Search for Double Higgs Production in the Final State with Two Photons and Two Bottom Quarks at the CMS Detector

Philip Robert Hebda

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Department of
Physics
Adviser: Daniel Robert Marlow

September 2015

Abstract

A search for the production of Higgs pairs in the decay channel $HH \to \gamma\gamma b\bar{b}$ is reported for both resonant and nonresonant cases. The data corresponds to an integrated luminosity of 19.7 fb^{-1} of proton-proton collisions at $\sqrt{s} = 8$ TeV collected by the CMS detector at the CERN Large Hardron Collider. The candidate events are selected by requiring two photons and two jets and are classified according to the number of jets tagged as coming from the hadronization of a b-quark. The search for resonance production of two Higgs bosons through a new particle X, as hypothesized in extensions to the Standard Model involving a Radion or KK-graviton from models with warped extra dimensions or involving a heavy Higgs from models with supersymmetry, is performed on the range $m_X \in [260, 1100]$ GeV. The search for Standard Model nonresonant production of two Higgs bosons is performed; in addition a theoretical framework is explored for the analysis of anomalous values of the couplings $t\bar{t}H$, HHH, and $t\bar{t}HH$. The observations are consistent with background expectations. Upper limits at the 95% confidence level are extracted on the production cross section of resonant and SM nonresonant production. In particular, the Radion with $\Lambda_R = 1$ TeV is observed (expected) to be excluded with masses below 0.97 TeV (0.88 TeV), while the analysis is not sensitive to the Radion with $\Lambda_R = 3$ TeV. The nonresonant double Higgs cross section is observed (expected) to be excluded at 1.91 fb (1.59 fb) or 72.9 (60.7) times the NNLO Standard Model value.

Acknowledgements

This thesis would not have been possible without the input of many along the way.

First, I extend my most sincere gratitude to my advisor Dan Marlow, with whom I have collaborated during my entire graduate studies. Dan has been a consistent source of valuable insight and vision for the bigger picture, and I have benefitted immensely from our conversations about physics and life. I have seen that performing research occasionally involves more than understanding the underlying physics, and Dan's advice in navigating administration has been refreshingly helpful.

I would like to thank the professors in the Princeton Physics Department. In particular I owe thanks to Chris Tully for discussions related to this thesis and for serving as a second reader, and to Igor Klebanov and Peter Meyers for serving on committees for my FPO and pre-thesis project.

From my time at Purdue as an undergraduate, I would like to thank my advisor Daniela Bortoletto and Artur Apresyan. Their influence and patience gave me a strong desire to see just how deep the rabbit hole goes.

At CERN, I have had the privilege to work with many talented physicists. In the HggHbb group, I am indebted to Olivier Bondu, Maxime Gouzevitch, Chiara Rovelli, Badder Marzocchi, Alexandra Oliveira, and Amina Zghiche. I have enjoyed defending our colors, and I hope that Run 2 brings about two Higgs.

I owe thanks to many other physicists with whom I have collaborated on other CMS projects, including the lumi, PLT, W′, and VHbb groups. Some of these folks include Nadia Adam, Adam Hunt, Paul Lujan, Andrzej Zuranski, Andres Delannoy, Dean Hidas, Andreas Kornmayer, Steve Schnetzer, Bob Stone, Sunghyun Chang, Christos Leonidopoulos, Michael Mooney, and Seth Zenz.

I would like to thank the friends over the years of my studies who have made life as or perhaps more interesting than my physics life, whether at CERN, Princeton, Purdue, or elsewhere. Some of these folks include Alex Tuna, Javier Duarte, Kurt

Brendlinger, Larry Lee, Sven Rohr, Aris Alexandradinata, Tim Lou, Eric Smith, Glen Klink, Cody Nelson, and Erik “Ocho Cinco” Stahoviak. Special recognition goes to Tuna, Kurt, Josh Hardenbrook, and Josh Kunkle for contributing to multiple versions of our CERN relay team – bestowed with the names Seal Team Six, ΔF, and Hostage Rescue Team – and, in turn, creating a true American dynasty. Heroes are forged in the fires.

La langue française occupait un endroit special pendant le temps que j’ai passé comme doctorat. Au moment que j’ai su que je déménagerais à Genève, j’ai commencé à l’apprendre avec l’aide de plusieurs personnes patientes et encourageantes d’abord à Princeton et puis à Genève, des professeurs ainsi que des amis. Je remercie Brian Jacobs, Emilie Coquelin, Roberto Daverio, la Croix-Rouge genevoise, et Liliane Munther. Et je ne veux pas oublier le Jet d’eau et les TPG qui me fascinent plus qu’ils devraient.

I owe thanks my family whom I have been very fortunate to have by my side through ups and downs. To my brothers, Craig and Jared, for many years of laughter and fighting, and to my parents, for always providing a place to call home, I appreciate the constant support and look forward to what the future has in store for us. As the great philosopher Vin Diesel once said in a movie I’d wished I’d missed, “Family is everything.”

Finally, to my fiancée Jigisha Darbha, thanks.

నేను వ్యక్తిగా ఎదగడానికి నీ సహచర్యం ఎన్నో అవకాశాలు ఇచ్చింది. కాలం పరీక్షిస్తున్నప్పుడు, నేను స్పష్టంగా ఆలోచించడానికి సహాయం చేసావు. మనం త్వరలో అధికారికంగా ఒకటి కొతున్నందుకు ఎంతో ఆనందంగా ఉంది. మీ పూర్వీకుల మాత్రు భాష నేను నెమ్మదిగా నేర్చుకుంటున్నాను - ఈ సమయం లో ఎప్పటిలానే నీ సహనం కోరుతున్నాను. అనువాదం కోసం శిరీషకు కృతజ్ఞతలు. ఐర్లాండ్, అమెరికా మరియు భారత దేశంలో, నిజానికి ప్రజా రవాణా మనల్ని ఎక్కడికి తీసుకెళితే అక్కడ నీతో మరెన్నో సాహసాలు చెయ్యటానికి ఎదురుచూస్తున్నాను.

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 1148900 and by the U.S. Department of Energy under Grant No. DOE DE-SC0007968.

To Dad,

for practicing spelling words with me

and driving me 770 miles to and from the GC,

among other things.

Contents

Chapter 1

Introduction

The Standard Model (SM) is the theoretical framework that best describes the experimentally observed phenomena of the fundamental particles and their interactions. In 2012, the last piece to the SM was put into place with the discovery of the Higgs boson. This discovery, the theory for which is described in Sections 1.1 and 1.2, is the foundation for the work based on this thesis, the goal of which is to perform the first search for double Higgs production, a process in which two Higgs bosons are produced. The motivations for this search are given in Section 1.4. Despite its many successes, the theory is not perfect, and it remains an overarching theme of particle physics to unify physical processes at all energy scales under one single framework. Some of these shortcomings are described in Section 1.3, giving a little taste for what could be discovered with the accumulation of more data at higher center-of-mass energies by the big experiments at the CERN Large Hadron Collider.

1.1 The Standard Model

The Standard Model of particle physics is a relativistic quantum field theory that describes how the known fundamental particles interact through the electromagnetic, weak, and strong forces. The theory was developed through the unification of the

electromagnetic and weak forces by Glashow in 1961 [1] and through the incorporation of this electroweak theory with the Higgs mechanism by Weinberg and Salam in 1967 [2, 3]. This theory explained the experimental observations of the day, and later experiments provided additional evidence as well as a means for measuring the free parameters of the theory. Some of this evidence is provided in Figure 1.1 in the form of the discoveries of the fundamental particles.

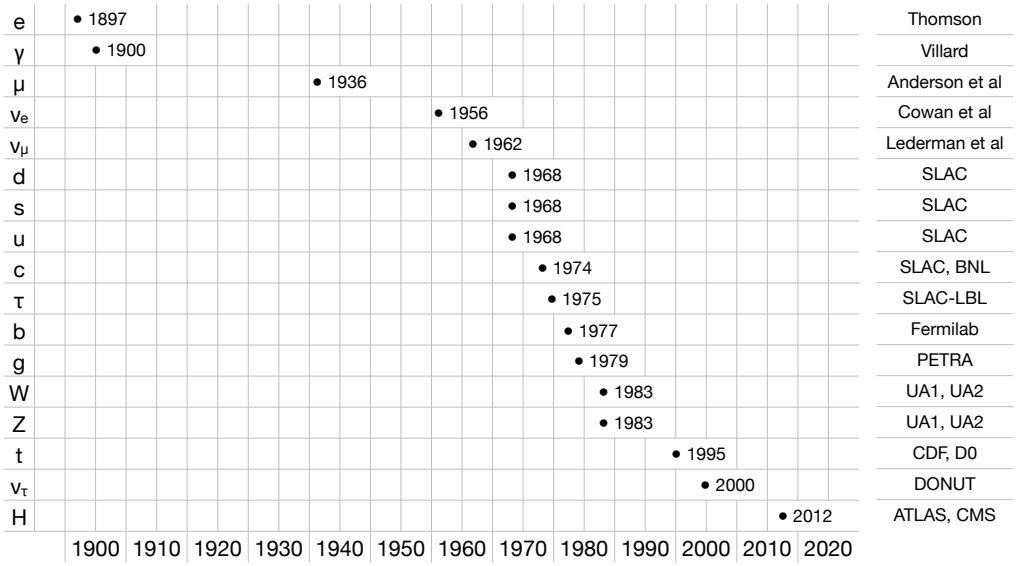

Figure 1.1: The discoveries of fundamental particles versus time [4].

In this theory, particles are treated as excitations of fields having half-integer spin or integer spin, and the forces are treated as interactions among excitations of these fields. The spin-$\frac{1}{2}$ particles, or fermions, can be divided into groups based on the ways in which they interact. The leptons, or those particles which only experience the electroweak force, are the electron e, muon μ, tau τ, electron neutrino ν_e, muon neutrino ν_μ, and tau neutrino ν_τ. The quarks, or those particles which experience both electroweak and strong forces, are the up u, down d, strange s, charm c, bottom b, and top t. The integer-spin particles, or bosons, are the spin-one photon γ, W, Z,

and gluon g and the spin-zero Higgs H. The particle content of the SM is summarized in Figure 1.2.

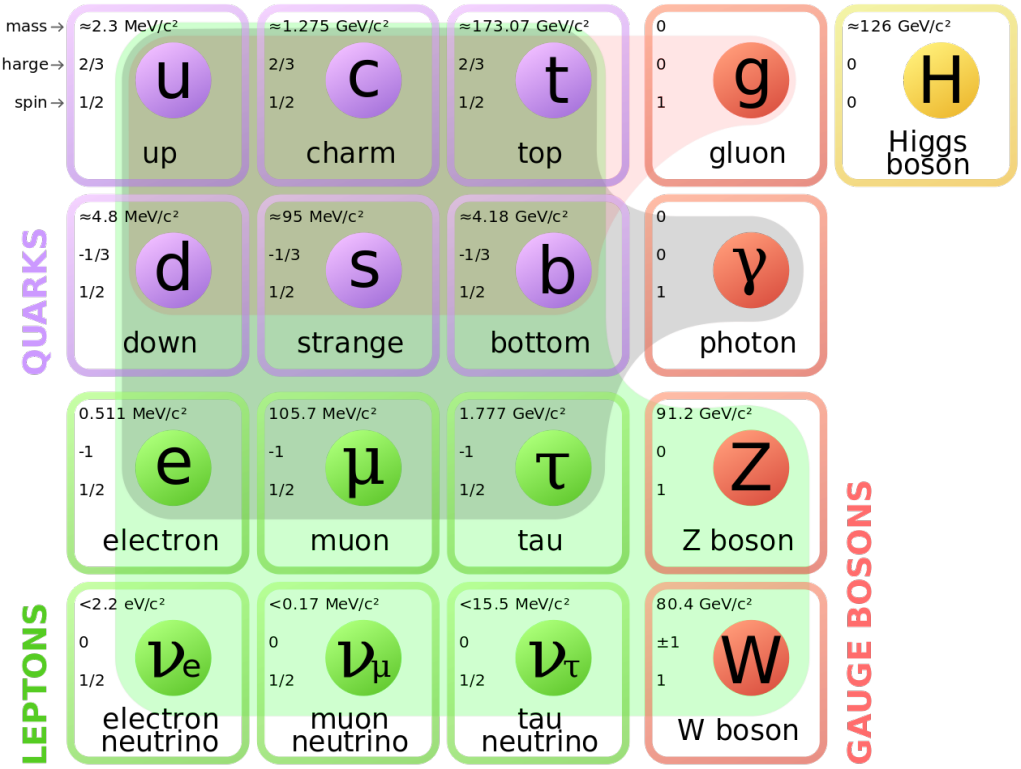

Figure 1.2: A diagram of the particle content of the SM [5]. The mass, electric charge, and spin is given for each, and the background color indicates how each fermion interacts with the bosons.

The dynamics of the SM are described through its Lagrangian, which is invariant under gauge transformations of the group $\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$. The strong force, also known as quantum chronodynamics (QCD), is associated with transformations under $\mathrm{SU}(3)_C$, which give rise to the conserved color charge C, denoted by red, green, or blue, and eight gauge fields. The group acts on 18 spinor fields corresponding to the quarks (six quark flavors in three colors) and the eight gauge fields corresponding to the gluons. The electroweak force is associated with transformations under $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$, the first part of which gives rise to the conserved left-handed chirality L and three gauge fields, and the second part of which gives rise to the

conserved weak hypercharge Y and one gauge field. This group acts on left-handed doublets and right-handed singlets of the quarks, leptons, and these four gauge fields. The quarks contribute nine doublets and 18 singlets, and the leptons contribute three doublets and three singlets (as right-handed neutrinos do not exist).

The symmetry group of the SM does not allow for gauge-invariant mass terms. Instead, the generation of particle masses is accomplished through the partial breaking of the symmetry group by the addition of the Higgs field and potential, after which gauge-invariant Yukawa interactions between fermions and the Higgs field spontaneously give fermion masses. The Higgs field ϕ is a doublet of $\mathrm{SU}(2)_L$, and its potential takes the form

$$V\left(\phi^\dagger\phi\right) = -\mu\phi^\dagger\phi + \lambda\left(\phi^\dagger\phi\right)^2 , \tag{1.1}$$

where $\mu, \lambda > 0$. The ground state of this potential is nonzero and degenerate. With the choice of a convenient gauge, the ground state can be expressed as

$$\langle\phi\rangle = \left\langle \frac{1}{\sqrt{2}}\begin{pmatrix}\phi^0 + i\phi^1 \\ \phi^2 + i\phi^3\end{pmatrix}\right\rangle = \begin{pmatrix}\mu/\sqrt{\lambda} \\ 0\end{pmatrix}, \tag{1.2}$$

where ϕ^1, ϕ^2, ϕ^3 are massless Goldstone bosons [6, 7] and the remaining degree of freedom represents fluctuations about this ground state, associated with the massive scalar Higgs boson H.

Recalling that the Higgs field is a doublet of $\mathrm{SU}(2)_L$, its covariant derivative in the Langrangian is

$$D_\mu = \partial_\mu - ig_1 T^a W_\mu^a - i\frac{g_2}{2}B_\mu , \tag{1.3}$$

where g_1 (g_2) and W_μ^a (B_μ) are the gauge couplings and three (one) gauge bosons associated with $\mathrm{SU}(2)_L$ ($\mathrm{U}(1)_Y$), respectively, and T^a are the generators of $\mathrm{SU}(2)_L$. The ground state of the Higgs field then gives mass terms to some of the gauge bosons. The mass eigenstates, in terms of the W_μ^a and B_μ, are the $W^\pm$, Z, and photon with

masses

$$m_{W^\pm} = \frac{g_1}{2}\frac{\mu}{\sqrt{\lambda}} \tag{1.4a}$$

$$m_Z = \frac{\sqrt{g_1^2 + g_2^2}}{2}\frac{\mu}{\sqrt{\lambda}} \tag{1.4b}$$

$$m_\gamma = 0\,. \tag{1.4c}$$

Thus, the gauge bosons acquire mass, and the symmetry of $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ is broken to $\mathrm{U}(1)_{EM}$ with gauge field being the photon and conserved quantity being the electric charge Q.

Masses for the fermions are generated through this symmetry breaking with the introduction of gauge-invariant interactions with the Higgs field into the Lagrangian. These take the form

$$\mathcal{L}_{\text{Yukawa}} = -g\bar{\psi}\phi\psi\,, \tag{1.5}$$

where g is the strength of the interaction. Here, the symmetry breaking caused by $\phi \to \langle\phi\rangle + H$ gives a mass term of the form $g\langle\phi\rangle$ and an interaction with the Higgs boson. Interactions with the other bosons arise from change in the covariant derivative of each fermion field due to the symmetry breaking.

These interactions can be represented pictorially with Feynman diagrams [8]. Figure 1.3 shows the interaction vertices of the SM in which the Higgs is not involved. These diagrams provide a powerful tool for the calculation of probability amplitudes of SM processes, which are directly related to the experimentally-measurable cross sections.

1.2 The Higgs Discovery

While each particle has its own story of discovery, as hinted briefly by Figure 1.1, this section will focus on the technical details related to that of the Higgs. What

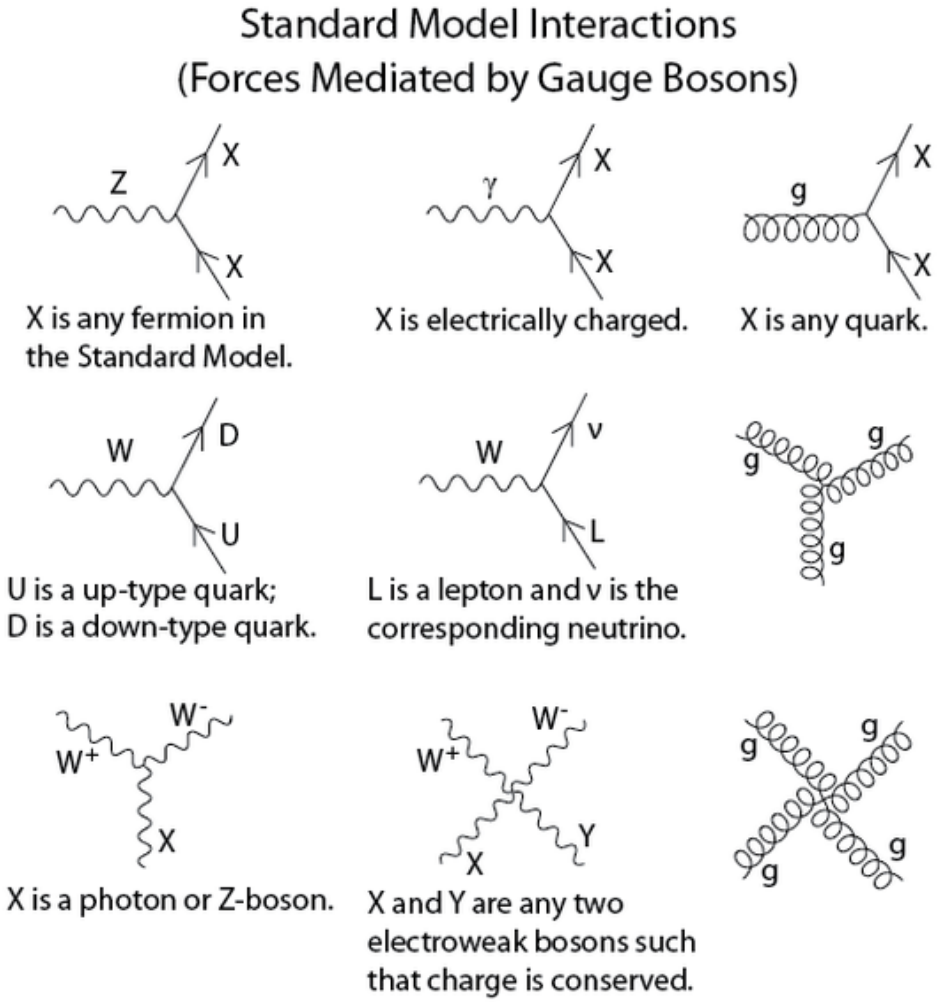

Figure 1.3: Interactions of the SM in which the Higgs in not involved [9].

makes the search especially difficult is the small couplings that the Higgs has with other particles, as shown in Figure 1.4. However, up until 2012, the mass of the Higgs boson was unknown, so big experimental searches had to accommodate a wide range of possible values.

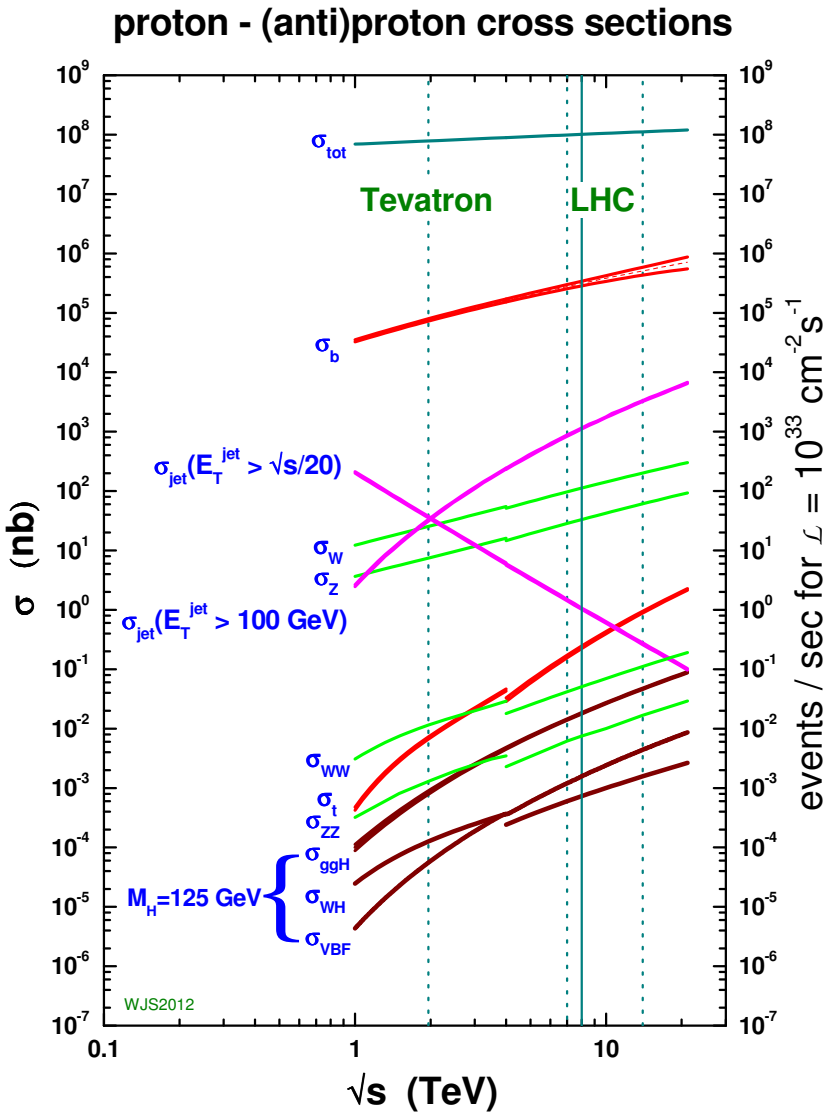

Figure 1.4: Production cross sections for various processes in the SM versus center-of-mass energy. The lower energies correspond to $p\bar{p}$ collisions from the Tevatron [10], whereas the higher energies correspond to pp collisions from the LHC [11].

The search for the Higgs can be broken down into production mechanism and decay mode. For proton-proton pp collisions at $\sqrt{s} = 8$ TeV, as realized at the LHC (see Section 2.2), Figure 1.5 shows the dominant production mechanisms and corresponding cross sections. The branching ratios for several of the dominant decay modes are given in Table 1.1.

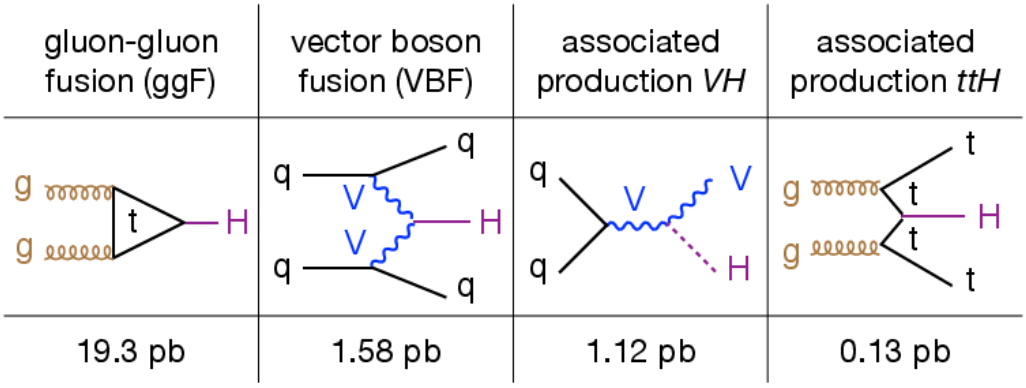

Figure 1.5: Production mechanisms and cross sections of the Higgs boson for the leading processes in the SM [4]. The cross sections correspond to pp collisions at $\sqrt{s} = 8$ TeV.

Since the announcement of discovery of a new Higgs-like particle in 2012 [12, 13], the properties of the Higgs have been narrowed through the accumulation of more data and more refined analyses [14]. The general conclusion of all these analyses is that the Higgs boson is, up to the sensitivity that the present dataset allows, consistent with that predicted by the SM. Figures 1.6 and 1.7 provide two highlights in μ for each of the production cross sections and decay modes, respectively, where μ is the ratio of the measured cross section to that predicted by the SM.

Table 1.1: Branching ratios for the Higgs boson [15].

	fermions			bosons			
Decay mode	bb	$\tau\tau$	$\mu\mu$	WW^*	ZZ^*	$\gamma\gamma$	$Z\gamma$
Branching fraction	58%	6.3%	0.022%	22%	2.6%	0.23%	0.15%

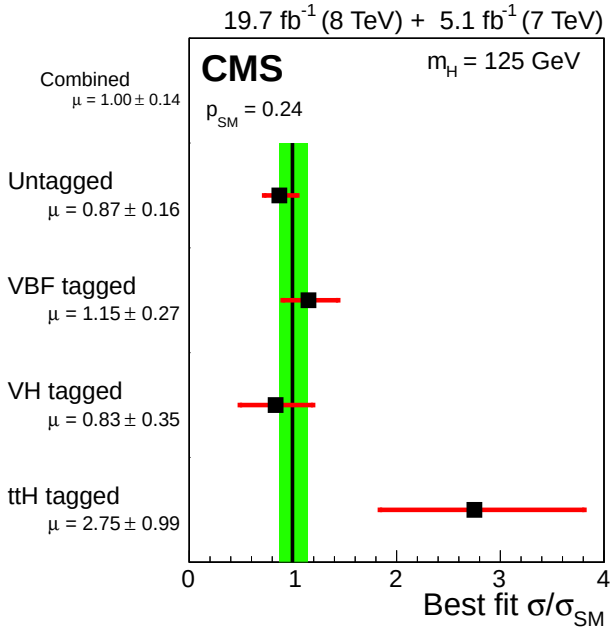

Figure 1.6: Values of the best-fit μ for the combination (solid vertical line) and for subcombinations by analysis tags targeting individual production mechanisms [16]. The vertical band shows the overall μ uncertainty. The μ ratio denotes the production cross section times the relevant branching ratios, relative to the SM expectation. The horizontal bars indicate the ± 1 standard deviation uncertainties in the best-fit μ values for the individual modes; they include both statistical and systematic uncertainties.

Perhaps the most important post-discovery measurement is that of the Higgs boson mass m_H. Figure 1.8 provides experimental evidence for the precision measurement of this value by considering only high-resolution channels for Higgs boson production and decay, and the measured value is around 125 GeV. This provides a means to probe other features of the SM or to search for signatures of physics beyond the SM (BSM). One class of those processes, which is the major topic of interest for this thesis, gives rise to the production of two Higgs bosons. The theoretical grounds for such processes are discussed in Section 1.4.

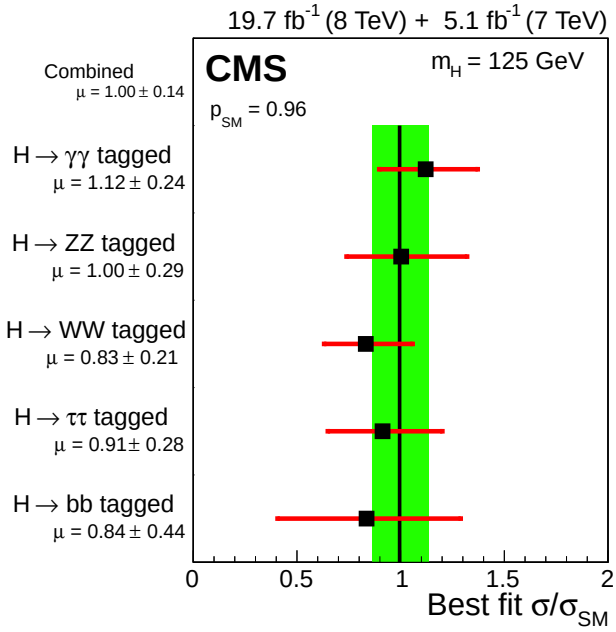

Figure 1.7: Values of the best-fit μ for the combination (solid vertical line) and for subcombinations by predominant decay mode [16]. The vertical band shows the overall μ uncertainty. The μ ratio denotes the production cross section times the relevant branching ratios, relative to the SM expectation. The horizontal bars indicate the ± 1 standard deviation uncertainties in the best-fit μ values for the individual modes; they include both statistical and systematic uncertainties.

1.3 Shortcomings of the SM

There are both experimental and theoretical reasons for why the SM is not a complete theory of particle physics. In some cases, the SM provides an ad-hoc description, and in other cases, there is no description at all. For the optimist, these provide compelling reasons to continue pushing the boundaries of our knowledge. Examples of such compelling questions concern gravity, dark matter, dark energy, and neutrinos, as detailed in Sections 1.3.1, 1.3.2, and 1.3.3, but this is not to omit other questions concerning, for example, matter-antimatter asymmetry, hierarchy, naturalness, and the generations of fermions.

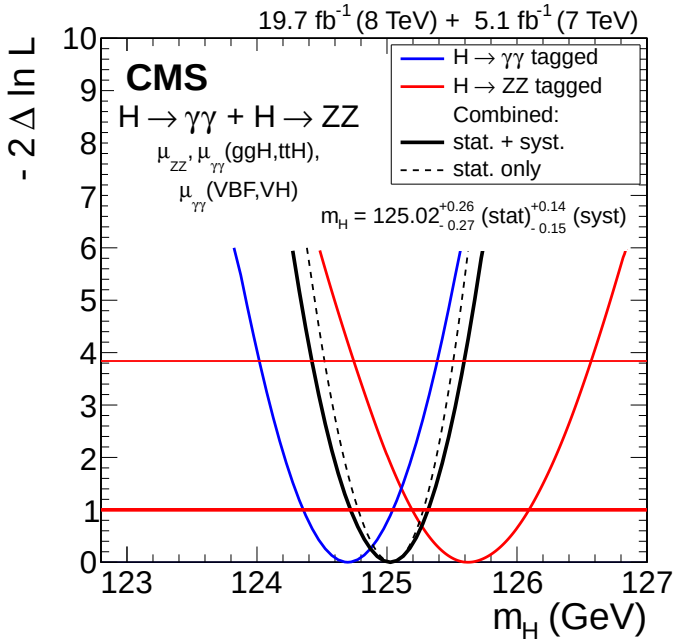

Figure 1.8: Scan of the test statistic $q(m_H) = -2\Delta \log L$ versus the mass of the boson m_H for the $H \to \gamma\gamma$ and $H \to ZZ \to 4\ell$ final states separately and for their combination [16]. Three independent signal strengths, (ggH,ttH)$\to \gamma\gamma$, (VBF,VH)$\to \gamma\gamma$, and $pp \to H \to ZZ \to 4\ell$, are profiled together with all other nuisance parameters. The solid curve is obtained by profiling all nuisance parameters and thus includes both statistical and systematic uncertainties. The dashed curve is obtained by fixing all nuisance parameters to their best-fit values, except for those related to the $H \to \gamma\gamma$ background description, thus including only statistical uncertainties. The crossings with the thick (thin) horizontal lines define the 68% (95%) CL interval for the measured mass.

1.3.1 Gravity

Gravity is well understood in the classical and relativistic limits, as first posited by Newton and Einstein, respectively. However, there is no complete quantum theory of gravity, and as such it is not accounted for by the SM. It is thought that the force is mediated by a spin-two particle called the graviton and that the gravitational force would unify with the other three forces in a very-high energy limit.

1.3.2 Dark Matter and Dark Energy

Ordinary matter, as built-up from the SM, only accounts for 4.9% of the mass-energy in the known universe [17, 18]. Most of the mass-energy is made up of dark matter at 26.8% and dark energy at 68.3%, the nature of which is largely unknown.

The presence of dark matter is inferred from observing the way ordinary matter behaves at astrophysical scales. In these observations, the behavior of certain objects cannot be explained by the ordinary matter alone. Rather, it is posited that these observations can be explained by the presence of some additional matter which interacts through gravity but not electromagnetism. Candidate dark matter particles provide motivation for many experimental searches at the LHC and other experiments.

The presense of dark energy is inferred from measurements of the expansion of the universe. These measurements indicate that the rate of expansion is increasing. Two candidates to account for the source of the accelerating expansion are a cosmological constant or scalar fields. The former gives a constant energy density in space and time, while the latter gives rise to an energy density that varies (perhaps very slowly) in space and time.

1.3.3 Neutrinos

In the version of the SM presented in 1.1, neutrinos are massless, electrically neutral leptons which come in three flavors. The observation of neutrino oscillation [19], in which the probability of measuring a particular flavor in a neutrino varies over its trajectory, implies that neutrinos have masses. These mass terms have been added to the SM, but it is unclear if their origin is from the same mechanism that provides the other fermions with their masses.

1.4 Double Higgs as a probe of SM and New Physics

With the mass of the Higgs known, searches involving the production of two Higgs bosons may be performed. There are two principle motivations for such a process. It provides a means for searching for an interaction predicted by the SM but not yet observed, namely that of the Higgs trilinear coupling, which provides further understanding of the Higgs potential. While searching for this rare SM process, similar topologies involving BSM physics may be sought. The goal of the latter is to either find a deviation from what the SM predicts or to place limits on the extent to which new theories may be excluded.

Double Higgs production at hadron colliders in the SM has been studied theoretically [20, 21], and the cross section for pp collisions at 8 TeV is predicted to be 9.96 fb, including next-to-next-to-leading order (NNLO) corrections. This is unfortunately out of reach of current experiments and is due, in part, to the destructive interference exhibited between contributions from the the Higgs self-coupling and from a top loop with two Higgs, as shown in Figure 1.9.

If the couplings from the diagrams in Figure 1.9 were not to take their SM values, the double Higgs cross section could be enhanced, leading to possible sensitivity at

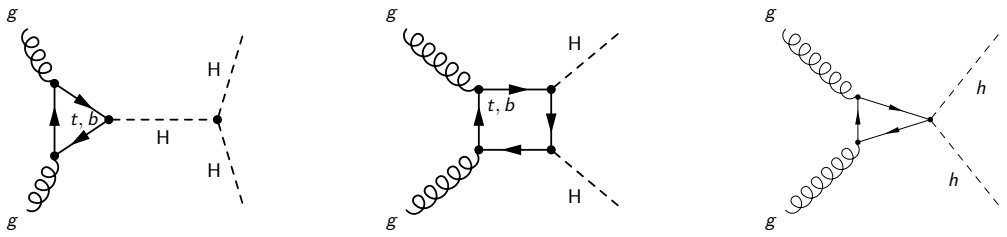

Figure 1.9: Feynman diagrams of the processes responsible for non-resonant gluon fusion production of two Higgs bosons in the final state. These diagrams involve the Higgs boson self-coupling (left), the top quark Yukawa coupling (middle, already measured at current experiments), and the anomalous coupling $ttHH$ (right) which is open in case of an effective parametrization of the Lagrangian by dimension 6 operators.

current experiments. (Although it could also be decreased.) These kinds of effects are present in models where the Higgs is interpreted as a composite state of some strong dynamics. The deviations of the SM couplings that could arise from the simpler composite Higgs case are already very constrained [22, 23]. However, under more general assumptions, higher order operators could in principle lead to sizable effects [24, 25, 26]. This possibility is discussed in Section 1.4.1.

Double Higgs production can also be used to search for the presence of a new, heavy particle with a spin of zero or two and with mass at least twice as large as m_H so that decay to HH is kinematically allowed. Such new particles can be motivated from BSM scenarios which typically have dominant decay modes to pairs of Ws, Zs, tops, and Higgs. Some of these scenarios are discussed briefly in Section 1.4.2.

1.4.1 Nonresonant Double Higgs Production from BSM

In this thesis, nonresonant double Higgs production is parametrized by the Higgs trilinear self-coupling λ, the Yukawa top coupling y_t, and an anomalous $t\bar{t}HH$ coupling c_2. The potential of the Higgs and its interactions with the top quark can be written

as

$$\begin{aligned}\Delta L_{H,t;\lambda,y_t,c_2} &= \partial_\mu H \partial^\mu H - m_H H^2 + \kappa_\lambda \lambda_{SM} v H^3 \\ &- \frac{\sqrt{2} m_t}{v} \left(v + \kappa_t H + \frac{\sqrt{2} c_2}{v} H H \right) (\bar{t}_L t_R + h.c.) \ ,\end{aligned} \tag{1.6}$$

where $\kappa_\lambda = \frac{\lambda}{\lambda_{SM}}$, $\kappa_t = \frac{y_t}{y_{t,SM}}$, $v = 246$ GeV, $\lambda_{SM} \equiv \frac{m_H^2}{2v^2} = 0.129$, and $y_{t,SM} = \frac{\sqrt{2} m_t}{v} = 0.995 \sim 1$. Possible contact interactions between the Higgs and gluons that could arise from heavy coloured scalars are neglected. Only dimension-six operators are relevant to BSM physics. The parameters c_2 and y_t are analytically related by the coefficients of two dimension-six operators, one being the derivative of the interaction among four Higgs doublets with coefficient c_H and the other involving the interactions between the Higgs doublet and top quarks with coefficient c_t.

For the interpretation of the BSM nonresonant search, the LO cross section for double Higgs production is expressed in a semi-analytical form [27] as

$$\begin{aligned}\sigma(pp \to HH) = \bar{\sigma} [& c_2^2 + (\alpha \kappa_t^2)^2 + (\beta \, \kappa_t \, \kappa_\lambda)^2 + A_1 c_2 (\alpha \, \kappa_t)^2 \\ & + A_2 (\alpha \, \kappa_t)(\beta \, \kappa_t \, \kappa_\lambda) + A_3 \, c_2 (\beta \, \kappa_t \, \kappa_\lambda)] \, ,\end{aligned} \tag{1.7}$$

where $\bar{\sigma}_{\mathrm{NNLO}} = 97.12$ fb, $\alpha = 0.475$, $\beta = 0.185$, $A_1 = -1.89$, $A_2 = -1.79$, and $A_3 = -1.70$. Over the range that κ_λ and κ_t are varied, as discussed in Section 3.3, the behavior of this cross section is in agreement with MC calculations used in this analysis. The cross section is normalized by $\bar{\sigma}_{NNLO}$, which is set such that the value at $\kappa_t = 1, \kappa_\lambda = 1$ and $c_2 = 0$ is equal to the SM prediction at NNLO. This approach is valid within the large m_t approximation, were the QCD corrections are dominated by soft emission which does not depend on the electroweak couplings. In addition, on the range of interest for κ_λ and κ_t, the k factors do not vary by more than 0.06% [28], so Equation 1.7 may be scaled to NNLO by applying the SM k factor.

The results are interpreted in a scenario where only dimension-six operators involving one Higgs doublet are relevant for nonresonant BSM physics. The dimension-six operators relevant to double Higgs production are [29]

$$\begin{aligned}\Delta L_{H,t;6D} = \frac{c_H}{2\Lambda^2}\partial_\mu(\phi^\dagger\phi)\partial^\mu(\phi^\dagger\phi) + \mu_H\phi^\dagger\phi - \lambda_6\,\phi^\dagger\phi\,\left(1+\frac{c_6}{\Lambda^2}\,\phi^\dagger\phi\right) \\ -c_t\left(1+\frac{c_t}{\Lambda^2}\phi^\dagger\phi\right)\left(\bar{Q}_L\phi^\dagger t_R + h.c.\right)\,.\end{aligned} \tag{1.8}$$

Together with the cut-off scale Λ, only c_6 and c_t parameters define double Higgs production. The coefficients of Equation 1.6 are related to the coefficients of the dimension-six operators of Equation 1.8 through

$$\kappa_t = \frac{v}{m_t\sqrt{2}}\left(1-\frac{1}{2}\frac{v^2}{\Lambda^2}c_H + \frac{v^2}{\Lambda^2}c_t\right) \tag{1.9a}$$

$$c_2 = \frac{v^2}{\sqrt{2}\Lambda^2}\left(c_t - \frac{1}{2}c_H\right)\,. \tag{1.9b}$$

The bare top quark mass and Higgs mass are

$$m_t = y_t\,v\left(1+c_t\,\frac{v^2}{\Lambda^2}\right) \tag{1.10a}$$

$$m_H^2 = 2\lambda_6\,v^2\left(1+c_6\frac{v^2}{\Lambda^2}\right)\,. \tag{1.10b}$$

1.4.2 Resonant Double Higgs Production from BSM

Double Higgs production can also used as a means to probe the existence of a heavy resonance of spin zero or two. In this thesis, this resonance is assumed to have negligible width compared to the experimental resolution and to be predominantly produced by gluon-gluon fusion. The selection was designed to limit the sensitivity to the angular distribution of the two Higgs bosons in order to keep the analysis spin-independent.

Theories with warped extra dimensions (WED) approach Planck scales allowing gravity to propagate in a compact extra dimension with a warp factor. Scenarios for the simplest choice of the warp function are known as Randall-Sundrum models (RS) [30], in which the metric for the case of single extra dimension is given by

$$ds^2 = e^{-2ky}\eta_{\mu\nu}dx^\mu dx^\nu - dy^2\,, \tag{1.11}$$

where y refers to the coordinate in the fifth dimension and k is related to its curvature. The so-called ultraviolet (UV) and infrared (IR) branes are introduced at $y = 0$ and $y = L$, respectively. Perturbations project in the four-dimensional effective theory as the graviton field. The massless zero mode of this field corresponds to the graviton.

More realistic descriptions of the WED picture requires a suitable mechanism to stabilize the distance between TeV and IR branes. The stabilization of the extra-dimension size in RS scenarios can be achieved by introducing a bulk scalar with a specific potential [31]. The four-dimensional projection of this scalar field is known as the Radion, and its mass is dependent on the format of the stabilization potential. Allowing matter fields to propagate in this extra dimension modifies its interactions between the gravity mediators [32]. Couplings between the gravity mediators and the matter fields occur by means of the energy momentum tensor, which gives rise to the Lagrangian [33, 34]

$$L = -\frac{c_i}{\Lambda_G}G^{\mu\nu(1)}T_{\mu\nu} - \frac{d_i}{\Lambda_R}RT_{\mu\nu}\,, \tag{1.12}$$

where $G^{\mu\nu(1)}$ is for the KK-graviton field, R is for the Radion field, c_i and d_i are constants which depend on the behavior of the matter fields on the bulk, and the scales Λ_G and Λ_R are the ultraviolet mass scale and Radion vacuum expectation, respectively. Both the scales can be interpreted as cut-offs of the theory.

Another way for BSM physics to manifest double Higgs production is through the existence of a second Higgs doublet from minimal supersymmetry (MSSM) and next to minimal composite scenarios [35, 36]. Possible manifestations of two Higgs doublet models (2HDM) can be based on the structure of its couplings to fermions [37]. Such models can have interactions in which

- all fermions couple to only one doublet (type I).
- up and down type quarks couple to different doublets (type II).
- fermions and quarks couple to different doublets (type III).
- up and down type quarks couple to different doublets while leptons couple to the doublet that couples to up quarks (type IV).

The most conventional parametrization of the potential is in terms of

- β, which is related to the ratio between the vacuum expectation values of the two doublets.
- α, which is the angle of the mass mixing of the neutral states.
- the five physical masses of the spectrum: two neutral scalars, one neutral pseudoscalar and two charged scalars.
- the coupling constants defining the allowed terms for the quartic potential.

The Higgs sector of the MSSM consists in a 2HDM type II. With the identification of the lightest neutral scalar with the SM Higgs boson and with reasonable assumptions about the potential, the parameter space for the search of a heavy neutral Higgs decaying to two SM Higgs bosons is reduced to the mass of the heavy Higgs and $\tan\beta$ [37, 38, 39, 40].

1.5 Double Higgs Decays

In the search for double Higgs production, the analysis presented here concerns the final state where one Higgs decays to two photons $H \to \gamma\gamma$ and the other decays to two b-quarks $H \to b\bar{b}$. The advantage of this final state from the experimental side is that the $H \to \gamma\gamma$ decay offers a high-resolution manner to tag one of the Higgs, while the $H \to b\bar{b}$ decay offers a high branching ratio. This disadvantages are that the $H \to \gamma\gamma$ decay has a low branching ratio and the $H \to b\bar{b}$ decay has poor resolution.

Table 1.2 gives the branching ratios for several of the leading double Higgs final states. Although the $\gamma\gamma b\bar{b}$ final state has a relatively low contribution at 0.26%, the excellent diphoton mass resolution puts it among one of the most sensitive final states in the search for nonresonant and low-mass resonant double Higgs production, as discussed in Chapter 8. On the other hand, the $b\bar{b}b\bar{b}$ has large QCD background that hurts its sensitivity in the low-mass regime, despite having a large branching ratio of 48%. Branching ratio is not the only part to the story, as explored in rest of this thesis.

Table 1.2: Branching ratios for decays of two Higgs bosons [15]. Note that ℓ stands for either e or μ.

Channel	Frequency (%)
$H(b\bar{b}, c\bar{c}, gg)H(b\bar{b}, c\bar{c}, gg)$	47.86
$H(b\bar{b})H(b\bar{b})$	33.30
$H(b\bar{b}, c\bar{c}, gg)H(VV^*)$	33.40
$H(b\bar{b}, c\bar{c}, gg)H(\tau^+\tau^-)$	8.77
$H(b\bar{b})H(\tau^+\tau^-)$	7.29
$H(VV^*)H(VV^*)$	5.83
$H(\ell^+\ell^-)H(VV^*)$	3.06
$H(\tau^+\tau^-)H(\tau^+\tau^-)$	0.40
$H(b\bar{b}, c\bar{c}, gg)H(\gamma\gamma)$	0.32
$H(b\bar{b})H(\gamma\gamma)$	0.26
$H(b\bar{b}, c\bar{c}, gg)H(\mu^+\mu^-)$	0.03
$H(\ell^+\ell^-)H(\gamma\gamma)$	0.03

Chapter 2

Experimental Facility

This chapter describes the experimental facility used to search for double Higgs production and other phenomena. Section 2.1 covers the European Organization for Nuclear Research (CERN), the organization that hosts the experimental facilities. Section 2.2 covers the Large Hadron Collider (LHC), the accelerator complex responsible for delivering pp collisions to each experiment. Section 2.3 covers the Compact Muon Solenoid (CMS), the detector which this analysis uses to study the pp collisions and search for double Higgs production.

2.1 CERN

CERN was established in 1954 by 12 countries in Western Europe with a mandate to establish a world-class research organization in fundamental physics [41]. Notable achievements in particle physics and computer science made through experiments at CERN include

- the discovery of neutral currents in 1973,
- the discovery of the W/Z bosons in 1984,
- the creation of the first website in 1991,

- the first creation of antihydrogen in 1995,
- and the discovery of the Higgs boson in 2012, as discussed in Section 1.2.

Today, CERN has grown to include 21 member states with plans to expand further. Many other states over the world participate in the experiments to varying degrees. Figure 2.1 shows a world map detailing each country's affiliation with CERN.

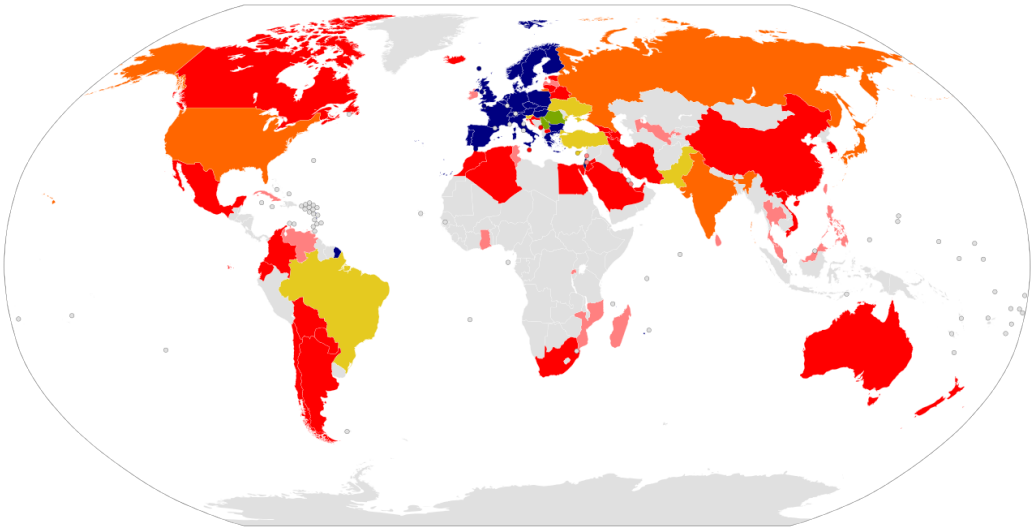

Figure 2.1: Map of the international relations each country has with CERN [42]. Member states are in blue. States for whom accession is in progress are in green. States who have declared intent to join are in yellow. Observers are in red. States with cooperation agreements are in orange. States with scientific contacts are in pink.

2.2 LHC

2.2.1 Overview

The LHC [11] is a *pp* accelerator that straddles the border between France and Switzerland. It was installed in a 27 km tunnel with diameter 3.7 m at an average depth of 100 m underground. The position of this tunnel with respect to nearby political and geographical points of interest is shown in Figures 2.2 and 2.3. The

tunnel was previously used by the Large Electron-Positron Collider (LEP) [43] from 1989 to 2000 for electron-positron collisions.

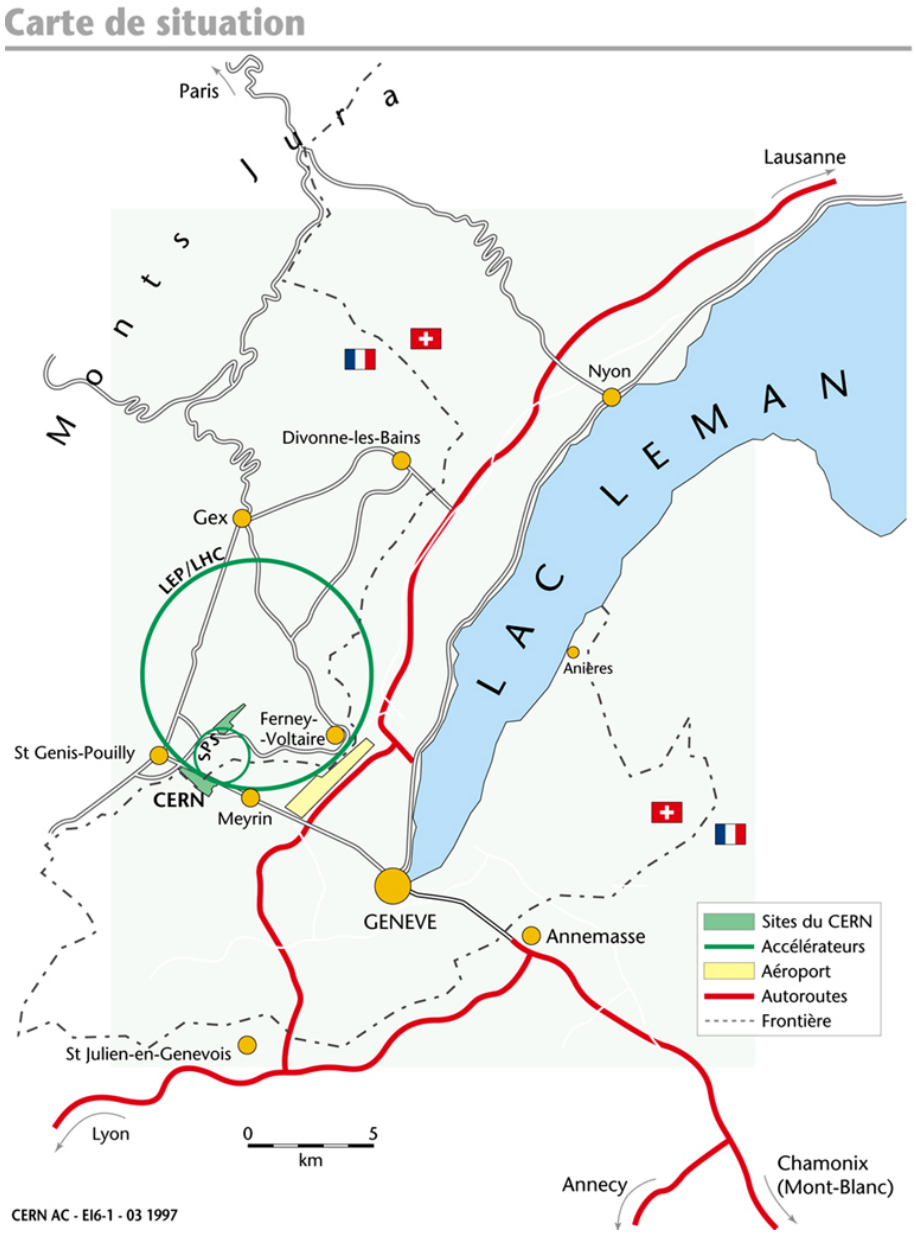

Figure 2.2: Map of the political and geographical highlights near the LHC [44].

The LHC is designed to accelerate protons or ions in two circular beam pipes, orbiting in opposite directions [45]. The pre-accelerators are described in Section 2.2.2, while the LHC itself is described in Section 2.2.3 These beams are made to cross in

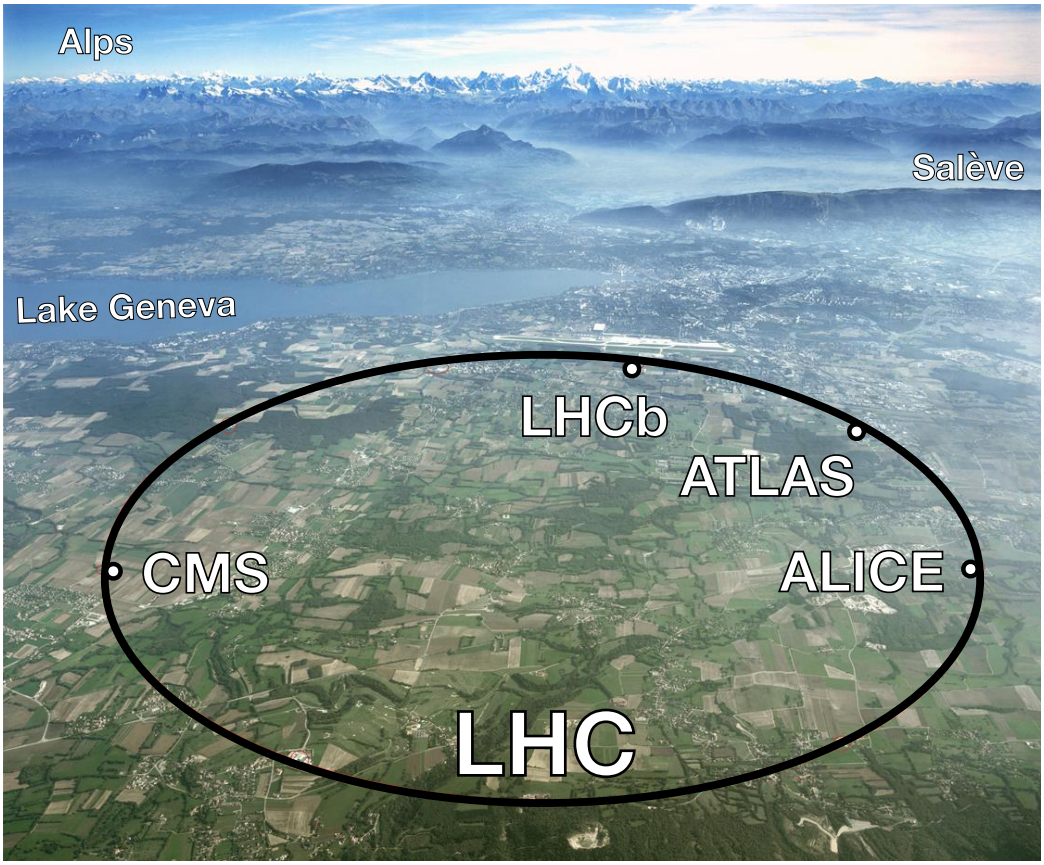

Figure 2.3: Aerial photo of the Geneva countryside with the LHC superimposed [4].

four designated points of the accelerator and create collisions between pairs of particles in opposite beams. Each of these points is home to a large detector, described in Section 2.2.4, for measuring the products of these collisions.

2.2.2 Accelerator Complex

The path that a proton takes before entering a collision in the middle of CMS is shown in Figure 2.4. To start, a hydrogen gas atom is stripped of its electron and accelerated in Linac2, a linear accelerator, to 50 MeV. Next is a series of three circular pre-accelerators that increases the kinetic energy of the protons and collects the protons into discrete bunches. These accelerators are the Proton Synchrotron Booster

(PSB), the Proton Synchrotron (PS), and the Super Proton Synchrotron (SPS), which increase the proton's energy to 1.4 GeV, 25 GeV, and 450 GeV, respectively.

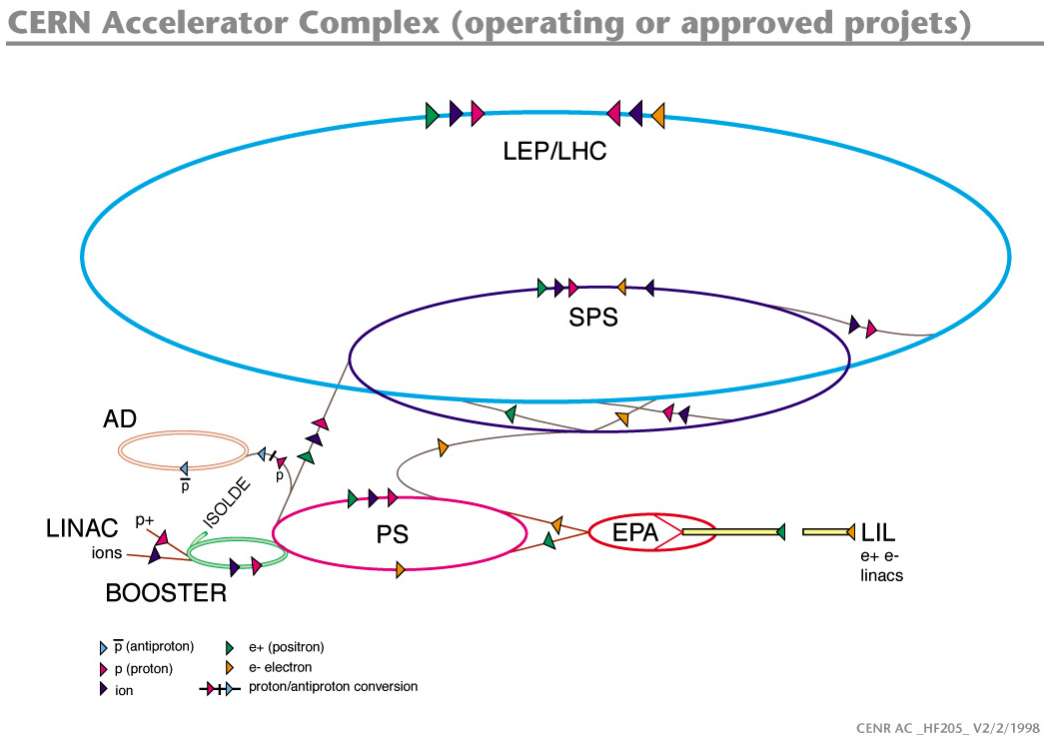

Figure 2.4: The flow of the CERN accelerator complex [46].

After the SPS, the protons are injected into the LHC, where they are accelerated up to 7 TeV, the designed energy of operation. The highest energy achieved to date is 6.5 TeV. The bunch structure is maintained during injection into the LHC, which is important for the timing of the collisions inside the detectors.

2.2.3 The Machine

The LHC is made up of 1232 dipole magnets divided by eight insertion points. A cross sectional view of a dipole is shown in Figure 2.5. The purpose of the dipoles, each of which has two holes for each beam, is to direct the proton beams on their circular path. To keep these high energy particles on their path, each dipole produces a magnetic field of 8.3 T using a superconductor cooled with liquid helium to 1.9 K,

which allows for a current of 12 kA. The length and mass of a dipole are 14 m and 35 tons, respectively.

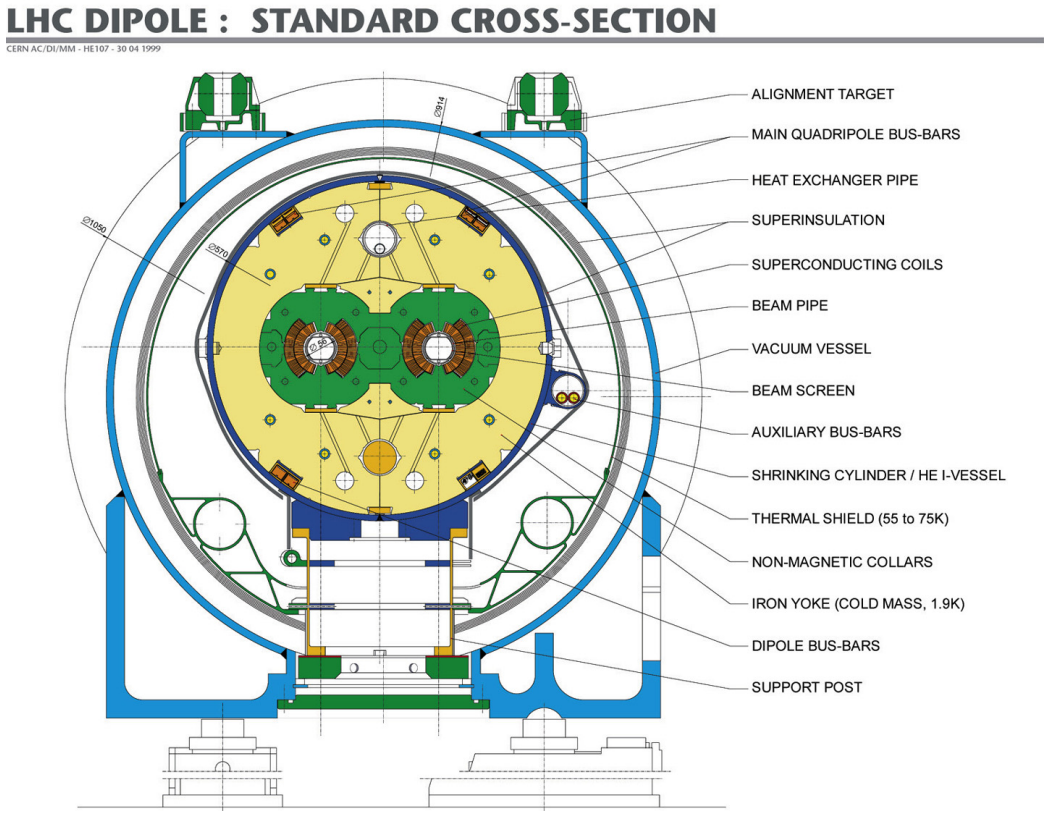

Figure 2.5: Diagram of a cross section of an LHC dipole [47].

Each insertion point serves a different purpose. Points 1, 2, 5, and 8 are the interaction points, where the beams overlap to produce collisions. Point 4 contains the radio frequency cavity system, which provides acceleration to the protons in the direction of their motion. Points 3 and 7 contain beam collimation systems. Point 6 contains the beam dump system for removing the beams from the LHC.

2.2.4 Detectors on the LHC

The general-purpose detectors on the LHC are A Toroidal LHC Apparatus (ATLAS) [48] and CMS, located at opposite sides of the ring at points 1 and 5, respectively. The physics goals of these two experiments are generally the same: to

search for and study the Higgs boson, to study and improve our understanding of SM processes, and to search for signs of BSM physics such as WED or SUSY.

The Large Hadron Collider beauty detector (LHCb) [49] is located at point 8 and is used to study processes related to the bottom quark. A Large Ion Collider Experiment (ALICE) [50] is located at point 2 and is used to solely to study collisions when one or both beams contain lead ions. This mode of operation of the LHC lasts about one to two months per year.

The relative positions of these four detectors on the LHC is shown in Figure 2.6. Although the figure makes it appear that all the experiments are at the same depth, the LHC tunnel is actually sloped at 1.5 degrees toward Lac Léman, causing the tunnel depth to vary between 50 and 150 m.

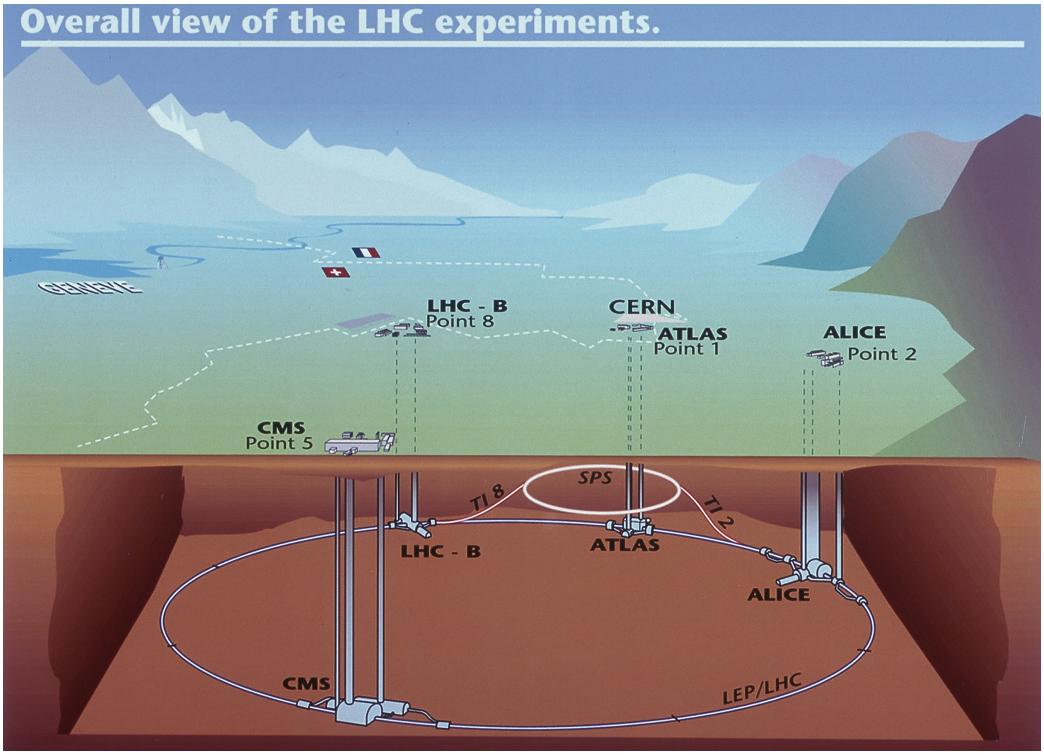

Figure 2.6: Underground view of the LHC and the relative positions of the four detectors [44].

2.2.5 LHC Operation in 2012

The principal physics result described in this work is an analysis of the data taken with the CMS detector during the operation of the LHC in 2012. During this time, the center-of-mass energy of the pp collisions was $\sqrt{s} = 8$ TeV, whereas the design center-of-mass energy of $\sqrt{s} = 14$ TeV will be achieved in future operation of the LHC.

Of the 10^{11} protons in each bunch, only a small number of pairs of protons interact during a bunch crossing inside CMS, where the beams are squeezed into a smaller transverse region. The number of proton-proton interactions per bunch crossing is called pileup. This number varies from fill to fill and even within a fill due to conditions in the accelerator. The distribution of the pileup during 2012 running is given in Figure 2.7, during which time the average was 21.

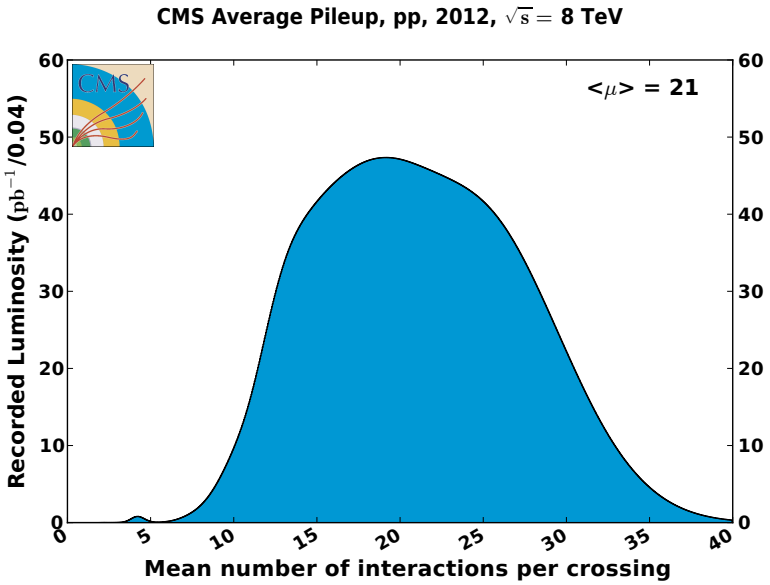

Figure 2.7: The pileup distribution in CMS during 2012 [51].

The total integrated luminosity delivered to CMS during this time was 23.3 fb $^{-1}$. The evolution of the integrated luminosity delivered to and recorded by CMS is shown in Figure 2.8. The delivered integrated luminosity is the amount of luminosity

delivered by the LHC to CMS, and ideally CMS would be able to same all of these collisions. Due to limitations of the data acquisition chain or availability of each subsystem, discussed in Sections 3.1 and 2.3, respectively, the recorded luminosity is slightly smaller than the delivered luminosity. In 2012, the total recorded luminosity was 21.8 fb $^{-1}$.

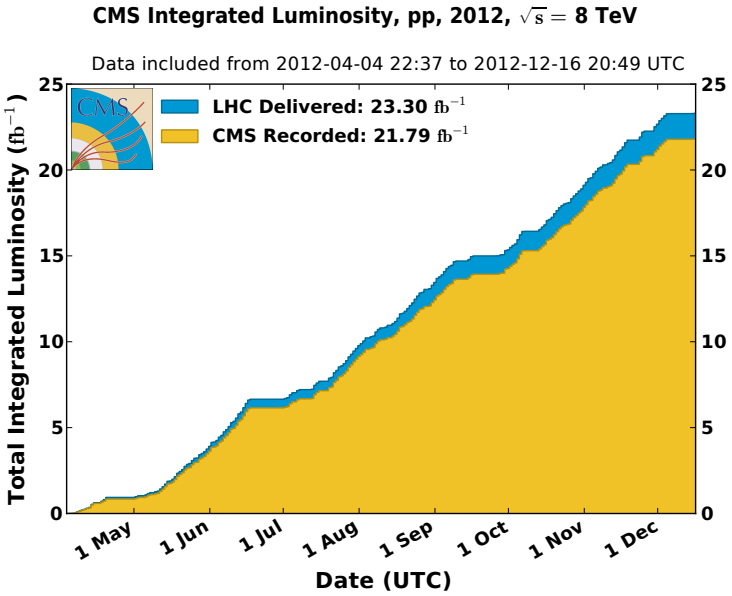

Figure 2.8: The integrated luminosity of pp collisions in CMS during 2012 [51].

The LHC delivered pp collisions in 2010 and 2011 at $\sqrt{s} = 7$ TeV, and these were also recorded by CMS. Figure 2.9 shows the total luminosity delivered to CMS in each of these years. Although the 6.1 fb $^{-1}$ of 2011 data is analyzed in the context of many measurements for the Higgs boson and other SM processes, this analysis of double Higgs production pertains principally to 8 TeV data.

2.3 CMS

CMS is one of two general purpose detectors on the LHC. It was built in part to search for the Higgs boson in the mass range not already excluded at the time by

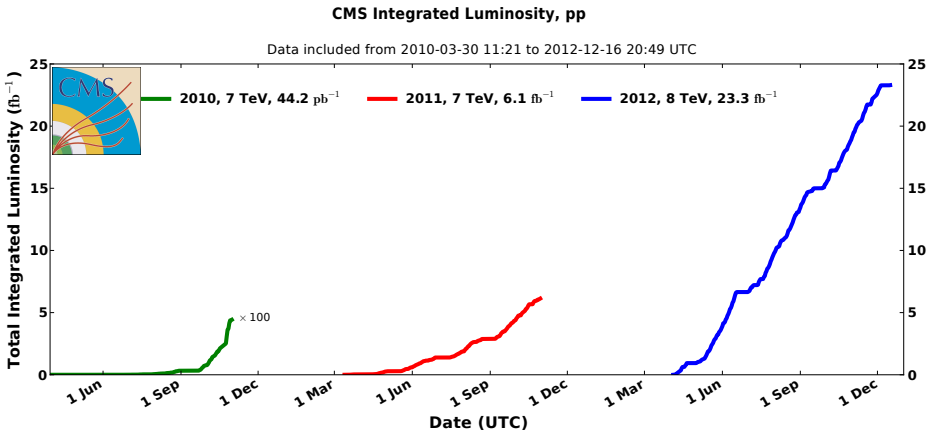

Figure 2.9: The integrated luminosity of pp collisions in CMS from 2010 to 2012 [51].

previous experiments and by theory. Other physics goals include the study of the SM, such as the top quark mass, and the search for BSM physics, such as supersymmetry, extra dimensions, and dark matter, as already discussed in Section 1.3.

The CMS coordinate system is defined with the origin at the center of the detector. The z-axis of this coordinate system points in the direction that the counterclockwise beam takes at this origin. The transverse plane is perpendicular to this axis, with the x-axis pointing to the center of the circle made by the LHC and the y-axis pointing vertically upward. The azimuthal angle ϕ gives the angle in the transverse plane as measured counterclockwise from the x-axis, and the polar angle θ is measured from the z-axis. In CMS, as the case with other collider experiments, a Lorentz-invariant pseudorapidity η is used in place of the polar angle. It is given by

$$\eta = -\log\left(\frac{\theta}{2}\right). \tag{2.1}$$

The central feature of CMS is its superconducting solenoid magnet [52]. At 12.5 m long and with an internal diameter of 6 m, it is the world's largest superconducting magnet. The superconducting material is niobium-titanium, which is cooled to 4 K and provided with a current of 18 kA. This current provides a uniform magnetic field

of 3.8 T inside the solenoid volume, which is referred to as the inner barrel. Outside of the solenoid, an iron return yoke returns the magnetic field that escapes the ends of the cylinder. The region radially outward from the cylinder, called the outer barrel, experiences a uniform field strength of 2 T due to this return yoke. The region beyond either end, called the endcap region, also has part of the return yoke, but the field strength and direction is quite complex here. The return yoke contributes 12.5 kton of the detector's 15 kton.

CMS is described in detail in [53] and [54]. Here an overview of each of the subsystems is provided, starting from the closest to the main interaction point (IP), where the proton beams are made to collide, and proceeding radially outward. Figure 2.10 gives the relative position and sizes of each of the subsystems. These subsystems are, in order of proximity to the IP, the silicon tracker, the electromagnetic calorimeter (ECAL), the hadronic calorimeter (HCAL), the solenoidal magnet, and the muon system.

2.3.1 Silicon Tracker

The silicon tracker [56] is responsible for measuring the positions of charged particles as they propagate from the IP. The aim is to reconstruct the paths of these charged particles by recording their position in successive layers of silicon planes. When coupled with the strong magnetic field provided by the solenoid magnet, a charged particle has a curved trajectory. Track reconstruction gives an estimate of the radius of curvature, which is directly related to the particle's transverse momentum p_{T}. A stronger field allows for better momentum resolution.

The choice of silicon technology is imperative to achieve this goal. When a charged particle passes through the material, electrons are liberated from the atoms of the semiconductor, and a bias voltage allows for the collection of the liberated electrons and holes. If the charged collected during a certain time window exceeds a specified

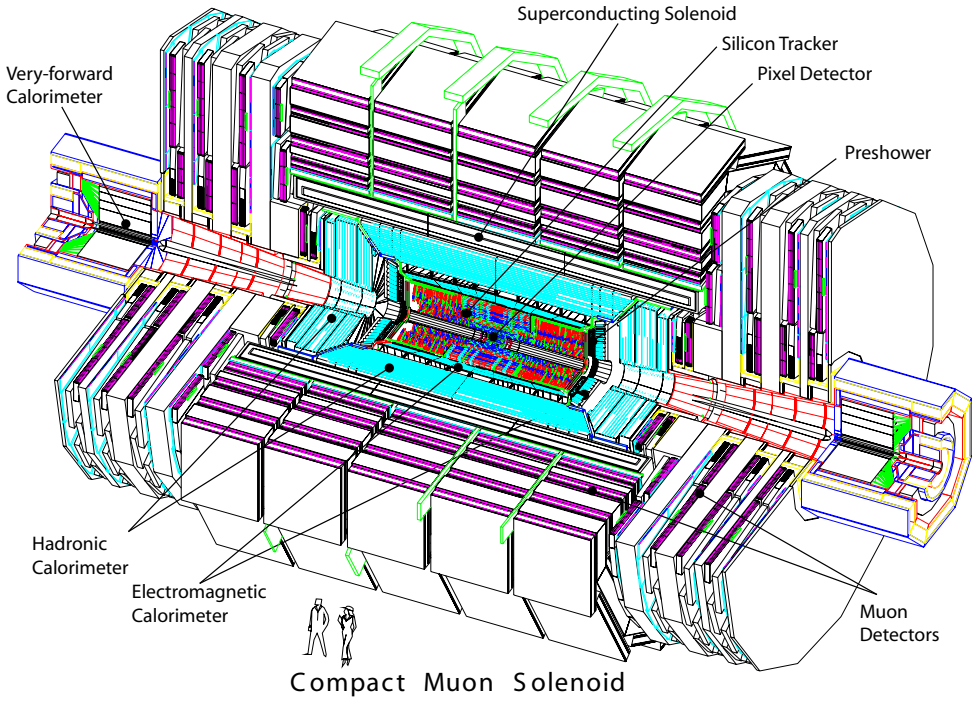

Figure 2.10: The CMS detector, with two people to give a sense of scale [55].

threshold, a hit is registered. The hits from successive layers are reconstructed offline for particle identification and momentum measurements. Section 4.1 discusses the Particle Flow algorithm used for performing this identification.

The tracker occupies a volume of length 5.4 m and radius 1.1 m and can detect particles with $|\eta|$ up to 2.5. It is divided into the pixel detector and strip detector, both of which are divided into barrel regions, or those where the layers are parallel to the z-axis, and endcap regions, or those where the layers are parallel to the transverse plane.

The pixel detector has three layers in the barrel region and two layers on either end in the endcap region, in total making up 66 million channels of size 100 μm $\times$ 150 μm. The strip detector surrounds the pixel detector and adds ten layers in the barrel and 12 layers in each endcap. The silicon strips vary in size from 10 cm $\times$ 80 μm to 25 cm $\times$ 180 μm and are oriented either parallel or perpendicular to the z-axis.

2.3.2 Electromagnetic Calorimeter

The ECAL [57] is a calorimeter whose principal function is to measure the energy of electromagnetic showers. It, like the tracker, is divided into barrel and endcap regions, and provides coverage up to $|\eta| < 3.0$. It accomplishes this by employing 75,000 lead tungstate crystals of size 25 mm $\times$ 25 mm $\times$ 230 mm in the barrel or 30 mm $\times$ 30 mm $\times$ 220 mm in the endcap. The ECAL is 25 radiation lengths long, a relatively large number that is exploited in particle identification. Electrons and photons are stopped by having their energy dissipated through many cascades, electrons through bremsstrahlung and photons through pair production. Hadrons and muons, also interacting electromagnetically, deposit some energy as a result of ionization, but as a consequence of their higher masses generally do not induce electromagnetic showers. The energy resolution provided by the ECAL, as measured from electron test beams, is

$$\frac{\sigma(E)}{E} = \frac{0.028}{\sqrt{E/\text{GeV}}} \oplus \frac{0.12}{E/\text{GeV}} \oplus 0.3\%. \tag{2.2}$$

2.3.3 Hadronic Calorimeter

The last subsystem within the volume of the solenoid magnet is the HCAL [58]. The HCAL has barrel and endcap divisions (HB and HE, respectively) similar to ECAL which covers up to $|\eta| < 3.0$. These components are made of plastic scintillator and brass absorber, segmented into 3700 towers. The towers vary in coverage based on their placement, where the most central towers span $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. The purpose of this subsystem is to measure the energy of hadrons, which usually appear in collimated sprays called jets j. In addition to energy and position measurement, nearly all hadrons are stopped and absorbed due to the high density and large material budget of brass. There is an additional component of HCAL in the barrel region

immediately outside of the magnet (HO) whose purpose is to serve as another layer of absorption for any hadrons that make it through the preceding layers.

A final component of HCAL in the forward region of the detector (HF) covers pseudorapidities $3.0 < |\eta| < 5.0$. In this region, the radiation is especially high, so a steel and quartz-fiber Cherenkov calorimeter is used instead. This component provides 860 such towers.

The energy resolution is improved by the incorporation of information from other subsystems via the Particle Flow algorithm, discussed in Section 4.1. The resolution is about 15% for the jets of interest to this analysis—i.e., those which are central in the detector with sufficiently high p_{T}. Section 4.3 discusses these requirements.

Figure 2.11 shows a quadrant of the inner subsystems of CMS. This gives a scale of the relative positions and sizes of the tracker, ECAL, and most of HCAL with respect to each other and the IP. Missing from the figure are the HF in the very forward region and the HO just to the outside of the magnet.

2.3.4 Muon System

The muon system [59] is the subsystem furthest from the IP. It measures the positions of muons through detection planes exploiting three different technologies: drift tubes (DT), resistive plate chambers (RPC), and cathode strip chambers (CSC). Together these components provide muon coverage up to $|\eta| < 2.4$. The DTs, located in the barrel, are gas tubes with a wires for collecting electrons liberated by a transversing muon. The RPCs, partially covering both barrel and endcap regions, consist of two parallel plates of opposite bias separated by a gas medium. The CSCs, covering the endcap region, consist of layers of anode, separated by layers of cathode strips oriented perpendicularly, all within a gas volume. Figure 2.12 shows the relative position of these three components with respect to each other and the subsystems inside the solenoid.

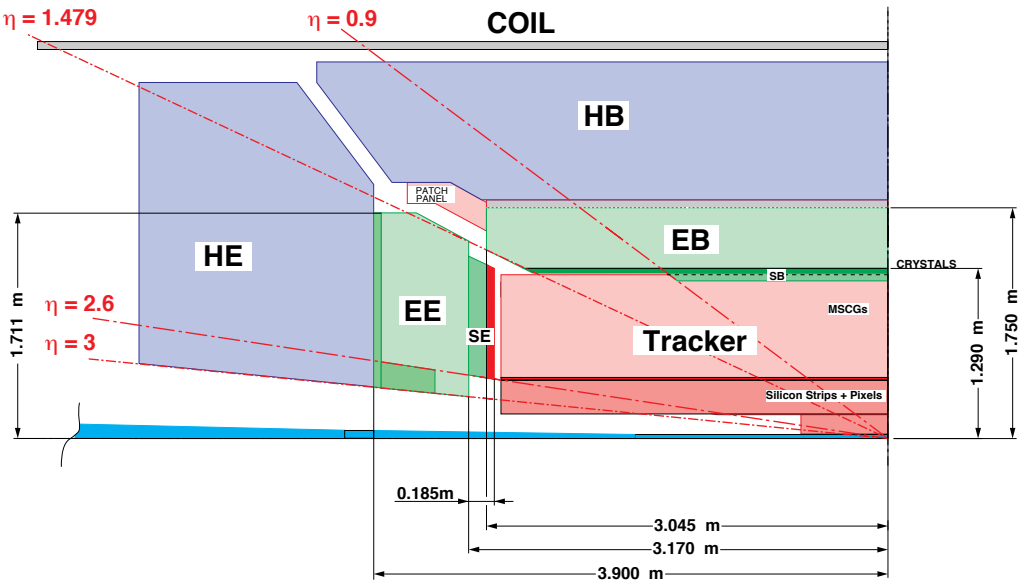

Figure 2.11: A schematic of one quarter of the inner CMS subsystems [57]. All components of the tracker are in red. The ECAL barrel (EB) and ECAL endcap (EE) are in blue. The HCAL barrel (HB) and HCAL endcap (HE) are in green. Coil refers to the superconducting solenoid. Dimensions and lines of constant η from the IP are superimposed.

The hits of the muon system are matched with hits from the tracker to form a single track with improved momentum resolution. Unlike other charged particles, which register significant activity in one or both calorimeters, the energy measurement of muons is based solely on their tracks. Higher energy muons are curved less by the magnetic field, causing the energy resolution to worsen as their tracks approach a straight line.

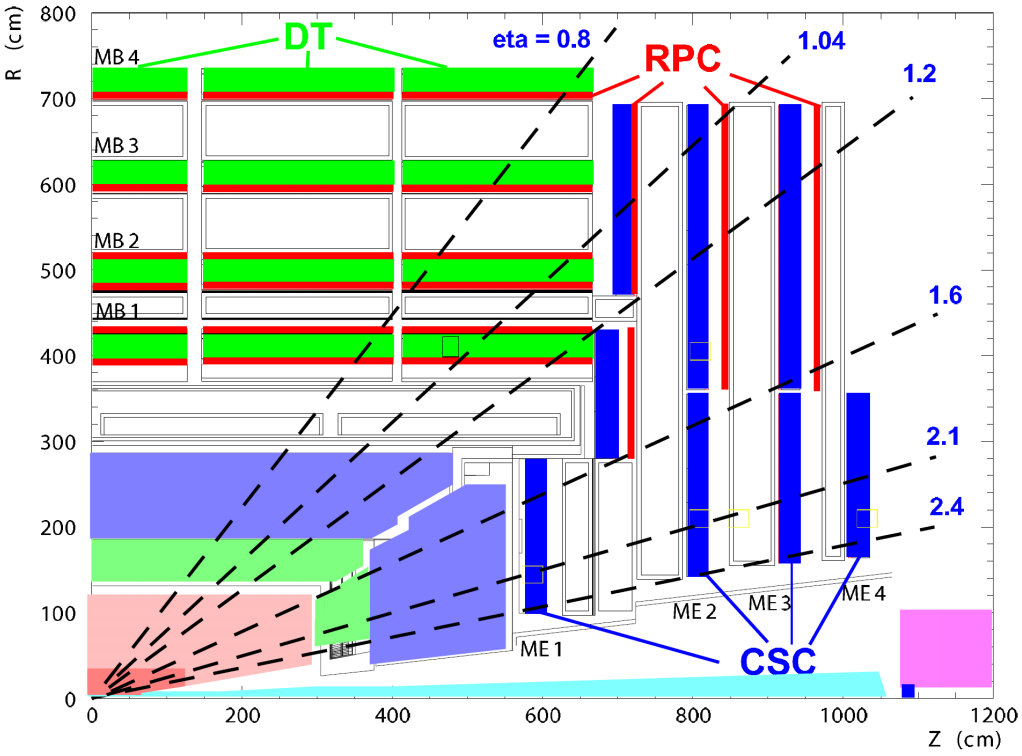

Figure 2.12: A schematic of one quarter of the muon systems [53]. Lines of constant η from the IP are superimposed.

Chapter 3

Big Data

The LHC delivers *pp* collisions at a high rate that is much higher than CMS can read out or store offline. From all these collisions, CMS selects a small fraction of events which are the most relevant for the physics processes of interest. This task is accomplished through the trigger system, which is discussed in Section 3.1 both in general and in the context of this analysis. The events selected by the trigger system compose the datasets used by all CMS analyses. The worldwide computing grid for storage and processing of both datasets and simulation samples is discussed in Section 3.2. Finally, the simulation samples employed by this analysis are discussed in Section 3.3.

3.1 The Trigger System

The gap in time between successive bunch crossings by the LHC is 25 ns, equivalent to a frequency of 40 MHz. This large rate is reduced by selecting only the most interesting events that pass a two-level trigger system. The first level, Level 1 (L1), is hardware-based and reduces the total event rate from 40 MHz to 100 kHz [60], whereas the second level, the High Level Trigger (HLT), is software-based and reduces

the total event rate from 100 kHz to 100 Hz [61]. From here, the events are processed further and stored for the physics analyses.

The L1 trigger attempts to identify basic physics objects based on coarse energy deposits in ECAL or HCAL or based on collections of hits in the muon chambers. Quantities derived from the sum of energy deposited, such as the missing transverse energy $\not{E}_{\mathrm{T}}$, provide other quantities on which to characterize events. The Level 1 global trigger (GT) combines these object candidates and derived quantities in order to select events to pass the the HLT with a processing time of 3 μs. Up to 128 separate trigger paths can be supported, 126 to 127 of which were used during 2012 data taking.

The HLT is able to access more information than the L1 trigger, and in doing so it can provide a better description of the event. At the HLT level, tracker information is used in conjunction with the full granularity of ECAL and HCAL. Information at this level is based on the presence of one or more candidate objects satisfying requirements based on their transverse momentum or energy and relative or absolute positions in the detector. With this added complexity and fewer events to process, the average processing time is 40 ms per processor.

3.1.1 The Trigger for $\gamma\gamma b\bar{b}$

The search for the $\gamma\gamma b\bar{b}$ final state can be viewed as an extension of the SM $H \to \gamma\gamma$ search [62]. The excellent diphoton mass resolution is the primary driver in the sensitivity, since it allows for low background contamination in the signal region, as will be shown in Chapter 6. Therefore, the trigger strategy centers on the ability to find two high-quality photon candidates.

During 2012 data taking, the LHC luminosity increased over time, and the triggers at both L1 and HLT had their thresholds increased in order to keep the event rates within limits. The requirement at L1 is for two e/γ candidates with E_{T} requirements

of 13 (7) GeV for the leading (subleading) candidate or for one e/γ candidate with an E_{T} requirement of 22 GeV. Note that at this level, there is no distinction between electron and photon candidates as the tracker information is not yet used.

Events are selected in the HLT through diphoton triggers with asymmetric E_{T} thresholds and complementary photon selections. One trigger selection requires a loose calorimetric identification based on the shape of the electromagnetic shower and loose isolation requirements on the photon candidates, while the other requires that the photon electromagnetic shower is primarily concentrated in a three-by-three crystal super-cluster. The trigger thresholds on the photon E_{T} are 26 (18) GeV and 36 (22) GeV on the leading (subleading) photon, depending on the acquisition period of LHC data taking in 2012. The path with the 26 (18) GeV thresholds is initiated by the L1 path with 13 (7) GeV thresholds, while the path with 36 (22) GeV thresholds is initiated by the L1 path with a single 22 GeV threshold.

In addition to keeping the trigger rate within its limits, it is also necessary to set thresholds low enough such that the selection of signal events, or trigger efficiency, remains as high as possible. The trigger efficiency is studied on Monte Carlo (MC) signal samples as well as on $Z \to e^+e^-$ data. According to the MC study, the efficiency is above 99.5% for all conditions of 2012 data taking.

Figures 3.1 and 3.2 show how the trigger efficiency varies on data with respect to the candidate photon p_{T} and the number of primary vertices in the event using the tag and probe technique. To account for differences in the shower shape between photons and electrons, the data sample was reweighted such that the shower shape distributions match. Within uncertainties, the trigger efficiency is higher than 99% for the 2012 data.

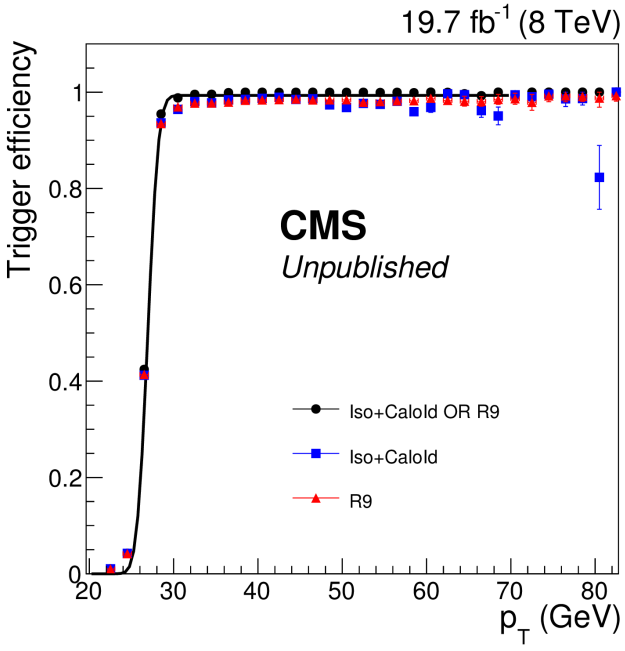

Figure 3.1: Efficiency of the trigger selection as a function of the photon candidate transverse momentum measured in $Z \to e^+e^-$ events [62].

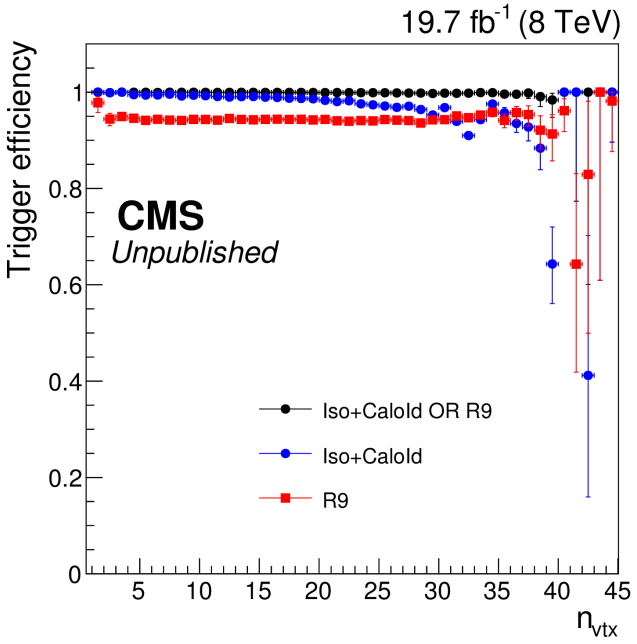

Figure 3.2: Efficiency of the trigger selection as a function of the number of primary vertex in the event measured in $Z \to e^+e^-$ events [62].

3.2 Data Storage Worldwide

All the data from the LHC and its experiments, including CMS, is processed, stored, and analyzed in a distributed global collaboration of computing centers [63]. The Worldwide LHC Computing Grid (WLCG) is the world's largest computing grid, comprising over 170 centers arranged in a structure of tiers. Part of this tier structure is shown in Figure 3.3

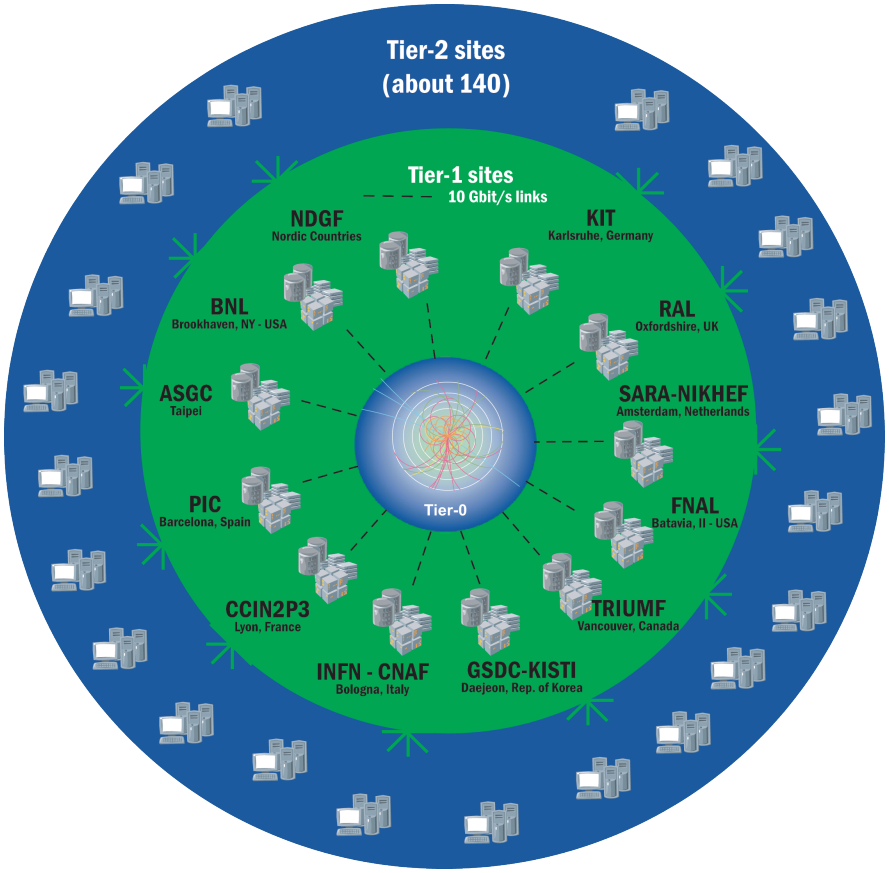

Figure 3.3: A schematic of the WLCG detailing the locations of the Tier-1 sites [64].

Tier 0 is the CERN Data Center, through which all LHC data passes for initial processing and reconstruction. For CMS, this comprises the events that pass the HLT. The raw and processed data from Tier 0 is pushed to one of the 13 Tier 1 sites, where later reprocessing and storage can be done. From here, the Tier 1 sites push the reconstructed datasets to Tier 2 sites for storage and processing by analysts.

This system handles the approximately 30 Petabytes of data generated per year by the LHC experiments in addition to the large number of MC samples generated and stored.

3.3 Simulation Samples

In a general search, MC simulation samples are employed to study the signal process of interest. They also allow for the study of background processes in more detail than would be offered by data alone. This analysis searches for a double Higgs final state by optimizing strategies separately between the resonant and SM nonresonant production mechanisms. The signal MC samples are discussed in Section 3.3.1. In addition, good agreement between data and background MC confirms that the major backgrounds are understood and that the data is behaving as expected. The background MC samples are discussed in Section 3.3.2.

Event generation begins with sampling the probability distribution estimated from the matrix element of the hard subprocess of interest (for example, $gg \to HH \to \gamma\gamma b\bar{b}$). The matrix element might have many contributions at tree level and, at higher orders, integrals involving loops, rendering an exact calculation computationally infeasible. From the hard subprocess, additional effects are added so that the simulation corresponds with the final state observed in the detector. These effects include the initial state radiation (ISR) and final state radiation (FSR) of a gluon or photon to the hard subprocess, multi-parton interactions (MPI) from other parton interactions in the proton-proton collision with their associated ISR and FSR, and beam remnants from the outgoing protons. Due to color confinement, the partons cannot exist individually, and hadrons are formed from the bare partons in a process called hadronization. Unstable hadrons decay further until more stable hadrons are produced. The clustering of the stable hadrons observed in the detector give an infer-

ence of the momentum of the parton produced in the hard subprocess. More detail on the generation of simulation samples is provided in [65, 66].

3.3.1 Signal Simulation

The resonant search is focused on a new resonance from a massive particle X with mass m_X between 260 GeV and 1.1 TeV. The lower bound is set by twice the Higgs mass since the search looks for decays $X \to HH$. The upper bound is set by the ability to reconstruct the $\gamma\gamma b\bar{b}$ final state: for very high resonance hypotheses, the decay products are highly boosted, causing the photons of the $H \to \gamma\gamma$ decay to fail identification and causing the jets of the $H \to b\bar{b}$ decay to overlap. These reconstruction methods are discussed in Chapter 4. For the nonresonant search, the final state is not sufficiently boosted to cause problems. The lack of boost in that search is due to the distribution of the four-body mass being lower than that for a high-mass resonance, where the spectrum is narrowly peaked about the resonance mass.

In the resonant search, the benchmark model is that of the Radion, which is simulated with MadGraph5 [67] and hadronized with Pythia6 [68]. The events are generated from the gluon-fusion production of an on-shell Radion where the decay width of the Radion is 10 MeV, much less than the experimental resolution. There are roughly 20k events generated per mass point, with twice the number at three points. This is summarized in Table 3.1 with the corresponding cross sections given in terms of parameters described in Chapter 1.

In order to verify the model-independence of the result, two additional resonant signals are generated. The spin-two graviton is considered as the angular distribution of the final state differs from a spin-zero scenario. However, due to the expected sensitivity of the analysis, as shown in Chapter 8, the analysis is not expected to be able to differentiate between these two scenarios. These samples were generated with a decay width of 1 GeV and 10 GeV, which allows for the assessment of the dependence

Table 3.1: Radion simulation samples and their corresponding cross sections.

$\Lambda_R = 3$ TeV, $kl = 35$, no $R - H$ mixing		
Mass m_X (GeV)	Number of events	$\sigma \cdot BR(R \to HH \to \gamma\gamma b\bar{b})$ (fb)
270	19,996	2.18
300	39,999	1.94
350	18,498	1.19
400	19,697	0.621
450	19,999	0.402
500	39,997	0.282
550	19,995	0.206
600	18,197	0.156
650	20,000	0.121
700	39,995	0.095
800	19,999	0.0626
900	19,996	0.0435
1000	19,996	0.0317
1100	19,400	-

of the decay width on the final result. The production mechanism is through gluon-gluon fusion, generated with MadGraph5 and hadronized with Pythia6. The sample generation is summarized in Table 3.2 with corresponding cross sections given in terms of parameters described in Chapter 1. Here, the RS1 KK-graviton refers to a scenario where $c_i = 1$ in Equation 1.12, whereas the bulk KK-graviton refers to a scenario in which the couplings between light quarks and the graviton are suppressed, causing the production cross section of the bulk KK-graviton to be lowered by a factor of 0.02 from the RS1 KK-graviton. The cross sections are given in Table 3.3.

A second alternative signal scenario to the Radion benchmark is the spin-zero heavy neutral Higgs from the MSSM. These samples are generated and showered with Pythia6. For this scenario, only low-mass resonance hypotheses are generated since, once the heavy Higgs mass rises above twice that of the top mass, it will

Table 3.2: Graviton simulation samples.

Mass m_X (GeV)	Width Γ_X (GeV)	Number of events
300	10	49,941
500	10	49,905
700	10	49,911
1000	10	49,921
270	1	19,798
300	1	19,996
350	1	19,999
400	1	19,999
450	1	19,999
700	1	19,999

overwhelming decay through that channel. The sample generation for the MSSM heavy Higgs is summarized in Table 3.4.

In the search for nonresonant HH production, the three parameters κ_λ, κ_t, and c_2 may be varied as discussed in Section 1.4.1. These samples are generated with MadGraph5 using suitable ranges for the parameters about their corresponding SM values. The values considered for κ_λ are -20, -15, -10, 0, 1, 2, 3, 5, 10, 15, and 20; for κ_t 0.75, 1, and 1.25; and for c_2 -3, -2, 0, 2, and 3. With these values, 124 scenarios are generated, each having about 20k events.

3.3.2 Background Simulation

The backgrounds processes for the search of double Higgs production are divided in resonant and nonresonant groups. The resonant backgrounds consist of SM Higgs production with $H \to \gamma\gamma$, meaning that for the diphoton mass, which is one of the most powerful discriminators between signal and background, these processes have the same distribution as the signal being sought. The nonresonant backgrounds consist

Table 3.3: Graviton cross sections. It is assumed that the maximal branching ratio is 25% for the KK-graviton to two Higgs for all masses. The cross section is the same for the 1 GeV and 10 GeV widths because the specific values of fermion localization leave some freedom in the partial width of the KK-graviton to two top quarks [69].

$kl = 35$, $k/M_{Pl} = 0.2$, elementary top		
	$\sigma \cdot BR(HH)(2 \cdot BR(\gamma\gamma) \cdot BR(b\bar{b}))$ (fb)	
Mass M_X (GeV)	RS1 KK-graviton	Bulk KK-graviton
260	3.408	0.0119
270	3.160	0.0251
300	2.560	0.0671
350	1.881	0.0975
400	1.440	0.0748
450	1.138	0.0487
500	0.921	0.0309
550	0.762	0.0197
600	0.640	0.0128
650	0.545	0.00850
700	0.470	0.00574
750	0.410	0.00393
800	0.360	0.00274
900	0.284	0.00138
1000	0.230	0.000727
1100	0.190	0.000398

Table 3.4: MSSM heavy Higgs simulation samples.

Mass m_X (GeV)	Number of events
260	300,000
300	299,142
350	299,571

of SM processes, which have final state objects reconstructed as $\gamma\gamma b\bar{b}$ but with no resonance in the region the signal processes have.

For the resonant background, six (five) processes are considered for the resonant (nonresonant) search. Four of the processes are SM single Higgs production with the $H \to \gamma\gamma$ decay: gluon-gluon fusion, vector boson fusion, associated production with a vector boson, and associated production with a pair of t-quarks, as shown in Figure 1.5. Gluon-gluon fusion mimics the final state when two jets are reconstructed from initial state or final state radiation (ISR or FSR) or from interactions from the underlying event (UE). The other production mechanisms produce two jets directly, whereby in associated production with a vector boson the vector boson can decay to two jets, and in associated production with a pair of t-quarks each t-quark always decays to bW. The first two prodcution mechanisms were generated with PowHeg [70], while the latter two were generated with Pythia6. All four were hadronized with Pythia6. In addition, a fifth mechanism, associated production with a pair of b-quarks, is considered for both searches and was generated with MadGraph5 and hadronized with Pythia6. For the resonant search, a final resonant background is considered, that of the SM double Higgs production (i.e., the signal process for the nonresonant search). This is summarized in Table 3.5, and its corresponding effect on the analysis sensitivity is given in Chapter 8.

Table 3.5: Resonant background simulation samples and their corresponding cross sections.

Process	Number of events	$\sigma \cdot BR(H \to \gamma\gamma)$ (fb)
ggF H	96,161	43.9
VBF H	99,671	3.60
VH	100,151	2.56
$t\bar{t}H$	93,183	0.295
$b\bar{b}H$	99,434	0.464
ggF HH	20,000	0.0257

The composition of the nonresonant background is dominated by the QCD production of two reconstructed photons, which can be characterized as prompt, corresponding to real photons, or as fake, for example from light neutral mesons mimicking the signature of a photon. The QCD component is therefore broken down into fake-fake, prompt-fake, and prompt-prompt. The first two are simulated QCD or photon-jet (QCD process requiring one photon) processes, generated and hadronized with Pythia6. The last is simulated with a diphoton-jet process, generated and hadronized with Sherpa [71]. In these three QCD processes, a fake photon can still be selected over a prompt one, so for the analysis, the QCD processes are broken down in terms of number of prompt photons, as will be shown in Chapter 5.

The nonresonant background has a small contribution from processes involving vector bosons and top quarks. The vector boson contribution is simulated from processes involving a Z boson or off-shell photon γ^* decaying to a pair of leptons (Drell-Yan process) or involving a W boson decaying leptonically with photon ISR or FSR. The top-quark contribution is simulated from processes involving the production of a top-quark pair with the addition of a photon-jet or diphoton pair or involving the production of a single top quark with a diphoton pair.

The nonresonant background samples used in this analysis and their corresponding cross sections are summarized in Table 3.6. Their effect on the analysis is studied in Chapters 5 and 6.

Table 3.6: Nonresonant background simulation samples and their corresponding cross sections In the QCD and photon-jet samples, the number denotes the scale on which the parton distribution function of the proton as QCD diverges as the scale goes to zero [72]. For example, 30-40 denotes a scale of 30 to 40 GeV, while 40 denotes a scale of 40 GeV and above.

Process	Number of events	σ (pb)
QCD 30-40	6,047,441	12,208
QCD 40	9,764,546	51,439
γj 20-40	5,901,106	150.34
γj 40	17,722,752	478.58
$\gamma\gamma j$	14,426,200	120.354
$Z/\gamma^* \to \ell^+\ell^-$	30,290,538	2,950.0
$Z(\ell^+\ell^-)\gamma$	6,588,161	132.6
$W(\ell\nu)\gamma\gamma$ (ISR)	1,003,920	0.319
$W(\ell\nu)\gamma\gamma$ (FSR)	1,000,310	1.84
$t\bar{t}\gamma\gamma$	122,040	0.14584
$t\gamma\gamma$	324,676	0.00337
$t\bar{t}\gamma j$	1,719,954	14

Chapter 4

Physics Objects

After data is selected by the trigger, the offline analyses begin with particle identification. There are no longer any timing limitations, so such identification can make use of all the detector information in the event. To this end, CMS uses the Particle Flow (PF) algorithm in most analyses, which is described in Section 4.1. As this search for double Higgs production centers around the identification of events with both $H \to \gamma\gamma$ and $H \to b\bar{b}$ decays, the identification and reconstruction of photons and jets are the first steps taken after the trigger selects potentially interesting events, and this chapter discusses the treatments needed at this stage for both data and MC samples. Recalling that the sensitivity in the separation between signal and background comes from the excellent diphoton mass resolution, the identification of two high quality photons is the starting point and discussed in Section 4.2. The next step is the identification of two jets coming from the hadronization of b-quarks and is discussed in Section 4.3.

4.1 Particle Flow

The PF algorithm reconstructs all stable particles in an event from the digitized electronic signals of all channels in all subsystems [73, 74]. These particles include

electrons, photons, charged hadrons, neutron hadrons, and muons, as shown in Figure 4.1. From these particles, derived objects are constructed, including jets and missing transverse energy. The algorithm itself links detector objects created from individual subsystems and groups them into blocks that are identified with a particle. The detector objects are discussed in Section 4.1.1, the linking of these objects is discussed in Section 4.1.2, and the identification of groups of these links with particles is discussed in Section 4.1.3.

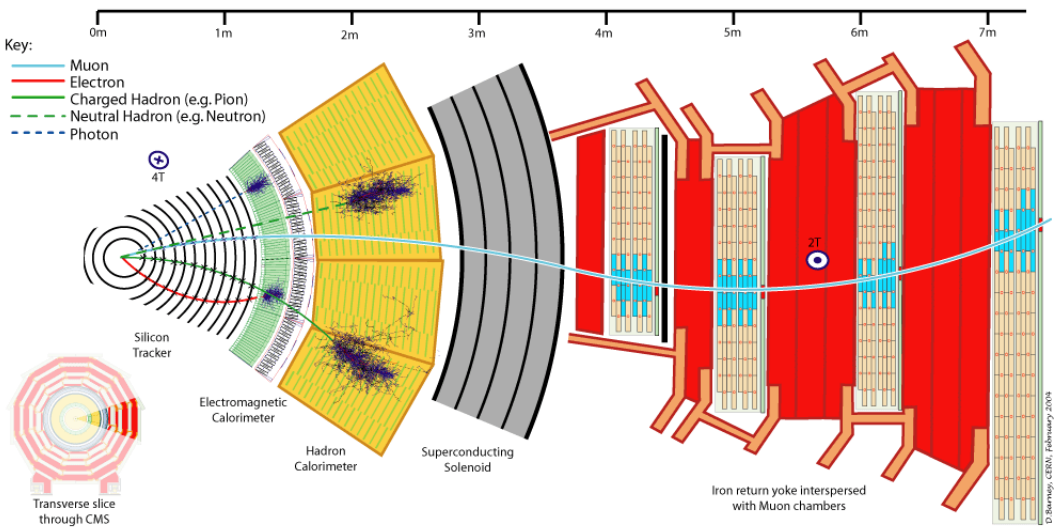

Figure 4.1: A slice of the CMS detector in the plane transverse to the beam line. The trajectories of an electron, photon, charged hadron, neutral hadron, and muon are superimposed with the interactions that each of these particles has with the various subsystems.

4.1.1 Detector Objects

The first step of the PF algorithm is to assemble detector objects created separately from individual subsystems. In the tracker, pixel and strip hits are associated into a track candidate through the iterative Combinatorial Track Finder algorithm [73]. In the muon chambers, standalone muon tracks (to distinguish from tracks formed by tracker hits) are assembled from hits in the three muon detectors, accounting for

the nonuniform magnetic field and large detector budget and constraining a track candidate to intersect the beamline. In the ECAL and HCAL, clustering of energy deposits is performed around a cluster seed, which is identified as a detector unit with an amount of energy exceeding a detector-dependent threshold and corresponding to a local maximum relative to neighboring units. Units are then added to adjacent seeds if their energy exceeds a noise threshold, and the PF clusters are formed by redistributing the the energy back to the cluster seeds, recalculating the position as a weighted average over the energy of each contributing cluster.

4.1.2 Linking

The next step of the PF algorithm is to link the detector objects to assemble PF candidates. Possible links are between a track and standalone muon track, between a track and a cluster, and between two clusters, each having their own associated linking parameters. A link is formed between a track and standalone muon track when the two can be merged into a global track with a fit having χ^2 below a certain threshold. A link is formed between a track and a cluster when the extrapolation of the track is within a certain distance of the cluster position. A link is formed between two clusters (either both in ECAL, both in HCAL, or one in each) when the two clusters are within a certain distance to each other.

4.1.3 Grouping and Identification

An ensemble of links creates a block, and identification proceeds iteratively through PF candidates. First, muons are identified from those blocks that contain a global track having a momentum sufficiently close to the momentum of the contained track. Then electrons are identified from blocks containing a track and an ECAL cluster where the track and energy cluster satisfy requirements consistent with the signature of an electron. Next photons and hadrons are identified from blocks containing a track

and a cluster from either ECAL or HCAL. If the calibrated energy in the clusters is greater than the sum of the momentum of the associated tracks, PF photons or PF neutrons are created from the difference. If the difference between the calibrated energy in the clusters of the block and the sum of the momentum from the tracks associated to the clusters is less than the energy in the ECAL clusters, a PF photon is created from the block; if the difference is greater than the energy in the ECAL clusters, a PF photon and a PF neutral hadron are made from the excesses ECAL and HCAL energy, respectively. Finally, if the calibrated energy in the clusters is less than the sum of the momentum of the associated tracks, a search for fake tracks and additional muons in the block is performed, and what remains in the block is a PF charged hadron.

From the list of PF candidates in an event, jets and taus are constructed by clustering nearby hadrons. In this way, the clustering of PF hadrons represents the original quark or tau from the underlying interaction. Missing transverse energy, which is a signature of one or more neutrinos in the event and/or the mismeasurement of the energy of PF candidate, is obtained by

$$\not{E}_{\mathrm{T}} = -\sum_{i} \vec{p}_{\mathrm{T},i}\,, \tag{4.1}$$

where the sum is over all PF candidates in the event.

The treatment for photons is discussed in more detail in Section 4.2. The construction of jets from PF candidates is discussed in more detail in Section 4.3.

4.2 Photons

PF photon candidates are reconstructed from clustering individual units in the ECAL and checking consistency with tracks. The calorimeter signals are calibrated for several detector effects [75, 76], providing the best energy resolution possible. The energy

scale is corrected in both data and simulation, while the photon energy is smeared in simulation in order to reproduce the same energy resolution that is observed in data.

With the PF photon candidates in hand, additional requirements are imposed in order to further separate prompt photons from fake photons originating from misidentified electrons or from jets. These additional requirements include an electron veto, an upper threshold on the energy deposited in HCAL in the region about the candidate, isolation, and the shower shape. The electron veto removes a photon if there is an electron candidate matching the photon ECAL cluster with no missing hits in the tracker and with no matching reconstructed conversion. Isolation requirements place thresholds on the amount of ECAL energy deposited in a region about the cluster; these are both detector based and PF based. Requirements on the electromagnetic shower shape include the width of the shower in terms of ECAL detector units and the ratio of the amount of energy in a 3×3 box around the cluster seed to that of a 5×5 box around the same seed.

After the quality cuts, the two photons candidates are requested to satisfy the sliding asymmetric cuts

$$p_{\mathrm{T},\gamma_1} > \frac{m_{\gamma\gamma}}{3} \tag{4.2a}$$

$$p_{\mathrm{T},\gamma_2} > \frac{m_{\gamma\gamma}}{4}\,, \tag{4.2b}$$

where p_{T,γ_1} and p_{T,γ_2} are the transverse momenta of the leading and subleading photons, respectively. These requirements on the transverse momentum are scaled by the diphoton invariant mass in order to prevent turn-on effects which could distort the shape of the $m_{\gamma\gamma}$ spectrum for low values of $m_{\gamma\gamma}$. The position of photons is required to be within the ECAL acceptance of $|\eta| < 2.5$ with an exclusion on the ECAL gap between the barrel and endcaps. If there are more than two photons passing the identification and kinematic requirements, the two with the largest p_{T} are chosen. After this choice, the diphoton mass is required to be between 100 and 180 GeV.

Figure 4.2 shows the resulting simulated distributions for the diphoton mass of the resonant signal and resonant backgrounds. For the resonant background, the sum of all SM $H \to \gamma\gamma$ production mechanisms is shown as a single contribution. Figure 4.3 shows the same distribution for data and the sum of backgrounds. (Note that these figures include requirements on jets, discussed in Section 4.3.) The resolution of the diphoton mass spectrum for the signal is a few GeV.

The two primary drivers in the diphoton mass resolution are the energy resolution of the individual photons and the direction of the photons from their origin, which is synonymous with the identification of the vertex from which they were produced. Due to the presence of the $H \to b\bar{b}$ decay in the search, the tracks from the jets allow for the efficient identification of the correct vertex. The criterion for the vertex choice is that which has the largest $\sum_i p_{\mathrm{T},i}$, where the sum is over all of the tracks associated with a particular vertex. With this criterion, the relative contribution to the diphoton mass resolution due to the vertex choice is negligible with respect to the energy resolution for individual photons.

4.3 Jets

PF jet reconstruction involves the clustering of all PF candidates except the two photons selected as the $H \to \gamma\gamma$ candidate [73, 74]. The anti-k_{T} algorithm [77] with a distance parameter of $R = 0.5$ is used to cluster the PF candidates into jets as an approximate attempt to reverse engineer the processes of hadronization and fragmentation after a b-quark was produced in the interaction. In this greedy algorithm, a list of particles and pseudojets is maintained, and at each iteration elements i and j are combined into one pseudojet if that pair has the minimum

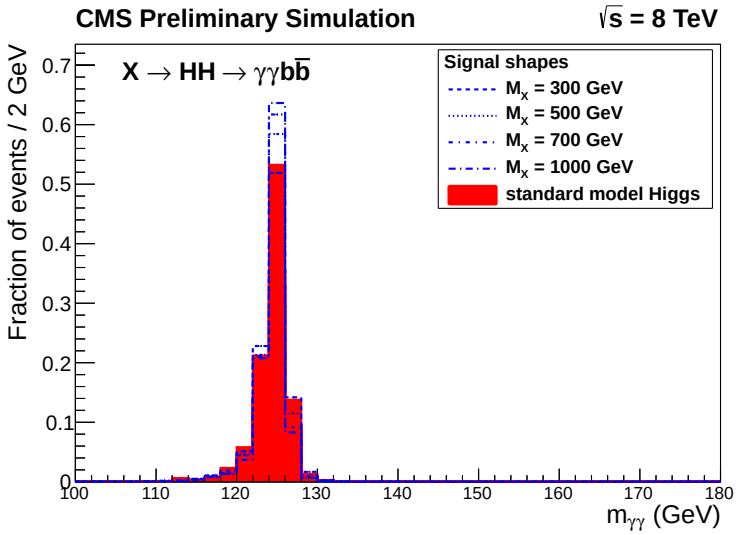

Figure 4.2: Simulated diphoton mass spectrum for the resonant signal and the sum of all production mechanisms of the SM Higgs boson after basic selections on photons and jets and requesting at least one loose b-tagged jet. The spectra are normalized to one.

distance measure in the list, where the distance measure d_{ij} is defined as

$$d_{i,j} = \min\left(\frac{1}{k_{\mathrm{T},i}^2}, \frac{1}{k_{\mathrm{T},j}^2}\right)\frac{\Delta R_{i,j}^2}{R^2}\,, \tag{4.3}$$

where $k_{\mathrm{T},i}$ is the transverse momentum of i and $\Delta R_{i,j} = \sqrt{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2}$ is the angular separation between i and j. If, on a particular iteration, $d_{i,B}$ is the minimum, defined as

$$d_{i,B} = \frac{1}{k_{\mathrm{T},i}^2}\,, \tag{4.4}$$

then element i is considered a jet and is removed from the list. The algorithm continues until the list is empty. The advantages of this clustering algorithm are that clusters tend to be around hard energy deposits and that the clustering is not sensitive to hadronization from pileup and the underlying event.

After the jet reconstruction, the energy is corrected for pileup contributions from both the same bunch crossing and neighboring bunch crossings. This is accomplished

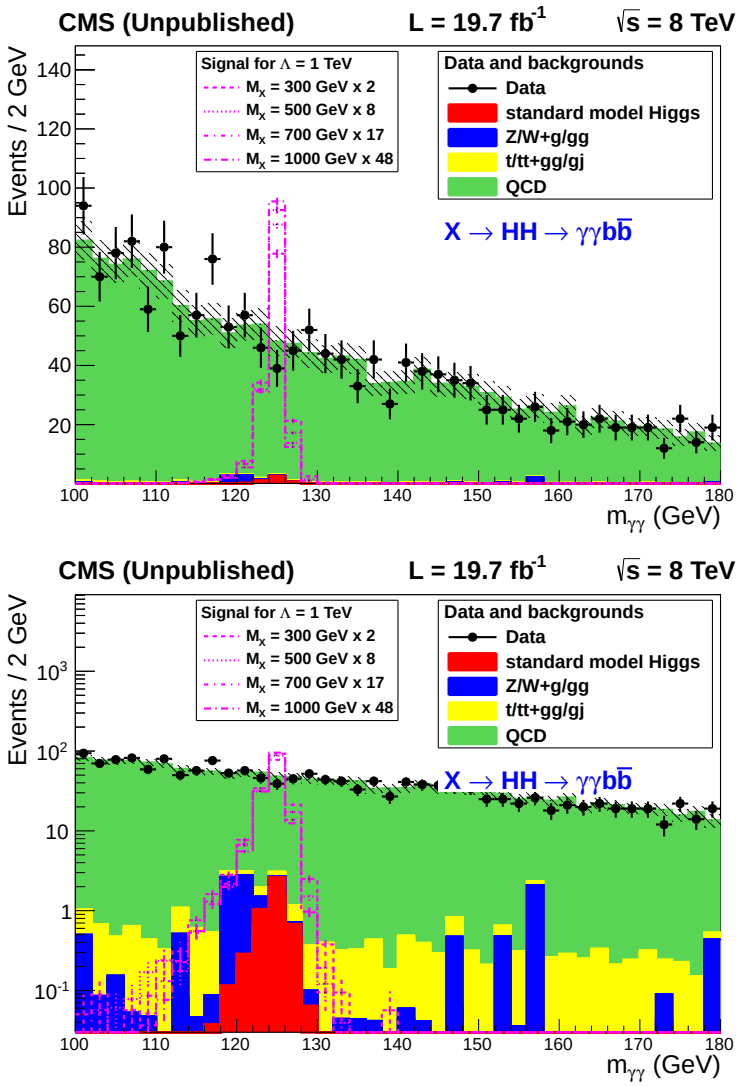

Figure 4.3: Control plots for the diphoton mass spectrum after basic photon and jet selections and requiring at least one loose b-tagged jet. The simulation is normalized to data and the statistical uncertainty on the number of simulated events is shown in dashed overlay. The top (bottom) figure is shown in linear (log) scale.

with the the jet area technique [78] carried out by the FastJet package [79] in which the jet p_{T} is corrected per unit area based on an event-by-event estimation of pileup activity. Jet energy is further corrected as a function of the jet p_{T} and η [80] for the effects of potential mismeasurement and the presence of neutrinos.

Additional requirements are imposed to reject jets that originate from detector noise or from the clustering of particles due to pileup in favor of those that originate from a parton in the primary interaction [81]. Two criteria used are based on the fraction of charged PF candidates attached to the primary vertex β^* and the jet width $\langle \Delta R^2 \rangle$, defined as

$$\beta^* = \frac{\sum_{i \in \text{other PV}} p_{\mathrm{T},i}}{\sum_{i \in \text{charged}} p_{\mathrm{T},i}} \tag{4.5a}$$

$$\langle \Delta R^2 \rangle = \frac{\sum_i \Delta R_i^2 p_{\mathrm{T},i}^2}{\sum_i p_{\mathrm{T},i}^2}\,, \tag{4.5b}$$

where "other PV" refers to the PF charged candidates associated with another primary vertex, the sum in the denominator of β^* is over the charged candidates in the jet, the sums in $\langle \Delta R^2 \rangle$ are over all constituents in the jet, and ΔR_i is the angular distance between constituent i and the jet axis. Upper thresholds are imposed on these quantities; in the case of β^* the threshold is a function of the number of reconstructed vertices.

The efficiency of the jet identification from the above criteria exceeds 95%. Additional kinematic requirements on the jets include $p_{\mathrm{T}} > 25$ GeV and $|\eta| < 2.5$ (within the tracker acceptance). If more than two jets pass identification and kinematic selections, the two jets with the highest $p_{\mathrm{T},jj}$ are chosen to make the dijet candidate, as will be discussed in more detail in Section 5.2. Figure 4.4 shows the resulting simulated distributions for the dijet mass of the resonant signal and resonant backgrounds. Figure 4.5 shows the same distribution for data and the sum of backgrounds. (Note that these figures include a requirement on the compatibility of one of the jets coming from a b-quark, discussed in Section 4.3.1.)

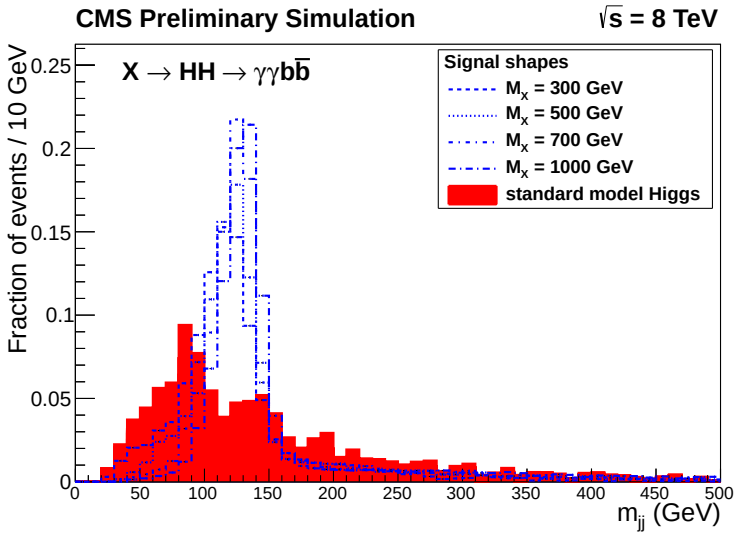

Figure 4.4: Simulated dijet mass spectrum for the resonant signal and the sum of all production mechanisms of the SM Higgs boson after basic selections on photons and jets and requesting at least one loose b-tagged jet. The spectra are normalized to one.

4.3.1 B-tagging

The relatively long lifetime of B mesons, coming from the production of a b-quark in the event, means that these mesons travel around 500 μm on average before decaying into more stable light-flavor hadrons. This property is exploited by looking for tracks consistent with a secondary vertex from the B-meson decay and allows for further discrimination between signal and background events. The Combined Secondary Vertex (CSV) b-tagger [82] combines information from track impact parameters and secondary vertices within a given jet, and provides a continuous output which serves as discrimination between jets coming from the hadronization of a b-quark against light-flavor and gluon jets. The b-jet tagging efficiency and the rate of misidentification depend on the threshold of the CSV output and are measured in data samples enriched in b-jets, for example in $t\bar{t}$ events. The simulation samples are corrected by applying event weights to account for the differences between data and simulation with respect

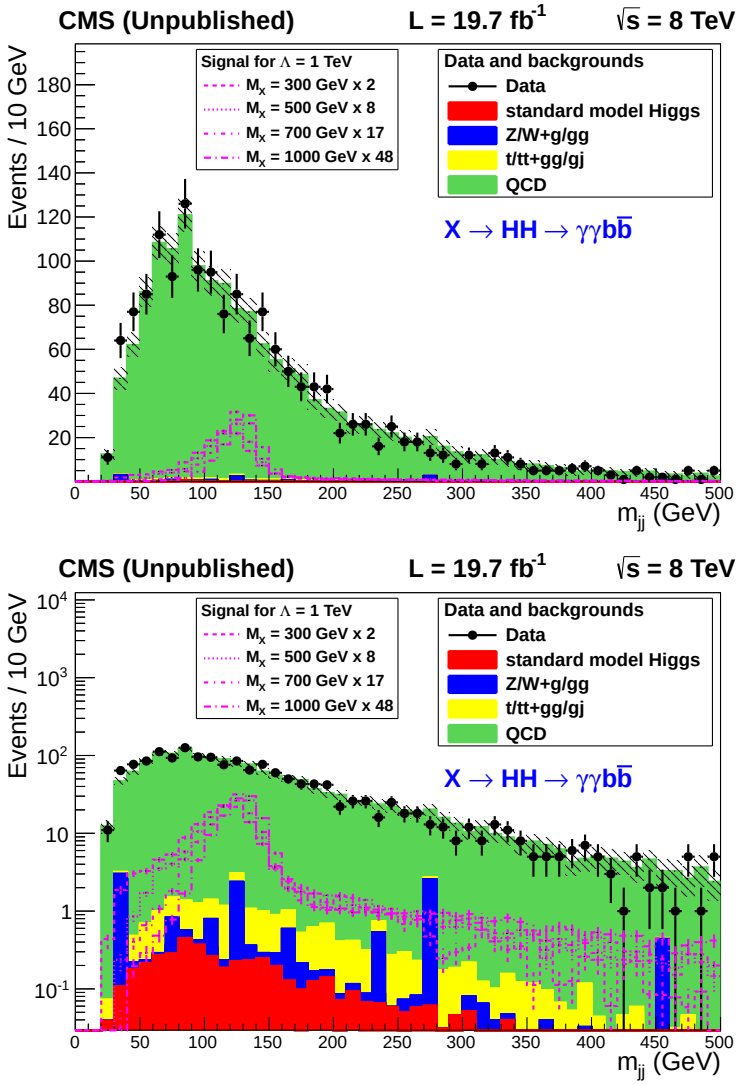

Figure 4.5: Control plots for the dijet mass spectrum after basic photon and jet selections and requiring at least one loose b-tagged jet. The simulation is normalized to data and the statistical uncertainty on the number of simulated events is shown in dashed overlay. The top (bottom) figure is shown in linear (log) scale.

to the b-tag efficiency. At preselection level, one loose b-tagged jet (CSV loose, or CSVL) is required in the event, which corresponds to a mistag rate of 10%.

4.3.2 Jet Energy Regression

The jet energy corrections are applied to all types of jets, whether originating from a gluon, a light-flavor quark, or a b-quark, whereas the jets in the signal of interest originate solely from the hadronization of b-quarks. Therefore, the correction can be improved by exploiting the properties unique to b-jets, which will in turn improve the resolution of the dijet mass. A motivation for implementing an additional correction is to improve separation between signal and background, particularly against the resonance background $Z(b\bar{b})H(\gamma\gamma)$, since the mass of the Z is very close to that of the Higgs relative to the dijet mass resolution.

Here a jet energy regression is presented, which acts as a multidimensional calibration tuned to the specific jet properties of the signal. It is not used in the final analysis, but it can serve as a 10–20% improvement in sensitivity of the low-mass resonant signal for future iterations of the analysis.

The correction comes from training a Boosted Decision Tree (BDT) [83] regression simultaneously on half of the MSSM signal samples at resonant masses of 260, 300, and 350 GeV; these samples were previously discussed in Section 3.3.1. These samples are generated with about 50k events each and are independent of the Radion samples on which the final limits are extracted.

The regression is trained on every jet passing the identification criterion and the following kinematic criteria:

- $p_T > 20$ GeV,
- $|\eta| < 2.5$, and

- $\Delta R(j, j_{\text{gen}}) < 0.4$ (the angular separation between the jet and its associated generator-level jet).

The p_T threshold is increased to 25 GeV for testing and implementation.

The BDT algorithm is implemented from the TMVA package [84] using gradient boost [85, 86]. The input variables are

- the jet transverse momentum p_T,
- the jet transverse mass m_T,
- the jet pseudorapidity η,
- the jet PF photon energy fraction,
- the jet PF neutral hadron energy fraction,
- the number of PF jet constituents (both charged and neutral),
- the lead track p_T associated with the jet,
- the jet secondary vertex flight distance error (if there is a secondary vertex),
- the jet secondary vertex mass (if there is a secondary vertex),
- the soft lepton p_T (if there is a soft lepton in the jet),
- the soft lepton relative p_T in direction of the jet (if there is a soft lepton in the jet),
- PF $\not\!\!E_{\text{T}}$ with $H \to \gamma\gamma$ specific corrections [87],
- $\Delta\phi(j, \not\!\!E_{\text{T}})$, and
- the median jet energy per jet area ρ.

The secondary vertex refers to that of the B-meson decay, if such a vertex was identified. The soft lepton refers to an electron or muon resonstructed inside the cone of the jet. The target of the training is the generator-level jet p_{T}, including energy from any neutrinos in the jet.

The figure of merit to quantify the effect of the regression is the resolution improvements measured in the dijet and four-body ($\gamma\gamma jj$) mass spectra of the signal sample. Possible overtraining, in which a model has poor predictive performance due to its tendency to treat noise in the training dataset as part of the relationship between the inputs and the output, has been studied and considered negligible. The resolution improvements in the m_{jj} and $m_{\gamma\gamma jj}$ spectra are estimated by fitting each spectrum with the sum of a Crystal Ball and third-order polynomial for events in both the medium and high-purity categories, which will be discussed in Section 5.1. The parameters of the Crystal Ball give estimates of the spread and central value of the distribution, and the ratio of these two gives the resolution. Table 4.1 shows in improvement in the resolution separately for the two spectra and for the two event categories. The improvement is shown visually in the m_{jj} spectrum in Figure 4.6 where both event categories are combined.

Table 4.1: Improvement from the regression on $m_{\gamma\gamma jj}$ and m_{jj} spectra, divided into medium or high-purity categories. (Categorization is discussed in Section 5.1.) All numbers are in units of percentage.

	$m_{\gamma\gamma jj}$ spectrum		m_{jj} spectrum	
m_X (GeV)	medium purity	high purity	medium purity	high purity
270	19.72	3.10	15.08	12.24
300	16.64	8.70	16.05	14.19
350	19.76	13.62	23.07	18.95
400	19.82	21.23	17.03	14.74

For validation of the technique, the effect of the regression is studied on data. Figure 4.7 shows the effect on the data control sample (a γjjj sample reweighted to

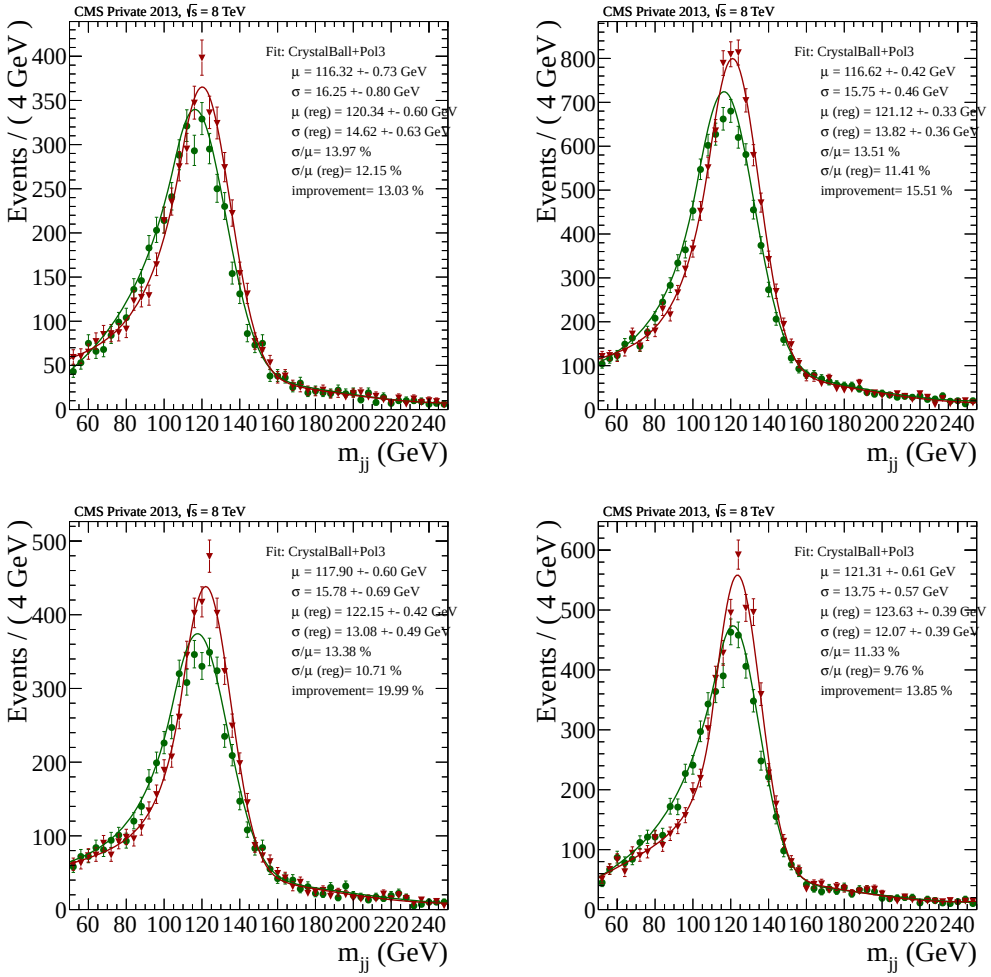

Figure 4.6: Resolution improvements for the m_{jj} spectrum at $m_X = 270, 300, 350, 400$ GeV mass points. The green distribution is before applying the regression, and the red one is after applying the regression. The fit model is the sum of a Crystal Ball and third-order polynomial.

the $\gamma\gamma jj$ data, discussed in Section 5.4.1). The peak is slightly shifted, and in the region about the Higgs mass, the yield increases about 10% without any local peaking structure. In addition, comparison between data and the sum of MC backgrounds was performed for the p_{T} balance between the dijet and diphoton candidates before and after regression. As shown in Figure 4.8, the ratio $\frac{p_{\mathrm{T},jj}}{p_{\mathrm{T},\gamma\gamma}}$ has a more narrow peak shifted closer to 1 after the regression is applied. Although the effect indicates the regression is doing its job, the data and background processes exhibit a slight overcorrection.

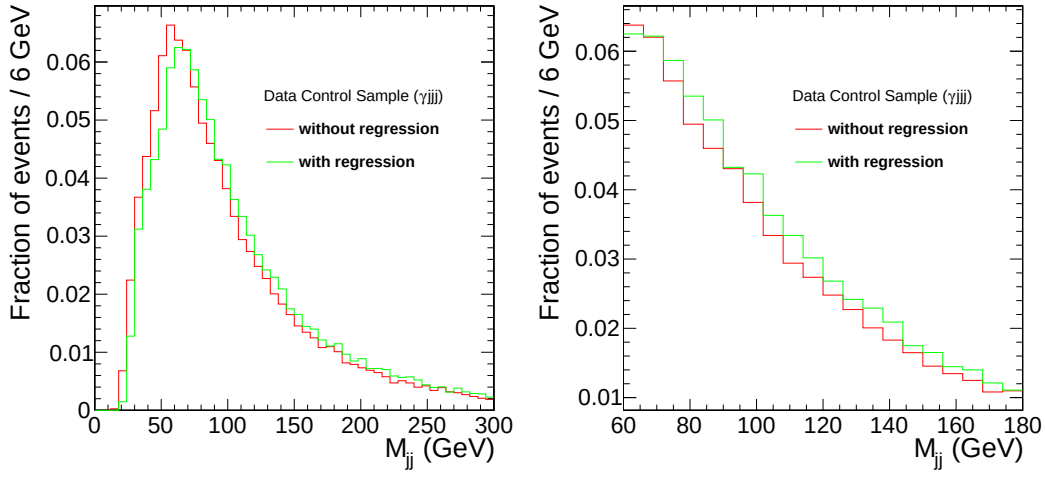

Figure 4.7: Effect of regression on the entire m_{jj} spectrum (left) and zoomed (right) to a range about the SM Higgs mass on the data control sample, discussed in Section 5.4.1.

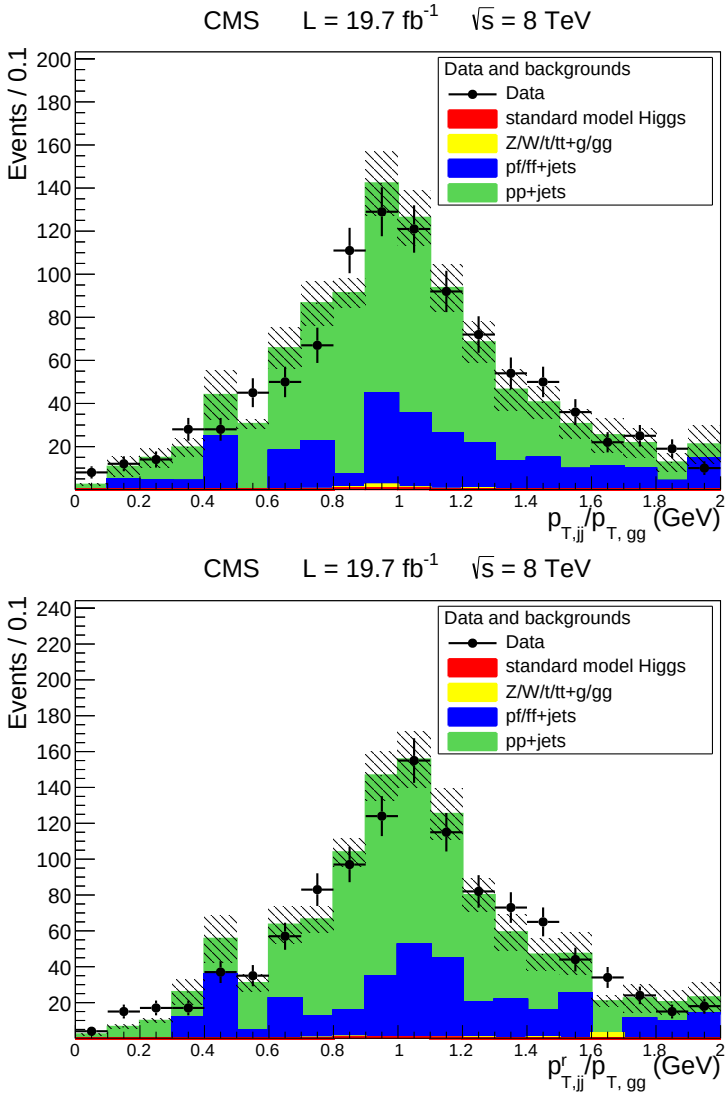

Figure 4.8: Improvement in the p_{T} balance before (top) and after (bottom) the regression is applied for a comparison between data and the sum of background processes. The events are a combination of the medium and high-purity categories with an additional requirement that there be no jets other than the two from the $H \to b\bar{b}$ decay.

Chapter 5

Event Selection

This chapter discusses additional event-wide requirements imposed to further improve the sensitivity of selecting the signal of interest over the background contributions beyond that of the preselection requirements covered in Chapter 4. Event classification is covered in Section 5.1, in which two classes of events are defined for extracting signal separately. Section 5.2 discusses how the photons and jets passing the preselection requirements are assembled into Higgs candidates. For the resonant search $X \to HH$, Section 5.3 discusses how the two Higgs come together to construct a resonant candidate. The construction of a data control sample, discussed in Section 5.4.1, allows for the study of additional event-wide requirements for improving sensitivity, discussed in Section 5.4. Finally, with all the event requirements in place, two event displays in Section 5.5 give a sense of what candidate events look like in the detector.

5.1 Event Classification

B-tagging, presented in Section 4.3.1, is used at the preselection level by requiring one CSVL b-tagged jet. To further improve the sensitivity, the working point used for signal extraction is the medium one (CSVM), which corresponds to an efficiency of around 60–70% and a mistag rate of 1–2%, depending on the jet p_{T}. There is

also a tight working point (CSVT) not used in this analysis, which corresponds to a mistag rate of 0.1%. With the CSVM working point, events are classified into three categories for the resonant search based on the number of CSVM b-tagged jets. Events with two or more b-tags, called high purity, drive the sensitivity of the search. Events with one b-tag, called medium purity, bring little to the sensitivity but allow for increased signal acceptance. Events without any b-tags, called low purity, are used only for cross-checks and are rejected from the main analysis. The preselection cuts and event categorization are summarized in Table 5.1. For the nonresonant search, the high and medium purity categories will be further divided based on the four-body mass spectrum, discussed in Section 5.4.2

Table 5.1: Summary of the selection applied to photons and jets and the event classification for the resonant search.

Photons	Jets	Classification
tight photon identification	loose jet identification	
$p_{T\gamma_1}/m_{\gamma\gamma} > 1/3$	pileup rejection	
$p_{T\gamma_2}/m_{\gamma\gamma} > 1/4$	$p_{Tj} > 25$ GeV	
$\lvert\eta_\gamma\rvert < 2.5$	$\lvert\eta_j\rvert < 2.5$	≥ 2 CSVM b-tags
$100 < m_{\gamma\gamma} < 180$ GeV	≥ 1 CSVL b-tag	exactly 1 CSVM b-tag

5.2 Higgs Reconstruction

From the lists of photons and jets passing the identification and kinematic requirements, Higgs candidates are constructed. When there are more than two photon or jet candidates, a choice must be made as to which pair is consistent with the decay of a Higgs. When choosing the diphoton candidate, the two photons with the highest p_{T} are chosen, since jets faking photons tend to be softer than prompt photons.

When choosing the jets for the dijet candidate, four choices were considered after b-tagging:

- the two jets with the highest p_{T},
- the pair of jets that maximizes $p_{\mathrm{T},jj}$,
- the pair of jets that maximizes $\frac{p_{\mathrm{T},jj}}{m_{jj}}$, and
- the pair of jets that minimizes $|m_{jj} - m_{\gamma\gamma}|$.

In the high-purity category, the dijet candidate is selected from all b-tagged jets, while in the medium-purity category, the candidate is selected from the pairing of the b-tagged jet with one of the non-tagged jets. The jet pair that maximizes $p_{\mathrm{T},jj}$ is selected. This choice selects the correct jets roughly 87% of the time, without producing any local peaking structure in the background. For the resonant search in the high-purity (medium-purity) category, the resolution, defined as the full-width at half-maximum, decreases from 40 (50) GeV to 30 (30) GeV as the resonance mass increases from 300 GeV to 1 TeV. This is also shown visually in Figure 4.4.

5.3 Resonance Reconstruction and Kinematic Fit

For the resonant search, the double Higgs system is constructed by combining the $H \to \gamma\gamma$ and $H \to b\bar{b}$ candidates. (This is also used in the nonresonant search where the resulting mass distribution does not correspond to a local peak.) In order to improve the resolution on the $m_{\gamma\gamma jj}$ distribution, an additional constraint is applied on the two selected jets. This constraint, called the kinematic fit [88], is performed by varying the four-momentum of each jet with respect to its energy and position resolutions in order to fix the dijet mass to the Higgs mass, i.e., $m_{jj} = 125$ GeV. The resulting distribution is denoted as $m^{\mathrm{kin}}_{\gamma\gamma jj}$. Distributions before ($m_{\gamma\gamma jj}$) and after ($m^{\mathrm{kin}}_{\gamma\gamma jj}$) the kinematic fit are shown in Figure 5.1 for the resonant signal and resonant backgrounds and in Figure 5.2 for data and the sum of backgrounds.

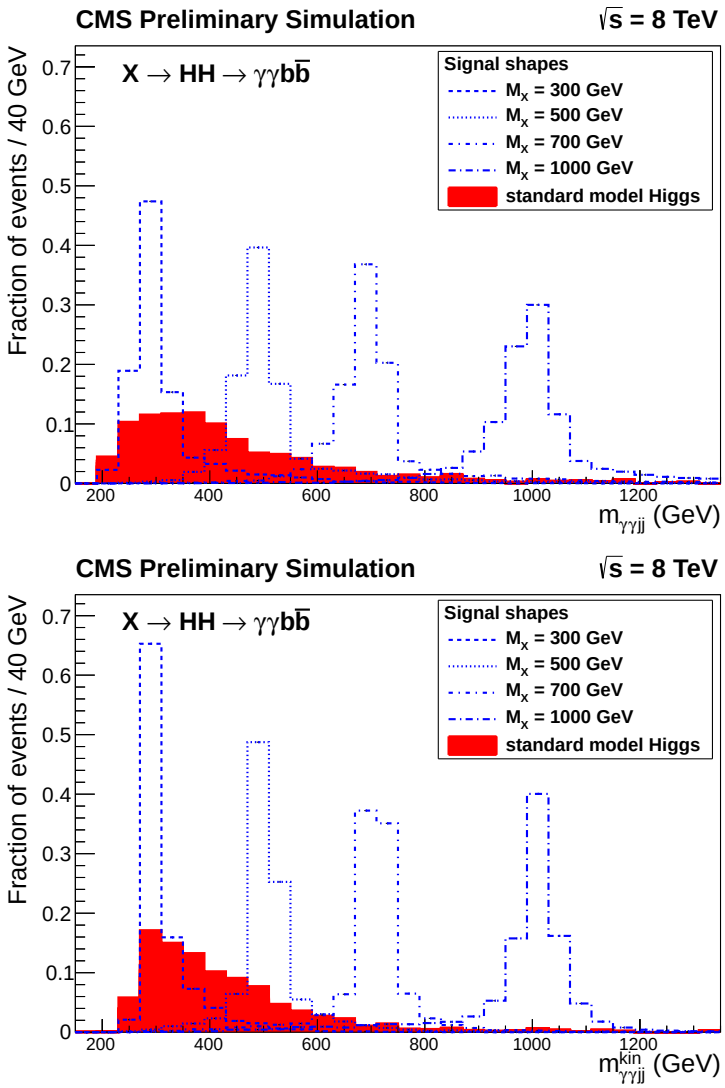

Figure 5.1: Simulated four-body mass spectrum for the resonant signal and the sum of all production mechanisms of the SM Higgs boson after basic selections on photons and jets and requesting at least one loose b-tagged jet. Before (after) the kinematic fit is shown in the top (bottom) figure. The spectra are normalized to one.

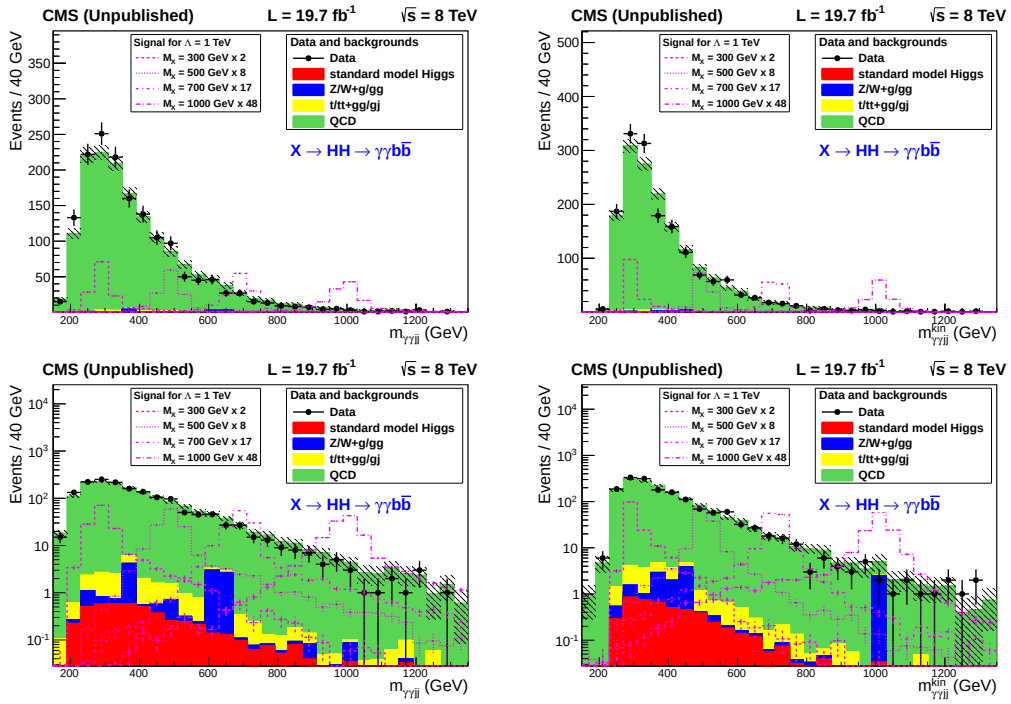

Figure 5.2: Control plots for the four-body mass spectrum after basic photon and jet selections and requiring at least one loose b-tagged jet. The simulation is normalized to data and the statistical uncertainty on the number of simulated events is shown in dashed overlay. The left (right) figures are before (after) the kinematic fit is imposed, and the top (bottom) figures are shown in linear (log) scale.

5.4 Further Optimization

The strategy for extracting a measurement of the signal yield centers on the existence of three spectra, $m_{\gamma\gamma}$, m_{jj}, and $m^{\text{kin}}_{\gamma\gamma jj}$, from which an excess can be measured. For the resonant search, an excess would appear as a localized bump on the $m^{\text{kin}}_{\gamma\gamma jj}$ spectrum giving a direct handle on m_X, but, recalling from Figure 5.2, there is a kinematic peak in the background around 300 GeV. In order to avoid looking for a peak on top of a peak, the resonant search is divided into two regimes:

- the high-mass analysis searches for an excess on the $m^{\text{kin}}_{\gamma\gamma jj}$ spectrum for $m_X \geq$ 400 GeV, and

- the low-mass analysis searches for an excess on the $m_{\gamma\gamma}$ spectrum for $m_X \in [260, 400]$ GeV.

In both cases, there is more discrimination to be gained by imposing requirements on the two spectra not being fit, as is described in Section 5.4.2.

The nonresonant search can be viewed as a special case of the low-mass resonant search in which the $m^{\text{kin}}_{\gamma\gamma jj}$ spectrum has no local peak. Here, an excess is sought on the $m_{\gamma\gamma} \times m_{jj}$ plane, with additional requirements imposed on $m_{\gamma\gamma jj}$ and the angle between the two Higgs candidates $|\cos\theta^{\text{CS}}_{HH}|$.

5.4.1 Data Control Sample

For optimizing the kinematic cuts imposed after preselection, a data control sample is constructed. The events for the sample are selected with the same preselection as data except for the inversion of one of the photon's identification and isolation requirements, giving a sample dominated by the contribution of γjjj rather than the $\gamma\gamma jj$ of the signal region.

The main kinematic variables related to the photons have a different shape between the control sample and the signal region, so reweighting of the control sample is needed to ensure that the distributions match those in the signal region. The event weights are derived by matching the shape of the control sample to data in the $p_{\text{T},\gamma_1} \times p_{\text{T},\gamma_1}$ plane, with a veto in signal region on the range $m_{\gamma\gamma} \in [115, 135]$ GeV. Afterwards, the normalization of the control sample is fixed to the normalization of the signal region. Figure 5.3 shows the effect of the reweighting on the three spectra of interest.

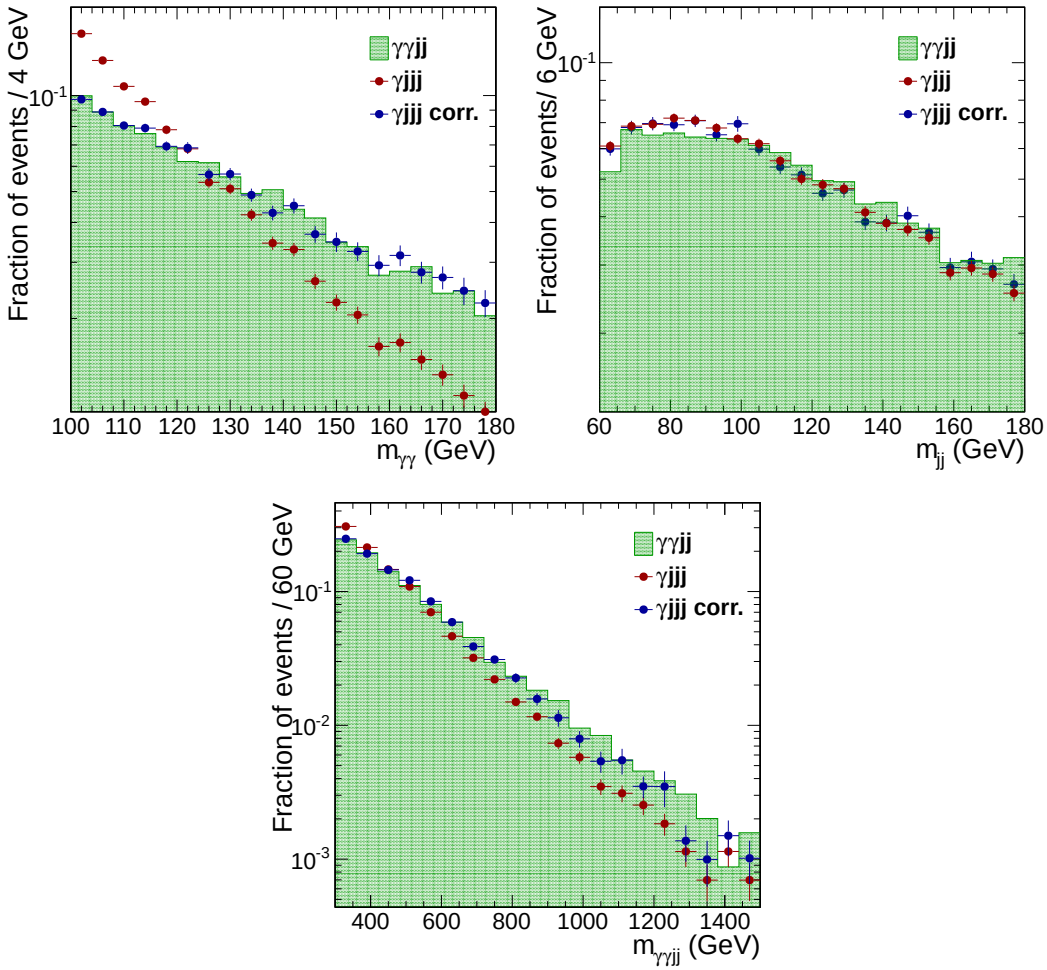

Figure 5.3: Effect of reweighting the γjjj control sample to match the kinematic distributions of the signal region. Double γ refers to the signal region, with a veto in $m_{\gamma\gamma}$ or m_{jj} on a range about the Higgs mass. Single γ refers to the γjjj sample before the weighting, and Single γ corr. refers to the reweighted γjjj sample. The shape of the reweighted γjjj sample is reasonably consistent with that of data for all three spectra.

5.4.2 Additional Discrimination

When searching for a high-mass resonance through the $m^{\text{kin}}_{\gamma\gamma jj}$ spectrum, the spectra not under consideration, namely the $m_{\gamma\gamma}$ and m_{jj} spectra, add discrimination as the double Higgs signal produces resonances on both corresponding to the decays $H \to \gamma\gamma$ and $H \to b\bar{b}$. In both the medium and high-purity categories, the requirements $m_{\gamma\gamma} \in [120, 130]$ GeV and $m_{jj} \in [90, 165]$ GeV are imposed. These ranges were found through optimization of the expected sensitivity with the data control sample.

In the search for a low-mass resonance through the $m_{\gamma\gamma}$ spectrum, the m_{jj} and $m_{\gamma\gamma jj}$ spectra provide additional discrimination through a similar optimization. The requirement imposed on the m_{jj} spectrum is the same for all resonance hypotheses, $m_{jj} \in [85, 155]$ GeV, while the requirement on the $m_{\gamma\gamma jj}$ spectrum varies with respect to the resonance mass hypothesis under consideration. Table 5.2 shows how the $m_{\gamma\gamma jj}$ mass window shifts with resonance mass hypothesis. In general the requirement is a small window about the m_X under consideration.

Table 5.2: $m_{\gamma\gamma jj}$ and m_{jj} requirements imposed in addition to the preselection in order to extract the signal on the range $m_X \in [260, 400]$ GeV.

m_X (GeV)	$m_{\gamma\gamma jj}$ selection (GeV)
260	$[225, 280]$
270	$[225, 295]$
300	$[255, 330]$
350	$[310, 395]$
400	$[370, 440]$
	m_{jj} selection (GeV)
all m_X	$[85, 155]$

For the SM nonresonant signal, excess is sought on the $m_{\gamma\gamma} \times m_{jj}$ plane with a loose selection on the dijet mass of $m_{jj} \in [60, 180]$ GeV. Additional requirements are imposed on $m^{\text{kin}}_{\gamma\gamma jj}$ and $|\cos\theta^{\text{CS}}_{HH}|$. For the former, the signal exhibits no narrow

resonance but rather a fairly wide peaking structure about 400 GeV, as shown in Figure 5.4. Note that the shape of this distribution varies widely among nonresonant scenarios based on the amount of interference between the box and tree diagrams in Figure 1.9. The optimal requirement for the SM nonresonant search is $m^{\text{kin}}_{\gamma\gamma jj} > 350$ GeV for both categories. To allow consideration of a range of scenarios, this requirement is implemented as a border between event categories, where both the high and medium purity categories are themselves divided into high-$m^{\text{kin}}_{\gamma\gamma jj}$ and low-$m^{\text{kin}}_{\gamma\gamma jj}$ subcategories. The optimal range on $|\cos\theta^{\text{CS}}_{HH}|$ is found to be category dependent, where the selection imposed is $|\cos\theta^{\text{CS}}_{HH}| < 0.90$ ($|\cos\theta^{\text{CS}}_{HH}| < 0.65$) in the high-purity (medium-purity) category. Table 5.3 summarizes these requirements for the nonresonant search.

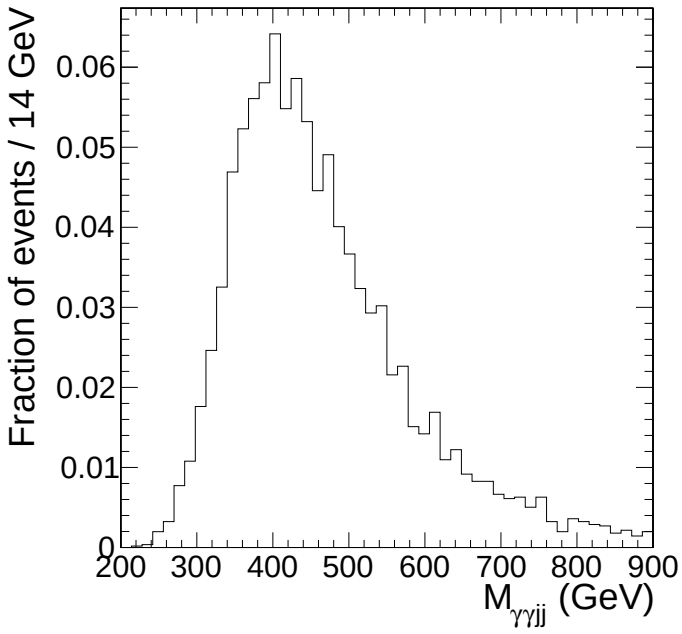

Figure 5.4: Simulated distribution in the $m_{\gamma\gamma jj}$ spectrum for the SM nonresonant signal.

Table 5.3: Additional discrimination for the nonresonant search beyond the preselection requirements in Table 5.1. Four event categories are formed from the number of b-tags and value of $m^{\text{kin}}_{\gamma\gamma jj}$.

b-tag category	High purity		Medium purity	
$m^{\text{kin}}_{\gamma\gamma jj}$ category (GeV)	< 350	> 350	< 350	> 350
$\lvert\cos\theta^{\text{CS}}_{HH}\rvert$	< 0.9		< 0.65	

5.5 Event Displays

Figures 5.5 and 5.6 show event displays of two events passing the identification and selection requirements for the resonant search at 300 GeV and the high-mass resonant search, respectively. These displays show the reconstructed tracks and calorimeter deposits. As the photon showers are much narrower, one ECAL cell contains more energy than the maximum HCAL unit corresponding to a jet, whose showers tend to be more uniform inside the jet cone. The transverse plane is shown to highlight the secondary vertices corresponding to the decay of a B meson inside a b-jet.

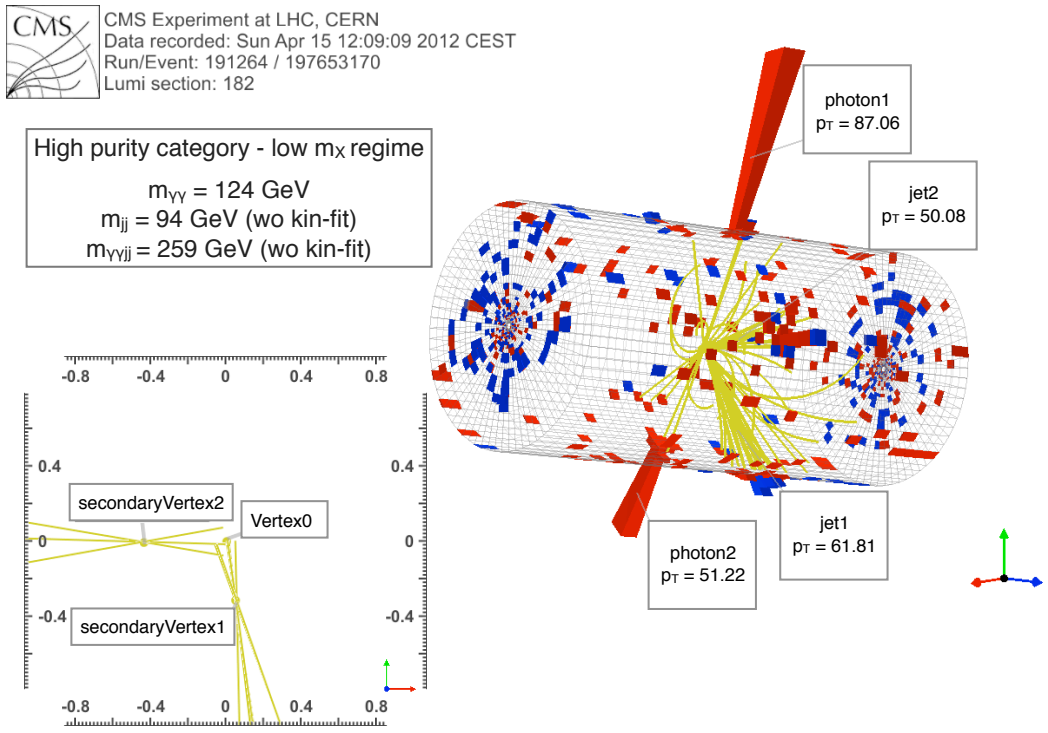

Figure 5.5: Event display for an event selected in the low-mass resonant search at 300 GeV.

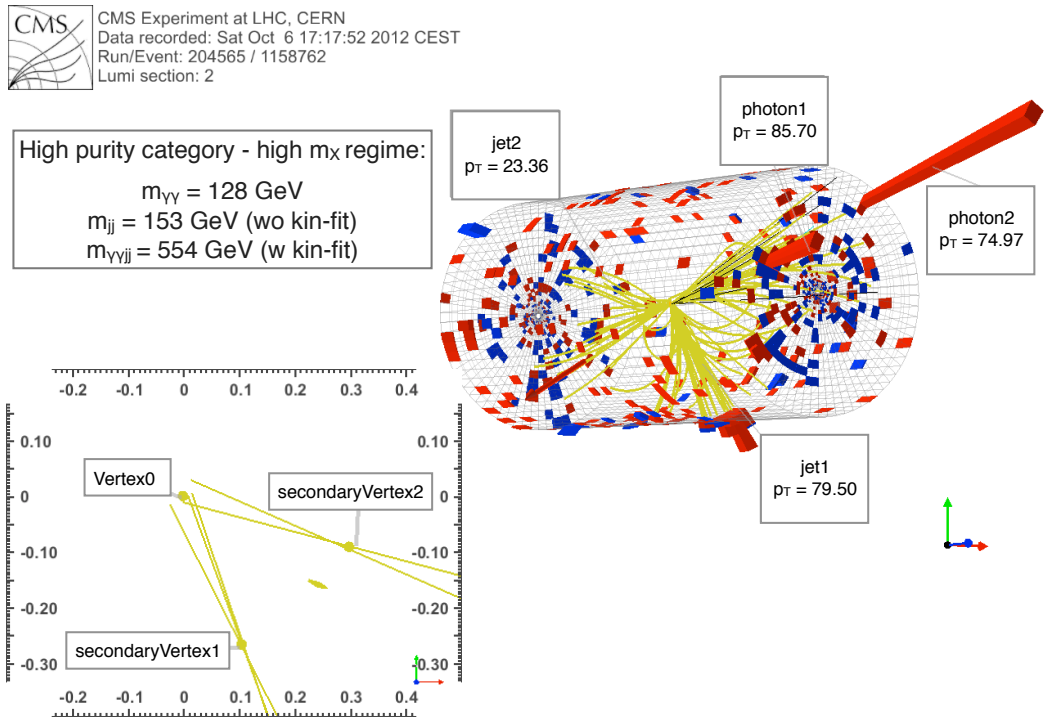

Figure 5.6: Event display for an event selected in the high-mass resonant search.

Chapter 6

Signal Extraction

After applying the preselection, event classification, and mass windows, the signal efficiency and event yields are examined, as discussed in Section 6.1. Then fits are performed on simulated signal spectra to estimate the expected signal shape and to the data for estimating the expected background shape. The fits are discussed in Sections 6.2 and 6.3 for the resonant and nonresonant searches, respectively.

6.1 Signal Efficiencies and Yields

The signal efficiency as a function of m_X for the resonant search is summarized in Figure 6.1. The signal efficiency increases from 260 GeV to 900 GeV due to improved photon and jet reconstruction. The signal efficiency peaks and then drops at 900 GeV due to the merging of the jets from the decay $H \to b\bar{b}$ into a single jet. For future consideration in extending the search above 1.1 TeV, jet substructure techniques would be necessary to resolve the merging of the two jets [89]. Both categories contribute approximately equally to the overall efficiency.

The relative yields for the nonresonant backgrounds in the low-mass resonant search at $m_X = 300$ GeV are summarized in Table 6.1. There is a normalization disagreement in the $\gamma\gamma j$ and γj contributions as the simulation has limitations in

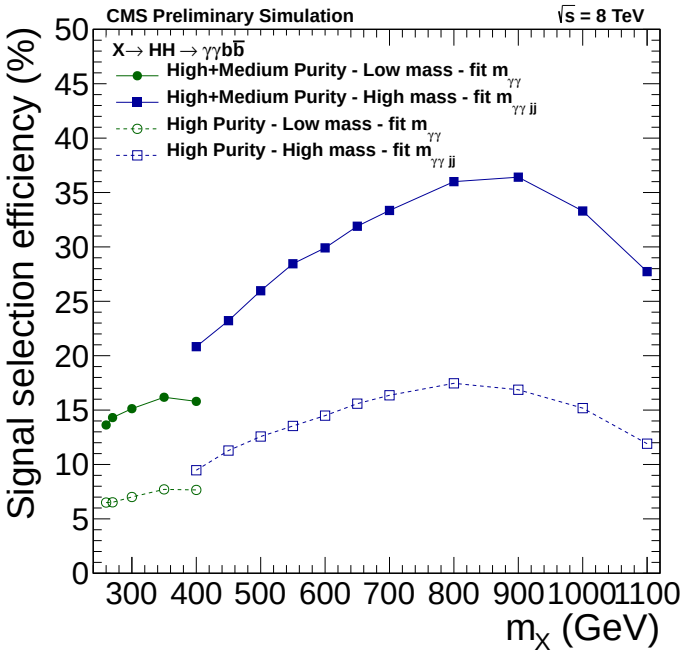

Figure 6.1: Signal efficiency for the resonant search for the final selection.

modeling QCD with one or two hard photons. Specifically, the total MC estimate is a factor of two smaller than the yield observed in data. As a result, the relative yields as a percentage of the total estimated nonresonant background are given. The relavitve yields for the nonresonant backgrounds in the high-mass resonant search are summarized in Table 6.2. For this search, the requirements are independent of the mass hypothesis, and again the relative yields as a percentage of the total estimated nonresonant background are given.

The relavive yields for the nonresonant backgrounds in the nonresonant search are summarized in Table 6.3. The yields are greater in the nonresonant search because the $m^{\text{kin}}_{\gamma\gamma jj}$ spectrum is less discriminating than in the low-mass resonant search and because the m_{jj} spectrum is fit on the range $m_{jj} \in [60, 180]$ GeV rather than selected on a narrower range. The yields for the SM nonresonant signal and resonant backgrounds as well as the counts for data are shown in Table 6.4.

Table 6.1: Relative event yields for the nonresonant backgrounds in the low-mass resonant search at 300 GeV. Note that there is a normalization disagreement coming from the shortcomings of simulating QCD with one or two hard photons, so percentages are given instead of numbers of events.

Sample	High purity	Medium purity
$\gamma\gamma j$	95.7%	94.7%
γj	$< 0.1\%$	4.6%
QCD	$< 0.1\%$	$< 0.1\%$
$Z/\gamma^*(\ell^+\ell^-) + Z(\ell^+\ell^-)\gamma + W(\ell\nu)\gamma\gamma$	$< 0.1\%$	0.1%
$t\bar{t}\gamma\gamma + t\gamma\gamma + t\bar{t}\gamma j$	4.3%	0.6%

Table 6.2: Relative event yields for the nonresonant backgrounds in the high-mass resonant search. Note that there is a normalization disagreement coming from the shortcomings of simulating QCD with one or two hard photons, so percentages are given instead of numbers of events.

Sample	High purity	Medium purity
$\gamma\gamma j$	95.2%	95.1%
γj	$< 0.1\%$	4.1%
QCD	$< 0.1\%$	$< 0.1\%$
$Z/\gamma^*(\ell^+\ell^-) + Z(\ell^+\ell^-)\gamma + W(\ell\nu)\gamma\gamma$	$< 0.1\%$	0.1%
$t\bar{t}\gamma\gamma + t\gamma\gamma + t\bar{t}\gamma j$	4.8%	0.7%

Table 6.3: Event yields for the nonresonant search. Expectations are given for the SM nonresonant signal, resonant background, and nonresonant background. Counts are given for data. Note that there is a normalization disagreement coming from the shortcomings of simulating QCD with one or two hard photons.

	High Purity		Medium Purity	
Sample	high $m^{\text{kin}}_{\gamma\gamma jj}$	low $m^{\text{kin}}_{\gamma\gamma jj}$	high $m^{\text{kin}}_{\gamma\gamma jj}$	low $m^{\text{kin}}_{\gamma\gamma jj}$
$\gamma\gamma j$	90.9%	76.4%	81.8%	82.6%
γj	$< 0.1\%$	15.6%	15.2%	16.3%
QCD	$< 0.1\%$	$< 0.1\%$	$< 0.1\%$	$< 0.1\%$
$Z/\gamma^*(\ell^+\ell^-) + Z(\ell^+\ell^-)\gamma + W(\ell\nu)\gamma\gamma$	$< 0.1\%$	$< 0.1\%$	1.2%	0.1%
$t\bar{t}\gamma\gamma + t\gamma\gamma + t\bar{t}\gamma j$	9.1%	8.0%	1.8%	1.0%

Table 6.4: Event yields for the nonresonant search. Expectations are given for the SM nonresonant signal and resonant backgrounds. Counts are given for data.

Sample	High Purity		Medium Purity	
	high $m_{\gamma\gamma jj}^{\text{kin}}$	low $m_{\gamma\gamma jj}^{\text{kin}}$	high $m_{\gamma\gamma jj}^{\text{kin}}$	low $m_{\gamma\gamma jj}^{\text{kin}}$
SM nonresonant HH	2.03	0.28	1.99	0.20
ggF $H \to \gamma\gamma$	0.05	0.04	0.29	0.32
VBF $H \to \gamma\gamma$	0.01	0.01	0.05	0.05
$WH(\gamma\gamma)$	< 0.01	< 0.01	0.12	0.09
$ZH(\gamma\gamma)$	0.04	0.02	0.07	0.05
$t\bar{t}H(\gamma\gamma)$	0.16	0.17	0.30	0.17
$b\bar{b}H(\gamma\gamma)$	< 0.01	0.01	0.01	0.04
Data	41	136	37	319

6.2 Resonant Fits

6.2.1 Low-mass Resonant Fits

For the low-mass resonant search, the signal yield is extracted by fitting the $m_{\gamma\gamma}$ spectrum. The signal model is built for each mass hypothesis by fitting the $m_{\gamma\gamma}$ spectrum in the simulation sample separately for the two categories. The functional form used is the sum of a Crystal Ball and a Gaussian, constrained to have the same mean, where the former models the core of the distribution and the latter models the tails. The position of the peak and the spread are independent of the resonant mass and the category. Figure 6.2 shows an example of the signal fit for a mass hypothesis of 300 GeV.

The background estimation is performed by fitting the same distribution in data for each category on the interval $[100, 180]$ GeV. This procedure is completely data-driven, and as such it is important to verify that the choice of the function does not bias the estimate of the signal strength obtained from the fit to data with the

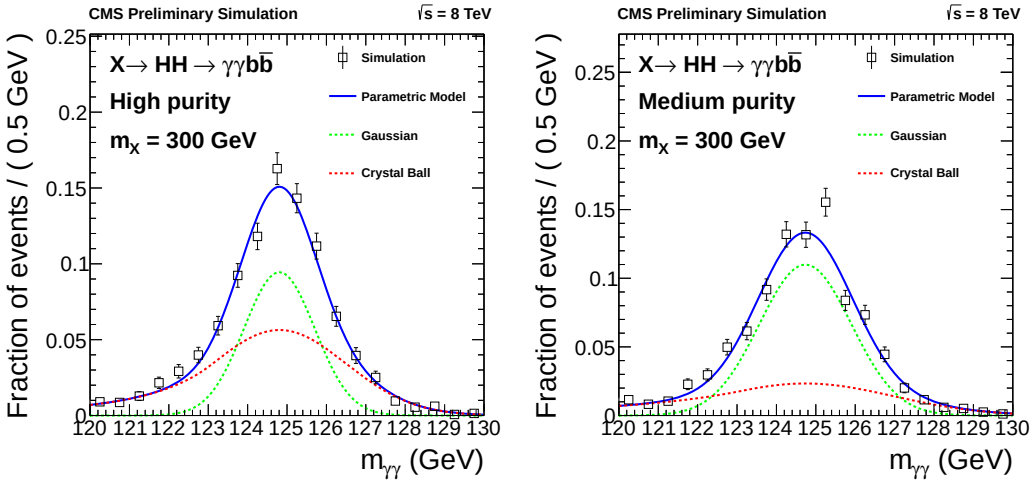

Figure 6.2: Simulated signal shape in the $m_{\gamma\gamma}$ spectrum for the high-purity (left) and medium-purity (right) categories for the Radion with mass 300 GeV. The open squares and corresponding statistical uncertainties represent the simulation. The blue line represents the signal model fitted to the simulation, while the green dashed line and the red dashed line represent the two components of the signal model.

sum of signal and background components. The bias is estimated by considering a set of truth models which approximately describe the background. For each truth model a large set of pseudo-data is generated and fitted by the sum of a candidate background model and the signal model. The bias is defined as the ratio of the extracted signal strength μ divided by the associated statistical uncertainty σ_μ and is considered negligible if

$$\left|\text{median}\left(\frac{\mu}{\sigma_\mu}\right)\right| < 14\%\,. \tag{6.1}$$

For both categories, more than one unbiased background candidate function is identified. For the background fit, a power law is chosen for both categories. Figures 6.3 and 6.4 shows the background fits to the data for four mass hypotheses.

The background fit estimates only the nonresonant contribution arising from diphoton production. The contribution from SM Higgs production with $H \to \gamma\gamma$ creates a resonance in the spectrum that mimics the signal process. Each resonant

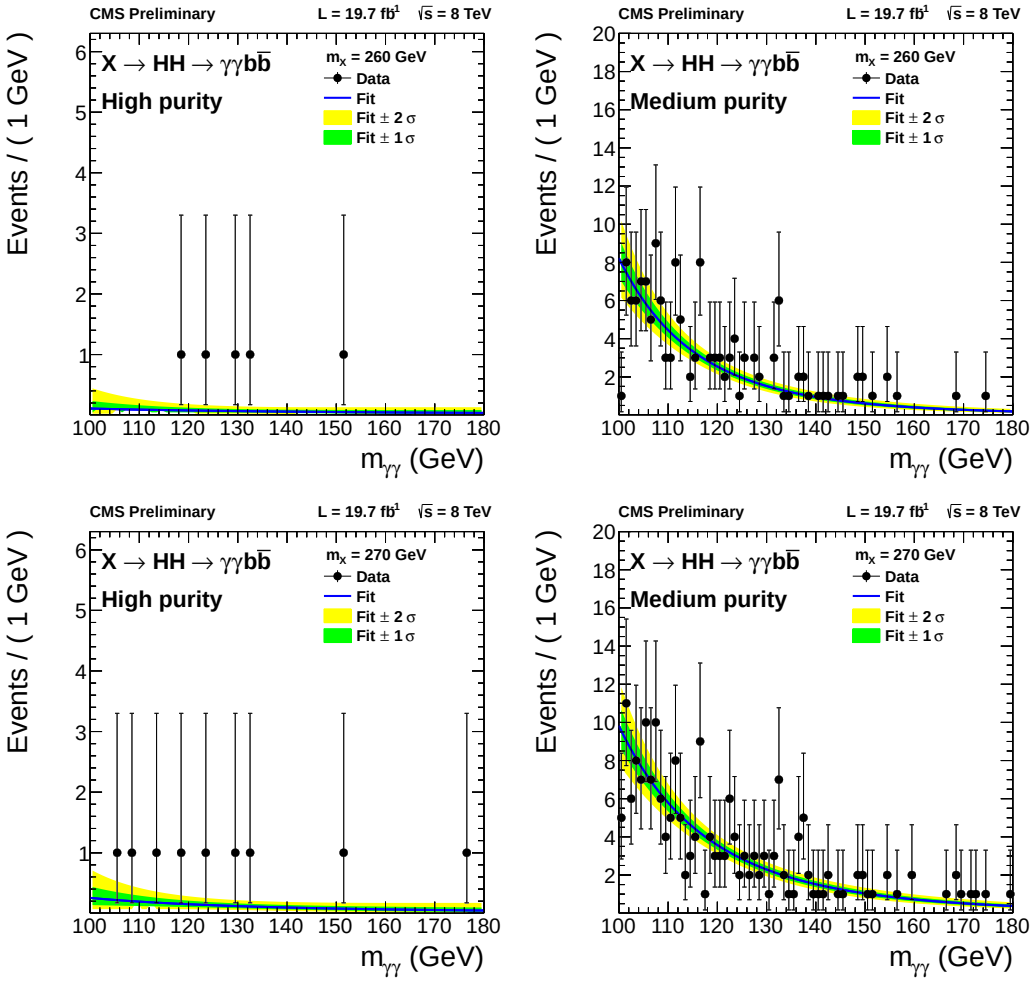

Figure 6.3: Events in the $m_{\gamma\gamma}$ spectrum in the high-purity (left column) and medium-purity (right column) categories for the resonance mass hypotheses 260 GeV (top row) and 270 GeV (bottom row). The nonresonant component of the background fit is shown in blue with its corresponding 1σ and 2σ confidence intervals.

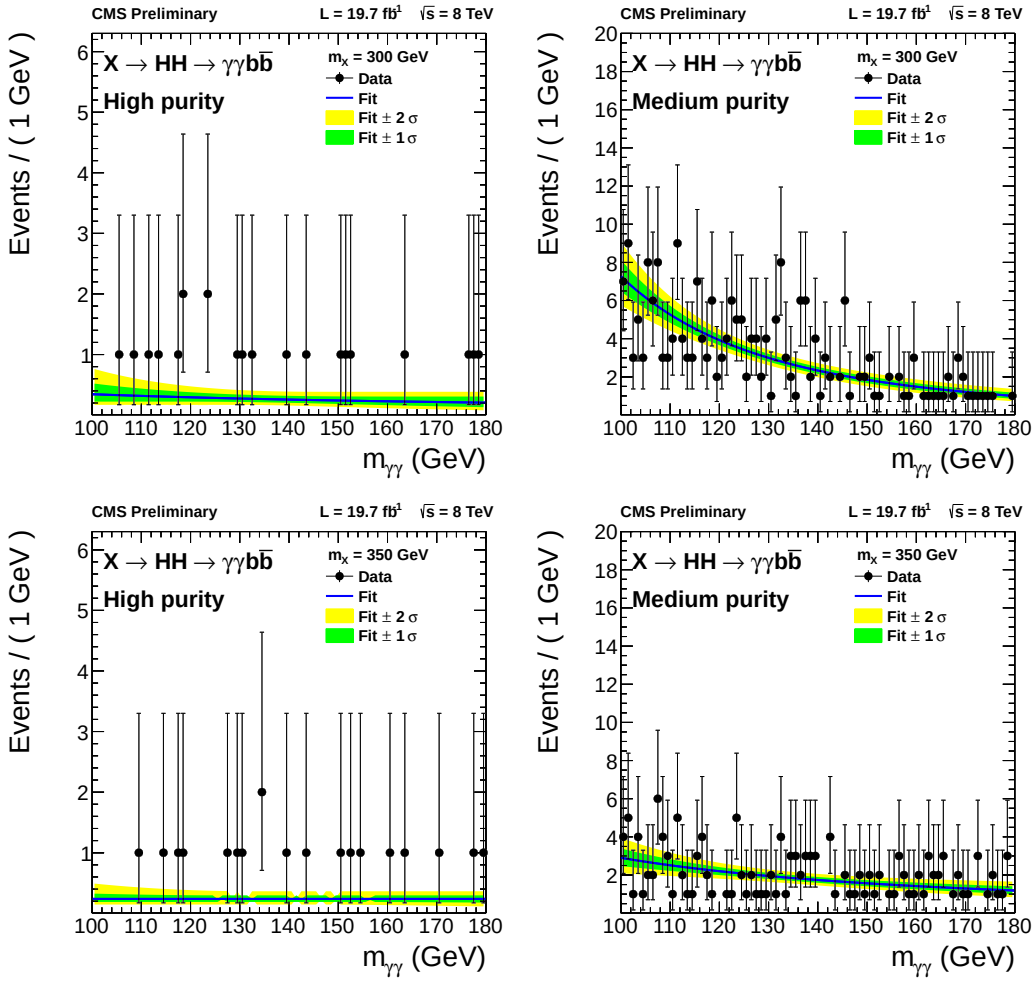

Figure 6.4: Events in the $m_{\gamma\gamma}$ spectrum in the high-purity (left column) and medium-purity (right column) categories for the resonance mass hypotheses 300 GeV (top row) and 350 GeV (bottom row) The nonresonant component of the background fit is shown in blue with its corresponding 1σ and 2σ confidence intervals.

background contribution is fit to simulation in the same way as the signal, and the normalization of each contribution fixed to its expected yield, which are summarized in Table 6.5. The expected yield for each contribution is small, and the effect on the expected sensitivity is found to be at most 2% for all low-mass resonant hypotheses.

Table 6.5: Expected yields for the resonant backgrounds of the low-mass resonant search at 300 GeV.

Sample	High purity	Medium purity
ggF $H \to \gamma\gamma$	0.02	0.19
VBF $H \to \gamma\gamma$	< 0.01	0.04
$WH(\gamma\gamma)$	< 0.01	0.05
$ZH(\gamma\gamma)$	< 0.01	0.03
$t\bar{t}H(\gamma\gamma)$	0.10	0.15

6.2.2 High-mass Resonant Fits

For the high-mass resonant search, the signal yield is extracted by fitting the $m^{\text{kin}}_{\gamma\gamma jj}$ spectrum in a procedure similar to the one for the low-mass resonant search. The signal model is built for each mass hypothesis by fitting the $m^{\text{kin}}_{\gamma\gamma jj}$ peak in the simulation sample separately for the two categories. The functional form used is the sum of a Crystal Ball and a Gaussian, with each constrained to have the same mean. The position of the peak follows the corresponding m_X hypothesis closely, and the resolution in the peak improves as m_X increases. Figure 6.5 shows examples of the signal fit for mass hypotheses of 500 GeV and 1 TeV.

The background estimation is done by fitting the same distribution in each category on the interval $[320, 1200]$ GeV. The lower edge is chosen to avoid the kinematic turn-on of the background while ensuring full containment of the 400 GeV signal. The same bias estimation procedure described for the low-mass resonant search is applied here. The chosen background function is a power law for both categories,

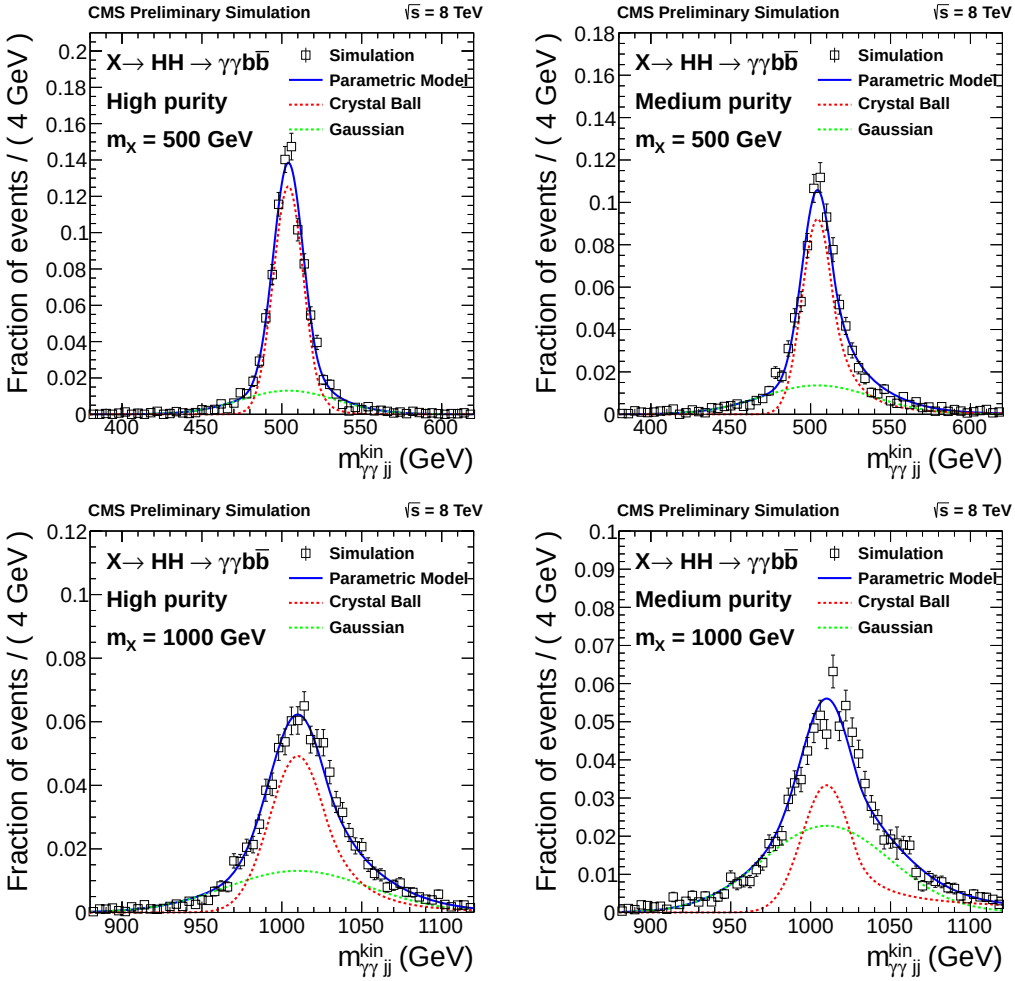

Figure 6.5: Simulated signal shape in the $m^{\text{kin}}_{\gamma\gamma jj}$ spectrum for the high-purity (left column) and medium-purity (right column) categories for the Radion with mass 500 GeV (top row) and 1 TeV (bottom row). The open squares and corresponding statistical uncertainties represent the simulation. The blue line represents the signal model fitted to the simulation, while the green dashed line and the red dashed line represent the two components of the signal model.

shown in Figure 6.6. Note that in this regime, the SM Higgs background does not have a resonance on the $m^{\text{kin}}_{\gamma\gamma jj}$ spectrum, so there is no resonant contamination from the background.

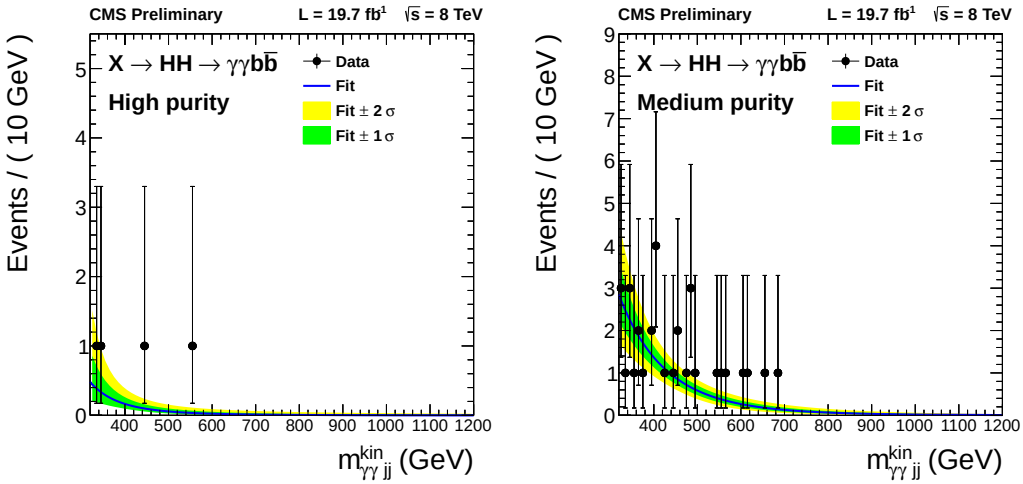

Figure 6.6: Events in the $m^{\text{kin}}_{\gamma\gamma jj}$ spectrum in the high-purity (left) and medium-purity (right) categories. The background fit is shown in blue with its corresponding 1σ and 2σ confidence intervals.

6.3 Nonresonant Fits

For the SM nonresonant search, the signal yield is extracted by fitting the $m_{\gamma\gamma} \times m_{jj}$ plane. The signal model is built by simultaneously fitting both dimensions for each of the four categories separately. The functional form used in both dimensions is the sum of a Crystal Ball and a Gaussian, with each constrained to have the same mean. The background estimation is done by fitting the same plane in each category on the interval $[100, 180]$ GeV for $m_{\gamma\gamma}$ and $[60, 180]$ GeV for m_{jj}. The same bias estimation procedure described for the low-mass resonant search is applied here. The chosen background function is a power law for both dimensions in all four categories. The SM Higgs production is treated as a resonant background, and each contribution is

fit in the same way as signal with the normalization of each contribution fixed to the expected yield, given in Table 6.4.

Chapter 7

Systematic Uncertainties

The expected signal yield is estimated through simulation with corrections for various effects. These corrections are related to the reconstruction and identification of photons and jets as well as and to the b-tagging of jets. The uncertainties associated with these corrections are applied to the reconstructed objects in the simulation through scaling and smearing the observables of interest. For the recorded luminosity, the normalization uncertainty is 2.6% [90]. In addition to the uncertainties related to photons and jets, covered in Sections 7.1 and 7.2, respectively, theory uncertainties are imposed on the resonant background, where the expected yield is also taken from simulation. These are covered in Section 7.3.

7.1 Photon Uncertainties

The photon-related uncertainties consist of those pertaining to the photon energy resolution (PER) and the photon energy scale (PES) [87]. As a function of the photon electromagnetic shower shape and η, an uncertainty between 0.23% and 0.93% is imposed on the PER, and an uncertainty between 0.12% and 0.88% on the PES. For hard photons, namely those with $p_{\mathrm{T}} > 100$ GeV, the uncertainty on PES is increased to 1%.

The photon preselection efficiency contributes a 1% normalization uncertainty to the $m_{\gamma\gamma}$ spectrum. The diphoton trigger efficiency contributes a 1% normalization uncertainty to all spectra. An additional normalization uncertainty of 5% is imposed in the high-mass resonant search to account for the differences in the p_{T} spectrum between photons of the signal and the electrons from $Z \to e^{+}e^{-}$ used to estimate the PES and PER and their corresponding uncertainties.

7.2 Jet Uncertainties

The jet energy scale (JES) uncertainty is found by varying the jet p_{T} by 1–2%, depending on the jet p_{T} and η [80]. The jet energy resolution (JER) uncertainty is found by varying the jet resolution by 10%. In the high-mass search, the jets tend to have a higher boost and be closer together, and effects related to their partial overlap are accounted for with an additional uncertainty of 1%. For the b-tagging efficiency uncertainty, the b-tagging scale factors are varied by one standard deviation in each category [82]. The uncertainty for the b-tagging efficiency between the two categories was found to have negative correlation.

7.3 Theory Uncertainties

No theory uncertainties are imposed for the resonant or nonresonant signal. For the resonant background, theory uncertainties are imposed on the SM Higgs contribution. These include contributions from missing order effects and the dependency on proton parton density functions [91, 15]. A systematic uncertainty is imposed on the Higgs mass for both for the resonant or nonresonant signal and for the resonant background. This uncertainty of 0.45 GeV is taken from the Higgs mass measurement performed at CMS in the $H \to ZZ \to 4\ell$ channel [92].

7.4 Summary and Impact on Analysis

The impact of the quoted systematic uncertainties is summarized in Table 7.1 for the low-mass resonant search and Table 7.2 for the high-mass resonant search. For the nonresonant search, uncertainties on the jets are associated with the m_{jj} shape (parametric shift and resolution) rather than acceptance. The analysis is statistically limited, and the systematic uncertainties worsen the expected sensitivity by at most 1.7% (3.8%) in the resonant (nonresonant) search.

Table 7.1: Systematic uncertainties for the low mass resonant search.

Normalization uncertainties	
Luminosity	2.6%
Diphoton trigger acceptance	1.0%
Photon selection acceptance	1.0%
b-tag efficiency, high (medium) purity	4.6% (1.2%)
m_{jj} and $p_{T,j}$ acceptance (JES & JER)	1.5%
$m_{\gamma\gamma jj}$ acceptance (PES $\oplus$ JES & PER $\oplus$ JER)	2.0%
Total, high (medium) purity	6.0% (4.1%)
Shape uncertainties	
Parametric scale shift (PES $\oplus$ m_H uncertainty)	$\frac{\Delta m_{\gamma\gamma}}{m_{\gamma\gamma}} = 0.45 \oplus 0.35\%$
Parametric resolution shift (PER)	$\frac{\Delta\sigma}{m_{\gamma\gamma}} = 0.25\%$
	$\frac{\Delta\sigma}{\sigma(m_{\gamma\gamma})} = 22\%$

Table 7.2: Systematic uncertainties for the high mass resonant search.

Normalization uncertainties	
Luminosity	2.6%
Diphoton trigger acceptance	1.0%
Photon selection acceptance	1.0%
b-tag efficiency, high (medium) purity	5.3% (1.8%)
m_{jj} and $p_{T,j}$ acceptance (JES & JER)	1.5%
$m_{\gamma\gamma}$ acceptance (PES & PER)	0.5%
Extra high p_{T}	5.0%
Total, high (medium) purity	8.0% (6.3%)

Shape uncertainties	
Parametric scale shift (PES $\oplus$ JES)	$\frac{\Delta m^{\mathrm{kin}}_{\gamma\gamma jj}}{m^{\mathrm{kin}}_{\gamma\gamma jj}} = 0.45 \oplus (0.8 \oplus 1.0) = 1.4\%$
Parametric resolution shift (PER $\oplus$ JER)	$\frac{\Delta\sigma}{\sigma(m^{\mathrm{kin}}_{\gamma\gamma jj})} = 10\%$

Chapter 8

Results

The signal yield is evaluated by fitting the data with a model that additively combines the signal and background shapes, where the normalization for each component is a free parameter. This is evaluated from a simultaneous fit to the $m_{\gamma\gamma}$, $m^{\text{kin}}_{\gamma\gamma jj}$, and $m_{\gamma\gamma} \times m_{jj}$ spectra for the low-mass resonant, high-mass resonant, and nonresonant searches, respectively. From the signal-plus-background fit, the confidence level (CL) for discovery or exclusion of double Higgs production is calculated. To compute the upper limits on the production cross section, the modified frequentist approach CL_s is used with an asymptotic approximation, taking the profile likelihood as a test statistic [93, 94]. This calculation is discussed in Section 8.1 for the resonant search and in Section 8.2 for the SM nonresonant search.

8.1 Resonant Results

8.1.1 Low-mass Resonant Results

In the low-mass resonant search, no excess above the expectation is observed, so upper limits on the signal cross section are calculated. The 95% CL for observed and expected upper limits is shown in Table 8.1 and Figure 8.1 for both categories

and for the high-purity category only. The latter result is provided to simplify the comparison with new physics models where the Higgs branching ratios for the $H \to \gamma\gamma$ and $H \to b\bar{b}$ decays can be modified with respect to their values in the SM. The green and yellow bands represent the 1σ and 2σ confidence intervals around the expected limit. Theory expectations for Radion, RS1 KK-graviton, and bulk KK-graviton are shown, where the Radion expectation assumes BR($R \to HH$) = 25% for all Radion masses above 300 GeV. Through comparison with the Graviton simulation, the search is verified to be spin-independent, so theory expectations for both spin-0 and spin-2 hypotheses may be overlaid together.

Table 8.1: Observed and median expected 95% CL upper limits for $m_X \leq 400$ GeV.

m_X (GeV)	Obs. limit (fb)	Exp. limit (fb)	Obs. limit (fb)	Exp. limit (fb)
			High-purity category only	
260	3.14	2.12	3.54	2.41
270	2.70	2.40	3.07	2.74
300	3.98	2.73	3.64	3.14
350	1.67	2.23	2.17	2.66
400	1.97	1.66	3.40	2.01

8.1.2 High-mass Resonant Results

In the high-mass resonant search, no excess above the expectation is observed, so upper limits on the signal cross section are calculated. The 95% CL for expected and observed limits is shown in Table 8.2 and Figure 8.2. The break at 400 GeV corresponds to the border between the two methods for signal extraction. As in the low-mass regime, the theory expectations for the Radion assumes BR($R \to HH$) = 25% for all Radion masses above 300 GeV. The result is again spin-independent, allowing for both spin-0 and spin-2 theory expectations to be overlaid together.

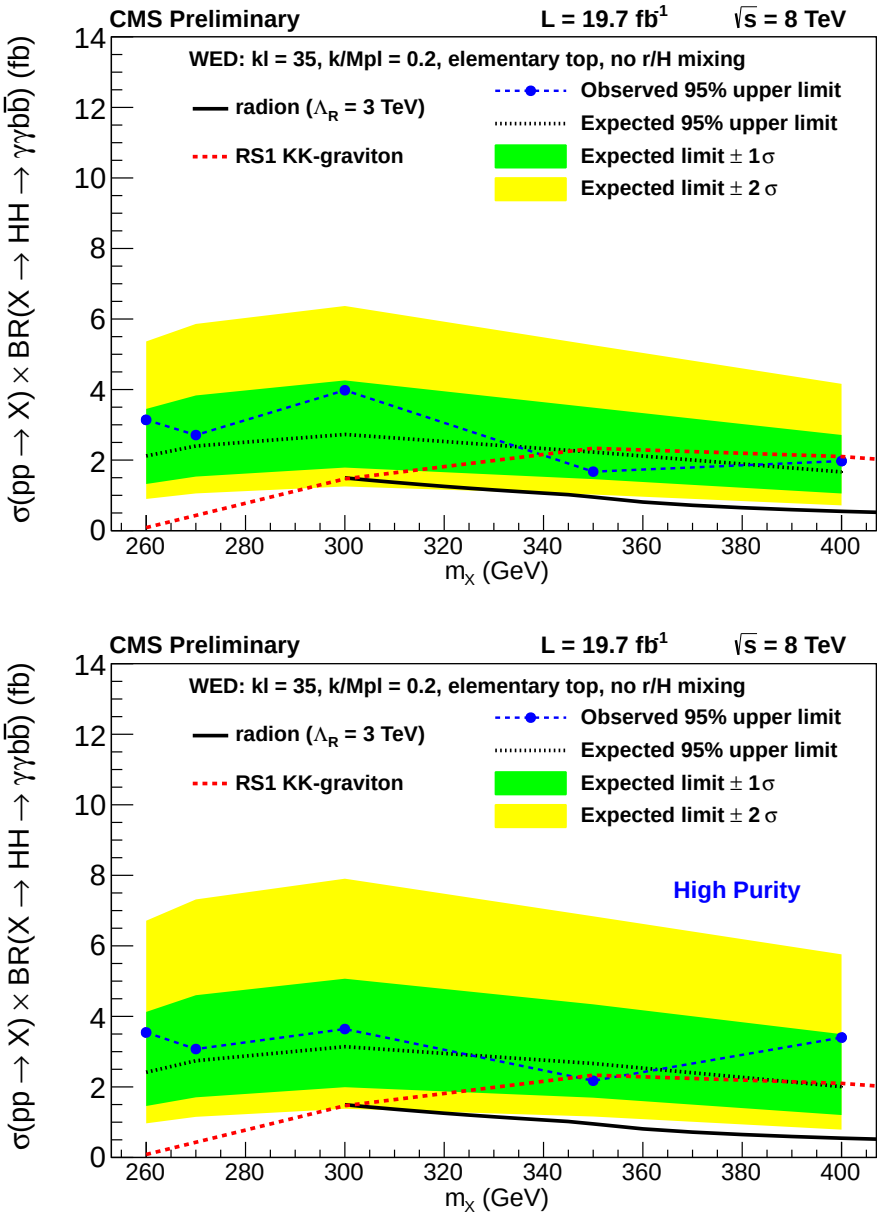

Figure 8.1: Expected 95% CL upper limits on the cross section times branching ratio $\sigma(pp \to X) \times \mathrm{BR}(X \to HH \to \gamma\gamma b\bar{b})$. Theory lines corresponding to WED models with Radion, RS1 KK-graviton, and bulk KK-graviton are overlaid. Limits from both categories (top) and high-purity category only (bottom) are shown. The results are obtained using the asymptotic CL_s approach.

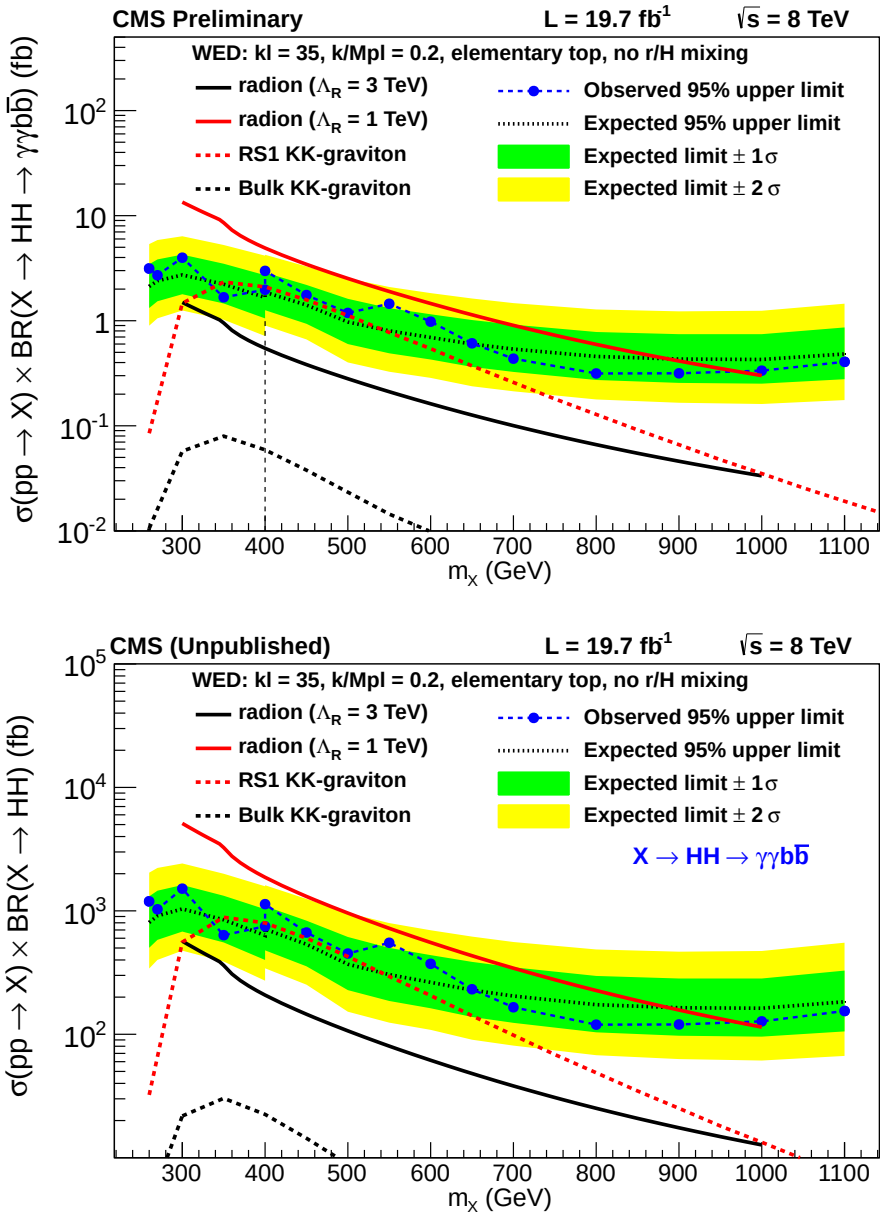

Figure 8.2: Expected 95% CL upper limits on the cross section times branching ratios $\sigma(pp \to X) \times \mathrm{BR}(X \to HH \to \gamma\gamma b\bar{b})$ (top) and $\sigma(pp \to X) \times \mathrm{BR}(X \to HH)$ (bottom). Theory lines corresponding to WED models with Radion, RS1 KK-graviton, and bulk KK-graviton are overlaid. The results are obtained using the asymptotic CL_s approach.

Table 8.2: Observed and median expected 95% CL upper limits for $m_X \geq 400$ GeV.

m_X (GeV)	Observed limit (fb)	Expected limit (fb)
400	2.98	1.87
450	1.76	1.42
500	1.19	0.97
550	1.45	0.80
600	0.98	0.69
650	0.61	0.60
700	0.44	0.54
800	0.31	0.46
900	0.32	0.43
1000	0.33	0.43
1100	0.41	0.48

8.1.3 Comparison of Resonant Results

Figure 8.3 provides a comparison of the observed and expected limits obtained among several final states in both the CMS and ATLAS Collaborations. The final states compared here are $\gamma\gamma b\bar{b}$, $b\bar{b}b\bar{b}$, $\tau\tau b\bar{b}$, and multileptons and photons. The comparison reveals that the $\gamma\gamma b\bar{b}$ is most sensitive to resonant double Higgs production for $m_X <$ 400 GeV, while the $b\bar{b}b\bar{b}$ final state is the most sensitive for $m_X > 400$ GeV.

8.2 Nonresonant Results

For the SM nonresonant search, no excess above the expectation is observed, so an upper limit on the signal cross section is calculated. The 95% CL is found using the same approach as in both regimes of the resonant search. The observed (expected) upper limit on the SM $pp \to HH \to \gamma\gamma b\bar{b}$ production cross section is 1.91 fb (1.59 fb). Assuming SM Higgs branching ratios, the observed (expected) upper limit on SM $pp \to HH$ production is 726 fb (604 fb). In terms of the SM signal strength

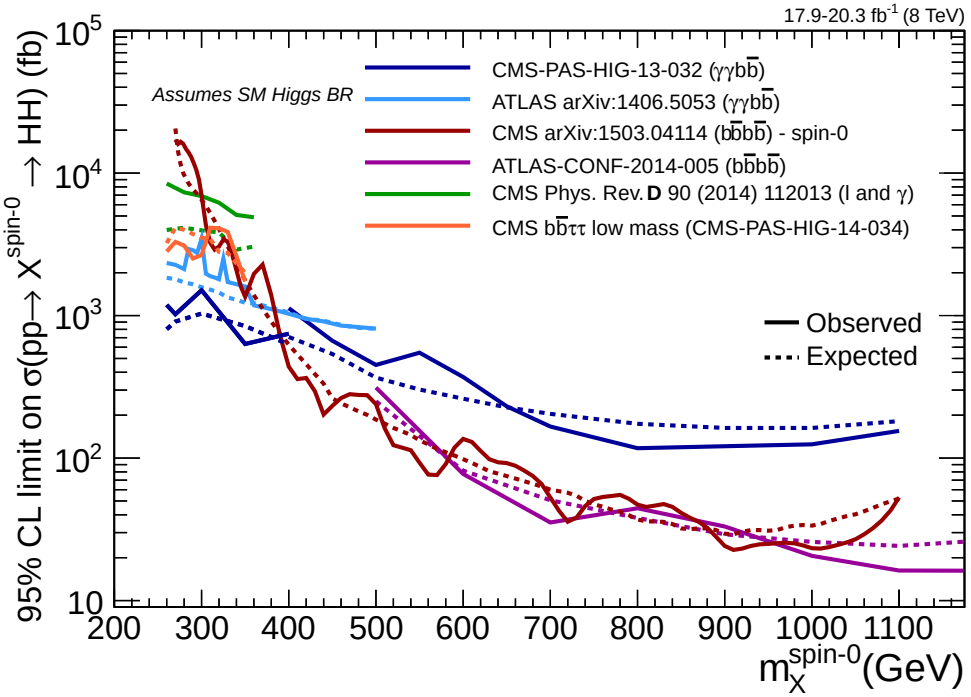

Figure 8.3: The observed and expected upper limits of $X_{\text{spin-0}} \to HH$ production at 95% CL are compared over various searches performed by the CMS and ATLAS Collaborations looking at the $\gamma\gamma b\bar{b}$ [95, 96], $b\bar{b}b\bar{b}$ [97, 98], $\tau\tau b\bar{b}$ [99], and multileptons and photons [100] final states. As the CMS $b\bar{b}b\bar{b}$ result is dependent on the resonance spin, the spin-0 result for that analysis is used.

modificator μ_{HH}, defined generally as

$$\mu = \frac{\sigma}{\sigma_{\text{SM}}}\,, \tag{8.1}$$

the observed (expected) limit is 72.9 (60.7). These calculations account for the theoretical uncertainty associated with the SM NNLO cross section.

Chapter 9

Conclusion

Double Higgs production is a possible key to discovering physics beyond the Standard Model. The process can be motivated either through the production of a new resonance decaying to a Higgs pair or through the couplings λ, y_t, or c_2, which effect the nonresonant production of a Higgs pair. The SM values for these couplings are $\kappa_\lambda = 1$, $\kappa_t = 1$, and $c_2 = 0$. The search also provides a means to test the SM through the measurement of the Higgs trilinear self-coupling, giving more insight into the nature of the Higgs potential. However, the SM rate is too low for current experiments to be sensitive to it. Preliminary estimates indicate that a factor of $\mathcal{O}(100)$ more data is needed at the relevant LHC experiments in order to achieve SM sensitivity.

With the current data, a search for the production of Higgs pairs is performed by the CMS Collaboration in the decay channel $HH \to \gamma\gamma b\bar{b}$ using 19.7 fb^{-1} of pp collisions collected at $\sqrt{s} = 8$ TeV. For the search of resonant production, masses under consideration are in the range between 260 GeV and 1100 GeV. The selected events are divided into two categories based on the number of jets consistent with the hadronization of b-quarks. A simultaneous fit is performed in each category to the $m_{\gamma\gamma}$ spectrum for resonant masses below 400 GeV and to the $m^{\text{kin}}_{\gamma\gamma jj}$ spectrum for resonant masses above 400 GeV, and the results in each category are combined. The

observations are consistent with expectations from Standard Model backgrounds, and upper limits at the 95% confidence level are placed on the production cross section. The limits are compared to the predictions from theories beyond the Standard Model, such as the Radion and KK-graviton from models with warped extra dimensions or the heavy Higgs from models with supersymmetry. The Radion with $\Lambda_R = 1$ TeV is observed (expected) to be excluded with masses below 0.97 TeV (0.88 TeV), and the RS1 KK-graviton is observed to be excluded with masses between 340 GeV and 400 GeV.

For the search of SM double Higgs production, a fit to the $m_{\gamma\gamma} \times m_{jj}$ plane is performed. The selected events are divided into four categories based on the number of jets consistent with the hadronization of b-quarks and the value of the four-body mass. The observations are consistent with expectations from the Standard Model backgrounds, and an upper limit at the 95% confidence level is placed on the production cross section. The observed (expected) upper limit on the production cross section is 1.91 fb (1.59 fb), or 72.9 (60.7) times the NNLO SM cross section. In addition, a theoretical framework is described for the future study describing how anomalous couplings change the production cross section.

Bibliography

[1] S. L. Glashow. Partial symmetries of weak interactions. *Nucl.Phys.*, 22:579–588, 1961.

[2] S. Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264–1266, 1967.

[3] A. Salam. Weak and electromagnetic interactions. *Conf.Proc.*, C680519:367–377, 1968.

[4] A. Tuna. *Evidence for Decays of the Higgs Boson to Tau Leptons at ATLAS*. PhD thesis, University of Pennsylvania, 2015.

[5] User:MissMJ. Standard model of elementary particles. `http://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles_modified_version.svg`, 2014.

[6] J. Goldstone. Field theories with superconductor solutions. *Il Nuovo Cimento (1955-1965)*, 19:154–164, 1961. 10.1007/BF02812722.

[7] J. Goldstone, A. Salam, and S. Weinberg. Broken symmetries. *Physical Review*, 127:965–970, 1962.

[8] R. P. Feynman. Space-time approach to non-relativistic quantum mechanics. *Rev. Mod. Phys.*, 20:367–387, 1948.

[9] User:Garyzx. Standard model feynman diagram vertices. `http://commons.wikimedia.org/wiki/File:Standard_Model_Feynman_Diagram_Vertices.png`, 2011.

[10] TeVI Group. Design Report Tevatron 1 project. Technical Report FERMILAB-DESIGN-1984-01, 1984.

[11] L. Evans and P. Bryant. LHC machine. *Journal of Instrumentation*, 3:S08001, 2008.

[12] ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys.Lett.*, B716:1–29, 2012.

[13] CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys.Lett.*, B716:30–61, 2012.

[14] CMS Collaboration. Observation of a new boson with mass near 125 GeV in pp collisions at sqrt(s) = 7 and 8 TeV. *JHEP*, 06:081, 2013.

[15] LHC Higgs Cross Section Working Group, S. Heinemeyer, C. Mariotti, G. Passarino, and R. Tanaka (Eds.). Handbook of LHC Higgs cross sections: 3. Higgs properties. *CERN-2013-004*, CERN, Geneva, 2013.

[16] CMS Collaboration. Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV. Technical Report arXiv:1412.8662. CERN-PH-EP-2014-288. CMS-HIG-14-009, CERN, Geneva, 2014. Comments: Submitted to Eur. Phys. J. C.

[17] D. Clowe, M. Bradac, A. H. Gonzalez, M. Markevitch, S. W. Randall, et al. A direct empirical proof of the existence of dark matter. *Astrophys.J.*, 648:L109–L113, 2006.

[18] Planck Collaboration. Planck 2013 results. I. Overview of products and scientific results. *Astron.Astrophys.*, 571:A1, 2014.

[19] Super-Kamiokande Collaboration. Evidence for oscillation of atmospheric neutrinos. *Phys.Rev.Lett.*, 81:1562–1567, 1998.

[20] J. Baglio, A. Djouadi, , R. Gröber, M. M. Mühlleitner, J. Quevillon, et al. The measurement of the Higgs self-coupling at the LHC: theoretical status. *JHEP*, 1304:151, 2013.

[21] T. Plehn, M. Spira, and P. M. Zerwas. Pair production of neutral Higgs particles in gluon-gluon collisions. *Nucl.Phys.*, B479:46–64, 1996.

[22] J. Ellis, V. Sanz, and T. You. Associated production evidence against Higgs impostors and anomalous couplings. *Eur.Phys.J.*, C73:2507, 2013.

[23] J. Ellis and T. You. Updated global analysis of Higgs couplings. *JHEP*, 1306:103, 2013.

[24] A. Belyaev, M. Drees, O. Eboli, J. K. Mizukoshi, and S. F. Novaes. Supersymmetric Higgs pair production at hadron colliders. *Phys.Rev.*, D60:075008, 1999.

[25] C. O. Dib, R. Rosenfeld, and A. Zerwekh. Double Higgs production and quadratic divergence cancellation in little Higgs models with T parity. *JHEP*, 0605:074, 2006.

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

[27] R. Contino, M. Ghezzi, M. Moretti, G. Panico, F. Piccinini, et al. Anomalous couplings in double Higgs production. *JHEP*, 1208:154, 2012.

[28] D. de Florian and J. Mazzitelli. Higgs boson pair production at next-to-next-to-leading order in QCD. *Phys.Rev.Lett.*, 111:201801, 2013.

[29] F. Goertz, A. Papaefstathiou, L. L. Yang, and J. Zurita. Higgs boson pair production in the D=6 extension of the SM. *JHEP*, 1504:167, 2015.

[30] L. Randall and R. Sundrum. A large mass hierarchy from a small extra dimension. *Phys.Rev.Lett.*, 83:3370–3373, 1999.

[31] W. D. Goldberger and M. B. Wise. Modulus stabilization with bulk fields. *Phys.Rev.Lett.*, 83:4922–4925, 1999.

[32] T. Gherghetta. TASI lectures on a holographic view of beyond the standard model physics. Technical Report arXiv:1008.2570, 2010.

[33] G. F. Giudice, R. Rattazzi, and J. D. Wells. Quantum gravity and extra dimensions at high-energy colliders. *Nucl.Phys.*, B544:3–38, 1999.

[34] C. Csaki, M. L. Graesser, and G. D. Kribs. Radion dynamics and electroweak physics. *Phys.Rev.*, D63:065002, 2001.

[35] R. Barbieri, D. Buttazzo, K. Kannike, F. Sala, and A. Tesi. One or more Higgs bosons? *Phys.Rev.*, D88:055011, 2013.

[36] E. Bertuzzo, T. S. Ray, H. de Sandes, and C. A. Savoy. On composite two Higgs doublet models. *JHEP*, 1305:153, 2013.

[37] G. C. Branco, P. M. Ferreira, L. Lavoura, M. N. Rebelo, M. Sher, et al. Theory and phenomenology of two-Higgs-doublet models. *Phys.Rept.*, 516:1–102, 2012.

[38] N. Craig and S. Thomas. Exclusive signals of an extended Higgs sector. *JHEP*, 1211:083, 2012.

[39] N. Craig, J. Galloway, and S. Thomas. Searching for signs of the second Higgs doublet. Technical Report arXiv:1305.2424. RU-NHETC-2013-07, 2013.

[40] A. Djouadi. The Anatomy of electro-weak symmetry breaking. II. The Higgs bosons in the minimal supersymmetric model. *Phys.Rept.*, 459:1–241, 2008.

[41] CERN. About CERN. `http://home.web.cern.ch/about`, 2015.

[42] User:U5K0. CERN international relations map. `http://en.wikipedia.org/wiki/File:CERN_international_relations_map.svg`, 2012.

[43] LEP Design Group. Design study of a 22 to 130 GeV electron-positron colliding beam machine (LEP). Technical Report CERN/ISR-LEP/79-33, CERN, 1979.

[44] S. Dailler. Map of the Geneva region and of the LHC. Technical Report LHC-PHO-1997-169, 1997.

[45] CERN. The accelerator complex. http://home.web.cern.ch/about/accelerators, 2015.

[46] J.-L. Caron. CERN accelerator complex (operating and approved projets). Technical Report LHC-PHO-1991-001, 1991.

[47] AC Team. Diagram of an LHC dipole magnet. Technical Report CERN-DI-9906025, 1999.

[48] ATLAS Collaboration. The ATLAS experiment at the CERN Large Hadron Collider. *Journal of Instrumentation*, 3:S08003, 2008.

[49] LHCb Collaboration. The LHCb detector at the LHC. *Journal of Instrumentation*, 3:S08005, 2008.

[50] ALICE Collaboration. The ALICE experiment at the CERN LHC. *Journal of Instrumentation*, 3:S08002, 2008.

[51] CMS Collaboration. CMS luminosity public results. https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults, 2014.

[52] CMS Collaboration. *The CMS magnet project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997.

[53] CMS Collaboration. The CMS experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

[54] CMS Collaboration. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical Design Report CMS. CERN, Geneva, 2006.

[55] CMS Collaboration. Detector drawings. Technical report, 2012. CMS Collection.

[56] CMS Collaboration. *The CMS tracker system project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997.

[57] CMS Collaboration. *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997.

[58] CMS Collaboration. *The CMS hadron calorimeter project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997.

[59] CMS Collaboration. *The CMS muon project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997.

[60] CMS Collaboration. *CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems*. Technical Design Report CMS. CERN, Geneva, 2000.

[61] CMS Collaboration. CMS high level trigger. Technical Report LHCC-G-134. CERN-LHCC-2007-021, CERN, Geneva, 2007.

[62] CMS Collaboration. Observation of the diphoton decay of the Higgs boson and measurement of its properties. *The European Physical Journal C*, 74(10), 2014.

[63] LHC Computing Grid Project. *LHC Computing Grid: Technical Design Report*. Technical Design Report LCG. CERN, Geneva, 2005.

[64] LHC Computing Grid Project. Worldwide LHC computing grid. `http://wlcg-public.web.cern.ch/`, 2015.

[65] M. A. Dobbs, S. Frixione, E. Laenen, K. Tollefson, H. Baer, et al. Les Houches guidebook to Monte Carlo generators for hadron collider physics. Technical Report arXiv:0403045. FERMILAB-CONF-04-183-T, 2004.

[66] P. Bartalini, E.L. Berger, B. Blok, G. Calucci, R. Corke, et al. Multi-parton interactions at the LHC. Technical Report arXiv:1111.0469. ANL-HEP-PR-11-65. CMS-CR-2011-048. DESY-11-185. KA-TP-32-2011. TTK-11-52, 2011.

[67] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer. MadGraph5 : Going Beyond. *JHEP*, 06:128, 2011.

[68] T. Sjostrand, S. Mrenna, and P. Skands. Pythia 6.4 physics and manual. *J. High Energy Phys.*, 0605:026, 2006.

[69] K. Agashe, H. Davoudiasl, G. Perez, and A. Soni. Warped gravitons at the LHC and beyond. *Phys.Rev.*, D76:036006, 2007.

[70] S. Frixione, P. Nason, and C. Oleari. Matching NLO QCD computations with parton shower simulations: the POWHEG method. *JHEP*, 0711:070, 2007.

[71] T. Gleisberg, Stefan. Hoeche, F. Krauss, M. Schonherr, S. Schumann, et al. Event generation with SHERPA 1.1. *JHEP*, 0902:007, 2009.

[72] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt. Parton distributions for the LHC. *Eur.Phys.J.*, C63:189–285, 2009.

[73] CMS Collaboration. Particle–flow event reconstruction in CMS and performance for jets, taus, and $\not\!\!E_{\mathrm{T}}$. Technical Report CMS-PAS-PFT-09-001, CMS Collaboration, 2009.

[74] CMS Collaboration. Commissioning of the particle–flow event reconstruction with the first LHC collisions recorded in the CMS detector. Technical Report CMS-PAS-PFT-10-001, CMS Collaboration, 2010.

[75] CMS Collaboration. Photon reconstruction and identification at sqrt(s) = 7 TeV. Technical Report CMS-PAS-EGM-10-005, CERN, 2010. Geneva, 2010.

[76] CMS Collaboration. Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at $\sqrt{s}$=7 TeV. *JINST*, 8:P09009, 2013.

[77] M. Cacciari, G. P. Salam, and G. Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008.

[78] M. Cacciari and G. P. Salam. Pileup subtraction using jet areas. *Phys. Lett. B*, 659:119, 2008.

[79] M. Cacciari, G. P. Salam, and G. Soyez. FastJet user manual. *Eur. Phys. J. C*, 72:1896, 2012.

[80] CMS Collaboration. Determination of jet energy calibration and transverse momentum resolution in CMS. *JINST*, 6:P11002, 2011.

[81] CMS Collaboration. Pileup jet identification. Technical Report CMS-PAS-JME-13-005, CERN, Geneva, 2013.

[82] CMS Collaboration. Identification of b-quark jets with the CMS experiment. *JINST*, 8:P04013, 2013.

[83] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and Regression Trees*. Chapman & Hall, New York, 1984.

[84] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss. TMVA: Toolkit for Multivariate Data Analysis. *PoS*, ACAT:040, 2007.

[85] J. H. Friedman. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29:1189–1232, 2000.

[86] J. H. Friedman. Stochastic gradient boosting. *Comput. Stat. Data Anal.*, 38(4):367–378, 2002.

[87] CMS Collaboration. Updated measurements of the Higgs boson at 125 GeV in the two photon decay channel. Technical Report CMS-PAS-HIG-13-001, CERN, Geneva, 2013.

[88] CMS Collaboration. Search for a standard-model-like Higgs boson with a mass in the range 145 to 1000 GeV at the LHC. *Eur.Phys.J.*, C73:2469, 2013.

[89] S. D. Ellis, C. K. Vermilion, and J. R. Walsh. Techniques for improved heavy particle searches with jet substructure. *Phys.Rev.*, D80:051501, 2009.

[90] CMS Collaboration. CMS luminosity based on pixel cluster counting - summer 2013 update. Technical Report CMS-PAS-LUM-13-001, CERN, Geneva, 2013.

[91] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka (Eds.). Handbook of LHC Higgs cross sections: 1. inclusive observables. *CERN-2011-002*, CERN, Geneva, 2011.

[92] CMS Collaboration. Measurement of the properties of a Higgs boson in the four-lepton final state. *Phys.Rev.*, D89(9):092007, 2014.

[93] A. L. Read. Presentation of search results: the CL_s technique. *J.Phys.G:Nucl.Part.Phys.*, 28:2693, 2002.

[94] T. Junk. Confidence level computation for combining searches with small statistics. *Nucl.Instrum.Meth.*, A434:435, 1999.

[95] CMS Collaboration. Search for the resonant production of two Higgs bosons in the final state with two photons and two bottom quarks. Technical Report CMS-PAS-HIG-13-032, CERN, Geneva, 2014.

[96] ATLAS Collaboration. Search For Higgs Boson Pair Production in the $\gamma\gamma b\bar{b}$ Final State using pp Collision Data at $\sqrt{s} = 8$ TeV from the ATLAS Detector. *Phys.Rev.Lett.*, 114(8):081802, 2015.

[97] CMS Collaboration. Search for resonant pair production of Higgs bosons decaying to two bottom quark-antiquark pairs in proton-proton collisions at 8 TeV. Technical Report CMS-HIG-14-013, CERN-PH-EP-2015-042, 2015.

[98] ATLAS Collaboration. A search for resonant Higgs-pair production in the $b\bar{b}b\bar{b}$ final state in pp collisions at $\sqrt{s} = 8$ TeV. Technical Report ATLAS-CONF-2014-005, CERN, Geneva, 2014.

[99] CMS Collaboration. Search for a heavy scalar boson decaying to a pair of 125 GeV Higgs bosons (hh) or for a heavy pseudoscalar boson decaying to Zh, in the final state with h$\rightarrow \tau\tau$. Technical Report CMS-PAS-HIG-14-034, CERN, Geneva, publication in progress.

[100] CMS Collaboration. Searches for heavy Higgs bosons in two-Higgs-doublet models and for $t \rightarrow ch$ decay using multilepton and diphoton final states in pp collisions at 8 TeV. *Phys. Rev. D*, 90:112013, 2014.