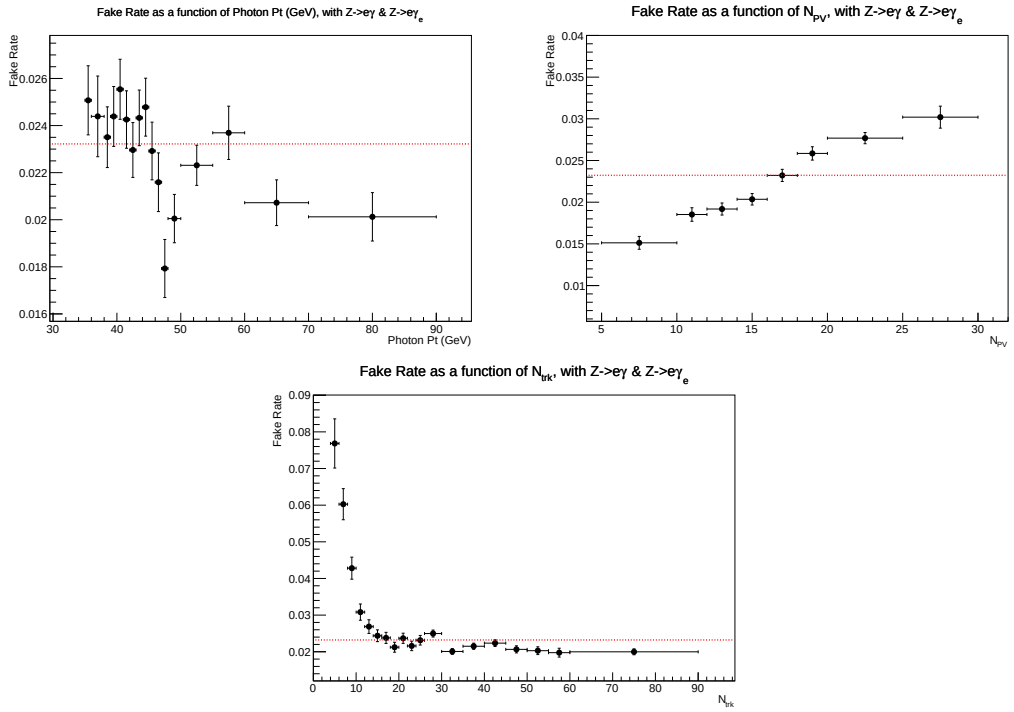

Figure 4.5: Fake rate vs. probe p_T (top left), number of primary vertices (top right) and track multiplicity (bottom).

As shown in the plots, there is a non-trivial dependency of the fake rate on the variables probed. There are different ways to solve this feature. The most complete one would be to achieve a multi-parameter description of the fake rate, including a 3-dimensional function with dependencies of the fake rate on each variable. That, however, demands a thorough study of these dependencies and a functional form to perform this task. A second way would be to choose one of the parametrizations, including the flat one, and assign a systematic error to that assumption. To know how much one choice impacts the fake rate, we can look at the final result and observe how much the yields change with each assumption.

In our case, the final result is the control sample of photon-like objects that fail the pixel seed veto normalized by the fake rate, which represents the estimation of our electron faking photon contamination in the signal region. The plots in Figure 4.6 show the shape and yields of this control sample when normalized by the different parametrizations of the fake rate. When checking the differences in number of events of each bin, we see that they are all within 5% of each other, being higher around the p_T peak and approaching zero for higher values.

Therefore, it is safe to choose any specific parametrization and assume a systematic error of 5% on the number of events to make sure that the different assumptions are compatible. With these assumptions, we choose to use the flat parametrization, since it only involves one fit and it has been shown to be much more stable with respect to different fit functions, for signal and background.

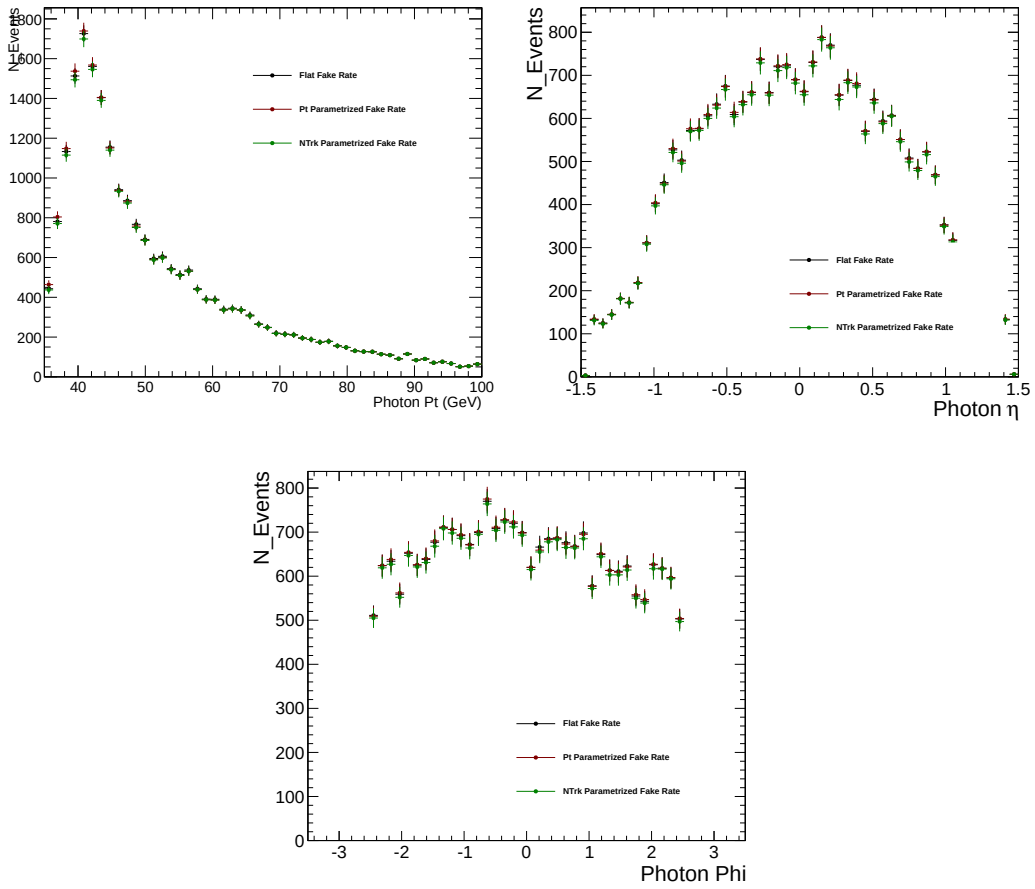

Figure 4.6: Control sample distribution of photon p_T (top left), η (top right) and ϕ (bottom).

4.3.3.2 Closure Test for Electron → Photon Fake Rate Measurment

As a closure test, we compare the generator-level fake rate to the fake rate as calculated by the method described above on MC (the reco-level fake rate) and check that they agree. For that, we used the Drell-Yan sample.

The generator-level fake rate is defined as:

$$F_{gen} = \frac{\#\text{Medium ID Photons \&\& Matched to Gen Electrons \&\& Pass the PSV}}{\#\text{Medium ID Photons \&\& Matched to Gen Electrons}}. \quad (4.17)$$

Here, the Medium photon ID is the one used in the analysis, including all the shower shape cuts to remove spikes and other contributions.

We compare the results of the two measurements, generator-level and reco-level, on the plots in Figure 4.7.

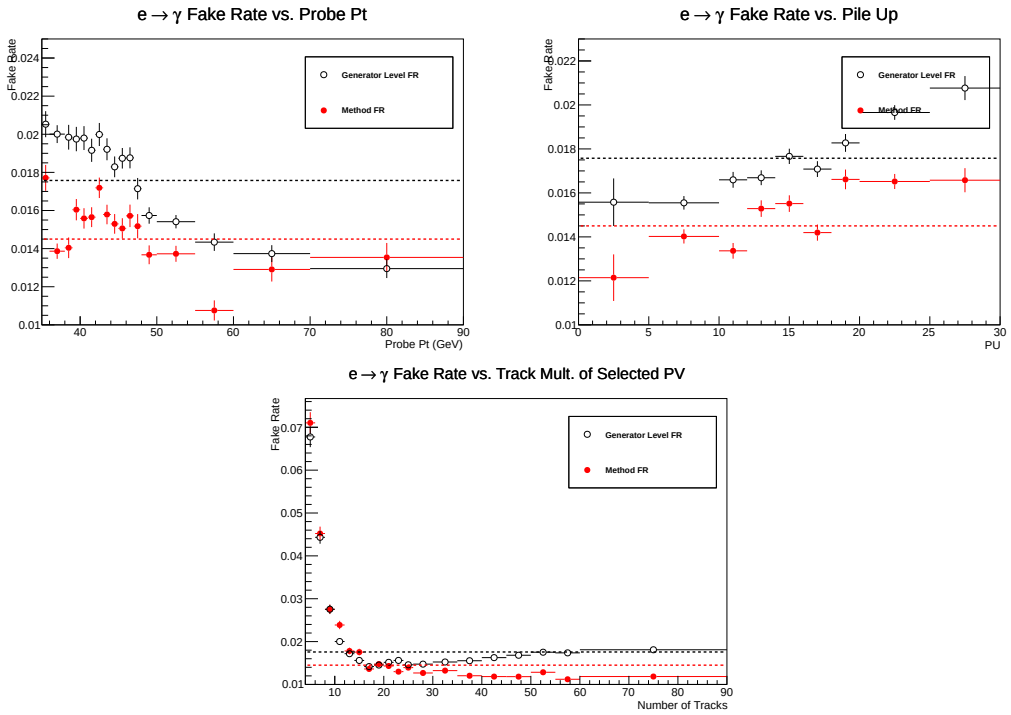

Figure 4.7: Closure test in probe p_T (top left) ,number of reconstructeed primary vertices (top right) and track multiplicity (bottom). The filled dots represent the fake rate calculated with the $Z \to ee$ method described previously. The open dots represent the fake rate calculated using the generator information described here. In order for the $Z \to ee$ method to close, those two distributions should be the same.

The lines represent the fake rate results from the flat assumption. For the generator-level fake rate, that is just the ratio from equation 4.17 on all events, while the parametrized fake rate is the ratio in each bin. As shown, there is a significant discrepancy between the two measurements. With those results, we can only say that the method is closed within 20%. The 20% comes from the difference between the flat fake rates from generator-level and reco-level results.

We investigated the cause of this result and noticed that it arises from a "hidden" primary vertex matching in the tag-and-probe method. When using the tag-and-probe for electrons as tag and photons as probe, the photon ends up being matched to the hardest primary vertex because the electron, in the electron ID requirements, is indeed matched. When we require the invariant mass of the pair to be close to the Z peak, the electron PV matching is indirectly passed to the probe photon, since they must come from the same source. This requirement

does not exist on the gen-level fake rate, which is basically a counting experiment.

The importance of the PV matching requirement comes about because of the nature of the DY→ *ee* process. Since it is a low multiplicity process, i.e., there will not be many objects naturally from the hard scatter process, the Z might not come from the hardest reconstructed primary vertex of the event. Therefore, when that happens, we measure the fake rate from a primary vertex that is not the hardest and, therefore, has fewer tracks than the primary vertex assumed. Since we know that the fake rate increases sharply when there are few tracks in the primary vertex, that explains why, in Figure 4.7, there is an increase in higher values of NTrack - those are actually hard scatter events with few tracks that were mistakenly matched to the hardest PV.

To overcome these problems, we match the gen-level particles used for the gen-level fake rate to the hardest reconstructed primary vertex with the following cuts (from the muon ID POG recommendations):

- $D_z < 0.5$;
- $D_{xy} < 0.2$.

With those requirements, we have the plots on Figure 4.8.

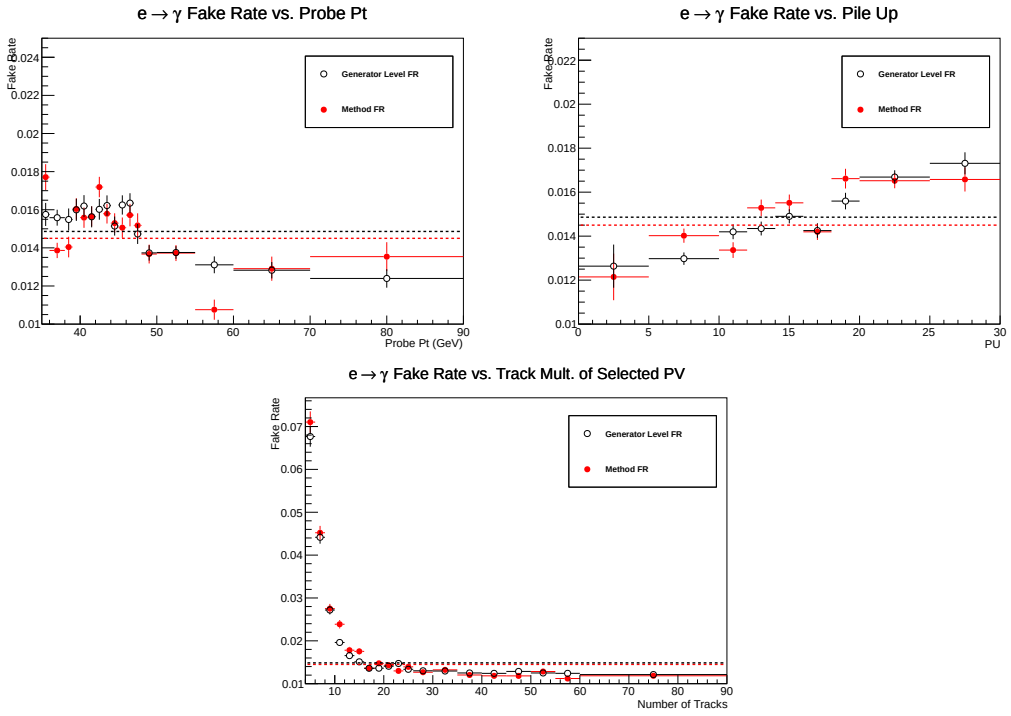

Figure 4.8: PV matched closure test in probe p_T,number of reconstructeed primary vertices and track multiplicity

We can see that the agreement between the gen-level and reco-level fake rates improves

and the method closes within 4%. This means that the fake rate measured in data and MC agree with an uncertainty of 4%.

4.3.3.3 Systematic Uncertainty on the Electron → Photon Fake Rate Measurement

Two of the main systematic errors have been discussed in the previous section:

- The flat assumption systematic error is found to be 5% on the number of events;
- The closure test systematic is found to be 4% on the value of the fake rate.

The remaining systematic uncertainty is related to the background estimation Fig. 4.4, i.e. the choice of the functional form to represent the backgroun composition. Two choices were made for that estimation: a simple decaying exponential and the RooCMSShape. Looking at the full mass spectrum of the invariant mass, without parametrization, the amount of expected background is much smaller than the number of signal events. Because of that, we don't expect the fake rate to be very dependent on the functional shape of the background. Indeed, the difference in the calculated fake rate for the two functions is about 4%.

These systematic errors are assumed independent and should be added as such in the final number for the fake rate. They are, however, very small compared to the other sources of systematic errors in the analysis.

4.3.4 Non-collision background estimates from data

The search is susceptible to contamination from non-collision backgrounds which arise from cosmic ray interactions, spurious signals in the ECAL, and accelerator-induced secondary particles (beam halo). These backgrounds have different arrival time distributions compared to prompt photons produced in hard scattering. To quantify the contamination due to these backgrounds a fit is performed to the candidate time distribution using background templates derived from the data. The contamination due to this out-of-time background is found to be negligible, and therefore not included in the final event yield.

4.3.5 Background modeling validation

The background modeling is examined in several control regions. A control sample enriched in $W(l\nu)\gamma$ events is defined with an inverted lepton-veto requirement in the preselection, thus selecting events with a loose e or μ. It is expected to be free of any signal contamination due to the presence of a lepton. Another control sample enriched in γ+jet events is constructed by just requiring no selection other than the preselection requirements. Figure 4.9 shows the data vs SM expectation in the two control regions. The observed data and estimated SM backgrounds are found to be consistent both in yield and shape.

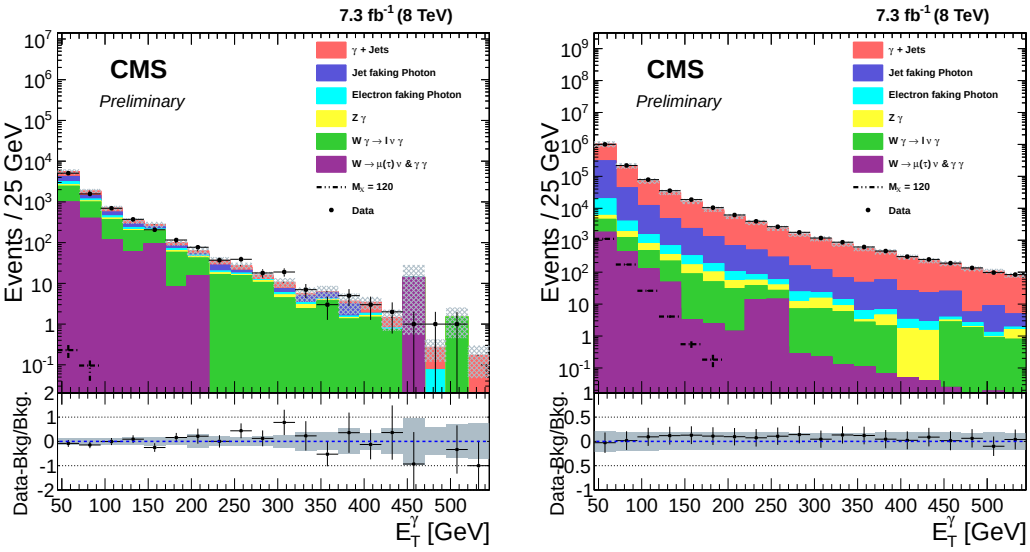

Figure 4.9: The E_T^γ distributions for data vs. SM expectation in a control region enriched by (left) W($l\nu$) events, and (right) γ+jet events.

4.4 Systematic uncertainties

The experimental systematic uncertainties considered in the analysis are listed in Table 5.7. Since the model-independent and model-specific selections differ significantly, e.g. the inclusion of the mismeasured $\not{E}_T$ reduction selection and E_T^γ window for the model specific results, the systematic uncertainties are evaluated separately for each selection. The systematic uncertainty associated with the measurement of the integrated luminosity is 2.6% [62]. The photon energy scale uncertainty [63] of about 1% affects the signal and background predictions by 4% for the model specific selection and by 0.5% for the model independent selection. Similarly, the jet energy scale uncertainty affects the signal and background predictions by $2-5\%$ depending on the process and selection. When varying the photon or jet energy scales, the $\not{E}_T$ is also recomputed. In addition, the systematic uncertainty associated with the jet energy resolution (0.5%) and unclustered energy (energy not contained within jets or leptons or photons) scale (2%) are propagated to the $\not{E}_T$ and effect the signal and background predictions by $2-4\%$. The uncertainty due to the choice of parton density functions (PDFs) is estimated following PDF4LHC recommendations [64, 65, 66] and are found to be 10% for SUSY Higgs boson signal, and 4% for $Z\nu\nu\gamma$ and the $W\gamma$. An additional uncertainty due to the choice of renormalization and factorization scales is evaluated using MCFM and is found to be 3% for $Z\nu\nu\gamma$ and the $W\gamma$. As described in the previous section, a 16% uncertainty is applied to the γ+jet normalization due to the difference in the jet multiplicity distribution between the data and background prediction in the γ+jet control region. The uncertainty due to the pileup modeling is found to be 1% and is estimated by shifting the central value of the total inelastic cross section from 69.4 mb to 73.5 mb. Finally, the systematic uncertainties associated with backgrounds estimated from data have been applied as discussed in previous

section.

Source	Signal	Jet→ γ	Electron→ γ	γ + jet	$Z\nu\nu\gamma$	$W\gamma$
PDF	10(0)	-	-	-	4(4)	4(4)
Luminosity	2.6(2.6)	-	-	2.6(2.6)	2.6(2.6)	2.6(2.6)
Photon energy scale ± 1 %	4(0.5)	-	-	4(0.5)	4(0.5)	4(0.5)
$\not{E}_T$ energy scale	4(2)	-	-	4(2)	4(2)	4(2)
Jet energy scale	3(2)	-	-	5(5)	3(2)	3(2)
Pileup	1(1)	-	-	1(1)	1(1)	1(1)
$Z\nu\nu\gamma$ MCFM NLO calculation	-	-	-	-	3(3)	-
γ+ jet normalization	-	-	-	16(16)	-	-
$W\gamma$ MCFM NLO calculation	-	-	-	-	-	3(3)
Jet→ γ unc.	-	35(35)	-	-	-	-
Electron→ γ unc.	-	-	6(6)	-	-	-

Table 4.3: Summary of all systematic relative uncertainties in percent on acceptance times efficiency calculation for SUSY Higgs model (Model independent) selection.

4.5 Results

To evaluate the 95% confidence level (CL) limits on the new physics production cross section, an asymptotic CL_S method [67, 68] is used where the systematic uncertainties in the signal and background predictions are treated as nuisance parameters with log-normal prior distributions.

4.5.1 Model-independent limits

Due to the variety of signals which can contribute to this final state, we present results for a generic signal using the model-independent selection described in Section 4.2. Although this selection does not have as strong discrimination power between signal-like and background-like events compared to the misreconstructed $\not{E}_T$ rejection selections, it does have less model dependence. This is due to $\not{E}_T$ significance and $\tilde{\not{E}}_T$ minimization requirements having a non-trivial efficiency dependence on the underlying event and observed $\not{E}_T$.

The total expected SM background and observed data events after the model-independent selection are found to be compatible within the systematic uncertainties. Table 4.4 shows a comparison of the event yields estimated for background processes and the observed data. Figure 4.10 shows the M_T and $\not{E}_T$ distributions after the model-independent selection has been applied.

Process	# of Events
γ+ jets	$(313 \pm 50) \times 10^3$
jet $\rightarrow \gamma$	$(906 \pm 317) \times 10^2$
e $\rightarrow \gamma$	$(1035 \pm 62) \times 10^1$
$W(\rightarrow \ell\nu) + \gamma$	2239 ± 111
$Z(\rightarrow \nu\bar{\nu}) + \gamma$	2050 ± 102
Other	1809 ± 91
Total background	$(420 \pm 82) \times 10^3$
Data	442×10^3

Table 4.4: Comparison of event yields for observed data and background, after the model-independent selection.

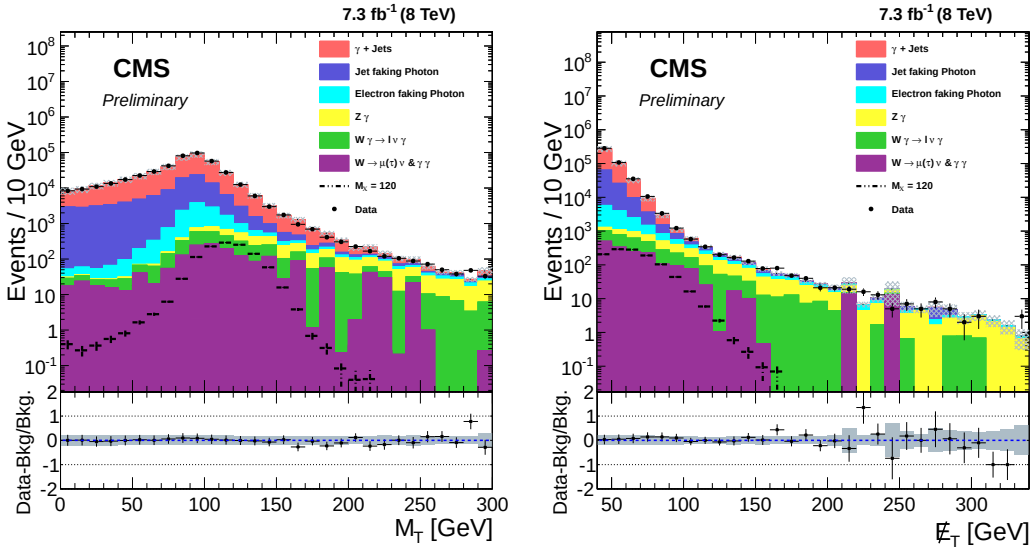

Figure 4.10: The M_T and $\not{E}_T$ distributions for data, background estimates, and signal after the model-independent selection. The bottom panels in each plot show the ratio of (data - background)/background and the gray band includes both the statistical and systematic uncertainty on the background prediction.

Figure 4.11 shows the observed and expected model-independent 95% CL upper limits on $pp \rightarrow h$ cross section times the exotic Higgs decay branching ratio times the acceptance and efficiency of selecting the signal ($\sigma \times BR \times A \times \epsilon$) for different $\not{E}_T$ and M_T thresholds (typical values for the signal efficiency ϵ, in the case where $M_\chi = 120$ GeV, are shown in Table 4.1). The observed and expected limits are also shown in Fig 4.11(c) at a 95% CL for $M_T > 100$ GeV and as a function of $\not{E}_T$.

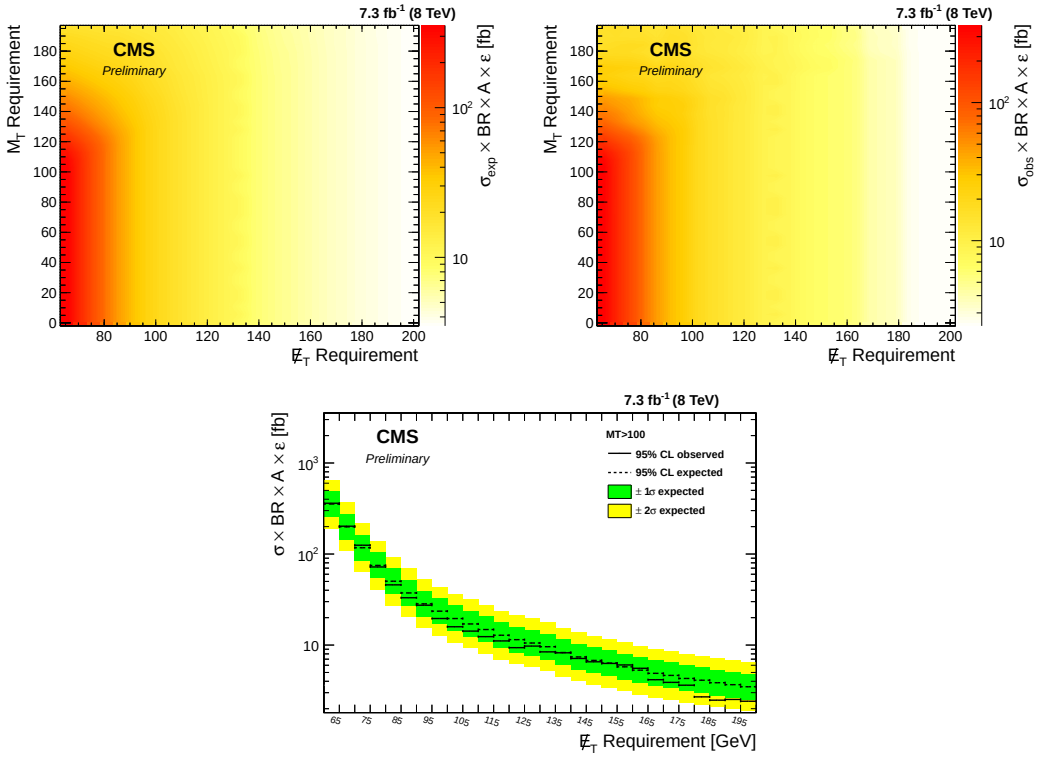

Figure 4.11: The expected (top left) and observed (top right) 95% CL upper limit on $\sigma \times BR \times A \times \epsilon$ for different M_T and $\not{E}_T$ thresholds and (bottom) for $M_T > 100$ GeV as function of the $\not{E}_T$ threshold.

4.5.2 Model-specific limits

The yields for supersymmetric decays of the Higgs boson ($h \to \widetilde{G}\widetilde{\chi}_1^0, \widetilde{\chi}_1^0 \to \widetilde{G}\gamma$) are calculated through imposing the model-specific selection described in Section 4.2. The yields for this selection are shown in Table 4.5. The 95% CL upper limits on the $\sigma\times$ branching ratio(BR) and $(\sigma \times BR)/\sigma_{SM}$, where σ_{SM} is the cross section for the standard model Higgs boson, are evaluated for different mass values of $\widetilde{\chi}_1^0$ ranging from 65 GeV to 120 GeV and are shown in Fig. 4.12.

Process	Estimate
γ+ jets	179 $\pm$ 28
jet $\rightarrow \gamma$	269 $\pm$ 94
e $\rightarrow \gamma$	355 $\pm$ 28
$W(\rightarrow \ell\nu) + \gamma$	154 $\pm$ 15
$Z(\rightarrow \nu\bar{\nu}) + \gamma$	182 $\pm$ 13
Other	91 $\pm$ 10
Total background	1232 $\pm$ 188
Data	1296
$M_{\widetilde{\chi}_1^0} = 65$ GeV	653.0 $\pm$ 77
$M_{\widetilde{\chi}_1^0} = 95$ GeV	1158.1 $\pm$ 137
$M_{\widetilde{\chi}_1^0} = 120$ GeV	2935.0 $\pm$ 349

Table 4.5: Expected (SM background) and observed event yields after the selection optimized for the supersymmetric decay of the Higgs boson ($h \rightarrow \widetilde{G}\widetilde{\chi}_1^0, \widetilde{\chi}_1^0 \rightarrow \widetilde{G}\gamma$) and the signal predictions correspond to BR($H \rightarrow$invisible + γ) =100%.

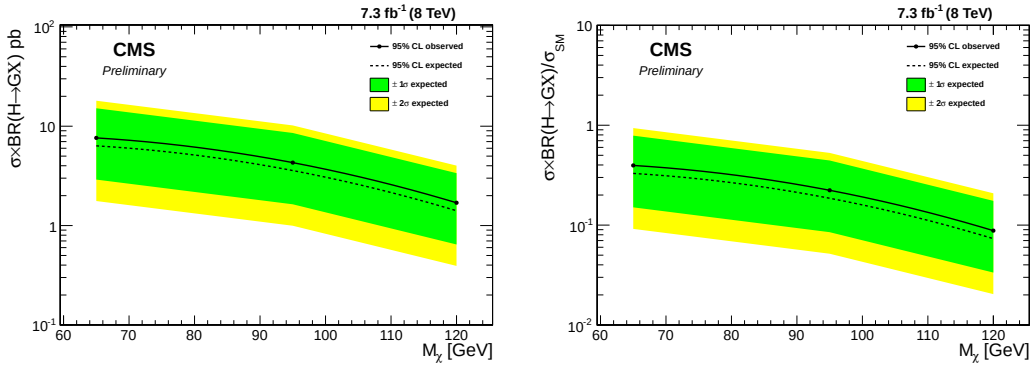

Figure 4.12: Expected and observed 95% CL upper limits on (a) $\sigma \times BR$ and (b) the ratio of this product over the SM Higgs production cross section as a function of different $M_{\widetilde{\chi}_1^0}$ values. The uncertainty on the expected limit at 1σ and 2σ levels are shown as green and yellow bands, respectively.

4.6 Higgs Exotic Decays to Photons and MET Projection Studies

4.6.1 Introduction

Exotic decays of the Higgs boson can produce the final state consisting of two photons and missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) in different topologies and theoretical benchmarks [69]. In the non-resonant case, the photons arise from opposite sides of the initial two-body decay: $h \rightarrow XX, X \rightarrow \gamma Y$, where Y is a stable neutral particle. For instance, such a decay can occur within general gauge mediation models of supersymmetry, in which the X corresponds

to a neutralino NLSP with mass less than half the Higgs mass, and the Y corresponds to a gravitino LSP [70, 71, 46]. In the resonant case, the photons are produced through an intermediate resonance: $h \to S_1 S_2$, with $S_1 \to \gamma\gamma$ on one side of the decay, while S_2 escapes detection, appearing as $E_{\mathrm{T}}^{\mathrm{miss}}$ in the detector. The resonant signal benefits from a peak in the diphoton invariant mass spectrum. The Feynman diagrams for the non-resonant and resonant decays can be seen in Figure 4.13.

Previous searches for the $\gamma\gamma + E_{\mathrm{T}}^{\mathrm{miss}}$ final state in the low energy regime include searches for the non-resonant decay in the supersymmetric scenario described above. CMS and ATLAS have set upper limits on the branching ratio of this decay, with the Higgs boson produced in association with a Z boson [72] and through vector boson fusion [73].

In this study, we devise a search strategy for the $\gamma\gamma + E_{\mathrm{T}}^{\mathrm{miss}}$ final state, motivated by the exotic decays of the Higgs described above. We estimate the sensitivity of this search for 100 fb^{-1} of $\sqrt{s} = 14$ TeV pp data from the LHC. This study was produced for the LHC Higgs Cross-Section Working Group's CERN Yellow Report 4, titled "Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector" [74].

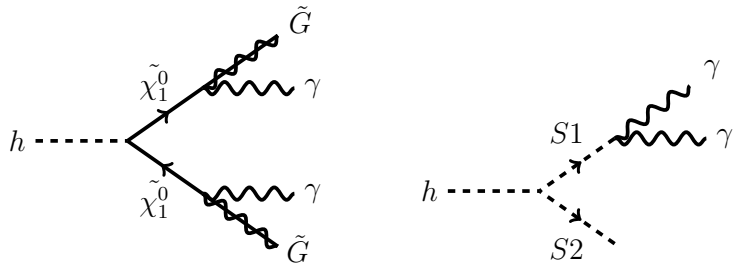

Figure 4.13: Feynman diagrams for the (left) the non-resonant and (right) resonant signal scenarios.

4.6.2 Methodology

4.6.2.1 Simulation Samples

Signal and background Monte Carlo (MC) samples were generated with Madgraph 5 [75] and hadronized with PYTHIA 8 [76], with the detector simulation provided by DELPHES 3 [77]. The samples were produced at $\sqrt{s} = 14$ TeV.

The object reconstruction and identification are performed with DELPHES, according to the information provided in the detector configuration card. For the photon reconstruction and identification, we assume an efficiency of 95% in the electromagnetic calorimeter barrel ($|\eta| < 1.5$) and 85% in the endcap ($1.5 < |\eta| < 2.5$). We also impose an isolation cut on the photons by requiring all particle flow candidates within a cone of $\Delta R < 0.3$ to have an energy ratio less than 0.1 with respect to the photon candidate. For muons, we assume an efficiency of 95% for the whole detector acceptance ($|\eta| < 2.5$). An isolation cut similar to the photons is also applied. Jets are reconstructed with the anti-k_t algorithm with $R = 0.4$, similar to the Run 2 analyses at CMS.

Signal MC Samples

The signal for the non-resonant case was based on the supersymmetric cascade decay of the Higgs boson into two neutralinos, which subsequently decay into two gravitinos and two photons (Figure 4.13, left). This class of models has been implemented in FeynRules [78] and generated via Madgraph 5. We assume a gravitino mass close to zero, which is consistent with gauge mediated low-scale SUSY breaking models with $\sqrt{f} \approx$ TeV [46]. We simulate neutralino masses in the range $[10, 60]$ GeV in steps of 5 GeV, with 100,000 events per mass point.

For the resonant case, we assume the Higgs boson decays into two scalar particles, S_1 and S_2 (Figure 4.13, right). One of the scalars then decays into two photons, while the other escapes detection. For this study, we assume the masses of these two particles are the same; this choice was made for simplicity, but for detailed studies, more combinations should be investigated. We generate samples with $M_1 = M_2 = [10, 60]$ GeV, in steps of 5 GeV, with 100,000 events per mass point.

We investigate the production of the Higgs boson through both gluon fusion and associated production with a Z boson (ZH), with the Z boson decaying to two muons. The inclusion of the dielectron decay of the Z can also be considered for future studies. A branching ratio of $B(H \to \gamma\gamma + E_{\mathrm{T}}^{\mathrm{miss}}) = 10\%$ is assumed for the signal. This value of the branching ratio was chosen to be within the current bounds on the Higgs boson width, yet close to the 8 TeV limits from the search for Higgs decays to the monophoton final state ($H \to \gamma + E_{\mathrm{T}}^{\mathrm{miss}}$) [72], when combining gluon fusion and ZH production mechanisms.

Background MC Samples

Although this analysis is not guaranteed to be entirely free from QCD multi-jet backgrounds, it has been shown in similar analyses (such as [72]) that it is possible to reduce the QCD to a sub-dominant contribution. As such, the remaining backgrounds for this analysis arise predominantly from the single boson ($\gamma/Z/W$) plus jets and diboson processes.

The background samples were modeled using the Snowmass LHE simulation samples [79]. They consist of single boson samples ($\gamma/Z/W$) with at least one jet and inclusive diboson ($\gamma\gamma/Z\gamma/W\gamma/\ WW/ZZ/WZ$) samples. No pure QCD sample was produced for this study. The samples include both hadronic and leptonic decays of the W and Z bosons. The cross sections used for normalizing the single boson samples were estimated with MCFM [59], assuming an efficiency of 15% for the 1 jet requirement (obtained with Madgraph). For the diboson samples, the cross sections used were estimated from the reference [80]. The cross sections and number of events in the samples are shown in Table 4.7.

4.6.2.2 Event Selection

Trigger Projections

For the ZH channel, the trigger strategy is expected to be straightforward and can be based on the decay of the Z to two muons. On the other hand, triggering is one of the main challenges for the gluon fusion channel, since the final state consists of two soft photons plus missing energy. The standard triggers used for $H \to \gamma\gamma$ analyses typically have a diphoton invariant mass cut which makes it incompatible with the low energy spectrum of this analysis. However, we have identified three possible trigger strategies for this channel, based on unprescaled triggers used by the CMS experiment in Run-2:

- Asymmetric Diphoton Trigger: This trigger requires two photons with different E_T and trigger-level identification requirements, plus a diphoton invariant mass cut. This type of trigger usually has a non-negligible turn-on curve in the leading and subleading photon E_T.
- Symmetric Diphoton Trigger: This trigger requires two photons with the same E_T requirement, without any other requirements.
- $\gamma + E_{\mathrm{T}}^{\mathrm{miss}}$ Trigger: This trigger requires only one barrel photon passing identification requirements and a E_T requirement that is usually higher than the previous two triggers. In addition, there is a calorimetric $E_{\mathrm{T}}^{\mathrm{miss}}$ requirement. We expect non-negligible turn-on curves with respect to both photon and $E_{\mathrm{T}}^{\mathrm{miss}}$ for this trigger.

The three triggers described here represent different selection strategies that were investigated and will be described below.

Offline Selection

The gluon fusion selection is based on the diphoton selection and must reflect the chosen trigger strategy, while maintaining a good signal efficiency. The ZH-produced signal events are tagged through the decay of the Z boson to muons, minimizing the largest backgrounds. The photon selection is chosen to maximize the signal acceptance in the ZH case, with E_T thresholds as low as possible. The final event selection requirements for the gluon fusion and ZH channels are summarized in Table 4.6. In this table, we use the following definitions for transverse mass:

$$M_T(\gamma\gamma, E_{\mathrm{T}}^{\mathrm{miss}}) = \sqrt{2E_T(\gamma\gamma)E_{\mathrm{T}}^{\mathrm{miss}}(1-\cos(\Delta\phi(\gamma\gamma, E_{\mathrm{T}}^{\mathrm{miss}}))}, \tag{4.18}$$

$$M_T(\gamma\gamma + E_{\mathrm{T}}^{\mathrm{miss}}, \mu\mu) = \sqrt{2E_T(\gamma\gamma + E_{\mathrm{T}}^{\mathrm{miss}})p_T(\mu\mu)(1-\cos(\Delta\phi(\gamma\gamma + E_{\mathrm{T}}^{\mathrm{miss}}, \mu\mu))}. \tag{4.19}$$

To exploit the topology of the resonant signature, we apply an additional requirement of a ± 10 GeV mass window, in the diphoton invariant mass distribution ($M(\gamma\gamma)$), around the signal mass (M_1). The efficiencies for each individual process and the different searches, after the full selection (without the $M(\gamma\gamma)$ mass window requirement), are shown in Table 4.7.

For the ZH case, we also explore the strategy performed by CMS in Run I [72], in which one or more photons are required in the event, instead of two or more. In this case, we gain back the efficiency that is lost due to the inefficiency in reconstructing the subleading photon, which can have very low E_T. The selection is similar to what is described in Table 4.6, but without the $M(\gamma\gamma)$ cut or the mass window requirement for the non resonant topology. The other variables that use the diphoton information are instead reconstructed with the leading photon in the event.

4.6.2.3 Background Estimation for Misidentified Photons

Background processes with mis-identified (or "fake") photons, such as jets and electrons misidentified as photons, that pass the final selection generally have very low efficiency at the LHC. Nonetheless, such backgrounds may be non-negligible since the production cross-subsections can be large. Such mis-identification rates are typically measured with data-driven

	gluon Fusion			ZH
Variable	Asymmetric $\gamma\gamma$	Symmetric $\gamma\gamma$	$\gamma + E_{\mathrm{T}}^{\mathrm{miss}}$	
Number of photons	> 1	> 1	> 1	> 1
$p_T(\gamma_1)$	> 45 GeV	> 40 GeV	> 55 GeV	> 20 GeV
$\lvert\eta(\gamma_1)\rvert$	< 2.5	< 2.5	< 1.4	< 2.5
$p_T(\gamma_2)$	> 30 GeV	> 40 GeV	> 20 GeV	> 20 GeV
$\lvert\eta(\gamma_2)\rvert$	< 2.5	< 2.5	< 2.5	< 2.5
$M(\gamma\gamma)$	$\in [15, 100]$ GeV	< 100 GeV	< 100 GeV	< 100 GeV
$E_{\mathrm{T}}^{\mathrm{miss}}$	> 90 GeV	> 90 GeV	> 90 GeV	> 60 GeV
$M_T(\gamma\gamma, E_{\mathrm{T}}^{\mathrm{miss}})$	< 140 GeV	< 140 GeV	< 140 GeV	< 140 GeV
$\Delta\phi(\gamma\gamma, E_{\mathrm{T}}^{\mathrm{miss}})$	< 1.5	< 1.5	< 1.5	< 1.5
Number of leptons	< 1	< 1	< 1	2 muons
$p_T(\mu_{1,2})$	-	-	-	> 20 GeV
$\lvert\eta(\mu_{1,2})\rvert$	-	-	-	< 2.5
$M(\mu\mu)$	-	-	-	$\in [75, 115]$ GeV
$M_T(\gamma\gamma + E_{\mathrm{T}}^{\mathrm{miss}}, \mu\mu)$	-	-	-	> 400 GeV

Table 4.6: Analysis selection for the gluon fusion channel (for each trigger scenario) and the ZH channel.

Process	σ (pb)	$N_{\mathrm{Generated}}$	gluon Fusion			ZH
			Asymmetric $\gamma\gamma$	Symmetric $\gamma\gamma$	$\gamma + E_{\mathrm{T}}^{\mathrm{miss}}$	
Backgrounds						
γ + Jets	1.0×10^5	5425448	1.9×10^{-6}	4.7×10^{-7}	8.9×10^{-7}	≈ 0
Z + Jets	0.94×10^4	1888446	5.6×10^{-4}	1.5×10^{-4}	5.0×10^{-5}	≈ 0
W + Jets	2.96×10^4	5263872	6.2×10^{-4}	1.9×10^{-4}	2.7×10^{-5}	≈ 0
$\gamma\gamma$	10.8×10^1	4268781	3.1×10^{-5}	1.0×10^{-5}	1.1×10^{-5}	≈ 0
$Z\gamma$	6.30×10^2	3406151	4.3×10^{-4}	1.4×10^{-4}	5.7×10^{-5}	≈ 0
$W\gamma$	1.03×10^3	5258034	1.4×10^{-4}	4.6×10^{-5}	5.4×10^{-5}	≈ 0
WW	1.24×10^2	8059829	2.6×10^{-1}	8.4×10^{-2}	9.8×10^{-5}	8.2×10^{-8}
ZZ	1.8×10^1	1101611	1.4×10^{-2}	4.7×10^{-3}	6.7×10^{-4}	7.3×10^{-6}
WZ	5.1×10^1	3319770	3.6×10^{-1}	1.2×10^{-1}	2.5×10^{-4}	2.9×10^{-6}

Table 4.7: Cross subsections, numbers of events generated per process, and selection efficiencies for background processes, for gluon fusion and ZH production mechanisms.

methods at the LHC. Although this study was limited by MC statistics in measuring fake photon backgrounds, a method was developed to mitigate this problem, which we describe below.

The object reconstruction and selection is done at DELPHES level, where, given the photon identification requirements described in subsection 4.6.2.1, we obtain an associated fake rate. These fake rates are accounted for in the overall efficiencies in Table 4.7. In order to bypass the efficiency loss due to the small fake rates, we select jets and electrons as fake photon candidates. For the background processes with one prompt photon (γ+jets, $W\gamma$ and $Z\gamma$), we select one fake photon candidate. For the processes with no prompt photons (W/Z+jets, WW, WZ and ZZ), we select two fake photon candidates. No fake photon selection is done for the $\gamma\gamma$+jets sample.

With the assumption of a flat fake rate for both jets and electrons, the fake photon candidates are randomly selected from the jets and electrons that passed the photon acceptance requirements. One extra assumption is that the electron-to-photon fake rate is set to be order of magnitude larger than the jets-to-photon fake rate. Therefore, electrons are set to have a probability of being selected to be a misidentified photon that is ten times higher than for jets.

After the choice of fake photon candidates, we calculate weights for the individual samples based on the E_T spectrum of the selected photons (prompt and fake) to match the spectrum found with the photon candidates reconstructed directly from DELPHES. This reweighting is done on the sum of E_T of the two leading photons for samples with at least one prompt photon, and on the E_T of the leading photon for samples with no prompt photon. An independent reweighting is also done in η. Both reweightings reflect the different reconstruction efficiencies and energy resolutions of objects that are not reconstructed as photons. After applying the weights, we observe a good agreement between the kinematic distributions of interest arising from photons reconstructed by DELPHES and from our fake photon candidates.

4.6.3 Results

We present the expected sensitivity of this search in terms of the necessary $h \to \gamma\gamma + E_{\mathrm{T}}^{\mathrm{miss}}$ branching ratio to reach a 5σ sensitivity for an assumed integrated luminosity of 100 fb^{-1} at $\sqrt{s} = 14$ TeV, with the frequentist sensitivity defined as:

$$\mathcal{S} = \frac{N_{\mathrm{Signal}}}{\sqrt{N_{\mathrm{Background}}}}. \tag{4.20}$$

In Figure 4.14 (left), we shown the signal efficiency for the different trigger selections presented and (right) the sensitivity plot for the gluon fusion case. This plot shows that, after the full selection, the performance of the triggers is comparable. Although it's safe to assume that a diphoton trigger with a low $M(\gamma\gamma)$ cut will be present in the future trigger menus of CMS and ATLAS, we choose to perform the analysis in the $\gamma + E_{\mathrm{T}}^{\mathrm{miss}}$ case. We make this choice as an effort to make the case for the existence of such a trigger strategy for the future LHC runs. While the diphoton triggers are designed with specific usages that are already well established, the $h \to \gamma\gamma + E_{\mathrm{T}}^{\mathrm{miss}}$ analysis could be viewed as a benchmark for the $\gamma + E_{\mathrm{T}}^{\mathrm{miss}}$ trigger for three reasons:

- It's a trigger that is already present at the LHC experiments, but can be retuned with a specific analysis as benchmark;
- A dedicated trigger for this analysis requiring two photons might not be as efficient at trigger level, given the soft spectrum of the second photon;
- This trigger can also be used for other exotic searches, such as the extension to low energies of the dark matter searches in the monophoton channel.

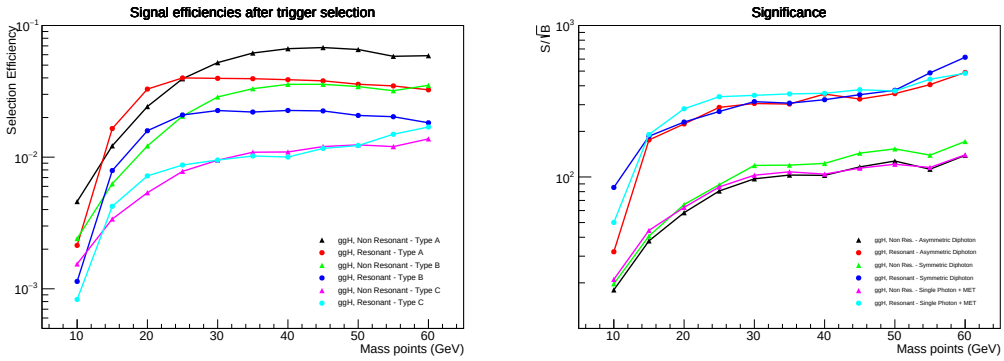

Figure 4.14: Signal efficiency (left) and significance (right) plots for different trigger scenarios in the gluon fusion analysis.

In Figure 4.15, on the left, we show the branching ratio of $h \to \gamma\gamma + E_{\mathrm{T}}^{\mathrm{miss}}$ needed for a significance of 5σ, assuming the Standard Model Higgs cross section, for the gluon fusion analysis (assuming the $\gamma + E_{\mathrm{T}}^{\mathrm{miss}}$ trigger strategy and selection). On the right, we show the branching ratio $h \to \gamma\gamma + E_{\mathrm{T}}^{\mathrm{miss}}$ needed for a significance of 2σ, which represents the 95% confidence level for exclusion, assuming SM ZH production. For the ZH case, we show the results for the strategies requiring at least one ($N(\gamma) \geq 1$) and at least two ($N(\gamma) \geq 2$) photons.

4.6.3.1 Systematic Uncertainties

While the uncertainties in the ZH channel is expected to be dominated by statistics, the gluon fusion channel is very sensitive to the systematic uncertainties associated with the background predictions. We estimate the effect of these uncertainties by parametrizing the sensitivity as:

$$\mathcal{S}_{sys} = \frac{N_{\mathrm{Signal}}}{\sqrt{N_{\mathrm{Background}}} + \sigma_{sys} \times N_{\mathrm{Background}}}, \tag{4.21}$$

with σ_{sys} representing a source of uncertainty that does not scale with the amount of statistics. Figure 4.15 shows the effect on the 5σ branching ratios due to the addition of a 10% systematic uncertainty according to Eqn 4.21.

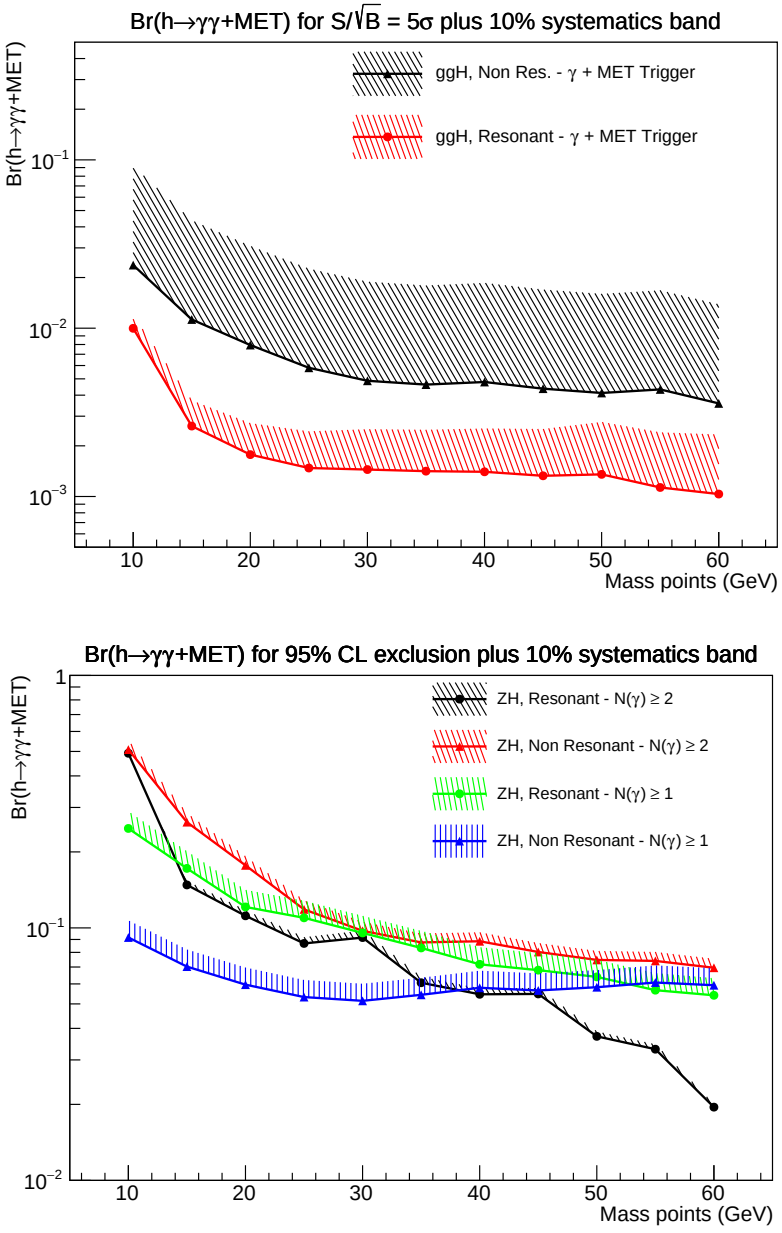

Figure 4.15: (Left) 5σ branching ratios for the gluon fusion channel, for resonant and non-resonant final states, using the $\gamma + E_{\mathrm{T}}^{\mathrm{miss}}$ trigger. (Right) Branching ratios for 95% confidence level exclusion in the ZH case, resonant and non-resonant topologies, requiring at least one photon ($N(\gamma) \geq 1$) and at least two photons ($N(\gamma) \geq 2$). The shaded areas correspond to a variation in systematics up to 10%.

Chapter 5

HH searches with photons and b-jets at CMS Run 2

5.1 Introduction

This chapter describes the inclusive search for the double Higgs production process in the decay mode HH $\to b\bar{b}\gamma\gamma$ at $\sqrt{s} = 13\,\text{TeV}$, with 35.87 fb^{-1} from the 2016 data taking period. This analysis is based on the search at 8 TeV performed in the same final state, recently published by CMS in Ref. [81] (the internal CMS documentation of this analysis can be found in Refs. [82, 83]), and on the earlier 13 TeV search using 2015 data [84].

Many theories beyond the SM (BSM) suggest the existence of new physics potentially manifested in the detection of a pair of Higgs bosons. The simplest signal we can look for is in the form of a resonant contribution to the invariant mass of the HH system (resonant search). If the new particles are not directly detectable (either too heavy or too light to be in the HH invariant mass spectrum), but still couple to HH, their virtual contribution to the non-resonant HH production can still be measured (as shown, e.g., in Refs. [85, 86]). Additionally, the fundamental couplings of the Higgs boson to other SM particles (including itself) can be modified as well in BSM theories (as shown, e.g., in Refs. [87, 88]). Both of these cases can be studied via the non-resonant HH process.

In order to study the HH production, the Higgs bosons' final states must be carefully chosen. On one hand, the overall cross section of the process must be kept high enough for good sensitivity. On the other hand, a good selection efficiency, online and offline, is important for a well performing analysis. This is achieved with the $b\bar{b}\gamma\gamma$ final state. The H $\to b\bar{b}$ leg provides a high branching ratio (57.7%), while the H $\to \gamma\gamma$ leg provides an efficient way to correctly identify the interesting events and have a high mass resolution. The total branching ratio of the HH $\to b\bar{b}\gamma\gamma$ channel is 0.26%.

The resonant analysis is dedicated to the search of a generic narrow width resonance (both spin-0 and spin-2). In this note, we will use a Warped Extra Dimensions (WED) theory (based on the Randall-Sundrum (RS) setup [89]) as a benchmark model for the resonant HH search. It provides candidates for both spin-0 (radion) and spin-2 (graviton[1]) resonances that decay

[1]The graviton can be interpreted either as the first Kaluza-Klein (KK) excitation, or the graviton in the bulk RS scenario [90, 91, 92]

into HH.

In the non-resonant search, we set limits in the Standard Model-like HH production, a process that has a cross section of $\sigma(pp \to HH)^{\rm SM}_{\rm NNLO} = 33.45$ fb at 13 TeV [93]. We also investigate explicitly the case of anomalous couplings in the Higgs boson potential, following the same model parametrization used in Ref. [81]. In the 13 TeV analysis however we study the parameter space of anomalous couplings using the approach suggested in [94], where physics benchmarks are defined based on basic signal kinematics.

5.1.1 Strategy Summary

The main strategy upgrades with respect to the Run-I analysis are:

- We make use of a multivariate analysis-based photon identification criterium, which improves the selection efficiency;
- Using $\tilde{M}_{\rm X}$ variable instead of 4-body mass, $M(jj\gamma\gamma)$ - this cancels the effects of the low dijet mass resolution (compared to $M(\gamma\gamma)$) uncertainties in the jet energy scale;
- An improved version of the Combined Secondary Vertex algorithm (CSVv2) for b-tagging was developed in CMS;
- A new categorization method was developed to deal with signal categories and phase spaces with not enough events for a reliable background description;
- A dedicated b-jet energy regression was developed for Run-II analysis;
- The main part of the analysis is performed in a framework based on the one developed by the H $\to \gamma\gamma$ group, so we benefit from the up-to-date photon selection tools available.

In the new version of the analysis we also profit from a better description of the signal in the MC samples (Section 5.2), and a larger set of MC events. The description of the simulated background is also improved. We observe a good agreement in the shapes of the basic distributions between data and MC in all control regions (Section 5.7).

There are however a few new challenges in the Run 2 analysis. We utilize the double-photon trigger to select event for the analysis. Compared to the 8 TeV data-taking period, the E_T thresholds of the L1 trigger requirements were increased, which reduces the selection efficiency of the signal. A smaller distance parameter in the jet clustering algorithm is used by CMS ($D = 0.4$ in Run-II vs $D = 0.5$ in Run-I), which introduces a larger bias and decreased resolution in the reconstruction of the $M(jj)$ variable. Another challenge of the CMS running conditions in Run-II is higher pile-up environment, especially during the 2016 data taking.

With respect to the 2015 version of the analysis, many improvements have been implemented focusing on maximizing S/B, given the larger amount of data available. New categorization schemes have been developed and the mass window selection has been re-optimized. A new training for the b jet energy regression has also been developed, with a better performance with respect to jet energy scale and resolution.

5.2 Samples

All the MC samples used were processed centrally. As default for CMS, the signal samples use PDF4LHC15_nlo_mc_pdfas set [95, 96, 97, 98, 99] in the four flavour scheme. The central value for the strong coupling is taken as $\alpha_s(m_Z) = 0.118$. The samples used in this analysis have been processed by the standard CMS chain, including a detailed detector simulation with GEANT4. These samples have also been pre-processed by the central H $\to \gamma\gamma$ analysis framework FLASHgg, in order to obtain the most up-to-date photon energy, resolution and identification algorithms used by the $H \to \gamma\gamma$ analysis.

5.2.1 Signal MC: resonant production

To simulate the generic resonances we use `MG5_aMC@NLO` [100] at leading order. For the gluon fusion produced spin-2 resonance, we use the model for a KK-graviton in the bulk described in [101], which is an adaptation of the RS1 model of Ref. [102], introducing the relevant coupling modifications. The model files can be found in [103] and it is found to agree at the level of cross sections and branching ratios with the bulk WED scenario implemented by the authors of [104] in the `CalcHep` [105] framework. To simulate the scalar resonance, we use the Higgs Effective Model [106] that can found in the `Feynrules` database [107].

The spin-0 and spin-2 resonances are simulated with masses: 250, 260, 270, 280, 300, 320, 340, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 900, 1000 GeV, with 50k events each, assuming resonance width of 1 GeV. The cross sections for the interpretations can be found here [108]. The `MG5_aMC@NLO` configuration cards used for the simulations are organized here [109].

5.2.2 Signal MC: nonresonant production

In the SM, Higgs boson pair production occurs predominantly by gluon-gluon fusion (GF) via an internal fermion loop. Since the Higgs boson couplings are defined by the particles masses, the top quark contribution is dominant, while couplings to light quarks are negligible [2]. In the absence of new light states, the GF Higgs boson pair production at the LHC can then be generally described (to leading approximation) by five parameters controlling the tree-level interactions of the Higgs boson. The Higgs boson trilinear coupling and the top Yukawa interaction exist in the SM Lagrangian, where the former is given by $\lambda_{SM} = m_h^2/2v^2$, with v the vacuum-expectation value of the Higgs field. Deviations from SM values are parametrized with the multiplicative factors κ_λ and κ_t, respectively. The contact interactions of the Higgs boson with gluons and those coupling two Higgs bosons with two gluons or a top-antitop quark pair, which could arise through the mediation of very heavy new states, are instead genuinely not predicted by the SM; they can be parametrized by the absolute couplings c_g, c_{2g}, and c_2. The relevant part of the Lagrangian then takes the form

$$\mathcal{L}_h = \frac{1}{2}\partial_\mu h \partial^\mu h - \frac{1}{2}m_h^2 h^2 - \kappa_\lambda \lambda_{SM} v\, h^3 - \frac{m_t}{v}(v + \kappa_t\, h + \frac{c_2}{v}\, h\, h)\,(\bar{t_L} t_R + h.c.) + \frac{1}{4}\frac{\alpha_s}{3\pi v}(c_g\, h - \frac{c_{2g}}{2v}\, h\, h)\, G^{\mu\nu} G_{\mu\nu}\,. \quad (5.1)$$

[2] This assumption is motivated also in BSM theories where the Higgs sector is minimal (see also [110])

The different Feynman diagrams contributing to a di-Higgs boson signal in pp collisions at leading order (LO) are shown in Fig. 5.1. The simulation setup used in this paper was produced by the authors of [111]. We use `MG5_aMC@NLO` as generator. The LO process is already at one-loop level; in the approach followed in [111], loop factors are calculated on an event-by-event basis with a `Fortran` routine on top of an $aMC@NLO$ [112, 100] effective model;

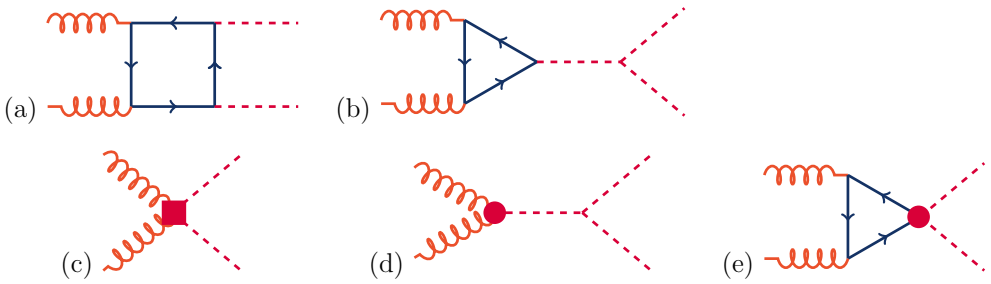

Figure 5.1: Feynman diagrams that contribute to Higgs boson pair production by gluon-gluon fusion at leading order. Diagrams (a) and (b) correspond to SM-like processes, while diagrams (c), (d), and (e) correspond to pure BSM effects: (c) and (d) describe contact interactions between the Higgs boson and gluons, and (e) exploits the contact interaction of two Higgs bosons with top quarks.

In the Ref. [94], it was designed a method to partition the 5 dimensional parameter space in regions with similar LO kinematics. When the simulation was launched in the CMS system this work was in a preliminary version, we had simulated the samples using to the recommendations of the first version. The list of relevant parameters used in each sample is in Table 5.2. The `MG5_aMC@NLO` configuration cards used for the simulations are organized here [109]. Since then, the metric used by the method was improved and the parameter space scan extended, resulting in a new set of benchmarks, that can be found in the last ArXiV version of the mentioned paper, and as well in [93]. The analytical formula that allows us to calculate cross sections for any point of the parameter space can be found in [113], and a handful script that calculates the cross sections point-by-point can be found in [114].

5.2.2.1 Reweighting

We are considering a $2 \to 2$ process at leading order. The two Higgs bosons are produced with identical transverse momenta (p_T^H), and they are back-to-back in azimuth at this order (before a parton shower). The final state can then be completely defined by three kinematic variables, if we ignore the irrelevant azimuthal angle of emission of the bosons. Furthermore, one of the three remaining variables can be used to isolate all the information related to the PDF of the colliding partons, which is also irrelevant to the physics of the production process once one focuses on a specific initial state (the gluon-gluon fusion process). The variable factorizing

Node	κ_λ	κ_t	c_2	c_g	c_{2g}
1	1.0	1.0	0.0	0.0	0.0
2	7.5	2.5	-0.5	0.0	0.0
3	15.0	1.5	-3.0	-0.0816	0.3010
4	5.0	2.25	3.0	0.0	0.0
5	10.0	1.5	-1.0	-0.0956	0.1240
6	1.0	0.5	4.0	-1.0	-0.3780
7	2.4	1.25	2.0	-0.2560	-0.1480
8	7.5	2.0	0.5	0.0	0.0
9	10.0	2.25	2.0	-0.2130	-0.0893
10	15.0	0.5	1.0	-0.0743	-0.0668
11	-15.0	2.0	6.0	-0.1680	-0.5180
12	2.4	2.25	2.0	-0.0616	-0.1200
13	-15.0	1.25	6.0	-0.0467	-0.5150

Table 5.1: Parameter values of the final benchmarks selected with $N_{clus} = 13$. The first cluster is the one that contains the SM sample.

out the PDF modeling can be taken as the magnitude of the boost of the center-of-mass frame as seen in the laboratory frame.

The two remaining variables, which provide direct information on the physics of GF HH production, can be chosen to be the invariant mass of the HH system (m_{HH}) and the modulus of the cosine of the polar angle of one Higgs boson with respect to the beam axis ($|cos\theta^*|$). Since we are using parton-level information, this last variable is equivalent to the polar angle in the Collins-Soper frame ($|cos\theta^*_{CS}|$) [115], which is commonly used in experimental analysis. The variables m_{HH} and $|cos\theta^*|$ can thus be used to fully characterize the final state kinematics produced by different choices of the value of anomalous Higgs boson coupling parameters.

By construction the full-simulated samples listed in Tab. 5.2 are good representatives of the kinematic space. Therefore, based in the generation level m_{HH} and $|cos\theta^*|$, those can be used to construct samples to any other parameter space point.

The procedure is made as follows:

- For each new parameter space point we simulate N events in `MG5_aMC@NLO`;
- We construct two dimensional histograms at generator-level of m_{HH} and $cos\theta^*$, with 20 GeV-wide bins in the m_{HH} and 0.2-wide bins in $cos\theta^*$ (without the moduli);
- We construct the same histogram with the sum of all the full simulated samples described in the last section (signal dataset);
- The new sample is constructed by weigthing the signal dataset event-by-event by:

$$W_e = \frac{New_{ij}}{D_{ij}}\,, \tag{5.2}$$

  where (ij) specify the bin in which the event e belongs and New_{ij} (D_{ij}) the number of events of the new signal sample (signal dataset) in that bin.

We produce reweighted samples for three types of theory scans. First, a plain scan in κ_λ, while all the other parameters are kept as in SM (κ_t, $c_2 = c_g = c_{2g} = 0$), where the gen-level histograms are made from 50,000 events. Then, a scan in the 2D plane defined by fixing $c_2 = c_g = c_{2g} = 0$, with varying κ_λ and κ_t.

Benchmark	κ_λ	κ_t	c_2	c_g	c_{2g}
1	7.5	1.0	-1.0	0.0	0.0
2	1.0	1.0	0.5	-0.8	0.6
3	1.0	1.0	-1.5	0.0	-0.8
4	-3.5	1.5	-3.0	0.0	0.0
5	1.0	1.0	0.0	0.8	-1
6	2.4	1.0	0.0	0.2	-0.2
7	5.0	1.0	0.0	0.2	-0.2
8	15.0	1.0	0.0	-1	1
9	1.0	1.0	1.0	-0.6	0.6
10	10.0	1.5	-1.0	0.0	0.0
11	2.4	1.0	0.0	1	-1
12	15.0	1.0	1.0	0.0	0.0
SM	1.0	1.0	0.0	0.0	0.0

Table 5.2: Parameter values of the twelve benchmarks and the Standard Model point.

5.2.3 Background MC

Even though the signal extraction and background modeling of this analysis are performed in a data-driven way, background simulated samples are used to validate the simulation of our signal description and to optimize the analysis. The main backgrounds for the $b\bar{b}\gamma\gamma$ final state come from either well-isolated photons coming from the hard scatter (prompt photons) or isolated photons reconstructed due to very collimated $\pi^0 \to \gamma\gamma$ decays fragmented from jets (fake photons). These backgrounds are the same as the ones to the SM $H \to \gamma\gamma$ analysis.

We can classify these backgrounds with respect to the number of prompt and fake photons in the selected diphoton candidate. The prompt-prompt background events are simulated with the Sherpa generator; it includes the Born processes with up to 3 additional jets at LO accuracy as well as the box processes at LO. The prompt-fake and the fake-fake contributions are simulated with PYTHIA8, with a "Double-EM Enriched" filter[3] applied during production to improve the samples selection efficiencies. Additionally, a sample of Drell-Yan events decaying into electrons, simulated at NLO with `MG5_aMC@NLO`, is used, although its final contribution to the event yield is minimal due to the $M(\gamma\gamma) > 100\,\text{GeV}$ selection cut.

5.2.4 Data

The data samples used in this analysis correspond to approximately 35.87 fb^{-1} of data collected in 2016.

[3]it requires a high p_T and well isolated photon-like signal (electromagnetic activity) coming from photons, electrons, or neutral hadrons)

5.3 Analysis objects and selection

This analysis uses general purpose reconstruction of photons and jets, which have been described in previous sections. We have brief descriptions below, pointing out specific choices made for the analysis.

5.3.1 Triggers And Pre-Selection

Exploiting the high online performance of the CMS ECAL to reconstruct photons and electrons, the dataset used in this analysis is constructed with a selection that requires two photons at High Level Trigger (HLT) level.

For the 2016 data taking period, the online strategy was based on a single HLT trigger path, based on selecting two photons, one with $E_T > 30$ GeV (leading) and one with $E_T > 18$ GeV (subleading). Quality criteria are required on these photons, such as on calorimeter isolation (the amount of calorimeter activity around the reconstructed photon), on the ratio between the energies deposited on ECAL and HCAL, and on R9 (the ratio between the energy deposited on a 3x3 ECAL crystal matrix around the most energetic crystal in the supercluster, and the supercluster energy). Additionally, it is required, at trigger level, that $M(\gamma\gamma) > 90$ GeV.

In order to achieve good data/simulation comparison, a pre-selection that is tighter than the online selection is applied on data and Monte Carlo. This pre-selection is described in Table 5.3. It is based on shower shape variables ($R9$), isolation variables (charged hadron isolation, CHI, the sum of all charged hadron particle flow candidates energies inside a cone of $\Delta R < 0.3$ around the photon axis), identification variables (H/E, the ratio between the photon's energy deposit in HCAL and in ECAL), and kinematic variables (E_T and the photon supercluster η). Only events that have at least one diphoton candidate passing the pre-selection requirements are considered in the analysis.

Requirements	Leading Photon	Subleading Photon
E_T	30 GeV	20 GeV
$\lvert\eta\rvert$	< 2.5 and outside $1.442 \lvert\eta\rvert < 1.566$	
Shower shape and Isolation	$R9 > 0.8$ or CHI < 20 or CHI$/E_T < 0.3$	
Identification	H/E < 0.08	

Table 5.3: Trigger based pre-selection applied on diphoton candidates.

The SM H $\to \gamma\gamma$ analysis provides scale factors and uncertainties related to this selection, which we also apply in the analysis and include in our systematical uncertainties.

5.3.2 Photons

The kinematic requirements applied on the photons, after the diphoton candidates have passed through the event pre-selection (see Section 5.3.1), are similar to the ones used in the SM H $\to \gamma\gamma$ analysis. The selection is as follows:

- Leading photon $E_T > 30$ GeV, trailing photon $E_T > 20$ GeV;
- Leading photon $E_T/M(\gamma\gamma) > 1/3$, trailing photon $E_T/M(\gamma\gamma) > 1/4$;

- $100 < M(\gamma\gamma) < 180\,\text{GeV}$.

Additionally, a photon identification requirement is applied to the photons.

The photon identification requirement is based on a multivariate algorithm that combines information from the photon cluster shape in ECAL, as well as isolation variables. This requirement provides a signal-like photon efficiency of 90% for photons both on the ECAL barrel and endcaps. The scale factors used to ensure data/MC agreement in the selection efficiency are also applied; these scale factors are calculated centrally at CMS. Additionally, an electron veto is applied to avoid background with electrons faking photons, with corresponding scale factors and uncertainties.

5.3.2.1 Fake Photon Control Region

A control region is created by selecting diphoton candidates with one photon that passed the ID requirement and one that didn't. All the other selections remain the same, and the procedure to select such diphoton candidates is the same as in the signal region. This control region is used in the analysis to perform closure tests on the background modeling.

5.3.2.2 Vertex

Inheriting from the main $H \to \gamma\gamma$ analysis, we use the vertex that gives the highest $H \to \gamma\gamma$ vertex MVA score. Because there are additional jets in the event, picking this vertex has a very small mismatch efficiency. Only in less than 0.1% of the events, the chosen vertex is different from the vertex associated to the true vertex in the simulation.

The $H \to \gamma\gamma$ vertex MVA score is based on three variables calculated for each reconstructed vertex: the sum of the squared transverse momenta of the charged particle tracks associated with the vertex, and two variables that quantify the vector and scalar balance of p_T between the diphoton system and the charged particle tracks associated with the vertex. In addition, if either photon is associated with any charged particle track that has been identified as resulting from conversion, the conversion information is also used. The variables are used as the inputs to a multivariate classifier based on a boosted decision tree (BDT) to choose the reconstructed vertex to be associated with the diphoton system. The average vertex finding efficiency of this algorithm in the SM $H \to \gamma\gamma$ analysis is 80%.

5.3.2.3 Gain Switch

In 2016, it was observed that high energy deposits in ECAL had their energies reconstructed with a certain bias, due to the shaping of the pulses in the ECAL VFE electronics. This shaping became a problem in Run 2 because we assume a certain shape for the signal amplitude from the crystals when measuring this amplitude, as described in the Multifit method. It has been observed that this bias occurs when an energy deposit is high enough to cause a gain switch in the ECAL multi-gain amplifier.

We investigated the fraction of selected events in our blinded signal region with photons that go through gain switches. The plots on Figure 5.2 show, in bins of $\tilde{M}_X$, the fraction of events with at least one of the photon candidates going through gain switches (to gain 1, gain 6 or both). These results show that, for the high mass region, around 20% of our events are

affected by gain switches. This non-negligible rate means that the analysis needs to use the new CMS re-processed data in which the issue has been fixed.

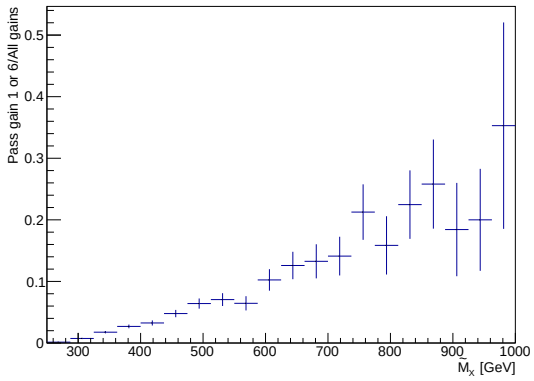

Figure 5.2: Fraction of events with photon candidates going through gain switch.

5.3.2.4 Regression

A new version of the photon energy regression has been trained with the Run 2 data taking conditions. This regression aims to correct the photon super cluster energy, as described in previous sections. We compared this new regression to the previous training, as seen in the plots of Figure 5.3 for three different resonance mass points. The difference observed is not large enough for this analysis to be affected, so the previous regression version is used (following main H $\rightarrow \gamma\gamma$ analysis).

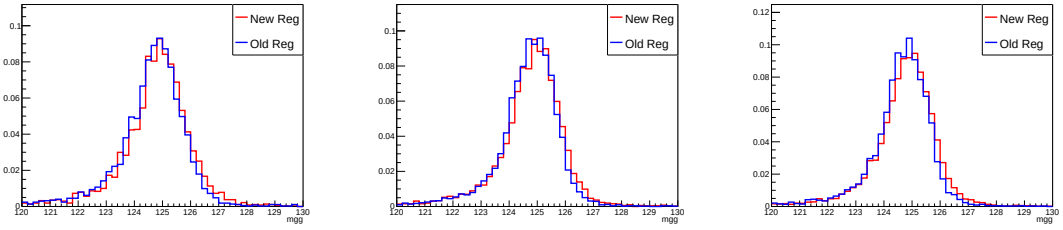

Figure 5.3: $M(\gamma\gamma)$ reconstructed with different versions of the photon energy regression for resonance masses of 300, 600 and 900 GeV, from left to right. Distributions normalized to unity.

5.3.3 Jets

The jets in Run 2 at CMS are reconstructed using anti-k_t algorithm with the distance parameter $R = 0.4$. This was a change with respect to the parameter distance of $R = 0.5$ in Run 1 due to increased pile-up in Run 2 data taking. This change results in smaller $M(jj)$ resolution but induces a bias towards lower energy of the signal $M(jj)$ peak, because less

energy is clustered in a jet. This effect can be seen in Figure 5.4 of the $M(jj)$ distribution for reconstructed jets matched to generator-level jets (that come from the Higgs) from the radion sample of $M = 300\,\text{GeV}$.

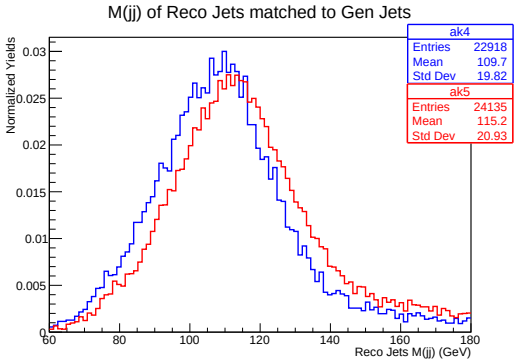

Figure 5.4: Difference between jet reconstruction used in Run I (red) and Run II (blue)

We use the *Loose ID* criteria to select the jets, which is described in Ref [116]. The jet candidates in the event, after passing the aforementioned ID, must have $p_{\text{T}} > 25\,\text{GeV}$ and $|\eta| < 2.4$ (so that they are within the tracker of CMS and can be tagged as coming from b quark). The jets must also be outside the photon cone with a $\Delta R(j,\gamma) > 0.4$. Dijet objects are then created and only dijets with $M(jj)$ between 60 and 180 GeV pass the selection. If more than one dijet has passed those criteria, the dijet with two jets with the highest b-tagging scores (see Sec. 5.3.4.1) are selected to form the dijet candidate.

5.3.4 Jet energy regression

In addition to the misfortune of a small distance parameter of the jet reconstruction algorithm, the energy of the jets coming from b-quarks cannot be fully reconstructed due to neutrinos (from the leptonic B meson decay) escaping the detector. In order to improve the $M(jj)$ resolution and S/B discrimination, we employ an energy regression technique on our signal jets. The regression technique will work by changing the jet 4-momentum based on the likelihood that this specific jet is a signal jet.

We use the TMVA package to implement the regression, and train it on $X \to \text{HH} \to b\bar{b}b\bar{b}$ MC samples, in order to ensure a statistically independent training. Input variables to the training include jet kinematics, energy deposited in the calorimeter, vertex information, and variables related to missing transverse energy of the event (MET) and the distance between the two jets, $\Delta R(j,j)$. A summary of all input variables is shown in Table 5.4.

The target for the regression training is $p_{\text{T}}^{gen}/p_{\text{T}}^{reco}$, where the generator-level jet (jet reconstructed with generator-level particles, "gen") contains neutrinos, and "reco" references the standard reconstructed jet. This means that the regression technique will aim to construct, in a piece-wise manner, a function $f(\bar{x}) = \langle p_{\text{T}}^{gen}/p_{\text{T}}^{reco}\rangle(\bar{x})$, where $\bar{x}$ are the regression input variables, which can be used to correct the jet's energy. The standard generator-level jet (gen-jet) in CMS does not include the neutrinos, so we add them manually from gen-particle collection, using ΔR cone of 0.4 for matching.

Input variables as in H $\to b\bar{b}$ analysis regression:	Jet kinematics: η, p_{T}, m_T Neutral hadron energy fraction, Photon energy fraction SecVtxdL, SecVtxdeL, SecVtxPt, SecVtxM, SecVtxNtrk Soft Lepton: p_{T}, p_{T} (rel), ΔR p_{T} (Lead Track), Number of verteces
Additional variables for our analysis:	$E_{\mathrm{T}}^{\mathrm{miss}}$, $\Delta\phi(Jet, E_{\mathrm{T}}^{\mathrm{miss}})$, ΔR(Leading jet, Trailing jet)

Table 5.4: Input variables used in TMVA regression. Upper part lists the variables that are also used in H $\to b\bar{b}$ analysis, and the lower part lists additional variables.

Adding neutrinos to the gen-jet brings the energy of the jet closer to the energy of original b-quark, which is illustrated in Fig. 5.5. Figure 5.5 also shows additional distributions, also comparing the gen-jets with and without neutrino additions: p_{T} of the leading and trailing jets, invariant mass of the Higgs boson candidates and the m_{jjjj}. Events from all mass samples are combined, which explains the shape of m_{jjjj} distribution. From these figures one can see the effects on the mass resolution of adding neutrinos to gen-jets.

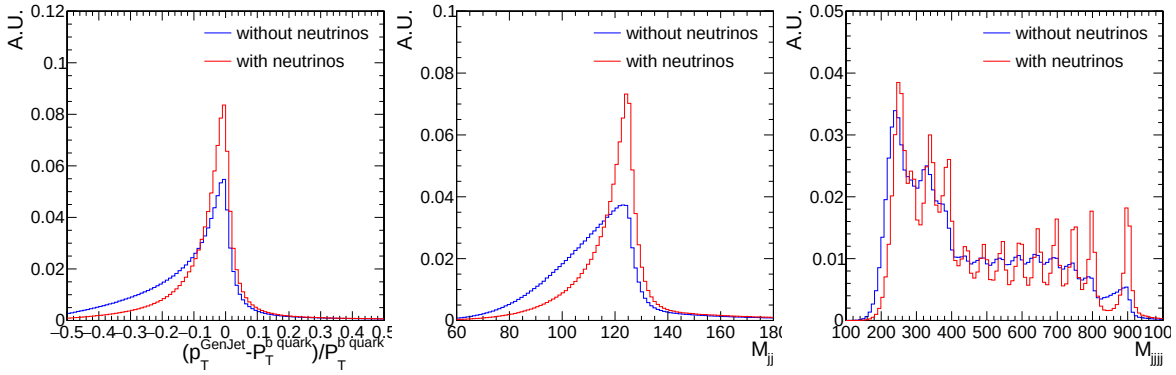

Figure 5.5: Relative p_{T} difference, m_{jj} and m_{jjjj} distributions of the b-quark and the corresponding gen-jet, obtained from HH $\to b\bar{b}b\bar{b}$ samples. Red histogram is for gen-jets containing neutrinos, blue is for jets without neutrinos.

For the training we select jets that satisfy the following criteria:

- $p_{\mathrm{T}} > 20\,\mathrm{GeV}$, $|\eta| < 2.4$
- Matched to the generated level jet within a cone $\Delta R < 0.4$ (this matching is done as part of MiniAOD reconstruction)
- Matched to a b-quark within a cone $\Delta R < 0.4$

We perform six different trainings to check the impact of additional variables:

- Using 15 variables based on H $\to b\bar{b}$ training, listed in Table 5.4. It is denoted as **Baseline** on the figures below.

- Using 15 variables plus $E_{\mathrm{T}}^{\mathrm{miss}}$ and $\Delta\phi(Jet, E_{\mathrm{T}}^{\mathrm{miss}})$.
- Using 15 variables plus $E_{\mathrm{T}}^{\mathrm{miss}}$, $\Delta\phi(Jet, E_{\mathrm{T}}^{\mathrm{miss}})$, and also ΔR(Leading jet, Trailing jet). This training is denoted as **full 15+3var** in the text.
- Using 15 variables as above but in addition, for each pair of jets from the Higgs boson, the training is performed separately for the leading and trailing jets. That is, two XML weight files are derived, one for the leading and one for the trailing jet in the event. This method is denoted as **js** on the figures and in the text.
- Using 15 plus $E_{\mathrm{T}}^{\mathrm{miss}}$ and $\Delta\phi(Jet, E_{\mathrm{T}}^{\mathrm{miss}})$ variables, and separating the training for leading and trailing jets as above (**15+2var js**).
- Using 15 plus $E_{\mathrm{T}}^{\mathrm{miss}}$, $\Delta\phi(Jet, E_{\mathrm{T}}^{\mathrm{miss}})$ and ΔR variables, and separating the training for leading and trailing jets as above (**15+3var js**).

The regression is performed with the TMVA package, using the Boosted Decision Tree technique with gradient boosting. 500 decision trees with depth 5 are created to estimate the target. The pruning technique is applied. The detailed TMVA modified options are as follows:

```
NTrees=500:MaxDepth=5:nCuts=500:BoostType=Grad:!UseBaggedGrad
Shrinkage=0.1:MinNodeSize=1:PruneStrength=5:PruneMethod=CostComplexity
NegWeightTreatment=IgnoreNegWeightsInTraining
```

After the training is done, its performance is checked in signal samples, $X \to \mathrm{HH} \to b\bar{b}\gamma\gamma$, at all mass points of m_X. The selection is the same as the training samples. All of our trainings are compared with the one done by $\mathrm{H} \to b\bar{b}$ analysis, which is denoted as **Hbb** on the figures. The p_{T}^{reco} of the reconstructed jet can be compared to the target p_{T}^{gen} of the generated jet with neutrinos. An example of such distributions is shown in Fig. 5.6 for the leading and trailing jets. From distributions such as in Fig. 5.6 we obtain the mean value for the **scale** and sigma (σ) for the **resolution**. The mean value and sigma are obtained from the Bukin function fit [117]. The scale and resolution of the leading and trailing jets versus their p_{T} are shown in Fig. 5.7. From this figure we can conclude that adding MET variables into the training improves significantly the resolution. As expected the **15+3var js** training gives the best per-jet resolution across the whole p_{T} range.

The purpose of the regression is to improve the Higgs boson mass resolution from the $\mathrm{H} \to b\bar{b}$ decay in $X \to \mathrm{HH} \to b\bar{b}\gamma\gamma$ signal. The distributions together with the fit to a Bukin function are shown in Figs. 5.9, where the reconstructed mass is shown after the **full 15+3var js** regression training, for $m_G = 300\,\mathrm{GeV}$ and $m_G = 900\,\mathrm{GeV}$.

After the fitting to the corresponding function is done, we obtain the mean and the width parameters of the fit. The width and mean (both in GeV) are shown on Fig. 5.10 versus the mass of the graviton particle. From this figure we arrive at the same conclusion as for single-jet plots of Fig. 5.7: the MET variables improve the resolution and the **full 15+3var js** training gives the best mass resolution. The training with **15** variables give similar results to the **Hbb** training. In all trainings the scale does not match the nominal value of the Higgs boson mass. This is expected because the jets (both at reco and gen levels) do not contain all the energy of the Higgs boson decay.

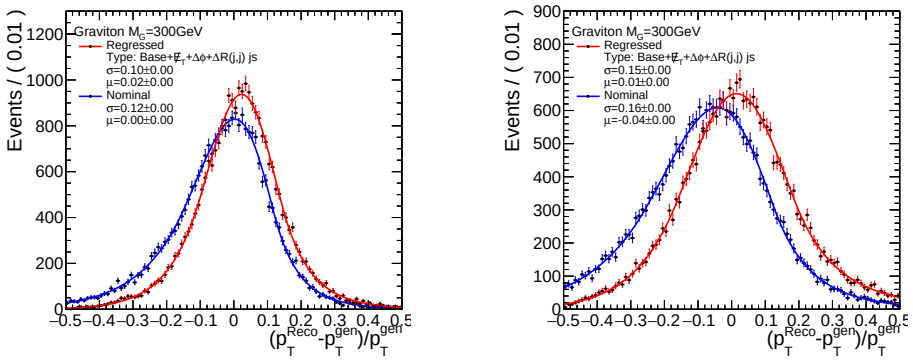

Figure 5.6: Relative p_{T} difference of the reconstructed and generated level jets after regression (red histograms) and without the regression (blue histograms), for the leading jet (left) and subleading jet (right).

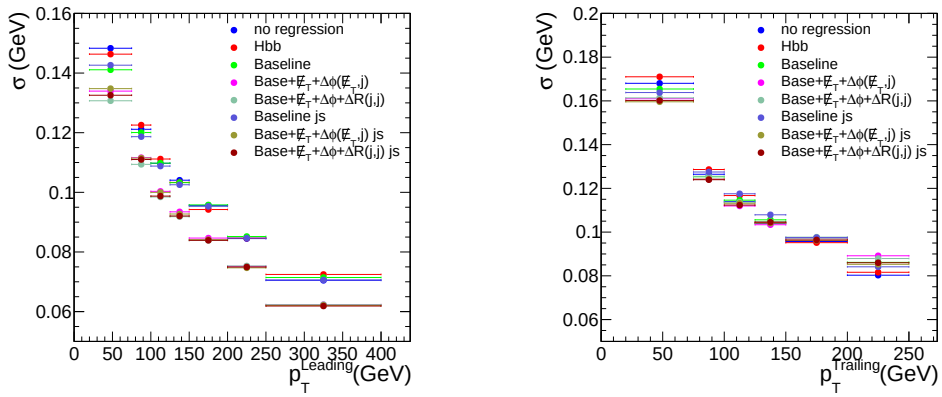

Figure 5.7: The resolution of the jet p_{T} for leading (left) and trailing (right) jets from the signal sample $G \to \mathrm{HH} \to b\bar{b}\gamma\gamma$.

In order to validate the developed regression in data, we use events with $\mathrm{Z} \to \ell\ell$ decays which also contain two b-tagged jets. It is assumed that a di-jet recoiled against a Z boson, and therefore the $p_{\mathrm{T}}(jj)$ must balance the $p_{\mathrm{T}}(\ell\ell)$ in the event. This check was done both in muon and electron channels of Z boson decay, analyzing `DoubleMuon` and `DoubleElectron` datasets correspondingly, using 2016 data.

Event selection requirements in both channels are: $p_{\mathrm{T}}(\ell_1) > 25\,\mathrm{GeV}$, $p_{\mathrm{T}}(\ell_2) > 15\,\mathrm{GeV}$, $p_{\mathrm{T}}(\ell\ell) > 50\,\mathrm{GeV}$, $75 < m_{\ell\ell} < 105\,\mathrm{GeV}$, $\Delta R^{min}_{\ell,jet} > 0.4$. The two jets in the event must pass Loose PF ID selection, Medium PU jet ID, tagged as b-jets with CSVv2 Medium WP, and have $p_{\mathrm{T}} > 20\,\mathrm{GeV}$, $|\eta| < 2.4$.

Figure 5.11 shows the mentioned p_{T}-balance distributions, $p_{\mathrm{T}}^{jj}/p_{\mathrm{T}}^{\ell\ell}$. The data is compared to the MC predictions. It can be seen that before any regression is applied (top left plot) the peak of the ratio distribution is below one. With the regression applied the peak moves to 1 for both our **full variables with js** training (top right) and the one from **Hbb** (bottom). The trend is the same in data and MC. This indicates that the regression does indeed brings

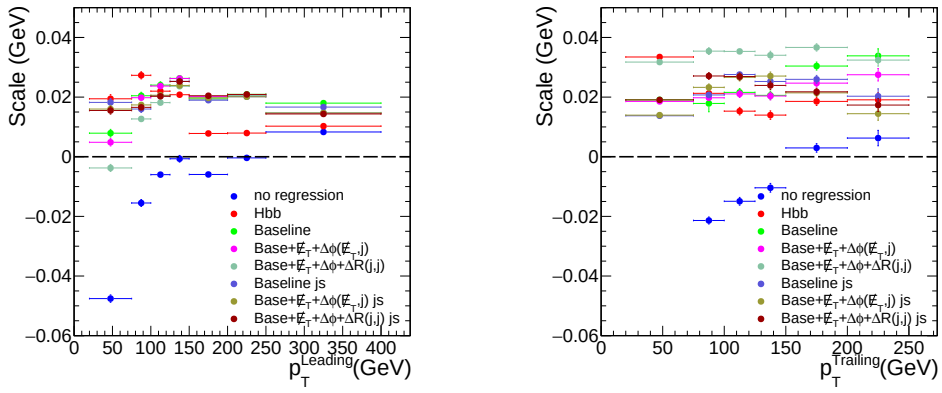

Figure 5.8: The scale of the jet p_{T} for leading (left) and trailing (right) jets from the signal sample $G \to \mathrm{HH} \to b\bar{b}\gamma\gamma$.

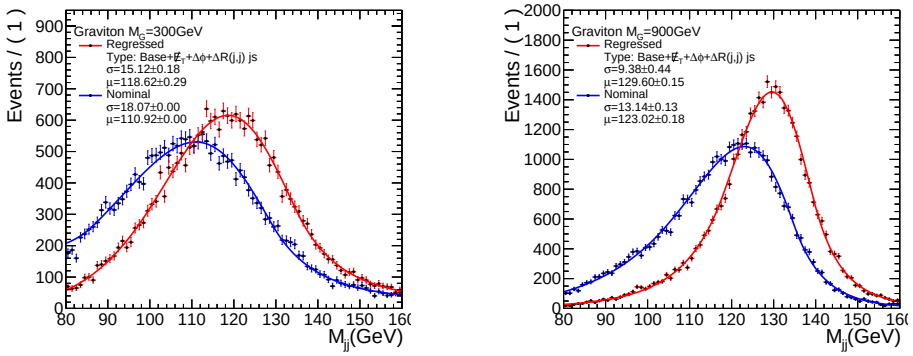

Figure 5.9: M_{jj} distributions from the reco-jets before and after the **full variables with js** regression for $m_G = 300\,\mathrm{GeV}$ signal sample (left) and $m_G = 900\,\mathrm{GeV}$ signal sample (right).

the p_{T} of the b-jets closer to their true values.

Similarly, figure 5.12 shows the mass distributions, m_{jj}, before and after regression. These figures indicate that m_{jj} is not distorted. This ensures us that the background distributions in our analysis signal region will not be distorted either.

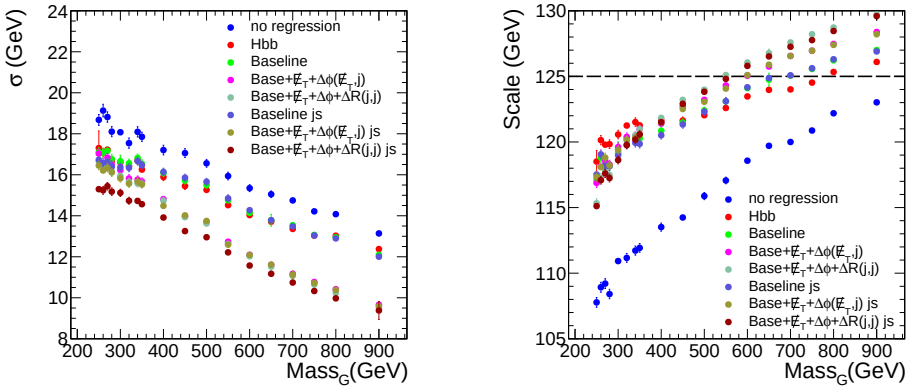

Figure 5.10: Performance plot comparing different regression trainings.

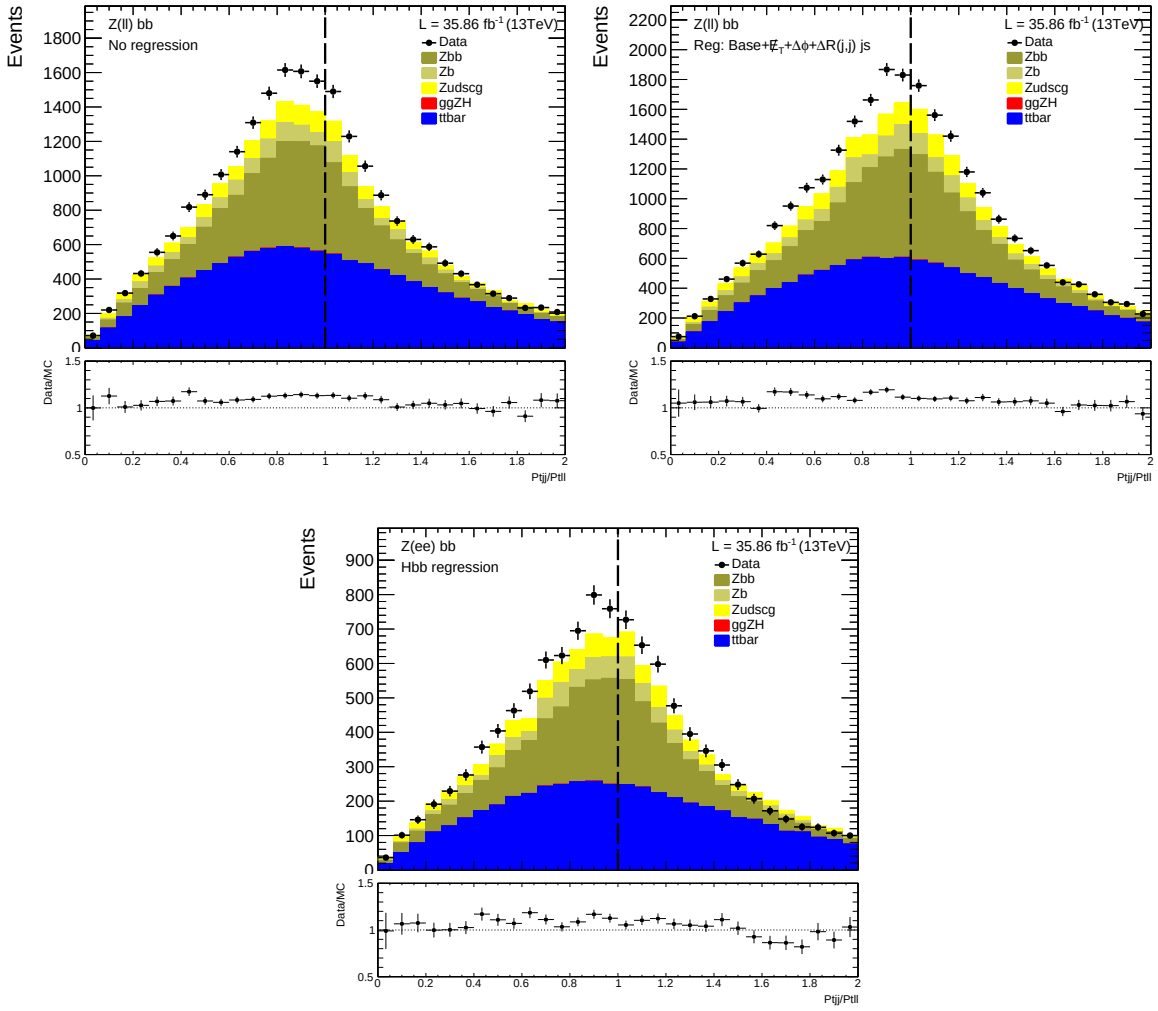

Figure 5.11: Pt balance (ratio) of the di-jet and di-lepton. (Top left) Distributions without regression, (top right) using **full 15+3var js** training and (bottom) using **Hbb** regression, on the combination of electrons and muons channels.

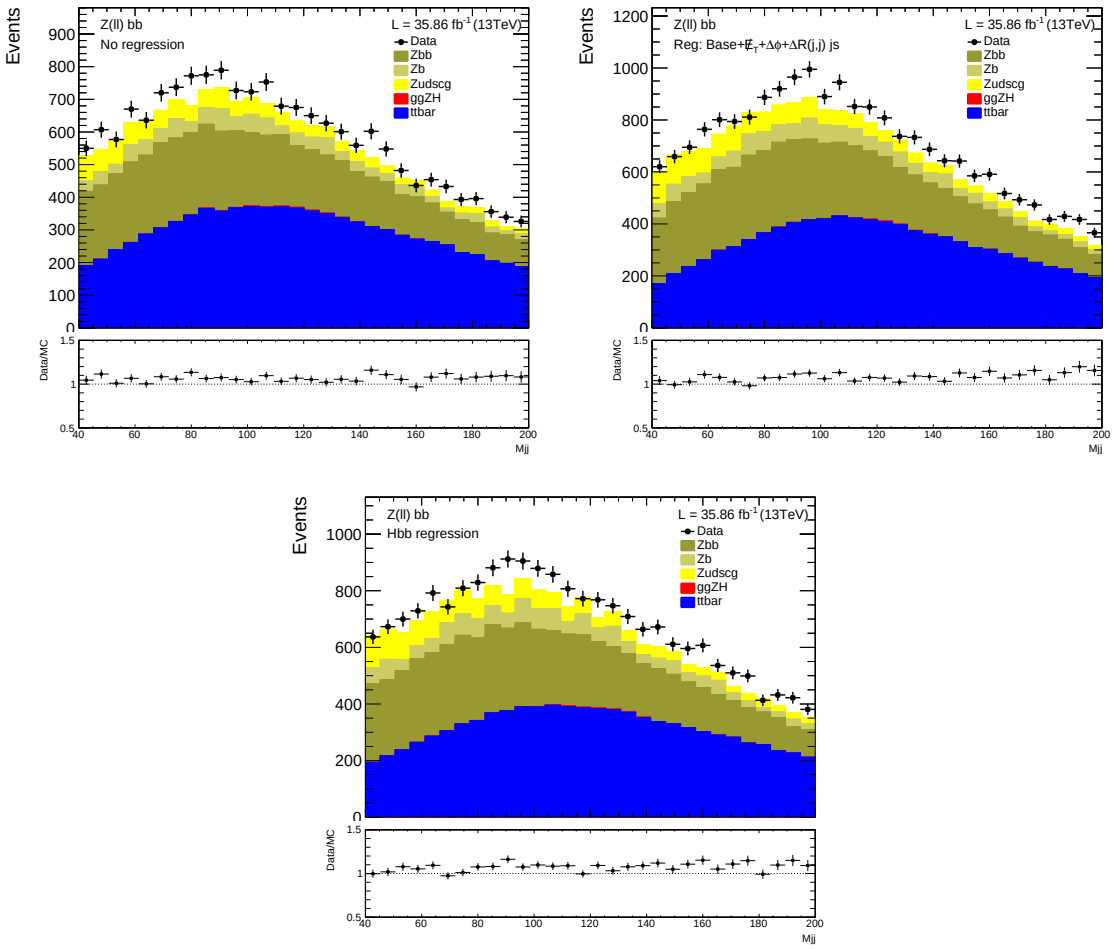

Figure 5.12: Distributions of the m_{jj}. (Top left) Distributions without regression, (top right) using **full 15+3var js** training and (bottom) using **Hbb** regression, on the combination of electrons and muons channels.

5.3.4.1 B-tagging

We utilize the *Combined Secondary Vertex* algorithm (CSVv2) for tagging b-jets, described in Ref. [118]. This b-tagging score for leading and subleading jets is then used in the MVA categorization.

The b-tagging scale factors have been calculated according CMS-recommended recipes, including the in situ calculation of signal efficiency. The signal efficiency has been calculated for all signal samples combined, in bins of p_T and $|\eta|$. The efficiency plots for tight WP, medium WP and loose WP can be seen in figure 5.13.

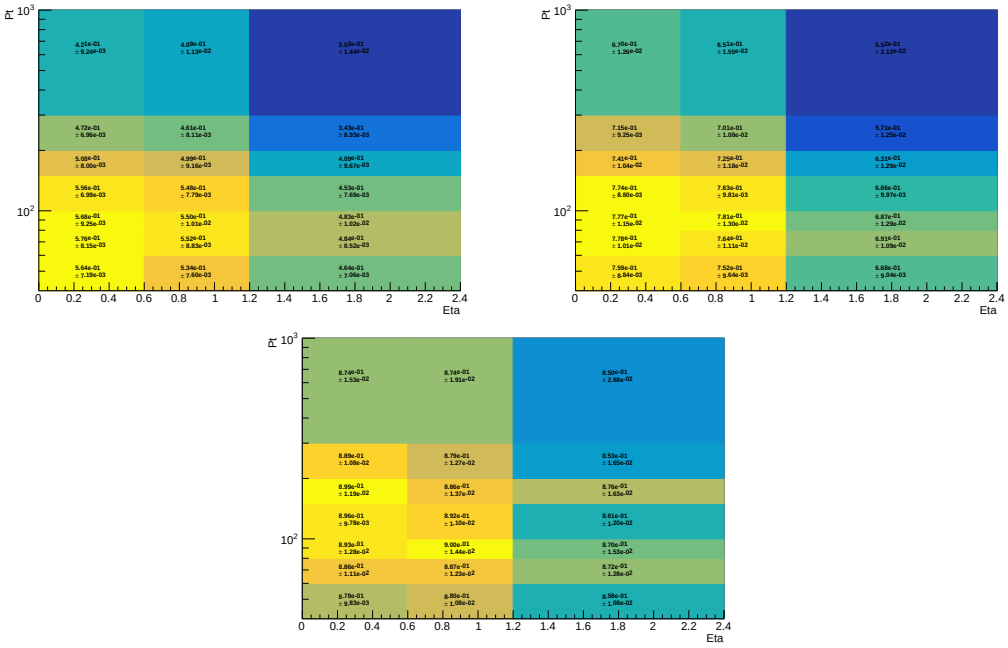

Figure 5.13: B-tagging efficiency for tight, medium and loose working points, as a function of jet p_T and $|\eta|$.

5.4 MVA Based Categorization

During the analysis, it has been noticed that different kinematic variables could potentially contribute to constraining the background contribution in the signal region without cutting too much on the signal efficiency. However, this large-dimensional optimization procedure (all investigated variables) was not optimal. Instead, we have developed a multivariate analysis (MVA), combining these different variables, into a single discriminant. This discriminant is used to categorize the events in High Purity, Medium Purity categories and a control region, similarly to the cut based categorization.

The input variables investigated for this MVA were:

- Leading and subleading jets b-tagging score;

- Helicity angles $|cos(\theta^*_{CS})|$, $|cos(\theta^*_{bb})|$ and $|cos(\theta^*_{\gamma\gamma})|$: $|cos(\theta^*_{CS})|$ is defined as the angle between the direction of the $H \to \gamma\gamma$ candidate to the Colin-Sopper reference frame (assumes each incoming particle in the scattering to have 6.5 TeV); $|cos(\theta^*_{xx})|$ is defined as the angle between the particle x and the direction defined by the $H \to xx$ candidate (randomly choosing between x's), where $x = \gamma$ or b;

- $p_T(\gamma\gamma)/M(jj\gamma\gamma)$ and $p_T(jj)/M(jj\gamma\gamma)$

The training was performed in the photon control region, as described in 5.8.2. Plots comparing the input variables in the photon control region and the blinded signal region are shown in Figure 5.14. As our signal in the training, we sum the 14 non-resonant HH samples available (box only, SM, and 12 BSM points). Thus, we have a training that is not specific to a single region in the parameter space, maintaining the sensitivities comparable between the benchmark points (as it is with the cut based categorization).

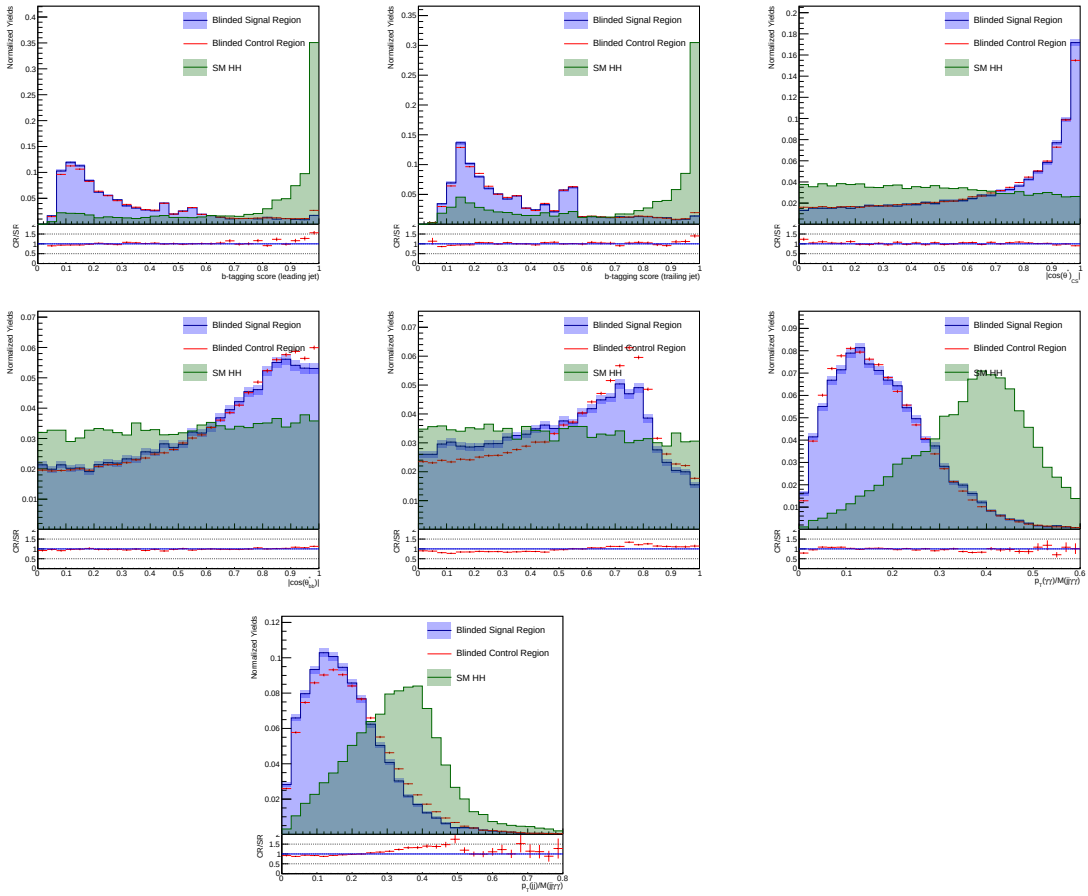

Figure 5.14: Distributions of input variables in the blinded photon control region, blinded signal region, and SM HH sample. All normalized to unity.

To improve the training, we split the training into two regions: low mass and high mass. The low mass training is performed with events with $\tilde{M}_X$ below 350 GeV, while the high mass

Mass Region	HPC	MPC
Low Mass	HHTagger > 0.96	$0.75 <$ HHTagger < 0.96
High Mass	HHTagger > 0.96	$0.6 <$ HHTagger < 0.96

Table 5.5: Non-resonant categorization with HHTagger discriminant.

training uses the complementary region. The training is based on a decision tree boosted with the gradient algorithm, with the trees randomized between iterations to decrease overtraining. To implement the training the TMVA package was used. From now on, we will refer to the training discriminant variable as HHTagger.

With the HHTagger discriminant, we build our two signal categories based on the optimization of the expected SM HH limit, separately in the high mass and in the low mass regions. These categories are called high purity category (HPC) and medium purity category (MPC); their names refer to their signal and background content. As signal, we use the SM HH sample to calculate the sensitivity. The outcome of this study was the categorization in Table 5.5. The expected number of background events, when comparing MPC and HPC between cut based and MVA approaches, is comparable and consistent, while for the number of signal events, the performance is better for the HHTagger categorization.

We also have to ensure that the HHTagger selection does not distort our variables of interest, $M(jj)$ and $M(\gamma\gamma)$. We demonstrate that there is no appreciable shaping by comparing the $M(jj)$ and $M(\gamma\gamma)$ shapes in different bins of the HHTagger discriminant. This can be seen in Figure 5.15.

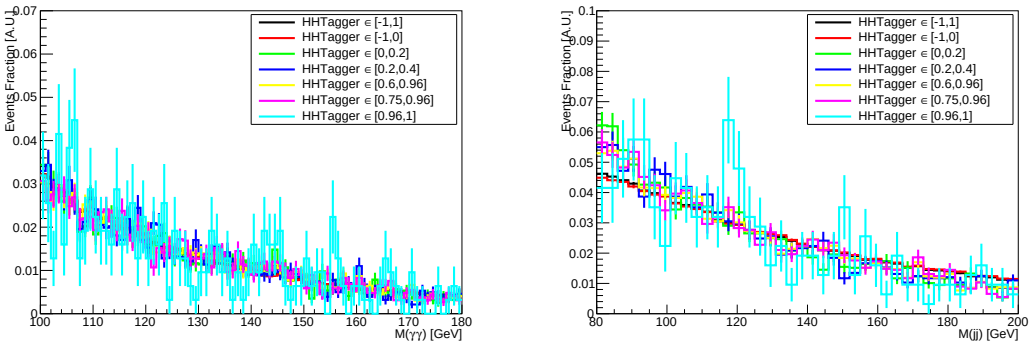

Figure 5.15: Photon control region distributions of $M(\gamma\gamma)$ (left) and $M(jj)$ (right) in bins of HHTagger. Although the slope changes between bins, this effect does not influence the limit setting.

5.4.0.1 Performance Cross-Checks

We have performed several cross checks to look for possible improvements on the MVA categorization.

- **Signal Hypothesis**

In the standard training, the sum of all non-resonant samples are used as the signal hypothesis. However, this might not be the optimal training for the SM HH case. To test this, we compare the performance of different trainings assuming the SM HH signal. The signal hypotheses tested are:

- All non-resonant (standard);
- SM HH;
- SM HH, with separate training for the high mass and low mass region (similar to standard);
- SM HH + Node 3 (this node, as defined in Table 5.1, contains the SM point);
- SM HH + Node 3, with separate training for the high mass and low mass region (similar to standard).

The background hypothesis for this test is the photon control region, in the high mass region.

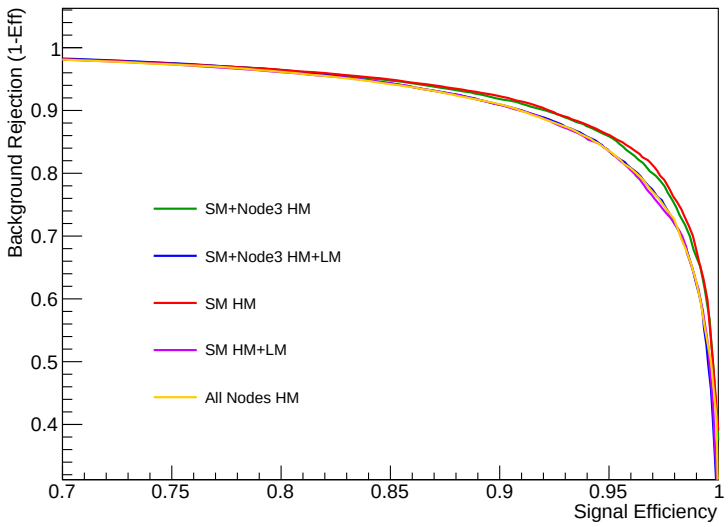

Figure 5.16: ROC curves with different signal hypotheses for training. The performance is evaluated in the high mass region, with the photon control region as background and SM HH as signal.

The ROC curves from the different trainings are shown in Figure 5.16. Since no significant improvement is seen in the high purity region (for background rejection larger than 95%, a typical value for the chosen WPs), the standard training method is kept in use.

- **Background Hypothesis**

In the standard training, the photon control region is used as a background model, avoiding MC reliance. However, this might not be optimal because the photon control region might

have different correlations between the MVA variables with respect to the signal region. To test this, we compare the performance of different trainings assuming different background hypotheses:

- Photon control region (standard);
- Blinded signal region;
- Blinded control region (to ensure that the difference between the two previous trainings does not come from blinding).

The background hypothesis for this test is the blinded signal region, in the high mass region, and the signal hypothesis is SM HH.

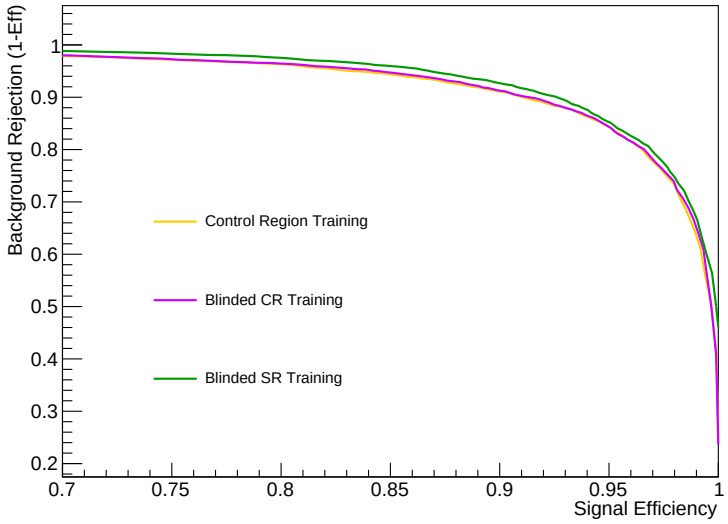

Figure 5.17: ROC curves with different background hypotheses for training. The performance is evaluated in the high mass region, with the blinded signal region as background and SM HH as signal.

The ROC curves from the different trainings are shown in Figure 5.17. While some improvement is seen, this training is not optimal for statistical reasons. The blinded signal region contains significantly less events than the photon control region. This limits the precision and accuracy of the multivariate analysis training. Specifically, it has been observed that the blinded signal region does not contain events in the high BDT region (signal-like phase space), which can cause over training. The second issue is that, optimally, training on a dataset that is statistically independent from the one to which it will be applied leads to a more robust procedure.

- **Resonant Hypothesis**

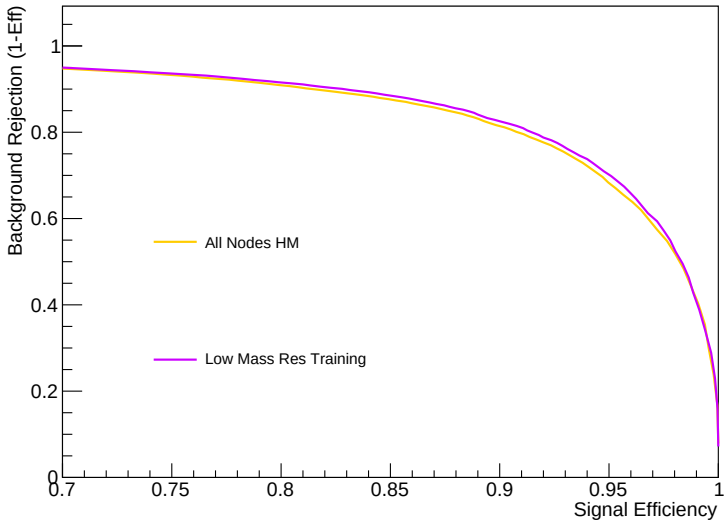

Figure 5.18: ROC curves with different signal hypotheses for training. The performance is evaluated in the low mass region, with the photon control region as background and the 300 GeV radion sample as signal.

While this MVA is trained with the non-resonant signal hypotheses, it can also be applied to the resonant search. We check, however, if a dedicated training with the resonant samples as signal hypothesis performs better, when applying the categorization to the resonant analysis. This is tested by comparing the categorization performance of the standard training versus a resonant training on a resonant signal point. The plot in Figure 5.18, these two trainings are shown and no significant difference is seen. Therefore the standard, non-resonant training will be used also for the resonant analysis.

5.5 $\tilde{M}_{\mathrm{X}}$ and Mass Window Selection

In order to increase the sensitivity of the resonant analysis, we perform a cut on the 4-body invariant mass before the signal extraction with the 2D fit. In the Run I analysis, the 4-body invariant mass was corrected with a kinematic fit, to mitigate the effects of the low mass resolution of the dijet system. However, it has been seen that this method is too reliant on the a-priori set of energy and spatial resolutions for the jets in that analysis (these resolutions must be measured in situ, since they are kinematic dependent). One solution for this was to use instead the variable $\tilde{M}_{\mathrm{X}}$, defined as:

$$\tilde{M}_{\mathrm{X}} = M(jj\gamma\gamma) - M(jj) - M(\gamma\gamma) + 250 \text{ GeV}. \tag{5.3}$$

This variable performs a kinematic fit "by hand", by effectively scaling the dijet and diphoton invariant masses to 125 GeV. In order to quantify the improvement of this variable with respect to other 4-body invariant mass reconstructions, we calculate the width of the

smallest interval that covers 68% of the signal shape in each reconstruction method. We compare $\tilde{M}_X$ with the standard $M(jj\gamma\gamma)$ and with $M(jj\gamma\gamma)_{KF}$, which is reconstructed with a kinematic fit in $M(jj)$, which tries to vary the jets within their uncertainties to achieve $M(jj) = 125$ GeV. The widths are compared in Figure 5.19.

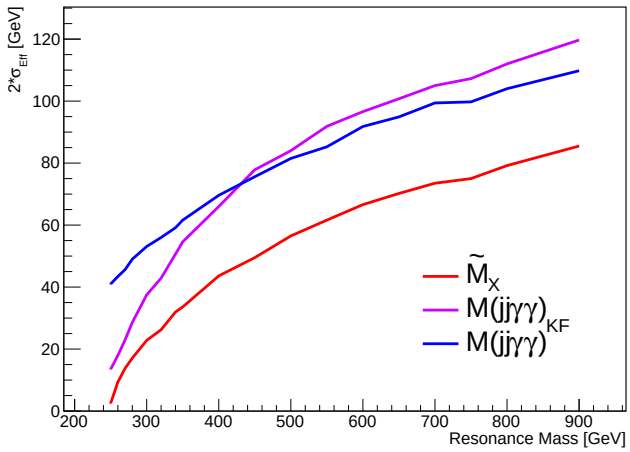

Figure 5.19: Different 4-body mass widths: $\tilde{M}_X$, $M(jj\gamma\gamma)$ with kinematic fit, and $M(jj\gamma\gamma)$ with no extra corrections.

The effect of reconstructing the 4-body invariant mass with $\tilde{M}_X$ on the signal shape can be seen in Figure 5.20. The figure shows that the $\tilde{M}_X$ reconstruction yields a better performing resolution for the 4-body invariant mass reconstruction, meaning that we can perform tighter cuts on it and increase the signal/background for each signal mass point.

One extra check we must do while using this variable is to make sure there are no unexpected effects on the background shape. The effect of $\tilde{M}_X$ is similar to the kinematic fitted $M(jj\gamma\gamma)$, but more pronounced, which is the sharp kinematic cut around the $\tilde{M}_X = 250$ GeV point. This can be seen in the figures in section 5.7.

It has also been observed that, for the non-resonant signal samples, the $\tilde{M}_X$ variable approaches the generator level HH invariant mass distribution more so than the kinematic fitted $M(jj\gamma\gamma)$ and the default 4-body invariant mass. For this, $\tilde{M}_X$ will also be used in the non-resonant analysis.

One desirable effect of the $\tilde{M}_X$ definition in this version of the analysis, compared to the 2015 definition (which only scaled the $M(jj)$ value as in $\tilde{M}_X = M(jj\gamma\gamma) - M(jj) + 125$ GeV), is that a $\tilde{M}_X$ selection does not bias $M(\gamma\gamma)$ and $M(jj)$. This can be seen in the 2D plots of $\tilde{M}_X : M(\gamma\gamma)$ and $\tilde{M}_X : M(jj)$ in Figure 5.21.

With $\tilde{M}_X$, we can improve the resonant analysis by tightening the signal region around the 4-body resonance mass (mass window). Through limit optimization, it has been checked that constructing a mass window with the smallest interval that covers 60% of the signal shape provides the best sensitivity. The size of these intervals, as a function of the hypothesis mass, is seen in Figure 5.22.

We implement this mass window by requiring that $W_- < \tilde{M}_X < W_+$. $W_\pm$ are calculated

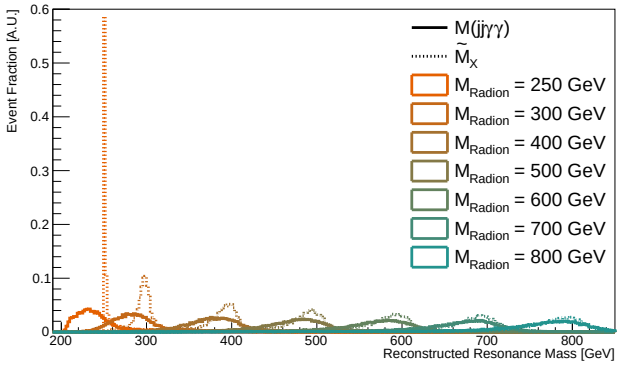

Figure 5.20: Different 4-body mass reconstructions: $\tilde{M}_X$ (dotted line), $M(jj\gamma\gamma)$ with kinematic fit (dashed line), and $M(jj\gamma\gamma)$ with no extra correction (full line). Signals for different radion masses are shown. This plot is made after full selection, in the b-tagging signal region (at least one medium b-tagged jet).

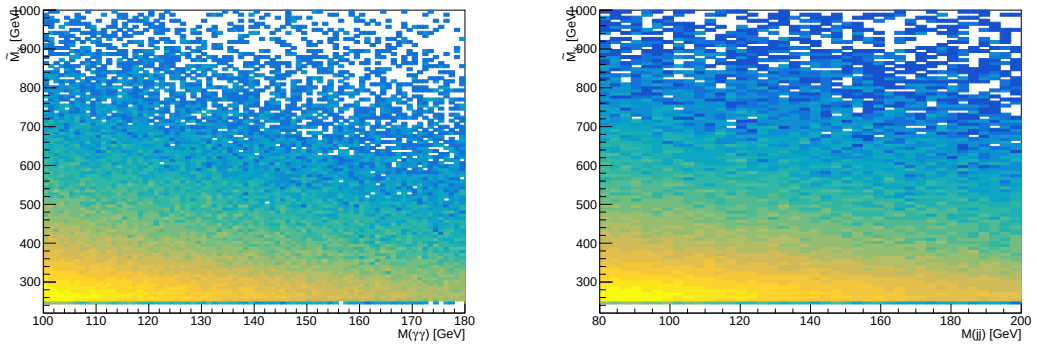

Figure 5.21: $\tilde{M}_X : M(\gamma\gamma)$ and $\tilde{M}_X : M(jj)$ in the photon control region, scaled to unity and Z-axis in log scale.

based on the widths defined in Figure 5.22. We fit $W_\pm$ with a 3rd degree polynomial so that the mass windows can be defined functionally based on the mass hypothesis. These fits, and therefore, the definition of $W_\pm$ can be seen in Figure 5.23. $W_\pm$ can be inferred through the y-axis of 5.23, W_- is defined by the blue curve and W_+ by the red.

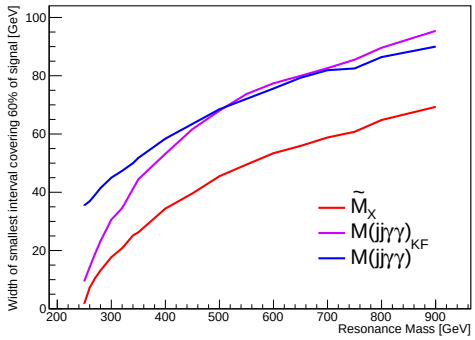

Figure 5.22: Mass window sizes as a function of the resonance mass.

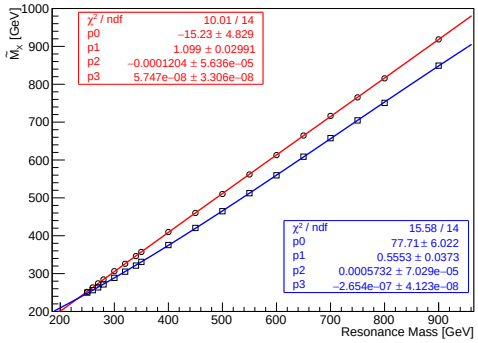

Figure 5.23: Mass window sizes as a function of the resonance mass.

5.6 Selection Efficiencies

As a summary of the previous sections, the cut flow of the analysis is as follows:

- At least two photons and two jets in the event
- At least two photons pass the trigger based pre-selection (see Sec. 5.3.1);
- At least two photons pass the kinematic and identification requirements (see Sec. 5.3.2) → select two highest E_T photons as diphoton candidate;
- At least two jets pass the kinematic selection (see Sec. 5.3.3) → select two jets with highest b-tagging score as dijet candidate;
- Event can be classified in either High Purity Category or Medium Purity Category;

The efficiency after each step above and taking into account the acceptance is estimated and shown for each of the signal samples considered in our analysis. In Figure 5.24, the signal efficiency for radions (spin-0) and gravitons (spin-2) from mass hypotheses of 250 up to 900 GeV. In Figure 5.25, the signal efficiency for the non-resonant nodes, including top-only contribution (Node 0) and SM (Node 1).

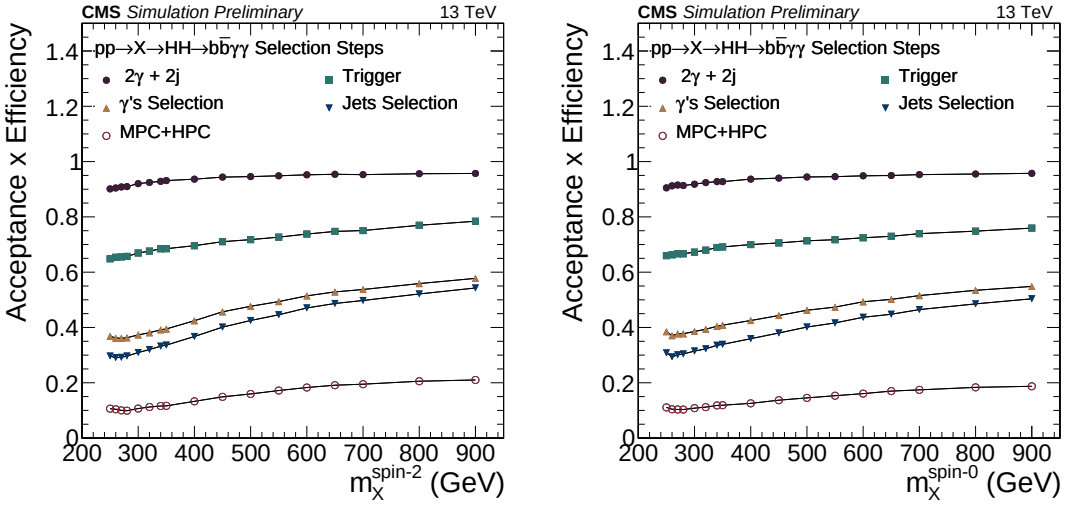

Figure 5.24: Graviton (left) and Radion (right) signal acceptance×efficiency for each cut (described in text).

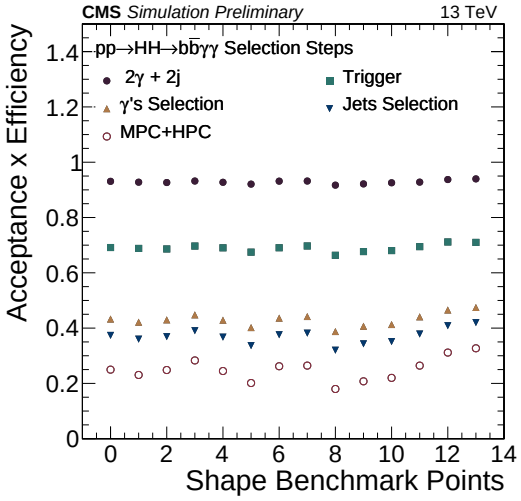

Figure 5.25: Non-resonant acceptance x efficiency. Jet energy regression applied in addition to MVA based categorization.

5.7 Control Plots

To validate the Monte Carlo simulations, we produce data/Monte Carlo comparison plots (control plots) of the signal region (SR), blinded in the $115 < M(\gamma\gamma) < 130$ GeV region. For background and data to match, the DiPhoton+Jets contribution has been scaled by a factor of 1.5, while the prompt-fake and fake-fake contributions from the GJets and QCD samples have been scaled to data. The signal in these plots is normalized for 500 fb.

It's a well known fact that modeling backgrounds with fake objects, such as jets identified as photons, leads to discrepancies when using MC simulation. This is one of the reasons why all background modeling performed in this analysis, both during limit extraction and MVA training, is done with data driven methods. The plots shown in this section aim to show that we understand the composition of the background, which is basically equally split between the contribution with two "real" photons, and the contribution with one "real" photon and one jet identified as a photon.

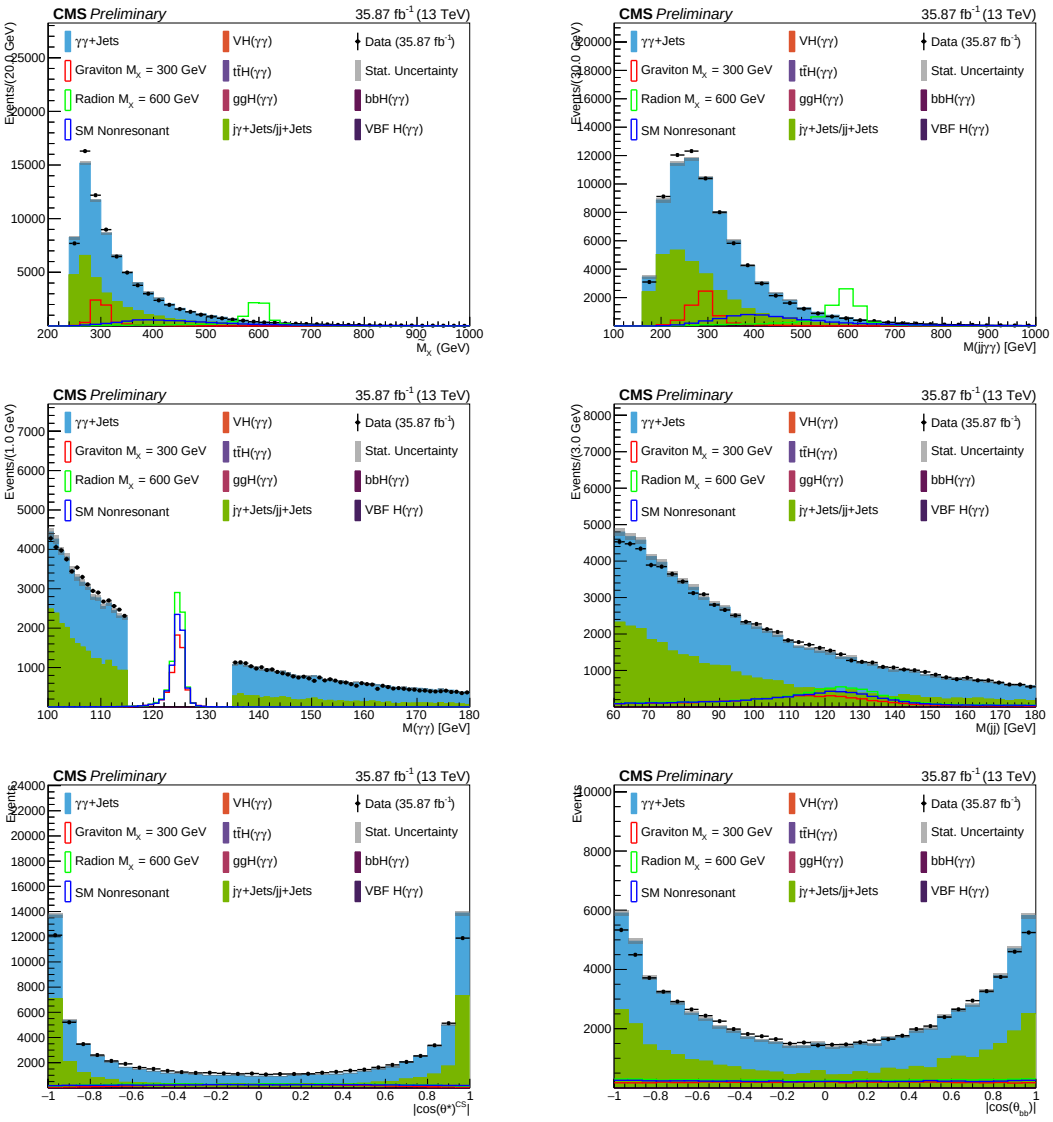

Figure 5.26: Distributions of data and MC for the signal region with the blinding region removed. Top left: $\tilde{M}_{\mathrm{X}}$. Top right: $M(jj\gamma\gamma)$. Center left: $M(\gamma\gamma)$. Center right: $M(jj)$. Bottom left: $|cos(\theta^*_{CS})|$. Bottom right: $|cos(\theta^*_{bb})|$.

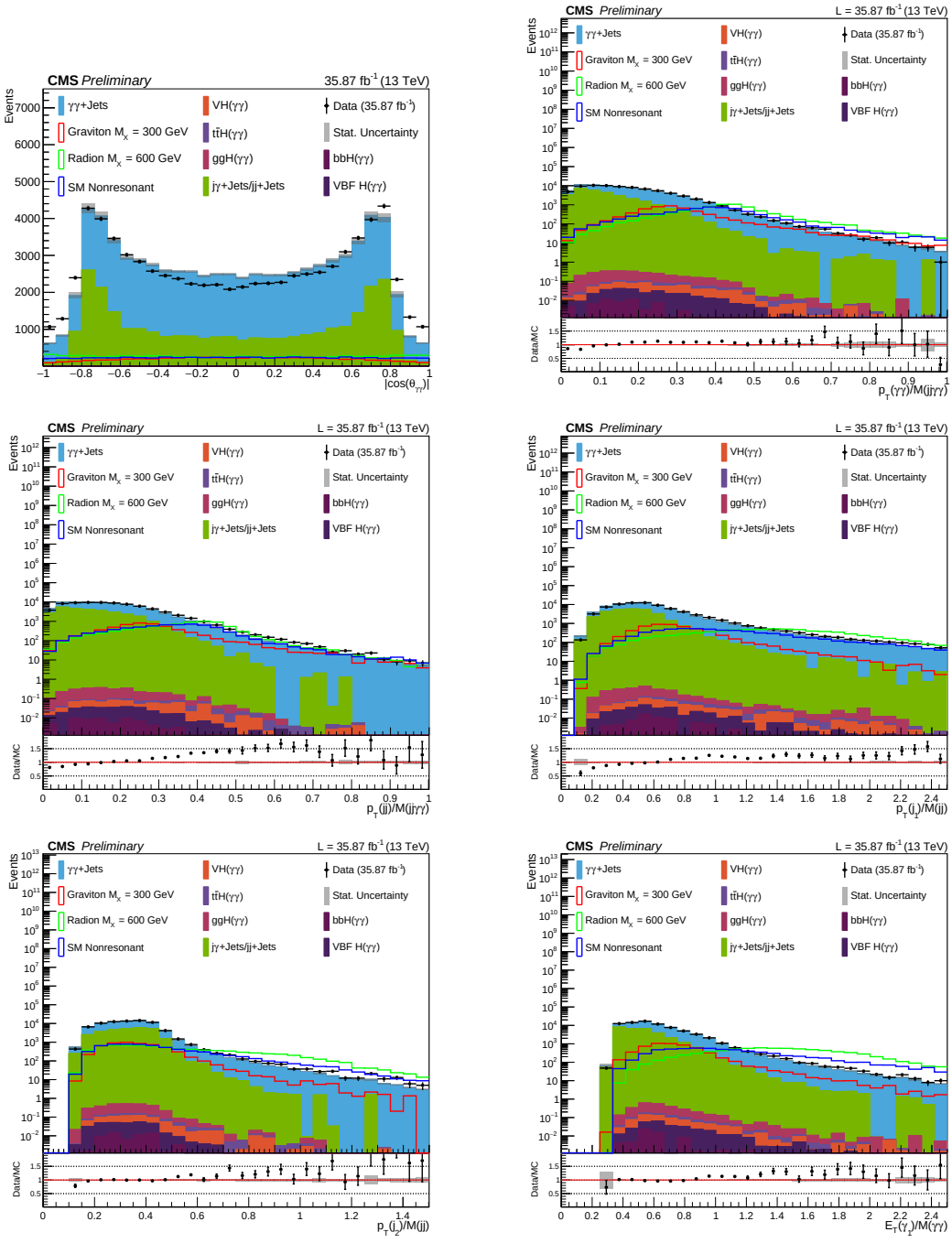

Figure 5.27: Distributions of data and MC for the signal region with the blinding region removed. Top left: $|cos(\theta^*_{\gamma\gamma})|$. Top right: $p_T(\gamma\gamma)/M(jj\gamma\gamma)$. Center left: $p_T(jj)/M(jj\gamma\gamma)$. Center right: leading jet $p_T(j)/M(jj)$. Bottom left: subleading jet $p_T(j)/M(jj)$. Bottom right: leading photon $E_T(\gamma)/M(jj)$

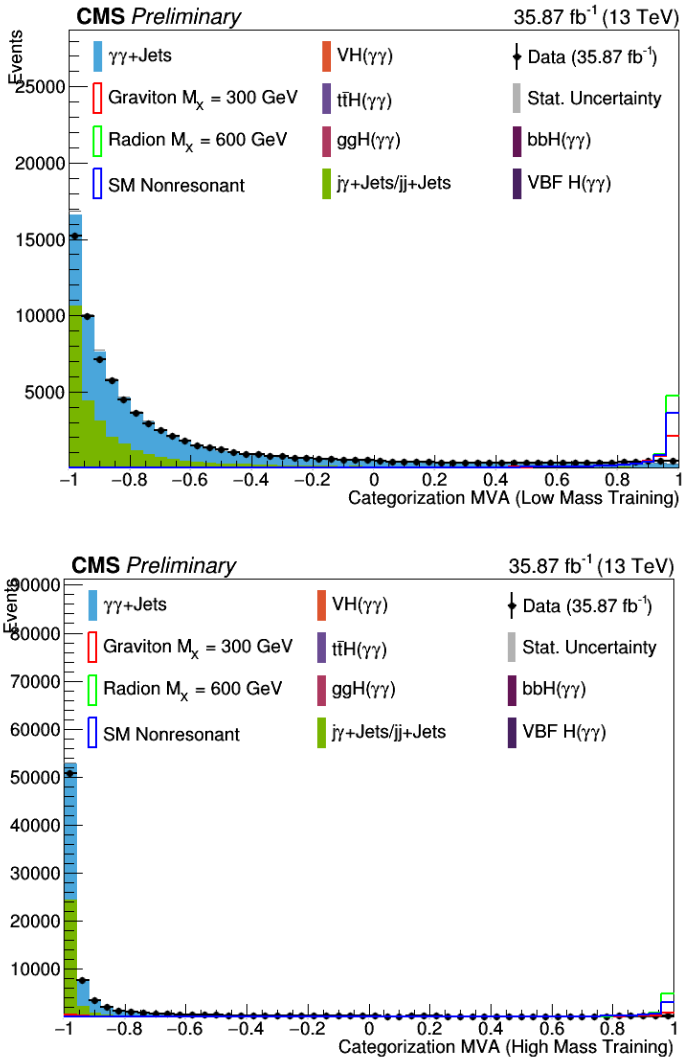

Figure 5.28: Distributions of data and MC for the signal region with the blinding region removed. Top: Categorization MVA output (low mass training). Bottom: Categorization MVA output (high mass training).

5.8 Statistical Modeling and Limit Extraction

The signal extraction and limit setting in this analysis is performed with a 2D fit in the $M(\gamma\gamma) : M(jj)$ plane, since we expect our signal to peak in both variables. For our background, they are expected to be uncorrelated given our statistical precision. With this last assumption, we can construct background function models as $f_\gamma(M(\gamma\gamma)) \times f_J(M(jj))$, where f_γ (f_J) are our functional choices to fit the diphoton (dijet) mass spectrum. A thorough explanation of the background uncorrelation hypothesis is given at the end of this section.

5.8.1 Signal Model

As a signal model in the limit extraction, we use parametric models fitted to the simulated samples after the full selection. Each fit is done in each different sample independently, i.e., for all resonance masses, spins and for all different non-resonant hypotheses. The choice of parametric model for $M(\gamma\gamma)$ and $M(jj)$ individually is a double-sided Crystal-Ball function. The double-sided Crystal-Ball function is defined as follows:

$$f(x;\mu,\sigma,\alpha_L,p_L,\alpha_R,p_R) = N \cdot \begin{cases} A_L \cdot \left(B_L - \frac{x-\mu}{\sigma}\right)^{-p_L}, & \text{for } \frac{x-\mu}{\sigma} > -\alpha_L \\ A_R \cdot \left(B_R + \frac{x-\mu}{\sigma}\right)^{-p_R}, & \text{for } \frac{x-\mu}{\sigma} > \alpha_R \\ e^{\frac{(x-\mu)^2}{\sigma^2}}, & \text{for } \frac{x-\mu}{\sigma} < -\alpha_L \text{ and } \frac{x-\mu}{\sigma} > \alpha_R \end{cases}, \tag{5.4}$$

where the A_L, A_R, B_L, B_R constants are defined by:

$$A_k = \left(\frac{p_k}{|\alpha_k|}\right)^{p_k} \cdot e^{-\frac{\alpha^2}{2}}, \tag{5.5}$$

$$B_k = \frac{p_k}{|\alpha_k|} - |\alpha_k|, \tag{5.6}$$

where k is either L or R. This definition is such that there are two independent tails, a left tail (L) and a right tail (R), and a gaussian core. This signal model is enough to model both the high mass resolution of $M(\gamma\gamma)$ and the lower resolution of $M(jj)$. With respect to the signal model chosen for previous versions of the analysis, such as the 2015 analysis, this choice is beneficial when comparing to a sum of a gaussian and a single sided Crystal-Ball because the left and right tails are made completely independent, while maintaining the same number of free parameters.

These signal fits can be seen in Figures 5.29, 5.30 and for the 300, and 600 GeV radion signals, and in figures 5.31 and 5.32 the signal fits for the non-resonant SM HH production in the high mass and low mass categories, respectively.

5.8.1.1 Correlation Studies

The choice of parametric signal model makes the assumption that the full 2D distribution can be modeled by a product of PDFs. This choice is not the most general one, as it does not model correlations between $M(jj)$ and $M(\gamma\gamma)$. One important question, therefore, is whether the analysis is sensitive to correlations that are not modeled by our choice of signal model.

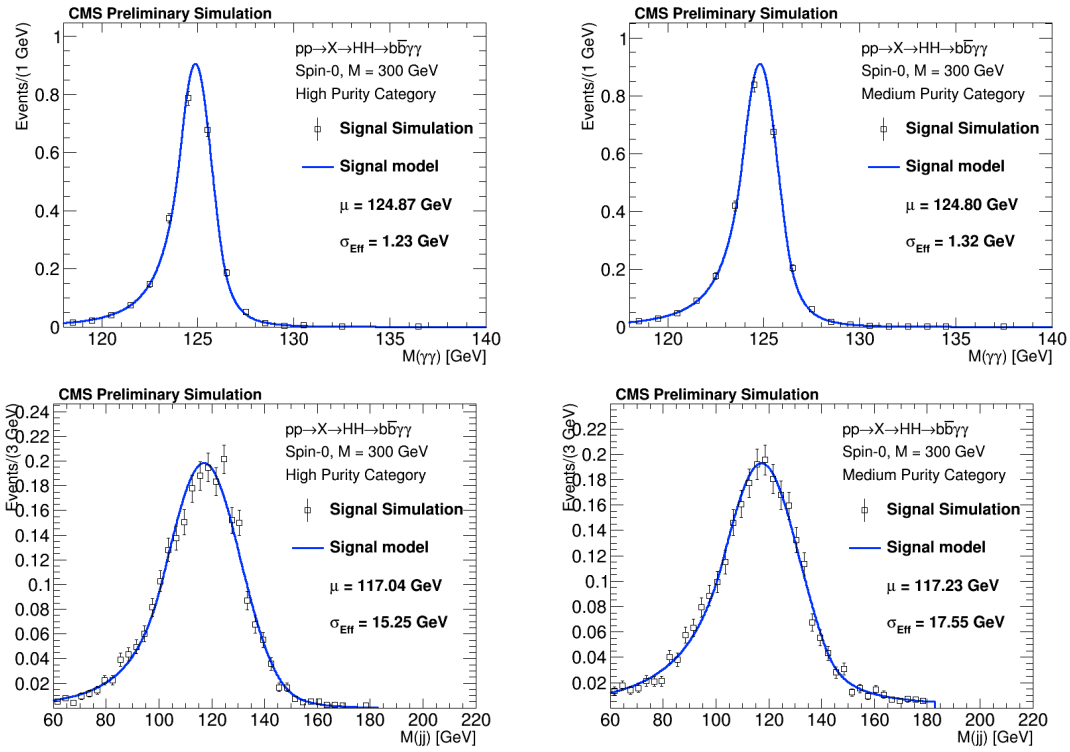

Figure 5.29: Signal fits for the radion 300 GeV mass sample after full analysis selection, in High (left) and Medium (right) purity categories. Top plots: $M(\gamma\gamma)$. Bottom plots: $M(jj)$.

To study this, we study the differences between the MC signal simulation and the 2D fitted PDF via residuals:

$$R_{ij} = \frac{N_{ij}^{\mathrm{PDF}} - N_{ij}^{\mathrm{MC}}}{\sigma_{N_{ij}^{\mathrm{PDF}}}^{\mathrm{Poisson}}}, \tag{5.7}$$

where ij refers to bin i in $M(\gamma\gamma)$ and bin j in $M(jj)$, and $\sigma_{N_{ij}^{\mathrm{PDF}}}^{\mathrm{Poisson}}$ is the Poissonian error of the expected (PDF) and observed (MC). These residuals are shown in Figures 5.33, 5.34, 5.35 and 5.36. The signal MC normalization for these plots are to 1/fb signal cross section. We see no structures in the residual plots in the region where the signal is expected. Therefore, we assume that the PDF product modeling is good enough given the statistical precision of our current dataset.

5.8.2 Background Model

To study the background fits, we use the fake photon control region (one photon in the diphoton candidate fails the identification requirements). From this control region, we randomly pick the number of events that is expected from the signal region under study according to the control sample described in section .

The functional choice to model the background in both fitting variables is the Bernstein family of polynomials. We also assume that the same order of polynomial is to be used in

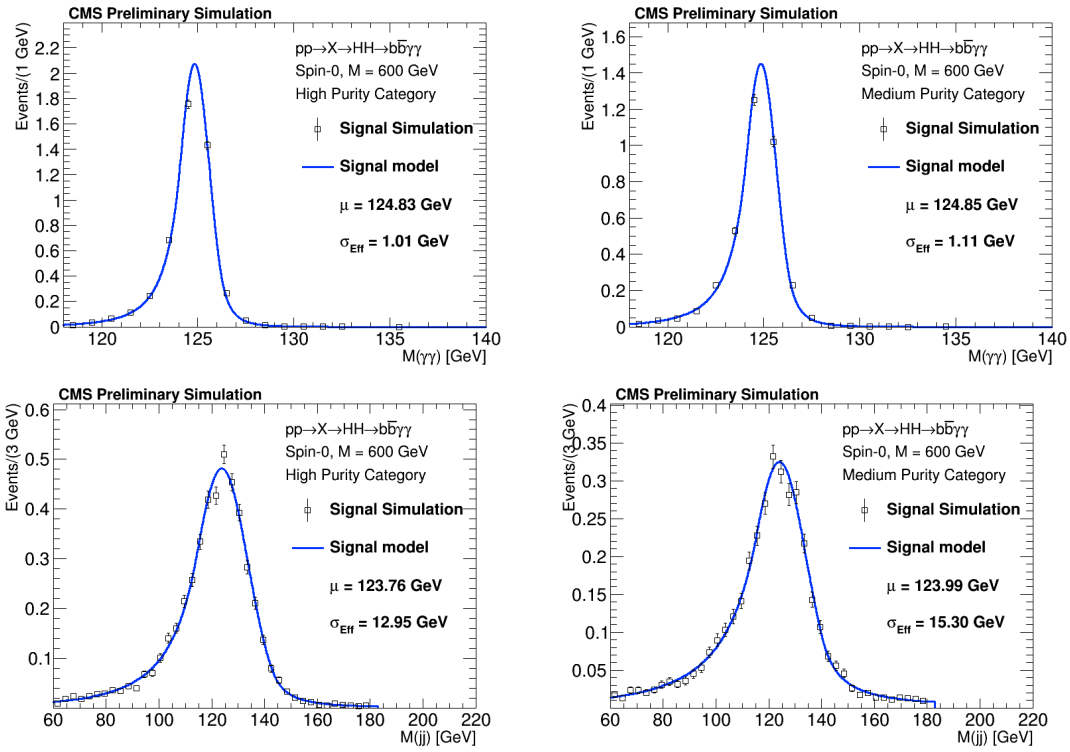

Figure 5.30: Signal fits for the Radion 600 GeV mass sample after full analysis selection, in High (left) and Medium (right) purity categories. Top plots: $M(\gamma\gamma)$. Bottom plots: $M(jj)$.

both variables. This comes from the fact that the order of the polynomial is related to the precision of the fit (degrees of freedom), which, in turn, is related to the number of events being fitted.

The first study performed is the order fixing. We fit consecutive orders of the three families of functions and check the difference between the negative log-likelihoods times two ($2\Delta NLL$) between the two consecutive fits. This $2\Delta NLL$ should be distributed as a $\chi^2(\alpha)$ distribution with the number of degrees of freedom equal to the difference in number of free parameters between the two consecutive orders (α). We then calculate the p-value of having a $2\Delta NLL$ higher than the one calculated before, given the $\chi^2(\alpha)$ distribution. If this p-value is lower than 0.05, we accept the higher order function, and continue the procedure for the next order. If this p-value is higher than 0.05, the higher order function is assumed to be too flexible given the data and the procedure terminates having found the lowest order suitable function.

Due to the different regimes of our signal regions after the mass window requirements and of the non-resonant selection, it is expected that our fits will involve very different background yields. For that, we perform the $2\Delta NLL$ test in all different signal regions. The results of this test are regions of validity in number of background events to be fit. This means that the choice of Bernstein order will depend on the number of events being fitted in a given signal region. The results of the study show that, for fits with less than 15 events, a first order Benstein passes the $2\Delta NLL$ test. For fits with 25 or more events, but less than 200,

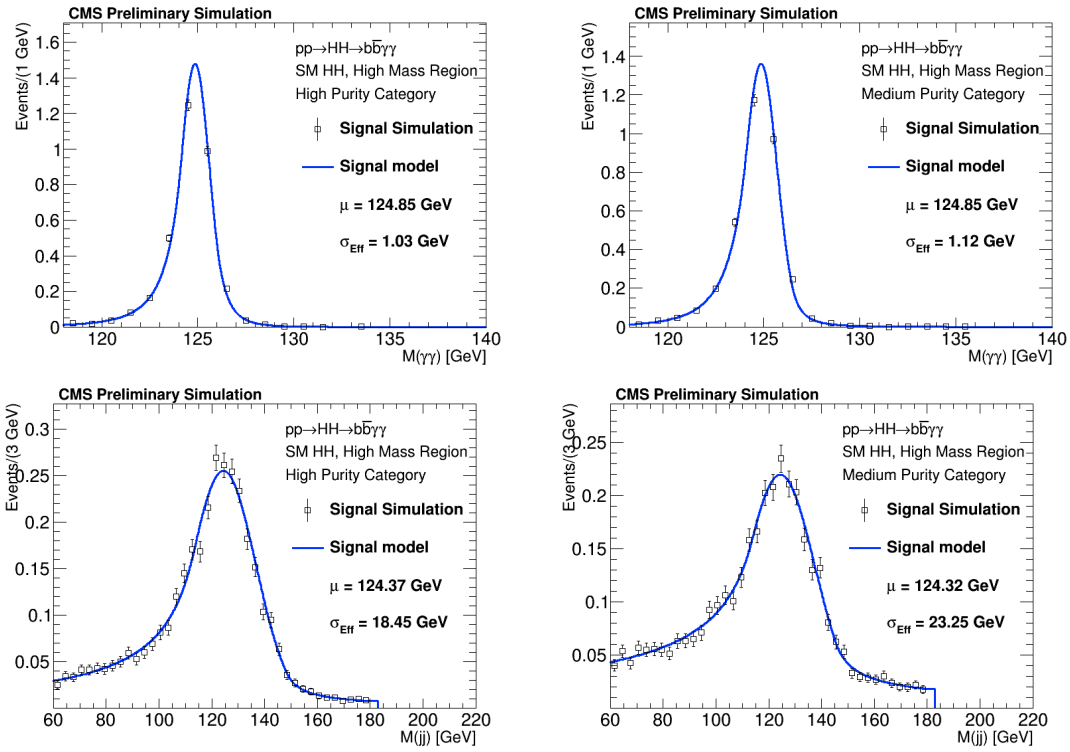

Figure 5.31: Signal fits for the SM HH non-resonant sample after full analysis selection, in High (left) and Medium (right) purity categories. Top plots: $M(\gamma\gamma)$. Bottom plots: $M(jj)$.

a second order Bernstein passes the $2\Delta NLL$ test. For fits with 200 or more events, a third order Bernstein passes the $2\Delta NLL$ test.

5.8.2.1 Bias Studies

After the order fixing, we must ensure that the functional form chosen does not bias a possible signal strength measurement in the analysis. This can happen because the real background shape that is being fitted might not be exactly the chosen functional form. Since we have no way of defining what this true shape is, we compare the signal strength measured (μ) from the background models with respect to different background shape hypotheses, as produced by a toy Monte Carlo.

The goal is to find at least one background model that can fit other background shapes without a statistically significant bias in the signal strength reconstruction. The goal of having a background model with a bias less than 0.14 for all assumed shapes is set. This is justified by investigating the effect that a signal strength bias can be correct by increasing the uncertainty on μ until the true value is within the 1σ coverage of μ.

We compare our 2D Bernstein model to models constructed with a Laurent series for both $M(\gamma\gamma)$ and $M(jj)$, and with sums of exponentials for both $M(\gamma\gamma)$ and $M(jj)$. We construct models with different Laurent and exponential sum orders.

The first step in the bias studies is to get pre-fit shapes of all background models. This

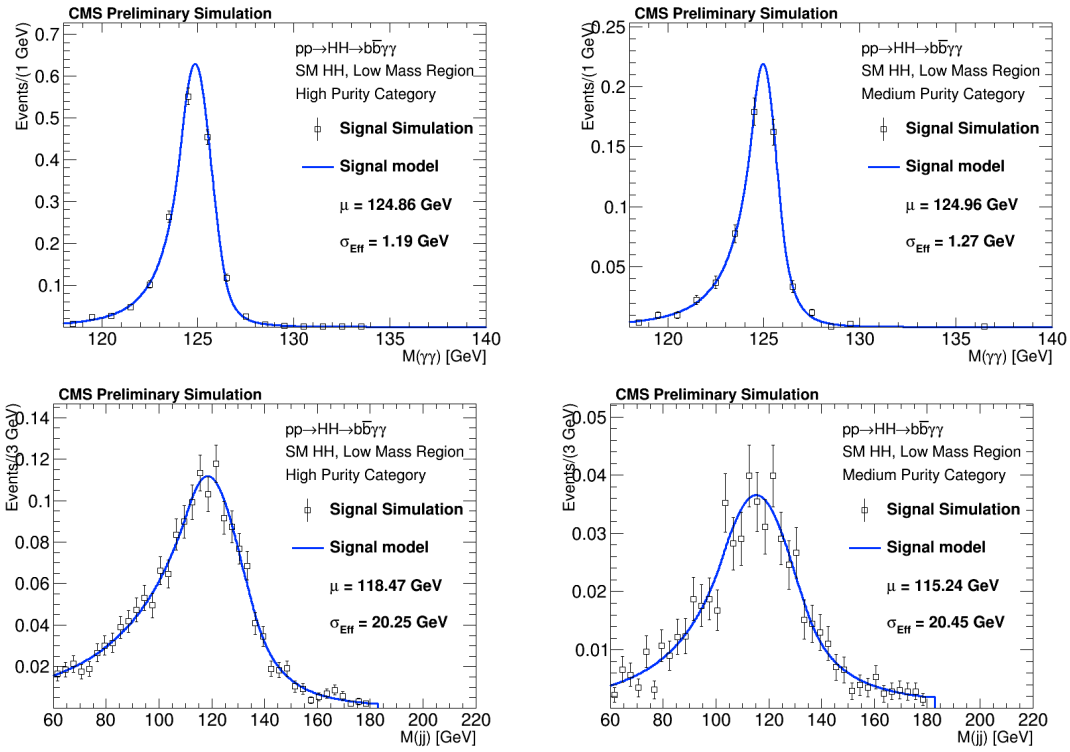

Figure 5.32: Signal fits for the SM HH non-resonant sample after full analysis selection, in High (left) and Medium (right) purity categories. Top plots: $M(\gamma\gamma)$. Bottom plots: $M(jj)$.

is done in the same datasets used for the order fixing procedure: fake photon control region scaled to match the statistics found in different data signal regions.

After the pre-fit shapes are constructed, toy Monte Carlo events are generated based on the pre-fitted background models. 2000 toy datasets are thrown for each background model. These toys are thrown injecting also signal events, according to the signal yields expected in each category. For that, we assume a signal cross section of 1 fb, for all resonant mass points and non-resonant benchmark points. Finally, the third step is fitting the 9 batches of toy datasets with the same background models and extracting the μ from each of the 2000 toy datasets.

Some examples of the measured biases can be seen in Figures 5.37-5.40. The x-axes on these plots represent a truth model with which toy MC was produced, while the y-axes represent the function used to fit the toy MC. The numbers on each bin are the absolute values of the biases on the signal strength measurements under a background hypothesis in the x-axis and a fit function in the y-axis.

5.8.2.2 Goodness-of-Fit

To check how well the background model fits the data, we perform a goodness-of-fit test in our blinded signal region. The Kolmogorov-Smirnov (KS) test was chosen for its good performance on unbinned datasets, which is the case of this analysis. Unfortunately, no 2-

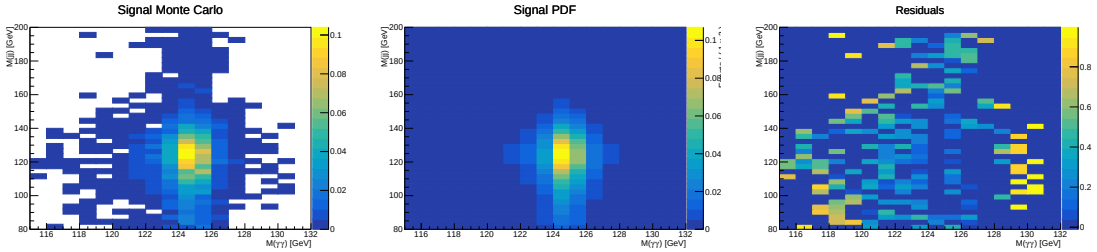

Figure 5.33: 2D distributions of the signal MC (left), fitted PDF model (center) and 2D residuals (right) for the High Mass-High Purity Category non-resonant selection.

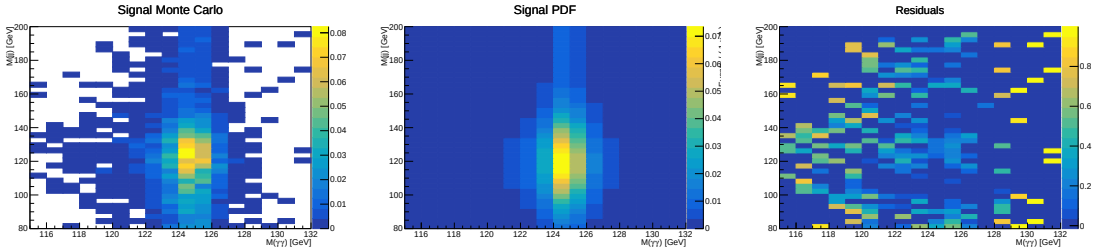

Figure 5.34: 2D distributions of the signal MC (left), fitted PDF model (center) and 2D residuals (right) for the High Mass-Medium Purity Category non-resonant selection.

dimension unbinned KS tests are available with the current tools used in CMS. The procedure taken was, then, to bin the 2D distribution with the analysis binning (40 bins in $M(jj)$ and 60 bins in $M(\gamma\gamma)$), making sure that the number of bins is much larger than the expected number of events (2400 bins is the case). For the blinding procedure, we set the bins of the 2D histograms to 0 in the blinding region ($120 < M(\gamma\gamma) < 130$ GeV). The requirement of the KS goodness-of-fit test is that the KS probability is » 0.05, which is achieved for all the categories and signal regions: all KS probabilities are larger than 0.45.

5.8.2.3 Correlation Studies

Assuming that the overall 2D shape can be modeled by a 2D second order polynomial, the most general function can be constructed as:

$$f(x,y) = \sum_{i=0}^{i=2} \sum_{k=0}^{k=2} c_{ik} x^i y^k, \tag{5.8}$$

where, in our case, $x = M(\gamma\gamma)$ and $y = M(jj)$ or vice-versa. However, in our modeling, we assume $M(\gamma\gamma)$ and $M(jj)$ to be independent, therefore, our choice of model takes the form of:

$$g(x,y) = \left(\sum_{i=0}^{i=2} a_i x^i \right) \left(\sum_{k=0}^{k=2} a_k y^k \right). \tag{5.9}$$

While the first equation has 9 degrees of freedom, the second only has 6. Therefore, by assuming our two parameters of interest to be independent, we lose three degrees of freedom

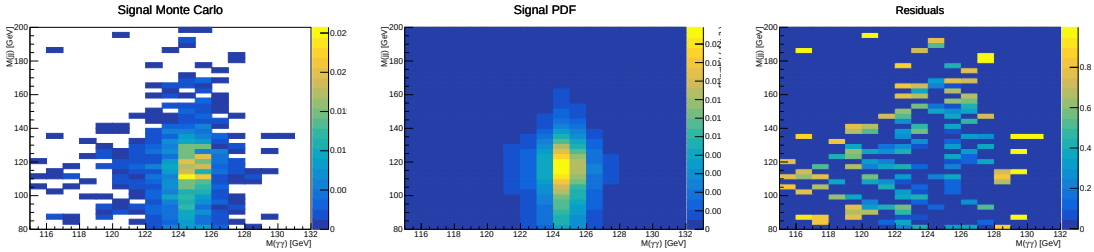

Figure 5.35: 2D distributions of the signal MC (left), fitted PDF model (center) and 2D residuals (right) for the Low Mass-High Purity Category non-resonant selection.

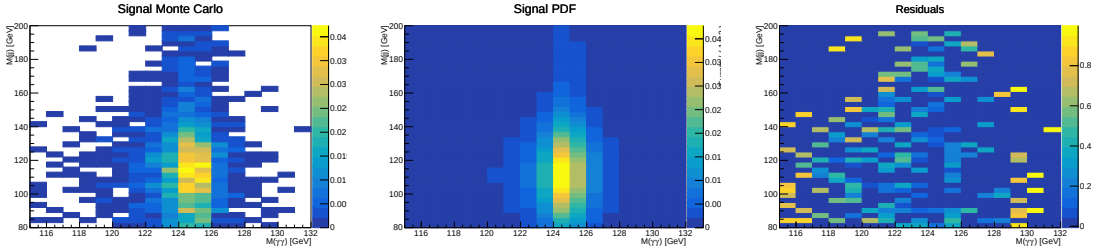

Figure 5.36: 2D distributions of the signal MC (left), fitted PDF model (center) and 2D residuals (right) for the Low Mass-Medium Purity Category non-resonant selection.

in our model PDF. To study our sensitivity to these missing degrees of freedom, we construct a new PDF adding back three new parameters, namely:

$$g_{corr}(x,y) = \left(\sum_{i=0}^{i=2} a_i x^i\right)\left(\sum_{k=0}^{k=2} a_k y^k\right) + \alpha \cdot M(\gamma\gamma) \cdot M(jj) \tag{5.10}$$

$$+ \quad \beta \cdot M(\gamma\gamma)^2 \cdot M(jj) + \omega \cdot M(\gamma\gamma) \cdot M(jj)^2. \tag{5.11}$$

We perform two tests with this PDF:

- We generate Asimov datasets [119] with $g_{corr}(x,y)$ for varying (α, β, ω) and then fit it with $g(x,y)$. Then we check the residuals comparing $g_{corr}(x,y)$ and $g(x,y)$ assuming different normalizations (i.e., different number of expected background events).
- We generate toy datasets with $g_{corr}(x,y)$ for varying (α, β, ω), with different values for the expected number of background events, with injected signal. Then we measure back the signal strength by using $g(x,y)$ and check the bias ($B = (\mu_{measured} - \mu_{true})/\sigma_\mu$).

In Figure 5.41, the 2D distributions of $g_{corr}(x,y)$ for different values of α, where the change in correlation between x and y can be seen. In Figure 5.42, the 2D distributions of $g(x,y)$ fitted to the Asimov datasets produced with $g_{corr}(x,y)$ with different values of α, where the change in correlation between x and y can be also be seen, albeit different from $g_{corr}(x,y)$. Therefore, we need to measure how sensitive we are to that difference. The first check is to calculate the 2D residuals, as was done for the signal correlation tests, between these two hypotheses,

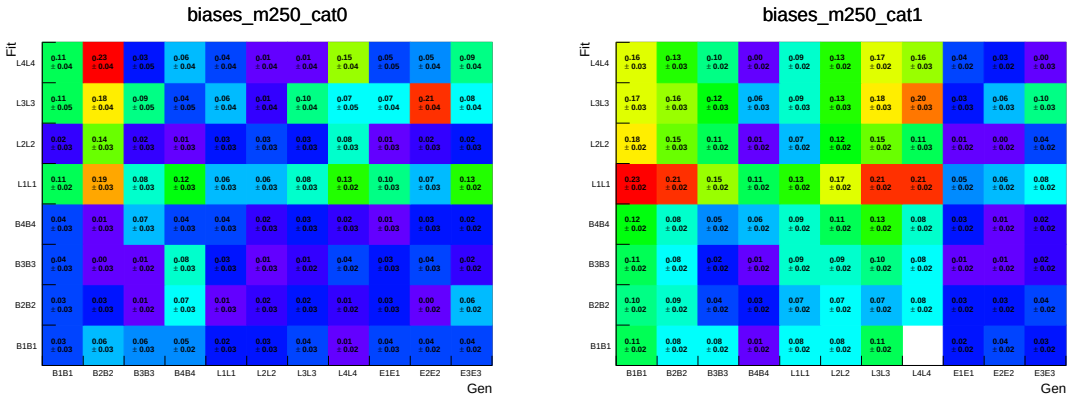

Figure 5.37: Biases measured in the 250 GeV resonant selection in the HPC (left) and MPC (right).

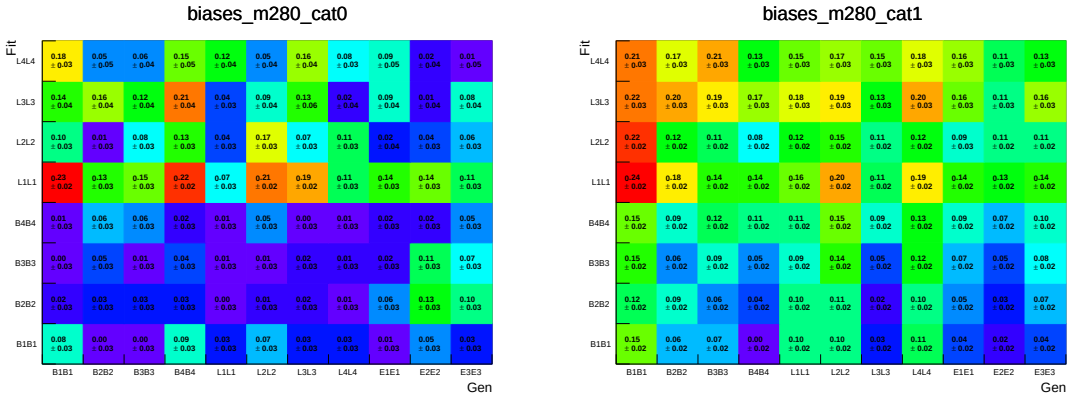

Figure 5.38: Biases measured in the 280 GeV resonant selection in the HPC (left) and MPC (right).

for different background normalizations. The residuals with the background normalized to 200 events can be seen in Figure 5.43, and for 100000 events in Figure 5.44. While very little statistically significant deviation is seen for 200 background events, structures do start to appear with 100k background events, which is an expected behavior. This test was further performed with 10, 100, 500, 1000, 5000 background events, with conclusions similar to the 200 case. The test was also performed with varying β and ω with similar conclusions.

For the second test, instead of generating Asimov datasets, we generate toy MC for the different normalizations and (α, β, ω) hypotheses. We then show the bias measurement for these different cases, in the hypothesis of varying α, in Figure 5.45. Since no bias larger than 14% is seen, we don't include any systematics on the signal strength due to possible background correlations that are not modeled by our choice of PDF.

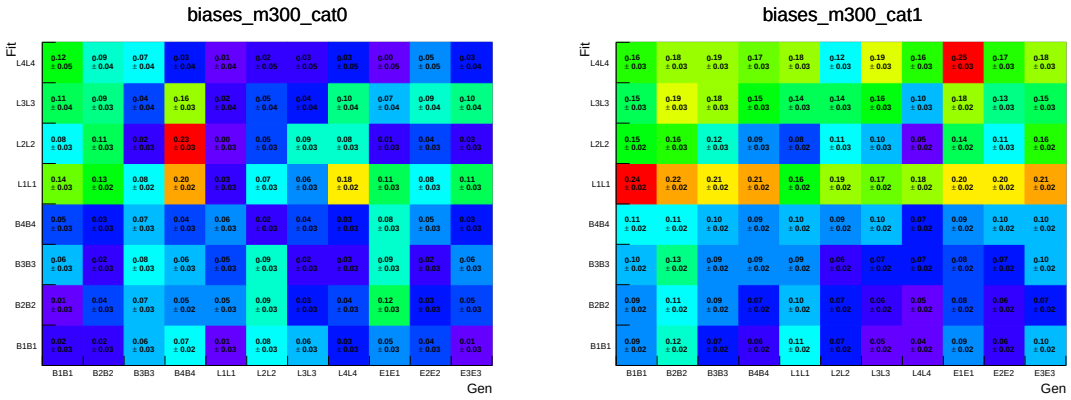

Figure 5.39: Biases measured in the 300 GeV resonant selection in the HPC (left) and MPC (right).

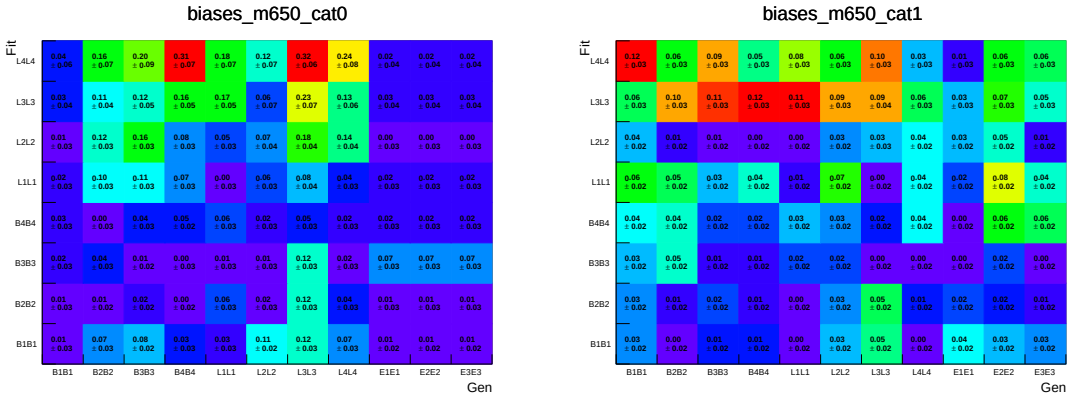

Figure 5.40: Biases measured in the 650 GeV resonant selection in the HPC (left) and MPC (right).

5.8.3 Single Higgs Background Modeling

Apart from the smoothly-falling background expected, depending on the integrated luminosity with which it is performed, the non-resonant analysis is also sensitive to the SM single Higgs production as a background process. In the resonant case, the mass window requirement reduces the single Higgs contributions to negligible levels, and therefore is not considered.

The SM single Higgs background consists of the Higgs resonance in $M(\gamma\gamma)$ and a $M(jj)$ shape that depends on the production mechanism. For single Higgs produced via gluon fusion and vector boson fusion, the two extra jets will constitute a smoothly falling background, therefore, we model this contribution in the 2D $M(\gamma\gamma) : M(jj)$ plane with a product of a double sided Crystal-Ball (similar to our signal model) and a second order Bernstein. For single Higgs produced in association with top quarks, bottom quarks and a vector boson, we are also able to model $M(jj)$ with a double-sided Crystal-Ball function, given the kinematic turn-on present in the first two cases, and the $V \to jj$ resonance in the latter. The Higgs model fits, in the High Purity category as an example, are shown in Figures 5.46, 5.47, 5.48,

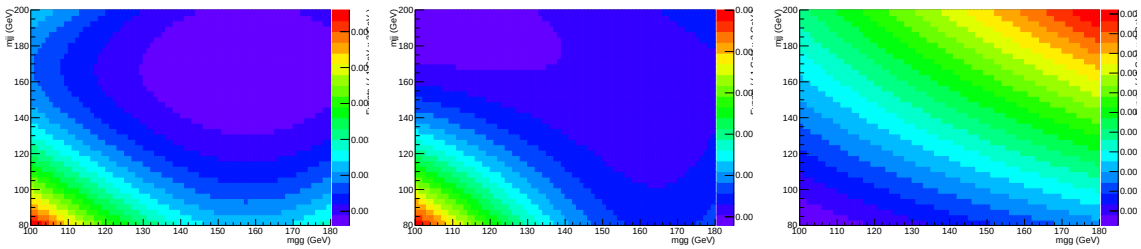

Figure 5.41: 2D distributions of $g_{corr}(M(\gamma\gamma), M(jj))$ with $\alpha = 0.0$, 0.5, 1.0, from left to right.

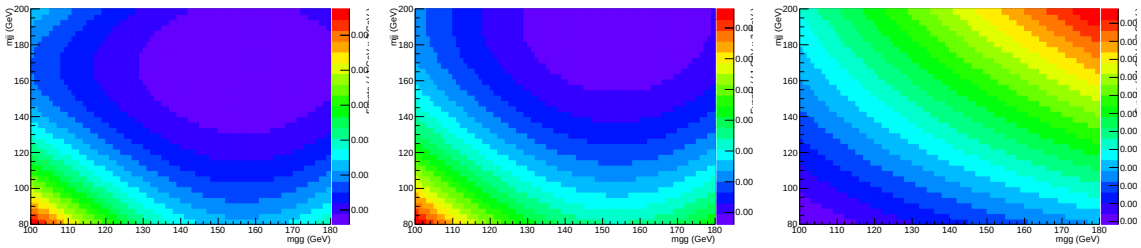

Figure 5.42: 2D distributions of $g(M(\gamma\gamma), M(jj))$ fitted to the Asimov datasets produced with $g_{corr}(x, y)$ with $\alpha = 0.0$, 0.5, 1.0, from left to right.

5.49 and 5.50. The cross sections used for the SM single Higgs estimations are listed in Table 5.8.3, along with their efficiencies in the four different non-resonant analysis categories.

	Cross section (pb)	HM-HPC (%)	HM-MPC (%)	LM-HPC (%)	LM-MPC (%)
ggH	44.14	0.029 ± 0.0017	0.148 ± 0.0038	0.033 ± 0.0018	0.151 ± 0.0039
VBF	3.7820	0.038 ± 0.001	0.239 ± 0.0025	0.048 ± 0.0011	0.242 ± 0.0025
VH	2.257	0.271 ± 0.0038	0.748 ± 0.0063	0.367 ± 0.0044	0.962 ± 0.0071
$b\bar{b}$H	0.488	0.0297 ± 0.0035	0.262 ± 0.010	1.02 ± 0.020	2.59 ± 0.032
$t\bar{t}$H	0.5071	3.41 ± 0.027	3.69 ± 0.029	8.38 ± 0.042	8.17 ± 0.042

Table 5.6: SM single Higgs cross sections at 13 TeV with their respective selection efficiencies for the four different non-resonant analysis categories: High Mass-High Purity, High Mass-Medium Purity, Low Mass-High Purity and Low Mass-Medium Purity categories.

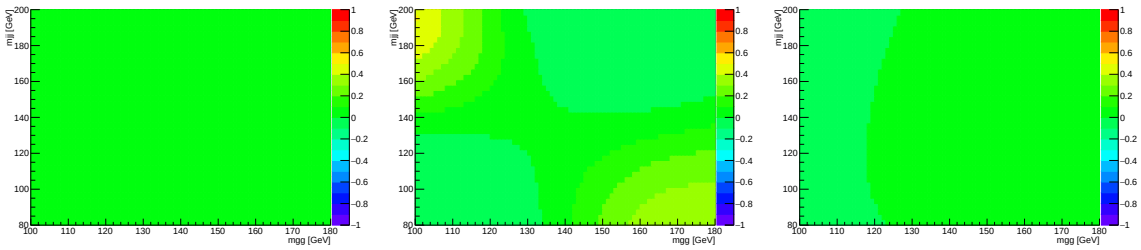

Figure 5.43: 2D residuals comparing distributions of $g(M(\gamma\gamma), M(jj))$ fitted to the Asimov datasets produced with $g_{corr}(x, y)$ with $\alpha = 0.0$, 0.5, 1.0 and the dataset, from left to right. The background normalization is 200 events.

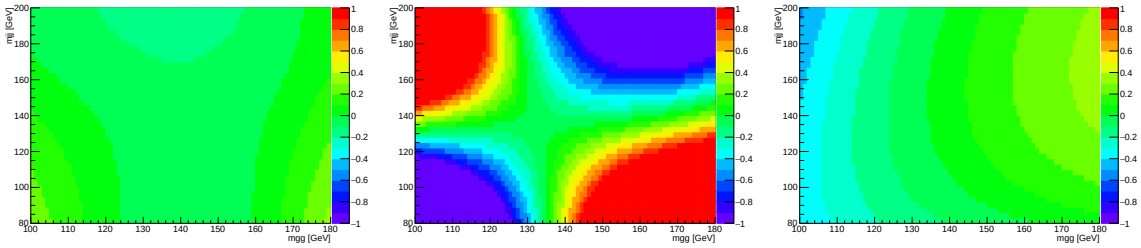

Figure 5.44: 2D residuals comparing distributions of $g(M(\gamma\gamma), M(jj))$ fitted to the Asimov datasets produced with $g_{corr}(M(\gamma\gamma), M(jj))$ with $\alpha = 0.0$, 0.5, 1.0 and the dataset, from left to right. The background normalization is 100k events.

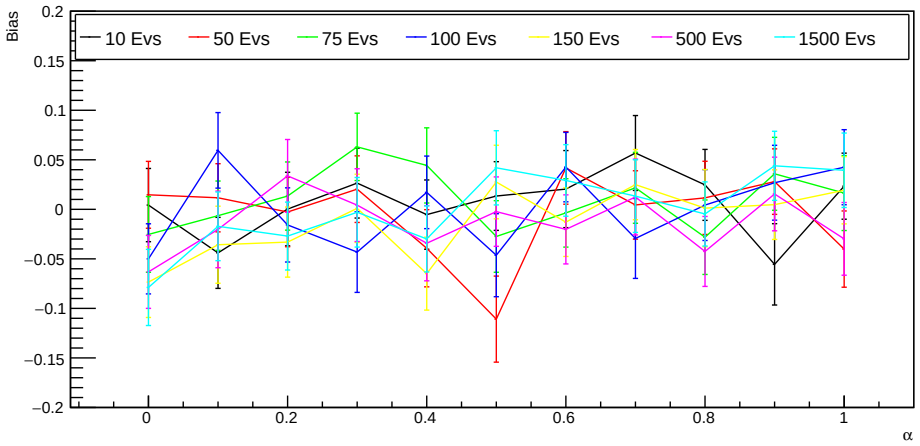

Figure 5.45: Relative bias on measuring the signal with $g(M(\gamma\gamma), M(jj))$ on toys created with $g_{corr}(M(\gamma\gamma), M(jj))$ with α from 0 to 1, for different background normalization hypotheses.

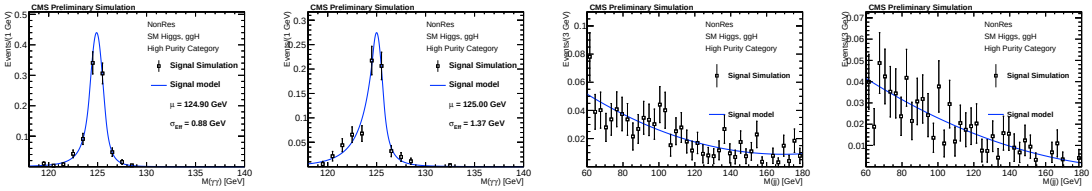

Figure 5.46: Higgs model fit to Higgs Monte Carlo (ggH) in the High Purity Category. From left to right: $M(\gamma\gamma)$ in high mass region, $M(\gamma\gamma)$ in low mass region, $M(jj)$ in high mass region, $M(jj)$ in low mass region.

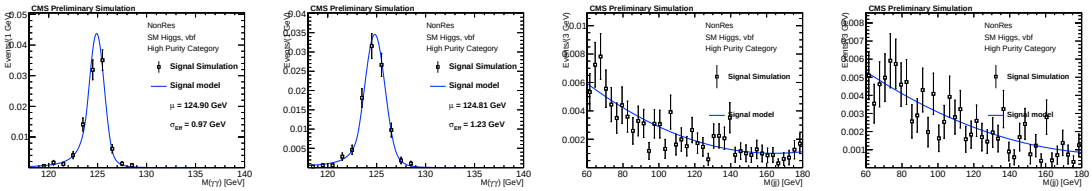

Figure 5.47: Higgs model fit to Higgs Monte Carlo (VBF) in the High Purity Category. From left to right: $M(\gamma\gamma)$ in high mass region, $M(\gamma\gamma)$ in low mass region, $M(jj)$ in high mass region, $M(jj)$ in low mass region.

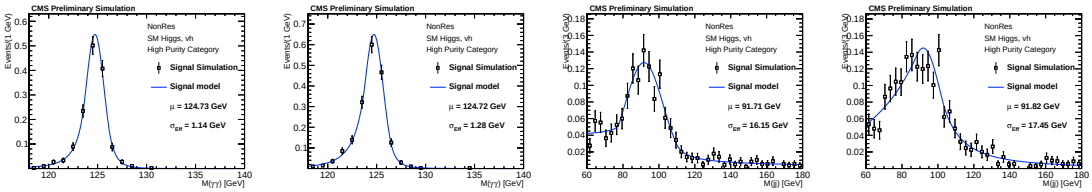

Figure 5.48: Higgs model fit to Higgs Monte Carlo (VH) in the High Purity Category. From left to right: $M(\gamma\gamma)$ in high mass region, $M(\gamma\gamma)$ in low mass region, $M(jj)$ in high mass region, $M(jj)$ in low mass region.

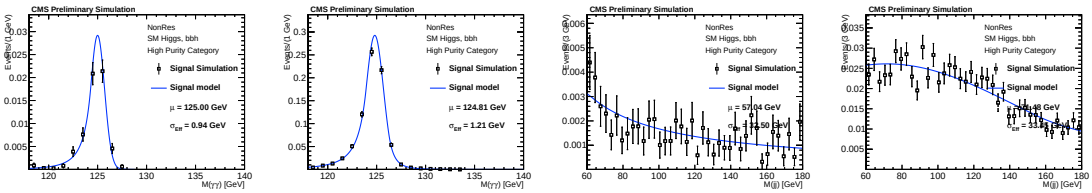

Figure 5.49: Higgs model fit to Higgs Monte Carlo (bbH) in the High Purity Category. From left to right: $M(\gamma\gamma)$ in high mass region, $M(\gamma\gamma)$ in low mass region, $M(jj)$ in high mass region, $M(jj)$ in low mass region.

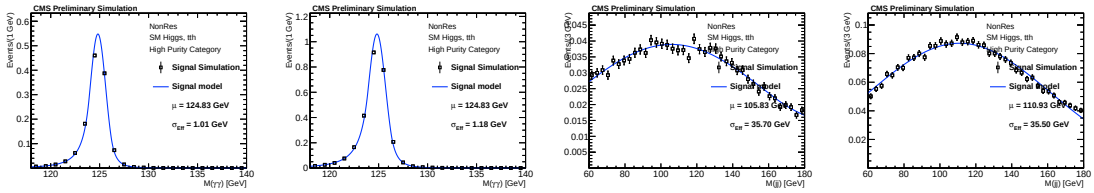

Figure 5.50: Higgs model fit to Higgs Monte Carlo (ttH) in the High Purity Category. From left to right: $M(\gamma\gamma)$ in high mass region, $M(\gamma\gamma)$ in low mass region, $M(jj)$ in high mass region, $M(jj)$ in low mass region.

5.9 Systematical Uncertainties

The expected number of signal events in this analysis is projected with Monte Carlo simulation. Possible differences between data reconstruction and Monte Carlo reconstruction are usually corrected with data/Monte Carlo scale factors. In this analysis, the systematics related to those differences come in two possible categories: normalization uncertainties and shape uncertainties. The normalization systematics are related to the uncertainty in the expected number of signal events. This is due to different efficiencies of the analysis selections in data and Monte Carlo reconstructions. Shape systematics are important in this analysis because the signal shapes enter the signal extraction procedure in the parametric fits. Therefore, uncertainties in the shape of the $M(\gamma\gamma)$ and $M(jj)$ distributions must be included as systematics.

Both normalization and shape uncertainties come from photons and jets. Since the photons in this analysis are selected with the same selection criteria as the SM $H \to \gamma\gamma$ analysis, we take the photon related systematics from the $H \to \gamma\gamma$ analysis. This includes the photon energy scale (PES) and photon energy resolution (PER). These two uncertainties are translated into two shape systematics ($\Delta M(\gamma\gamma)/M(\gamma\gamma)$ and $\Delta\sigma_{M(\gamma\gamma)}/\sigma_{M(\gamma\gamma)}$), and into the photon selection selection acceptance uncertainty (which inclues the trigger pre-selection requirements). The PES also has to cover effects of linearity in the energy scale for high E_T. For this, $\Delta M(\gamma\gamma)/M(\gamma\gamma)$ is kept at 0.05% for 2016.

For jets, the jet energy scale (JES) and jet energy resolution (JER) are important ingredients in the list of systematics. As for photons, they enter in the analysis in two shape systematics ($\Delta M(jj)/M(jj)$ and $\Delta\sigma_{M(jj)}/\sigma_{M(jj)}$), and in the jet selection acceptance uncertainty (related to the jet kinematic requirements). An extra jet related systematic is related to the b-tagging requirements. The analysis has defined four b-tagging regions in total (two for resonant and two for non-resonant); for each one, the uncertainty of the b-tagging efficiency must be taken into account.

An extra set of normalization systematics are needed because of the mass window requirement in the resonant analysis. This systematic is related to the change in signal efficiency after variations of PES/PER/JES/JER.

A systematic due to the uncertainty in the integrated luminosity measurement in CMS is included.

No theory systematics are applied to our BSM signals.

The values of those quantities are shown in Table 5.7 for 2015 data and will be updated with 2016 values.

5.9.1 Signal shape smearings

One important ingredient when applying the analysis systematics is the smearing of the signal shapes. After the signal model is fitted to the signal simulation, all PDF parameters are fixed. Following, the signal mean and width are then multiplied by smearing factors related to scale and resolution uncertainties, respectively. While this is not problematic for the mean, since it merely causes a scaling of the x-axis, this might not be the case for the resolution smearing. In the latter case, the tail parameters of the signal modeling can be affected by the smearing and not correspond to the frozen parameters from the pre-smearing fit. To test this, we fit the

Sources of Systematical Uncertainties	Type	Value
General uncertainties		
Integrated luminosity	Normalization	2.7%
Photon related uncertainties		
Photon energy scale ($\frac{\Delta M(\gamma\gamma)}{M(\gamma\gamma)}$)	Shape	1.0%
Photon energy resolution ($\frac{\Delta \sigma_{\gamma\gamma}}{\sigma_{\gamma\gamma}}$)	Shape	1.0%
Diphoton pre-selection (with trigger uncertainties)	Normalization	2.0%
Photon Identification	Normalization	1.0%
Jet related uncertainties		
Jet energy scale ($\frac{\Delta M(jj)}{M(jj)}$)	Shape	2.0%
Jet energy resolution ($\frac{\Delta \sigma_{jj}}{\sigma_{jj}}$)	Shape	8.0%
Resonant specific uncertainties		
Mass window selection (with jet selection uncertainty)	Normalization	5.0%
b tagging efficiency (Low Mass, high purity)	Normalization	2.5%
b tagging efficiency (Low Mass, medium purity)	Normalization	1.0%
b tagging efficiency (High Mass)	Normalization	1.0%
Nonresonant specific uncertainties		
Jet Selection plus $\tilde{M}_{\mathrm{X}} > 350$ GeV	Normalization	3.0%
b tagging efficiency (high purity)	Normalization	4.5%
b tagging efficiency (medium purity)	Normalization	1.0%

Table 5.7: Summary of systematic uncertainties. The uncertainty in the b tagging efficiency is anticorrelated between the b tag categories.

smeared MC (MC with photon and jet energy resolution smearing uncertainty values applied) with the signal PDF with the tails fixed to their values from the un-smeared MC (MC with central values of smearings) and we compare with the smeared MC fit with the signal PDF without fixing the tail parameters - these different fits are shown in Figures 5.51 and 5.52. We have also compared the 2D residuals (as defined in the Section 5.8) of the fixed tails PDF vs the smeared MC and with the floating tails PDF vs the smeared MC, shown in Figure 5.53. Additionally, we compare the fixed and floating fit shapes in Figure 5.54. Since no issue has been seen, we continue using the procedure described in the previous section for applying the signal smearing.

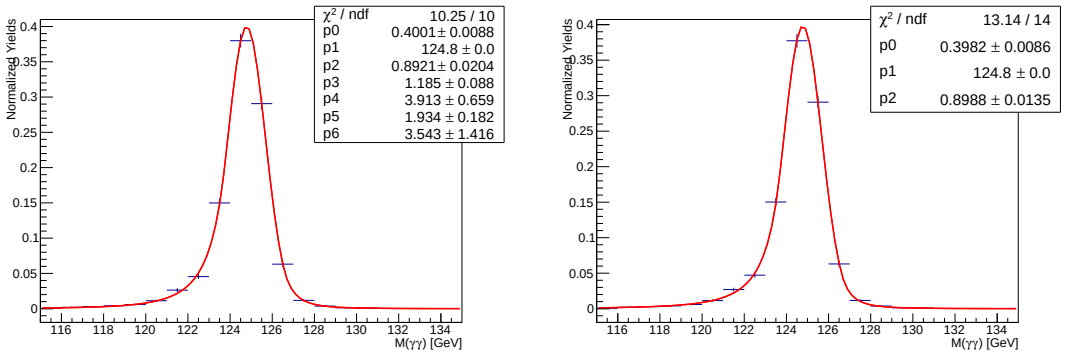

Figure 5.51: Signal fits for the High Mass-Medium purity categories, in the un-smeared MC with floating tails (left) and in the smeared MC with fixed tails (right). The tail parameters on the right are fixed to their values on the left.

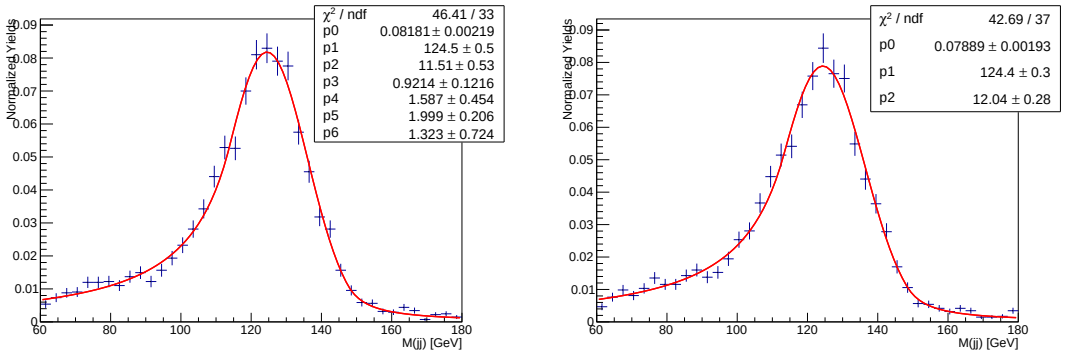

Figure 5.52: Signal fits for the High Mass-Medium Purity category, in the un-smeared MC with floating tails (left) and in the smeared MC with fixed tails (right). The tail parameters on the right are fixed to their values on the left.

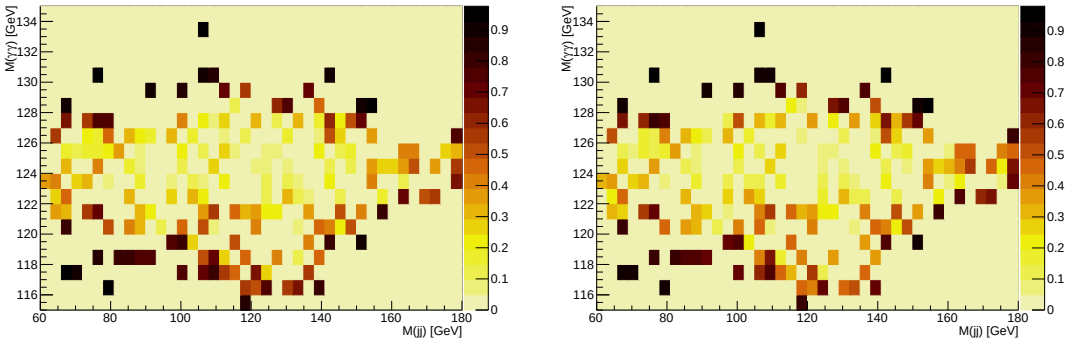

Figure 5.53: Residuals comparing the floating signal PDF fit to the smeared MC (left) and the fixed tails PDF fit to the smeared MC (right) in the High Mass-Medium Purity category.

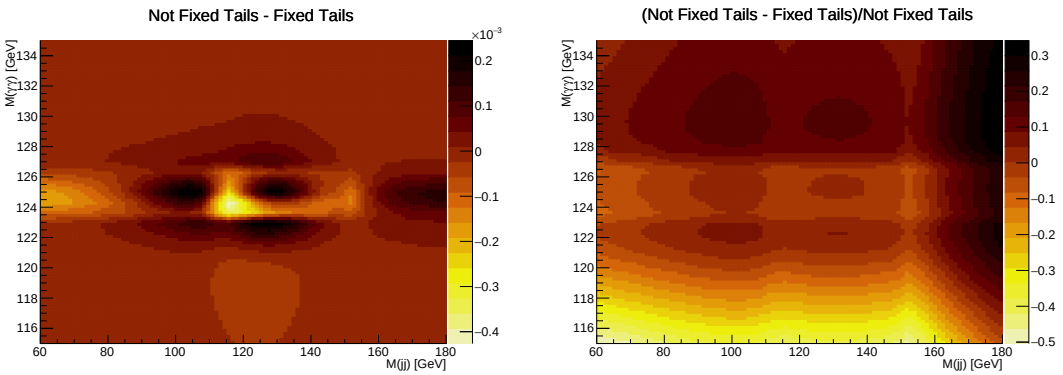

Figure 5.54: Comparison between fixed and floating PDF shapes when fitting the smeared MC. Left: absolute difference between fixed and floating PDF shapes when fitting the smeared MC. Right: difference normalized by floating PDF.

5.10 Results

The expected limits shown in this section are obtained with the Higgs Combination tool, with the Asymptotic method. Figure 5.55 shows the results on spin-0 resonances. Figures 5.56 (in fb) shows the SM-like non-resonant limit and its breakdown in the different analysis categories: LM (Low Mass), HM (High Mass), MPC (Medium Purity Category), and HPC (High Purity Category).

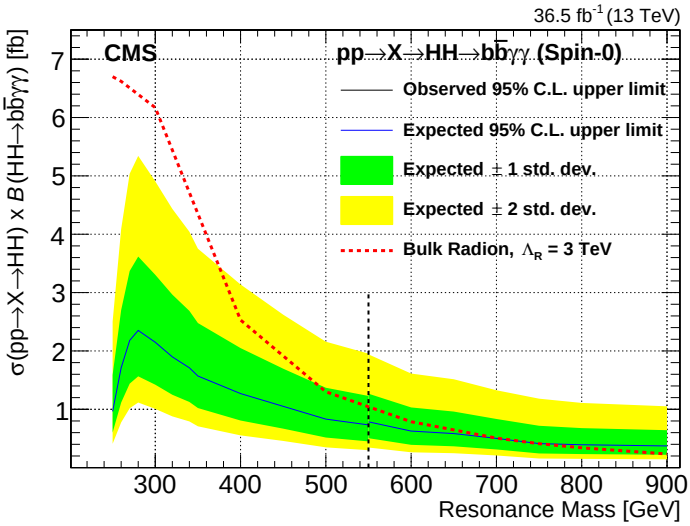

Figure 5.55: Limits on spin-0 resonances.

As a comparison, expected and observed limits obtained with 2.7 fb^{-1} are shown in Figures 5.57 and 5.58. The SM-like non-resonant result using 2015 data is a limit about 90 times the SM cross-section. Scaling this result using a frequentist approach, with the square root of the integrated luminosity ratio of 2.7/35.87, the expected limit with the 2016 dataset is about 25 times the SM cross section. The expected limit obtained with the 2016 analysis is, on the other hand, 16.7 times the SM. This represents an improvement on the baseline analysis sensitivity of over 30%.

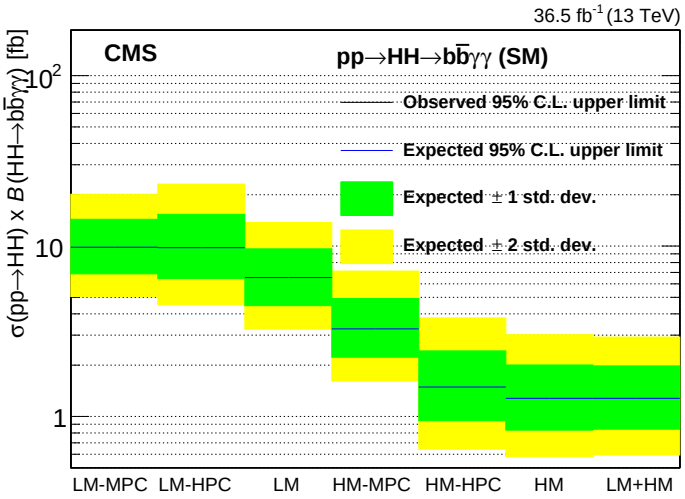

Figure 5.56: SM-like non-resonant limits.

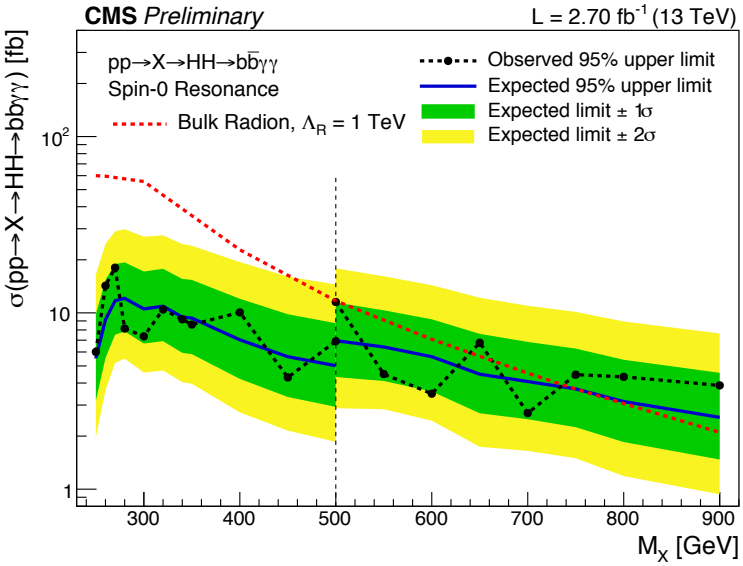

Figure 5.57: Limits on spin-0 resonances with the 2015 dataset.

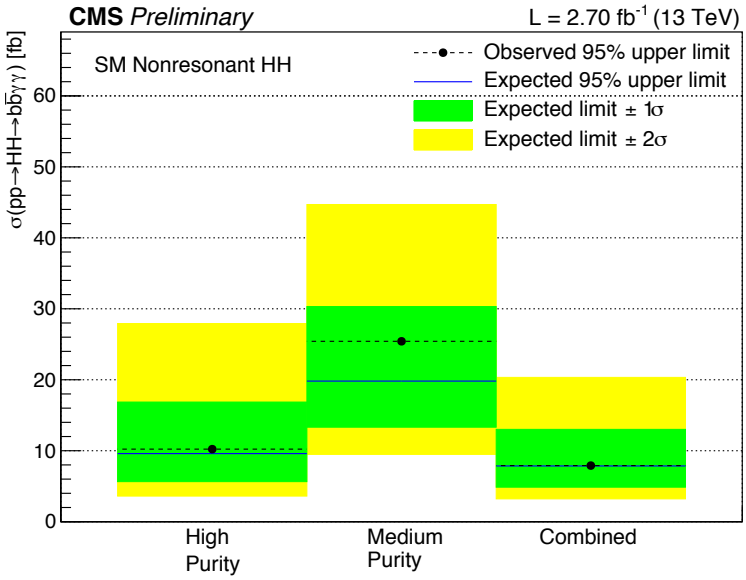

Figure 5.58: SM-like non-resonant limits with the 2015 dataset.

Chapter 6

Conclusions

This thesis is devoted to the study of the Higgs boson as a portal to beyond the Standard Model physics, a theoretically and experimentally rich pursuit at the LHC experiments. The capability of coupling the Higgs boson to new physics ensures the Higgs as an invaluable tool in this new era of particle physics.

The analyses presented in this document look for new physics coupling with the Higgs both in its decay (Higgs decays with photons and MET) and in its production (Higgs pair production). The Higgs decays with photons and MET analysis at CMS was the first of its kind in this exotic final state, but also in exploring a new kinematic regime for photons and MET production. The di-Higgs search in the $b\bar{b}\gamma\gamma$ final state, with CMS Run 2 data, was a continuation of a long effort, dating back to Run 1, which culminated in the strongest limit on SM-like HH production to date (as of April 2017). While built on solid ground, many new features were added to the 2016 $HH \to b\bar{b}\gamma\gamma$ analysis, such as the novel categorization scheme and the dedicated b-jet energy regression. These changes improved the analysis sensitivity to SM-like HH production by over 30% from the 2015 analysis baseline.

Both analyses presented in this thesis profit enormously from the performance, online and offline, of the CMS electromagnetic calorimeter. A large part of the work performed during the research period of this thesis was dedicated to ensuring this performance, particularly online, as a part of the ECAL DAQ team responsible for maintaining and upgrading the online ECAL software. This included taking leadership roles during the Run 2 recommissioning period in re-integrating ECAL in the upgraded CMS central DAQ framework and interfacing with the newly installed TCDS. From 2014 to 2016, several DAQ tools were also integrated in the ECAL online software, such as back-end electronics monitoring tools and the ability to perform different and independent types of runs.

Even though it's been running for almost a decade now, the LHC program is still in its infancy. Given the programmed upgrades beyond 2020, with 3000 fb^{-1} expected to be delivered in total, we are still at about 1% of the overall amount of data to be analyzed by the LHC experiments. The full LHC dataset will provide an invaluable opportunity to study the SM at high energies with an unparalleled precision. Even the most remote corners will be available then, such as the elusive Higgs self-coupling, explored in this thesis with the LHC Run 2 dataset. While physics beyond the Standard Model remains elusive, the search goes on - it might just be right around the next corner.

Bibliography

[1] S. Weinberg, *A model of leptons*, *Phys. Rev. Lett.* **19** (Nov, 1967) 1264–1266.

[2] A. Salam, *Weak and Electromagnetic Interactions*, *Conf.Proc.* **C680519** (1968) 367–377.

[3] S. Glashow, *Partial Symmetries of Weak Interactions*, *Nucl.Phys.* **22** (1961) 579–588.

[4] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, *Phys. Rev. Lett.* **13** (Aug, 1964) 321–323.

[5] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global conservation laws and massless particles*, *Phys. Rev. Lett.* **13** (Nov, 1964) 585–587.

[6] P. W. Higgs, *Broken symmetries and the masses of gauge bosons*, *Phys. Rev. Lett.* **13** (Oct, 1964) 508–509.

[7] H. Fritzsch, M. Gell-Mann, and H. Leutwyler, *Advantages of the Color Octet Gluon Picture*, *Phys. Lett.* **B47** (1973) 365–368.

[8] K. G. Wilson, *Confinement of quarks*, *Phys. Rev. D* **10** (Oct, 1974) 2445–2459.

[9] D. J. Gross and F. Wilczek, *Ultraviolet behavior of non-abelian gauge theories*, *Phys. Rev. Lett.* **30** (Jun, 1973) 1343–1346.

[10] H. D. Politzer, *Reliable perturbative results for strong interactions?*, *Phys. Rev. Lett.* **30** (Jun, 1973) 1346–1349.

[11] Beringer, J. at all, *Review of Particle Physics*, *Phys. Rev. D* **86** (2012).

[12] **LHC Higgs Cross Section Working Group** Collaboration, J. R. Andersen et al., *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, `arXiv:1307.1347`.

[13] **CMS** Collaboration, S. Chatrchyan et al., *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, *Phys. Lett.* **B716** (2012) 30–61, [`arXiv:1207.7235`].

[14] **ATLAS** Collaboration, G. Aad et al., *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett.* **B716** (2012) 1–29, [`arXiv:1207.7214`].

[15] **CMS** Collaboration, V. Khachatryan et al., *Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV*, *Phys. Rev.* **D92** (2015), no. 1 012004, [arXiv:1411.3441].

[16] **ATLAS** Collaboration, G. Aad et al., *Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector*, *Eur. Phys. J.* **C75** (2015), no. 10 476, [arXiv:1506.0566]. [Erratum: Eur. Phys. J.C76,no.3,152(2016)].

[17] **ATLAS, CMS** Collaboration, G. Aad et al., *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*, *JHEP* **08** (2016) 045, [arXiv:1606.0226].

[18] **CMS** Collaboration, V. Khachatryan et al., *Measurement of differential cross sections for Higgs boson production in the diphoton decay channel in pp collisions at $\sqrt{s} = 8$ TeV*, *Eur. Phys. J.* **C76** (2016), no. 1 13, [arXiv:1508.0781].

[19] P. Fayet, *Mixing between gravitational and weak interactions through the massive gravitino*, *Phys. Lett.* **B70** (1977) 461.

[20] H. Baer, M. Brhlik, C. H. Chen, and X. Tata, *Signals for the minimal gauge-mediated supersymmetry breaking model at the fermilab tevatron collider*, *Phys. Rev.* **D55** (1997) 4463.

[21] H. Baer, P. G. Mercadante, X. Tata, and Y. L. Wang, *Reach of tevatron upgrades in gauge-mediated supersymmetry breaking models*, *Phys. Rev.* **D60** (1999) 055001.

[22] S. Dimopoulos, S. Thomas, and J. D. Wells, *Sparticle spectroscopy and electroweak symmetry breaking with gauge-mediated supersymmetry breaking*, *Nucl. Phys.* **B488** (1997) 39.

[23] J. R. Ellis, J. L. Lopez, and D. V. Nanopoulos, *Analysis of lep constraints on supersymmetric models with a light gravitino*, *Phys. Lett.* **B394** (1997) 354.

[24] M. Dine, A. Nelson, Y. Nir, and Y. Shirman, *New tools for low energy dynamical supersymmetry breaking*, *Phys. Rev.* **D53** (1996) 2658.

[25] G. F. Giudice and R. Rattazzi, *Gauge-mediated supersymmetry breaking*, in *Perspectives on Supersymmetry*, p. 355. World Scientific, Singapore, 1998.

[26] G. C. Branco, P. M. Ferreira, L. Lavoura, M. N. Rebelo, M. Sher, and J. P. Silva, *Theory and phenomenology of two-Higgs-doublet models*, *Phys. Rept.* **516** (2012) 1–102, [arXiv:1106.0034].

[27] J. R. Espinosa, *Implications of the top (and Higgs) mass for vacuum stability*, *PoS* **TOP2015** (2016) 043, [arXiv:1512.0122].

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

[29] **CMS** Collaboration, S. Chatrchyan et al., *The CMS Experiment at the CERN LHC*, *JINST* **3** (2008) S08004.

[30] **CMS** Collaboration, F. Beaudette, *The CMS Particle Flow Algorithm*, in *Proceedings, International Conference on Calorimetry for the High Energy Frontier (CHEF 2013): Paris, France, April 22-25, 2013*, pp. 295–304, 2013. arXiv:1401.8155.

[31] **CMS** Collaboration, S. Chatrchyan et al., *Description and performance of track and primary-vertex reconstruction with the CMS tracker*, *JINST* **9** (2014), no. 10 P10009, [arXiv:1405.6569].

[32] **CMS** Collaboration, V. Khachatryan et al., *Performance of Photon Reconstruction and Identification with the CMS Detector in Proton-Proton Collisions at sqrt(s) = 8 TeV*, *JINST* **10** (2015), no. 08 P08010, [arXiv:1502.0270].

[33] **CMS Collaboration** Collaboration, *Jet algorithms performance in 13 TeV data*, Tech. Rep. CMS-PAS-JME-16-003, CERN, Geneva, 2017.

[34] **CMS Collaboration** Collaboration, *Identification of double-b quark jets in boosted event topologies*, Tech. Rep. CMS-PAS-BTV-15-002, CERN, Geneva, 2016.

[35] G. Bauer et al., *The cms data acquisition system software*, *Journal of Physics: Conference Series* **219** (2010), no. 2 022011.

[36] M. Bellato et al., *Run control and monitor system for the CMS experiment*, *CoRR* **cs.DC/0306110** (2003).

[37] P. Adzic et al., *Reconstruction of the signal amplitude of the CMS electromagnetic calorimeter*, *Eur. Phys. J.* **C46S1** (2006) 23–35.

[38] M. Anfreville et al., *Laser monitoring system for the CMS lead tungstate crystal calorimeter*, *Nucl. Instrum. Meth.* **A594** (2008) 292–320.

[39] **CMS** Collaboration, S. Chatrchyan et al., *Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in pp Collisions at* $\sqrt{s} = 7$ *TeV*, *JINST* **8** (2013) P09009, [arXiv:1306.2016]. [JINST8,9009(2013)].

[40] **CMS** Collaboration, V. Khachatryan et al., *Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at ?s = 8 TeV*, *JINST* **10** (2015), no. 06 P06005, [arXiv:1502.0270].

[41] P. Adzic, *Energy resolution of the barrel of the cms electromagnetic calorimeter*, *Journal of Instrumentation* **2** (2007), no. 04 P04004.

[42] **CMS Collaboration** Collaboration, *First results on Higgs to gammagamma at 13 TeV*, Tech. Rep. CMS-PAS-HIG-15-005, CERN, Geneva, 2016.

[43] D. Contardo, M. Klute, J. Mans, L. Silvestris, and J. Butler, *Technical Proposal for the Phase-II Upgrade of the CMS Detector*, Tech. Rep. CERN-LHCC-2015-010. LHCC-P-008. CMS-TDR-15-02, Geneva, Jun, 2015. Upgrade Project Leader Deputies:

Lucia Silvestris (INFN-Bari), Jeremy Mans (University of Minnesota) Additional contacts: Lucia.Silvestris@cern.ch, Jeremy.Mans@cern.ch.

[44] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, *Physics Letters B* **716** (2012) 30.

[45] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Physics Letters B* **716** (2012) 1.

[46] C. Petersson, A. Romagnoni, and R. Torre, *Higgs Decay with Monophoton + MET Signature from Low Scale Supersymmetry Breaking*, *JHEP* **10** (2012) 016, [arXiv:1203.4563].

[47] T. Asaka, K. Hamaguchi, and K. Suzuki, *Cosmological gravitino problem in gauge mediated supersymmetry breaking models*, *Phys. Lett.* **B490** (2000) 136–146, [hep-ph/0005136].

[48] CMS Collaboration, *Search for Dark Matter and Large Extra Dimensions in pp Collisions Yielding a Photon and Missing Transverse Energy*, *Phys.Rev.Lett.* **108** (2012) 261803, [arXiv:1204.0821].

[49] ATLAS Collaboration, *Search for dark matter candidates and large extra dimensions in events with a photon and missing transverse momentum in pp collision data at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *Phys.Rev.Lett.* **110** (2013) 011802, [arXiv:1209.4625].

[50] A. Apresyan on behalf of CMS Collaboration, *Identification and mitigation of anomalous signals in CMS HCAL*, CMS Conference Report CMS-CR-2012-238, 2012.

[51] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet User Manual*, *Eur.Phys.J.* **C72** (2012) 1896, [arXiv:1111.6097].

[52] W. Bialas, D.A. Petyt on behalf of CMS Collaboration, *Mitigation of anomalous apd signals in the CMS ECAL*, *JINST* **8** (2013) C03020.

[53] **CMS** Collaboration, CMS Collaboration, *Performance of electron reconstruction and selection at the cms detector at $\sqrt{s}$=8 TeV*, CMS Physics Analysis Summary CMS-PAS-EGM-13-001, 2010.

[54] CMS Collaboration, *Pileup jet identification*, CMS Physics Analysis Summary CMS-PAS-JME-13-005, 2013.

[55] CMS Collaboration, *Missing transverse energy performance of the CMS detector*, *JINST* **6** (2011) P09001, [arXiv:1106.5048].

[56] **CMS Collaboration** Collaboration, V. Khachatryan et al., *Search for new phenomena in monophoton final states in proton-proton collisions at $\sqrt{s}$ = 8 TeV*, arXiv:1410.8812.

[57] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, *MadGraph 5 : Going Beyond*, *JHEP* **06** (2011) 128, [arXiv:1106.0522].

[58] T. Sjöstrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 physics and manual*, *JHEP* **05** (2006) 026, [hep-ph/0603175].

[59] J. M. Campbell and R. K. Ellis, *Update on vector boson pair production at hadron colliders*, *Phys. Rev. D* **60** (Nov, 1999) 113006.

[60] **GEANT4** Collaboration, S. Agostinelli et al., *Geant4–a simulation toolkit*, *Nucl. Instrum. Meth.* **A 506** (2003) 250.

[61] CMS Collaboration, *Measurement of the Inclusive W and Z Production Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV*, *JHEP* **10** (2011) 132, [arXiv:1107.4789].

[62] CMS Collaboration, *CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update*, CMS Physics Analysis Summary CMS-PAS-LUM-13-001, 2013.

[63] CMS Collaboration, *Performance of photon reconstruction and identification in proton-proton collisions at $\sqrt{s}$ = 8 tev*, paper in preparation CMS-EGM-14-001.

[64] M. Botje, J. Butterworth, A. Cooper-Sarkar, A. de Roeck, J. Feltesse, S. Forte, A. Glazov, J. Huston, R. McNulty, T. Sjöstrand, and R. Thorne, "The PDF4LHC Working Group Interim Recommendations." 2011.

[65] R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, et al., *Impact of Heavy Quark Masses on Parton Distributions and LHC Phenomenology*, *Nucl. Phys.* **B 849** (2011) 296–363, [arXiv:1101.1300].

[66] A. Martin, W. Stirling, R. Thorne, and G. Watt, *Parton distributions for the LHC*, *Eur. Phys. J. C* **63** (2009) 189, [arXiv:0901.0002].

[67] A. L. Read, *Presentation of search results: The CL(s) technique*, *J. Phys.* **G 28** (2002) 2693.

[68] T. Junk, *Confidence level computation for combining searches with small statistics*, *Nucl. Instrum. Meth.* **A 434** (1999) 435–443, [hep-ex/9902006].

[69] D. Curtin et al., *Exotic decays of the 125 GeV Higgs boson*, *Phys. Rev.* **D90** (2014), no. 7 075004, [arXiv:1312.4992].

[70] A. Djouadi and M. Drees, *Higgs boson decays into light gravitinos*, *Phys. Lett.* **B407** (1997) 243–249, [hep-ph/9703452].

[71] J. D. Mason, D. E. Morrissey, and D. Poland, *Higgs Boson Decays to Neutralinos in Low-Scale Gauge Mediation*, *Phys. Rev.* **D80** (2009) 115015, [arXiv:0909.3523].

[72] **CMS** Collaboration, V. Khachatryan et al., *Search for exotic decays of a Higgs boson into undetectable particles and one or more photons*, *Phys. Lett.* **B753** (2016) 363–388, [arXiv:1507.0035].

[73] **ATLAS** Collaboration, T. A. collaboration, *Search for exotic Higgs-boson decays in events with at least one photon, missing transverse momentum, and two forward jets produced in $\sqrt{s}$ = 8 TeV **pp** collisions with the ATLAS detector*, .

[74] D. E. De Florian Sabaris, C. Grojean, F. Maltoni, C. Mariotti, A. Nikitenko, M. Pieri, P. Savard, M. Schumacher, and R. Tanaka, *Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector*, .

[75] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079, [arXiv:1405.0301].

[76] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, *JHEP* **0605** (2006) 026, [arXiv/0603175].

[77] **DELPHES 3** Collaboration, J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi, *DELPHES 3, A modular framework for fast simulation of a generic collider experiment*, *JHEP* **02** (2014) 057, [arXiv:1307.6346].

[78] N. D. Christensen, P. de Aquino, N. Deutschmann, C. Duhr, B. Fuks, C. Garcia-Cely, O. Mattelaer, K. Mawatari, B. Oexl, and Y. Takaesu, *Simulating spin-$\frac{3}{2}$ particles at colliders*, *Eur. Phys. J.* **C73** (2013), no. 10 2580, [arXiv:1308.1668].

[79] J. Anderson et al., *Snowmass Energy Frontier Simulations*, in *Community Summer Study 2013: Snowmass on the Mississippi (CSS2013) Minneapolis, MN, USA, July 29-August 6, 2013*, 2013. arXiv:1309.1057.

[80] J. M. Campbell, R. K. Ellis, and C. Williams, *Vector boson pair production at the LHC*, *JHEP* **07** (2011) 018, [arXiv:1105.0020].

[81] **CMS** Collaboration, V. Khachatryan et al., *Search for two Higgs bosons in final states containing two photons and two bottom quarks*, arXiv:1603.0689.

[82] t. p. We, *The search in run-i*, CMS Note 2013/075, 2013.

[83] t. p. We, *The search in run-i*, CMS Note 2014/118, 2014.

[84] **CMS Collaboration** Collaboration, *Search for H(bb)H(gammagamma) decays at 13TeV*, Tech. Rep. CMS-PAS-HIG-16-032, CERN, Geneva, 2016.

[85] S. Dawson, A. Ismail, and I. Low, *What is in the loop? The anatomy of double Higgs production*, *Phys. Rev.* **D91** (2015), no. 11 115008, [arXiv:1504.0559].

[86] J. Cao, Z. Heng, L. Shang, P. Wan, and J. M. Yang, *Pair Production of a 125 GeV Higgs Boson in MSSM and NMSSM at the LHC*, *JHEP* **1304** (2013) 134, [arXiv:1301.6437].

[87] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, *JHEP* **0307** (2003) 001, [hep-ph/0206293].

[88] R. Grober, M. Muhlleitner, M. Spira, and J. Streicher, *NLO QCD Corrections to Higgs Pair Production including Dimension-6 Operators*, arXiv:1504.0657.

[89] L. Randall and R. Sundrum, *A Large mass hierarchy from a small extra dimension*, *Phys.Rev.Lett.* **83** (1999) 3370–3373, [hep-ph/9905221].

[90] H. Davoudiasl, J. Hewett, and T. Rizzo, *Experimental probes of localized gravity: On and off the wall*, *Phys.Rev.* **D63** (2001) 075004, [hep-ph/0006041].

[91] C. Csaki, M. L. Graesser, and G. D. Kribs, *Radion dynamics and electroweak physics*, *Phys.Rev.* **D63** (2001) 065002, [hep-th/0008151].

[92] K. Agashe, H. Davoudiasl, G. Perez, and A. Soni, *Warped Gravitons at the LHC and Beyond*, *Phys.Rev.* **D76** (2007) 036006, [hep-ph/0701186].

[93] B. Mellado Garcia, P. Musella, M. Grazzini, and R. Harlander, *CERN Report 4: Part I Standard Model Predictions*, .

[94] M. Dall'Osso, T. Dorigo, C. A. Gottardo, A. Oliveira, M. Tosi, and F. Goertz, *Higgs Pair Production: Choosing Benchmarks With Cluster Analysis*, arXiv:1507.0224.

[95] S. Carrazza, J. I. Latorre, J. Rojo, and G. Watt, *A compression algorithm for the combination of PDF sets*, *Eur. Phys. J. C* **75** (2015) 474, [arXiv:1504.0646].

[96] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, arXiv:1510.0386.

[97] S. Dulat, T. J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, D. Stump, and C. P. Yuan, *The CT14 Global Analysis of Quantum Chromodynamics*, arXiv:1506.0744.

[98] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne, *Parton distributions in the LHC era: MMHT 2014 PDFs*, *Eur. Phys. J. C* **75** (2015) 204, [arXiv:1412.3989].

[99] **NNPDF** Collaboration, R. D. Ball et al., *Parton distributions for the LHC Run II*, *JHEP* **04** (2015) 040, [arXiv:1410.8849].

[100] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **1407** (2014) 079, [arXiv:1405.0301].

[101] A. Oliveira and R. Rosenfeld, *Hidden sector effects on double higgs production near threshold at the LHC*, *Phys.Lett.* **B702** (2011) 201–204, [arXiv:1009.4497].

[102] P. Aquino, *http://feynrules.irmp.ucl.ac.be/wiki/RSmodel* , .

[103] https://twiki.cern.ch/twiki/bin/view/Sandbox/DiHiggsModels.

[104] T. Hapola and O. Antipin, *http://cp3-origins.dk/research/units/ed-tools* , .

[105] A. Belyaev, N. D. Christensen, and A. Pukhov, *CalcHEP 3.4 for collider physics within and beyond the standard model*, *Comput.Phys.Commun.* **184** (2013) 1729–1769, [arXiv:1207.6082].

[106] https://feynrules.irmp.ucl.ac.be/wiki/HiggsEffectiveTheory.

[107] A. Alloul, N. D. Christensen, C. Degrande, C. Duhr, and B. Fuks, *FeynRules 2.0 - A complete toolbox for tree-level phenomenology*, *Comput. Phys. Commun.* **185** (2014) 2250–2300, [arXiv:1310.1921].

[108] https://github.com/CrossSectionsLHC/WED.

[109] https://github.com/syuvivida/DibosonBSMSignal_13TeV.

[110] F. Goertz, *Indirect Handle on the Down-Quark Yukawa Coupling*, *Phys.Rev.Lett.* **113** (2014), no. 26 261803, [arXiv:1406.0102].

[111] B. Hespel, D. Lopez-Val, and E. Vryonidou, *Higgs pair production via gluon fusion in the Two-Higgs-Doublet Model*, *JHEP* **09** (2014) 124, [arXiv:1407.0281].

[112] S. Frixione, F. Stoeckli, P. Torrielli, and B. R. Webber, *NLO QCD corrections in Herwig++ with MC@NLO*, *JHEP* **1101** (2011) 053, [arXiv:1010.0568].

[113] A. Carvalho, M. Dall'Osso, M. Gouzevitch, F. Goertz, M. Tosi, P. De Castro Manzano, and T. Dorigo, *Analytical Parametrisation and shape classification of anomalous HH production in EFT approach*, .

[114] https://github.com/kenmimasu/Rosetta/tree/master/dihiggs.

[115] J. C. Collins and D. E. Soper, *Angular distribution of dileptons in high-energy hadron collisions*, *Phys. Rev. D* **16** (1977) 2219–2225.

[116] "Jet identification." https://twiki.cern.ch/twiki/bin/view/CMS/JetID. Revision r94 of the twiki page.

[117] A. D. Bukin, *Fitting function for asymmetric peaks*, *ArXiv e-prints* (Nov., 2007) [arXiv:0711.4449].

[118] "Cms b-tagging offline guide." https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideBTagging. Revision r63 of the twiki page.

[119] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *Eur. Phys. J.* **C71** (2011) 1554, [arXiv:1007.1727]. [Erratum: Eur. Phys. J.C73,2501(2013)].