Evidence for Higgs boson decays to a low mass lepton pair and a photon with the ATLAS detector at the LHC

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
an der
Fakultät für Mathematik, Informatik und Naturwissenschaften
Fachbereich Physik
der
Universität Hamburg

vorgelegt von
Artem Basalaev

Hamburg 2021

Gutachter/innen der Dissertation:	Dr. Sarah Heim Prof. Dr. Peter Schleper
Zusammensetzung der Prüfungskommission:	Dr. Sarah Heim Prof. Dr. Peter Schleper Prof. Dr. Elisabetta Gallo Prof. Dr. Gudrid Moortgat-Pick Prof. Dr. Dieter Horns
Vorsitzender der Prüfungskommission:	Prof. Dr. Dieter Horns
Datum der Disputation:	3 November 2021
Vorsitzender Fach-Promotionsausschusses PHYSIK:	Prof. Dr. Wolfgang Hansen
Leiter des Fachbereichs PHYSIK:	Prof. Dr. Günter H. W. Sigl
Dekander Fakultät MIN:	Prof. Dr. Heinrich Graener

Eidesstattliche Erklärung

Hiermit versichere ich an Eides statt, die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen benutzt zu haben. Die eingereichte schriftliche Fassung entspricht der auf dem elektronischen Speichermedium. Die Dissertation wurde in der vorgelegten oder einer ähnlichen Form nicht schon einmal in einem früheren Promotionsverfahren angenommen oder als ungenügend beurteilt.

Hamburg, den 1. September 2021.

Artem Basalaev

Abstract

This thesis presents the first evidence of Higgs boson decay to two leptons and a photon. Rare Higgs boson decays predicted in the Standard Model (SM) presently come within experimental reach thanks to the large amount of data collected by the experiments at the Large Hadron Collider (LHC) at CERN. Given a central role played by the Higgs boson in the SM, studying its rare decays opens up a possibility for more stringent tests of the SM.

A very rare decay of Higgs boson to a low mass electron or muon pair and a photon is explored in this thesis using 139 fb^{-1}of pp collision data at a center-of-mass energy $\sqrt{s}$ = 13 TeV recorded with the ATLAS detector at the LHC during Run 2 (2015-2018). In the phase space with lepton pair invariant mass $m_{\ell\ell} < 30$ GeV ($l = e, \mu$) the expected SM branching ratio is $\mathcal{B}(H \to \ell\ell\gamma) \approx 10^{-4}$. The low branching ratio presents an experimental challenge, with the expected number of background events vastly exceeding the number of signal $H \to \ell\ell\gamma$ events.

Another experimental challenge of the analysis is that in the low lepton pair mass regime the two leptons tend to be highly collimated. For the electrons, their energy deposits in the calorimeter of the ATLAS detector often remain unresolved by the standard reconstruction algorithms. To overcome this, a new identification algorithm is used, based on multivariate discriminant. The new algorithm more than doubles the number of selected $ee\gamma$ events compared to the standard reconstruction algorithms. In addition, collimated leptons often do not satisfy standard criteria on additional activity in their vicinity (isolation), which is imposed to reduce backgrounds. Isolation criteria are therefore corrected taking into account contribution of leptons to each other's isolation.

A combined statistical model is constructed using parametric functions describing signal and background $m_{\ell\ell\gamma}$ distributions in each analysis category. The observed signal yield is extracted from the fit of this model to data. An excess is observed over the background-only hypothesis with a significance of 3.2 standard deviations. The best-fit ratio of the observed event yield to the SM expectation is 1.5 ± 0.5. The fiducial cross-section times branching ratio in the $m_{\ell\ell} < 30$ GeV region is measured at $\sigma(H) \times \mathcal{B}(H \to \ell\ell\gamma) = 8.7^{+2.8}_{-2.7}$ fb. The analysis is still limited by the small number of expected events and the systematic uncertainties constitute only 35% of the statistical uncertainty. Among systematic uncertainties, the uncertainty associated with a bias in background function choice is dominating. The impact of systematic uncertainties on the results is studied in detail and the results are additionally verified using pseudo-experiments.

Zusammenfassung

Das Thema dieser Arbeit ist die erstmalige experimentelle Evidenz für den Zerfall des Higgs-Bosons in zwei Leptonen und ein Photon. Dank der enorm großen Zahl der aufgezeichneten Proton-Proton-Kollisionen am Large Hadron Collider (LHC) am CERN gelangen vom Standardmodell der Teilchenphysik vorhergesagte, seltene Higgs-Boson-Zerfälle in Reichweite der experimentellen Sensitivität. Die Untersuchung solcher seltenen Higgs-Boson-Zerfälle erlaubt eine detaillierte Überprüfung der Vorhersagen des Standardmodells.

In dieser Arbeit wird ein seltener Zerfall des Higgs-Bosons in ein Elektron- oder Myonpaar mit niedriger invarianter Masse und ein Photon untersucht. Der verwendete Datensatz entspricht 139 fb^{-1} Proton-Proton-Kollisionen bei 13 TeV Schwerpunktsenergie, aufgezeichnet mit dem ATLAS-Detektor am LHC während des Run 2 (2015-2018).

Im betrachteten Phasenraum, mit einer invarianten Masse des Leptonpaares, $m_{\ell\ell} < 30$ GeV ($l = e, \mu$), beträgt das vom Standardmodell vorhergesagte Verzweigungsverhältnis dieses Zerfalls $\mathcal{B}(H \to \ell\ell\gamma) \approx 10^{-4}$. Das niedrige Verzweigungsverhältnis stellt eine besondere Schwierigkeit für den experimentellen Nachweis des Zerfalls dar, denn die erwartete Zahl der Ereignisse aus Untergrundprozessen ist sehr viel höher als die erwartete Zahl von Signalereignissen des $H \to \ell\ell\gamma$ Zerfalls.

Eine weitere Herausforderung ist die Tatsache, dass bei niedriger invarianter Masse des Leptonpaares die beiden Leptonen stark kollimiert sind. Das bedeutet im Fall der Elektronen, dass die Standard-Algorithmen zur Rekonstruktion von Schauern im elektromagnetischen Kalorimeter nicht in der Lage sind, beide Schauer getrennt aufzulösen. Um dies zu verbessern, wurde ein neuer Identifikations-Algorithmus verwendet, basierend auf einer multivariaten Diskriminante. Mit dem neuen Algorithmus werden im Vergleich zum Standard-Algorithmus mehr als doppelt so viele Elektronpaar-Ereignisse als solche erkannt. Weiterhin erfüllen kollimierte Leptonpaare häufig nicht die Standardkriterien für Isolation, also Abwesenheit von zusätzlicher Aktivität im Detektor in der Nähe der Leptonen. Die Kriterien für die Lepton-Isolation werden verwendet, um Untergrund-Ereignisse zu reduzieren. In der hier präsentierten Analyse wird die Isolation für kollimierte Paare korrigiert, indem der Beitrag des einen Leptons zur Isolation des jeweils anderen Leptons berücksichtigt wird.

Zur statistischen Analyse wird ein sogenanntes Kombiniertes Statistisches Modell verwendet, welches die Signal- und Untergrundverteilungen in mehreren Ereigniskategorien durch Funktionen parametrisiert. Die beobachtete Zahl von Signalereignissen wird durch eine Parameteranpassung des statistischen Modells an Daten ermittelt. Dabei wurde ein statistischer Exzess von 3.2 Standardabweichungen im Vergleich zur Untergrundhypothese beobachtet. Das Verhältnis der gemessenen Zahl von Signal-Ereignissen zur erwarteten Zahl aus der Standardmodellvorhersage beträgt 1.5 ± 0.5. Das Produkt des fiduziellen Wirkungsquerschnitts und des Verzweigungsverhältnisses in der Region $m_{\ell\ell} < 30$ GeV wurde zu $\sigma(H) \times \mathcal{B}(H \to \ell\ell\gamma) = 8.7^{+2.8}_{-2.7}$ fb bestimmt.

Die präsentierte Analyse ist statistisch limitiert, die systematischen Unsicherheiten betragen etwa 35% der statistischen Unsicherheiten. Die dominierende systematische Unsicherheit stammt aus einer möglichen Verzerrung durch die Wahl der Funktion zur Beschreibung des Untergrunds. Der Einfluss der systematischen Unsicherheiten auf das Ergebnis der Analyse wurde im Detail studiert und das Resultat zusätzlich durch Pseudo-Experimente verifiziert.

Acknowledgements

Foremost, I would like to thank Sarah Heim, my research advisor, for the opportunity to work with her on this thesis project and for her continued support throughout the years. Also I would like to thank Kurt Brendlinger for providing his expert advice and guidance for my work on the analysis. I would like to thank both of them, as well as Jan and Daniela Küchler, for helping to review this thesis. Thank you for all your suggestions and your patience!

I am grateful to all my colleagues in the DESY-ATLAS group for all the fruitful discussions we had. In particular I would like to thank Kerstin Tackmann for the discussions about theory uncertainties for the analysis. It has been a pleasure to work in this group, and I learned a lot.

I would also like to thank all the ATLAS collaborators I worked with in all these years, it has been a wonderful time! Thanks to my colleagues in the analysis team, Anthony Morley and Mateusz Dydnal. Thanks to all my colleagues in Simulation group, and especially to John Chapman, for the valuable experience outside of the physics analysis. My special thank you goes to Peter Jenni for his support, advice, and being a source of inspiration for me, which brought me necessary confidence to continue pursuing my degree.

Finally I would like to thank Ruth for her suggestions for this thesis, and most of all for her moral support and confidence in me, which helped me to see this work through to its completion.

Contents

CHAPTER 1

Introduction

The field of particle physics studies the world on the smallest of scales, focusing on the indivisible building blocks of matter – elementary particles. The current state-of-the-art combines several theoretical models in the Standard Model of particle physics, postulating the existence of the following elementary particles: three generations of matter constituents (fermions) and the force carriers (bosons) – Z, W bosons and the photon for the electroweak, the gluon for the strong interaction, and the Higgs boson.

In our everyday experience, most matter consists of three fermions of the first generation (up and down quarks and electrons), as well as photons for electromagnetic radiation. With the exception of neutrinos, which are abundant but hard to detect, other elementary particles are short-lived and typically require high energies (above the particles' mass) to be produced. Such conditions are met by particle accelerators and colliders. Equipped with sensitive detectors, these experiments enable us to infer the presence of the short-lived particles by analyzing their decay products. Thanks to advances in accelerator and detector technology, all particles from the other generations have been successfully detected. Among the force carriers, gluons were detected at DESY at the DORIS storage ring in 1978 [1], and Z and W bosons were discovered at CERN at the Super Proton Synchrotron in 1983 [2, 3]. The Higgs mechanism [4–6] provides a compelling way to explain the observed values of the Z and W masses through electroweak symmetry breaking. But its particle manifestation – the Higgs boson – remained elusive for a long time.

Due to its large mass, production and detection of the Higgs boson necessitated the construction of the largest particle accelerator ever built: the Large Hadron Collider (LHC) at CERN with 27 km circumference situated on the border between France and Switzerland. A new particle consistent with the Standard Model Higgs boson properties with a mass of approximately 125 GeV was discovered by the ATLAS and CMS collaborations [7, 8] using data gathered during the first run of the LHC. Thus the last remaining piece of the Standard Model was experimentally confirmed.

Although a very successful theory, the Standard Model is known to be incomplete. For example, it does not provide a suitable particle candidate for dark matter, whose existence is firmly established from astrophysical observations (summarized in Ref. [9]). The Standard Model also does not include gravitational interaction, nor does it provide a satisfying explanation for the differences of the characteristic scales of the interactions.

There are extensions of the Standard Model aimed to resolve these problems, and due to its relationship with the mass of the particles, the Higgs boson can be uniquely sensitive to potential

deviations from the Standard Model. Some theories propose the Higgs boson to be a portal to the dark matter sector, other theories predict additional Higgs bosons. Most of the proposed extensions would manifest themselves in interactions of the known Higgs boson deviating from the Standard Model prediction. Therefore it is very important to study the properties of the Higgs boson, including its different decay channels.

This thesis presents a search for the Higgs boson decaying to a photon and a low-mass lepton pair (e^+e^- or $\mu^+\mu^-$) with the ATLAS detector at the Large Hadron Collider. In the explored lepton pair invariant mass range, $m_{\ell\ell} < 30$ GeV ($l = e, \mu$), the decay is almost exclusively happening via γ^* pole Feynman diagrams [10], i.e. through an intermediate state with a virtual photon $H \to \gamma^*\gamma \to \ell^+\ell^-\gamma$. Therefore the presented search is complimentary to a $H \to Z\gamma \to \ell^+\ell^-\gamma$ search in the same final state. There are three particles in the final state, and future measurements of the forward-backward asymmetry of the decay could provide a test for some of the core aspects of the Standard Model, such as charge-parity (CP) symmetry violation [11]. The decay of a Higgs boson to a photon and a low-mass lepton pair is very rare, and features a challenging event topology. In particular, the virtual photon is typically boosted, resulting in the leptons being very close together. This often leads to overlapping of the energy deposits from the electrons in the electromagnetic calorimeter and requires special triggers during data taking and a special off-line identification algorithm to reconstruct these electrons. The new algorithm more than doubled the number of selected $ee\gamma$ Higgs boson events compared to the standard reconstruction.

The main result of the presented search is the first evidence for Higgs boson decay to a photon and a lepton pair at 3.2σ significance. Additionally, the Higgs boson production cross-section times the $H \to \ell\ell\gamma$ branching ratio for $m_{\ell\ell} < 30$ GeV is measured to be $8.7^{+2.8}_{-2.7}$ fb.

The thesis has the following structure. In Chapter 2, basic concepts of the Standard Model are described, with a focus on the Higgs mechanism. In Chapter 3, the experimental setup is described: the Large Hadron Collider and the ATLAS detector. In Chapter 4, the data analysis strategy is described in detail, along with the results and their interpretation. Chapter 5 concludes the thesis with a summary and outlook.

CHAPTER 2

Theory overview

2.1 The Standard Model of particle physics

The Standard Model (SM) of particle physics is a quantum field theory (QFT) incorporating all known particles and their interactions. The particle content of the SM is shown in Figure 2.1. In this chapter, a basic theoretical description of the SM is given, focusing on the Higgs mechanism and other aspects relevant for this thesis. Information given here is based on Ref. [13], and much more detailed and pedagogical description can be found there.

In the QFT, particles are represented as excitations of all-permeating quantum fields. How these fields change over time is defined by their Lagrangian density, and interactions between particles are governed by the interaction terms of the combined Lagrangian density of the underlying fields. The equations of motion of the QFT are derived from the so-called *least action principle*, which in the QFT context can be written as follows:

$$\frac{\delta S}{\delta \phi_i} = 0, \tag{2.1}$$

where ϕ_i are values of the i^{th} field, and S is called *action* functional. The action is an integral over Lagrangian density $\mathcal{L}$, which itself is a function of values of fields ϕ_i, their derivatives $\frac{\partial \phi_i}{\partial x^k}$, $\frac{\partial \phi_i}{\partial t}$, coordinates x^k ($k = 1, 2, 3$) and time t:

$$S[\phi_i] = \int \mathcal{L}\left(\frac{\partial \phi_i}{\partial x^k}, \frac{\partial \phi_i}{\partial t}, x^k, t\right) \mathrm{dt}\, \mathrm{d}^3\mathrm{x} \tag{2.2}$$

The SM is a *gauge* theory, which means that its action S is locally invariant under internal symmetry transformations. The gauge postulate naturally gives rise to *gauge fields* and emergence of additional particles – the force carriers (gauge bosons). Symmetry transformations are represented by matrices denoting possible rotations in the space of related quantum numbers[1]. Specifically, the SM is postulated to be invariant under the following transformations:

$$\mathrm{SU(3)} \times \mathrm{SU(2)} \times \mathrm{U(1)}, \tag{2.3}$$

[1] An example of a symmetry transformation could be relabelling quantum numbers by the opposite ones. E.g. replacing all positive charges with negative charges and vice versa does not affect the resulting electromagnetic force.

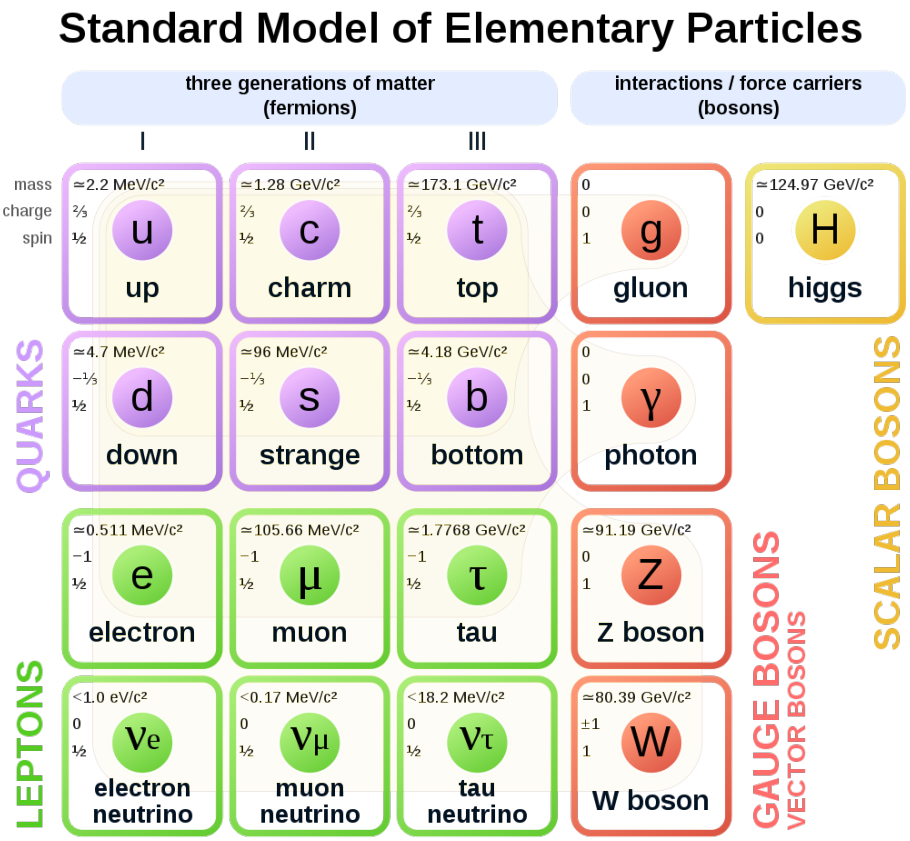


Figure 2.1: Elementary particles of the Standard Model. Quarks and leptons, collectively referred to as fermions (spin quantum number is 1/2) constitute matter. Interactions between particles are mediated by bosons: gluons for strong interaction, Z and W bosons and the photon for the electroweak interaction, and the Higgs boson. All known bosons are vector particles (spin 1), except the Higgs, which is the only scalar boson (spin 0). Taken from Ref. [12].

where U(1) is a group of complex numbers with absolute value 1 under multiplication; SU(2) and SU(3) are groups of two-dimensional (three-dimensional) unitary matrices with determinant 1.

As follows from Noether's theorem, internal symmetries of the theory are associated with conservation laws. The SU(3) symmetry is related to the strong interaction and conservation of the color charge, described by quantum chromodynamics. The SU(2) × U(1) symmetry is associated with the electroweak interaction and conservation of the weak hypercharge, described by the electroweak theory.

2.2 Quantum chromodynamics

The Lagrangian density for quantum chromodynamics (QCD) can be written as follows:

$$\mathcal{L}_{\text{QCD}} = \sum_q \left(\bar{\psi}_{qi} i\gamma^\mu \left[\delta_{ij}\partial_\mu + ig_s\left(G^a_\mu t_a\right)_{ij} \right] \psi_{qj} - m_q \bar{\psi}_{qi}\psi_{qi} \right) - \frac{1}{4} G^a_{\mu\nu} G^{\mu\nu}_a, \tag{2.4}$$

where ψ are 4-component (Dirac bispinor) quark fields, and the index q denotes different quark types ($q = u, d, c, s, t, b$). The indices i and j reflect the fact that quark fields can have different *color* charges. The underlying SU(3) symmetry allows for 3 different color charges. Inspired by the RGB color encoding, color charges for quarks are called red r, green g and blue b (antired $\bar{r}$, antigreen $\bar{g}$ and antiblue $\bar{b}$ for antiquarks). The Greek letter indices in the equation are used for compact (Einstein's) notation to refer to all spacetime components. Each index takes values ranging from 0 to 3; 0 for time, and 1, 2, 3 for space coordinates. For each index variable appearing twice, sums spanning all combinations of index values are implied. The quark mass is denoted by m_q, and $g_s = \sqrt{4\pi\alpha_s}$ corresponds to the coupling (strength) of the strong interaction α_s. The γ^μ are 4×4 Dirac matrices, which act on the four components of the Dirac spinor ψ, representing Lorentz transformations; t_a are 3×3 Gell-Mann matrices, representing rotations in the SU(3) color space, with $a = 1, \ldots, 8$. $G^{\mu\nu}_a$ is the *gluon* field strength tensor, and G^μ_a is the gluon field potential. It is the gluon field that mediates strong interaction between quarks. From the structure of SU(3), it follows that there are 8 types of gluons, which can carry the following linear combinations of color charges:

$$r\bar{g},\ g\bar{r},\ r\bar{b},\ b\bar{r},\ g\bar{b},\ b\bar{g},\ \frac{1}{\sqrt{2}}(r\bar{r} - g\bar{g}),\ \text{and}\ \frac{1}{\sqrt{6}}(r\bar{r} + g\bar{g} - 2b\bar{b}). \tag{2.5}$$

For example a "red" quark can transform into a "green" quark by exchanging a $r\bar{g}$ gluon.

There is one more aspect of the strong interaction which has profound consequences for the experiment: since gluons themselves carry a color charge, they exhibit *self-interaction*. The gluon field tensor has the following form:

$$G^{\mu\nu}_a = \partial_\mu G^\nu_a - \partial_\nu G^\mu_a - g f^{abc} G^\mu_b G^\nu_c. \tag{2.6}$$

The third term describes the self-interaction of the gluons, and f^{abc} are structure constants of the SU(3) color group. Self-interaction of gluons leads to the phenomenon called *color confinement* under normal (low-energy) conditions. From experimental perspective, it manifests itself in the absence of freely propagating color states. Any freely propagating state should be *colorless*, or in other words have a color charge combination resulting in a "white" color. For example a meson – a two-particle

composite state – may consist of the r and $\bar{r}$ quarks, while a baryon – a three-particle composite state – may consist of the r, g, b quarks.

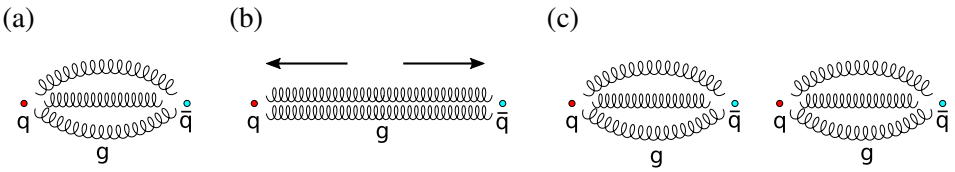


Figure 2.2: Color confinement in QCD: for interacting quarks, gluon field forms a characteristic tube due to gluon self-interaction (a); quarks are being pulled apart, the gluon tube is stretched, the energy stored in the gluon field grows (b); when the energy exceeds $2m_q$, a quark-antiquark pair appears (c).

The origin of color confinement can be qualitatively illustrated by imagining two interacting quarks being pulled apart (see Figure 2.2). The interaction between quarks is mediated by virtual gluons, and since they attract each other, the field lines are being "squeezed" into a narrow tube. The energy density stored in such a tube is constant, while the total energy increases linearly with distance. At some point, the energy grows high enough to produce a quark-antiquark pair out of vacuum, resulting in two colorless mesons.

The color confinement also results in the coupling of the strong interaction at long distances or, equivalently, at low interaction energies, being larger than one. This introduces serious calculation difficulties: it is not possible to write down a series expansions in the powers of the coupling α_s, since such series will not converge. QCD is said to be *non-perturbative* in this regime.

Conversely, at high energies, or at very close distances, the coupling of the strong interaction diminishes, eventually leading to a regime called *asymptotic freedom*, where with high precision the quarks in question can be regarded as non-interacting.

2.2.1 Important aspects for proton-proton collisions

At the LHC, particles are produced in proton-proton collisions. Therefore to make a prediction for any process (such as the expected cross-section of Higgs boson production) one has to understand the proton-proton collision in which the particle was produced. Protons at low energies consist of three quarks: two u quarks and one d quark, which are called *valence* quarks. However at high collision energies at the LHC, the structure of the proton becomes more complicated. The gluons mediating interactions between valence quarks produce many quark-anti-quark pairs (called *sea* quarks). A further complication arises from the fact that cross-section calculation includes the regime in which QCD is non-perturbative. A schematic view of a proton collision is shown in Figure 2.3.

In the non-perturbative regime the *parton density functions* (PDFs) are used. They describe the proton structure on average, as probability densities to find a point-like constituent (quark or gluon, together referred to as parton) of a certain momentum fraction x at the energy scale q^2. Using the fact that the PDFs are universal, measurements from different experiments can be used. For the presented analysis, the `PDF4LHC15` [15] PDF set was used, which is based on a combination of `CT14` [16], `MMHT14` [17], and `NNPDF3.0` [18] PDF sets. The PDFs introduce a source of systematic uncertainty associated with the uncertainty in their measurement.

The choice of a cut-off scale between the perturbative and non-perturbative QCD regimes (*factorization* scale, μ_F), introduces another source of systematic uncertainty. But even in the perturbative

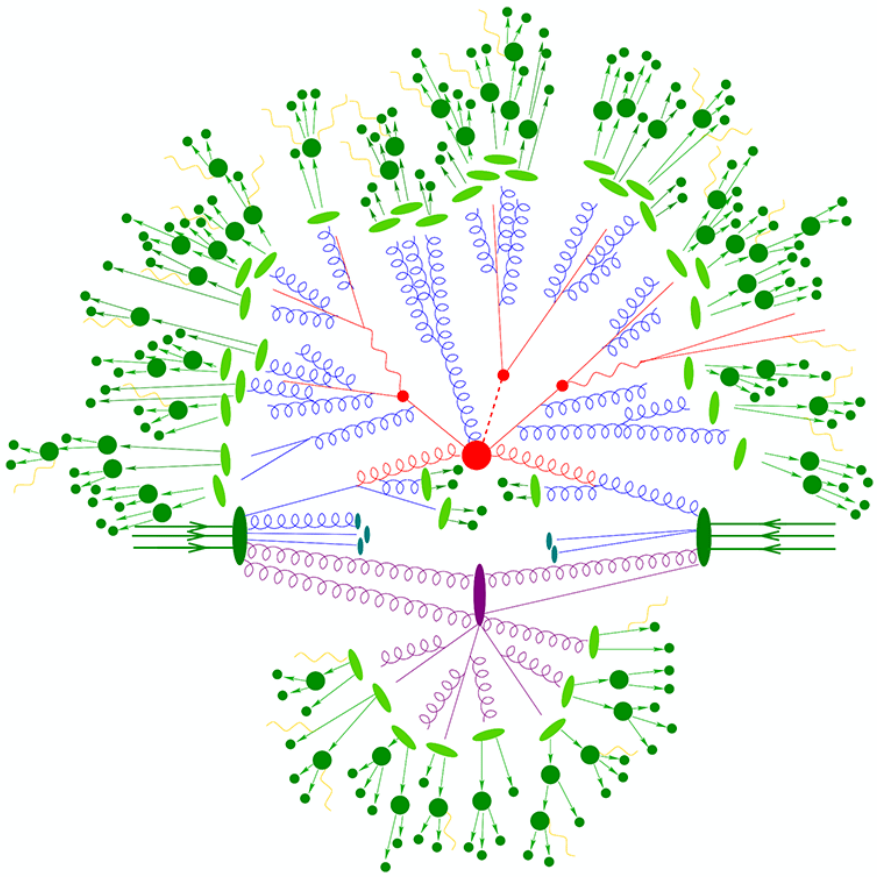


Figure 2.3: Schematic structure of a *pp* collision with a *hard-scatter* component (inelastic collision of partons in which heavy particles such as Higgs boson can be produced). Red circles and gluon lines at the center represent the initial hard-scatter component. Surrounding it are blue gluon lines of the subsequent QCD radiation. At the periphery *hadronization* (oval light green blobs) happens: color states turn into colorless hadrons, which produce more hadrons in downstream cascades (dark green). At the bottom another interaction (violet) originating from the same proton is shown – an *underlying event*. Finally the particles originating from other sources than the hard-scatter proton-proton collision (including previous and subsequent collisions) are contributing to the *pile-up*. Taken from Ref. [14].

regime, the calculation would be divergent due to large momentum in the loops of the Feynman diagrams representing the amplitude. This is solved by introducing another cut-off scale: *renormalization* scale, μ_R, and hence one more source of systematic uncertainty.

Finally, even in the perturbative regime, calculating at the fixed order in α_s misses the contribution of QCD *radiation*. Because of their self-interaction, gluons tend to produce large parton radiation cascades whose contribution is significant. Contribution of the QCD radiation is taken into account by modeling the *parton showers* (radiation cascades) by iteratively adding parton emissions to the event up to the scale of the non-perturbative QCD. This technique introduces yet another source of systematic uncertainty.

2.2.2 Monte-Carlo event generation

The calculation of the expected cross-section of some process (e.g., of Higgs boson production) can be factorized in two parts. The first part depending only on the universal PDFs, features QCD in the non-perturbative regime, but the PDFs can be experimentally measured as explained above. The second part, depending on the hard-scatter component, features QCD in the perturbative regime and can therefore be calculated. The cross-section calculation is then given by an integral of the following form:

$$\sigma_{pp\to X} = \sum_{a,b} \int dx_a dx_b f_a f_b \int d\Phi_X \frac{1}{2\hat{s}} |M_{ab\to X}|^2, \tag{2.7}$$

where a and b denote different partons, f_a and f_b are their corresponding PDFs. The $\frac{1}{2\hat{s}}$ is the parton flux factor, relating the LHC center-of-mass energy with the partonic center-of-mass energy. M denotes the matrix element, corresponding to the "amplitude" for the process in Feynman diagrams, and $d\Phi$ is the phase space element.

The integral in Equation 2.7 has to be calculated numerically. It is highly multi-dimensional due to integration over the whole phase space. Monte-Carlo (MC) methods are very well suited for such integrals. They are based on random sampling of the phase space, and converge to the true integral value with increasing number of sampled points. For the highly multi-dimensional integrals, MC methods typically converge faster than any other integration methods.

Due to particle physics processes being inherently stochastic, using the MC methods has an additional advantage: each sampled point in the phase space can be interpreted as one *generated event*. This means that it is possible, for example, to use phase space cuts on the generated event samples to simulate detector acceptance.

The MC generated event samples are obtained using dedicated software, called MC *event generators*. Different event generators use different approaches in the modelling of the pp interaction, and, as a result, some perform better in certain situations than the others (also different generators are often used as a cross-check).

It should be noted that the MC event generators do not simulate detector effects. The generated event samples contain products of the collision and decays of the short-lived particles at time scales where interaction with the detector material does not yet occur. Simulation of the ATLAS detector is described in Section 3.2.6.

2.3 Electroweak theory

Historically the electromagnetic and weak interactions were treated separately. Such separate treatment is made possible due to vastly different ranges of these interactions. The electromagnetic interaction, mediated by a massless photon, has infinite effective range. The weak interaction, mediated by massive bosons, is limited to the short ranges by the short lifetimes of these massive virtual bosons. The resulting effective range of the weak interaction is limited to subatomic distances.

The electromagnetic interaction is described by quantum electrodynamics (QED). The QED Lagrangian density, invariant under the U(1) transformation, can be written as follows:

$$\mathcal{L}_{\text{QED}} = \bar{\psi} i\gamma^{\mu} \left[\partial_{\mu} + iqA_{\mu}\right]\psi - m\bar{\psi}\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \tag{2.8}$$

where ψ is a 4-component (Dirac bispinor) field corresponding to a quark or a lepton, m is the mass of the particle in question, q – its electric charge, A_{μ} is the potential of electromagnetic field and $F_{\mu\nu}$ is the electromagnetic field tensor:

$$F_{\mu\nu} = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu}. \tag{2.9}$$

A_{μ} is a *vector* potential, as follows from the requirement to be invariant under the U(1) transformation. It corresponds to a single massless particle mediating the electromagnetic interaction, in which the electric charge is conserved. The mediating particle itself does not carry electric charge, hence there is no self-interaction term in Equation 2.9. These requirements are fulfilled by the photon, and therefore A_{μ} potential is often understood to refer directly to the photon.

The electric charge takes integer values ± 1 and 0 for the leptons, and fractions $\pm\frac{1}{3}$ and $\pm\frac{2}{3}$ for the quarks. However since freely propagating states are necessarily colorless in QCD, free states with non-integer electric charges are not observed as well.

The QED is a very precise approximation for electroweak interaction at low energies (large distances), while in the high energy (short distance) regime the combined theory for electroweak interactions is used, invariant under the $\text{SU}(2)_L \times \text{U}(1)_Y$ transformation.

Here the subscripts denote two important aspects of the electroweak theory. The subscript Y in $\text{U}(1)_Y$ refers to the conserved *weak hypercharge*, to stress that it is different from the U(1) symmetry group of QED conserving electric charge. The subscript L in $\text{SU}(2)_L$ refers to the fundamental fact of charge parity (CP) symmetry violation in electroweak interactions, which results in differences in the interaction depending if particles are left- or right-handed.

Approximately, left- or right-handedness of a particle can be understood as the following: "right-handed" if the direction of particle's spin vector is the same as the direction of particle's motion and vice versa. This definition is called *helicity*, but it is not Lorentz-invariant for massive particles, since one can always change the reference frame such that the helicity will flip. There is a related Lorentz-invariant quantity called *chirality*, and it is that quantity which is referred to when discussing left- or right-handed particles. The same applies for massless particles.

To describe the structure of electroweak interactions, fermions should be regarded in left-handed doublets and right-handed singlets. Focusing on the first generation of fermions, fields can be written as:

$$\psi_1 = \begin{pmatrix} u \\ d \end{pmatrix}_L, \ \psi_2 = u_R, \ \psi_3 = d_R \text{ (for quarks)}, \tag{2.10}$$

or

$$\psi_1 = \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \ \psi_2 = \nu_{eR}, \ \psi_3 = e_R \text{ (for leptons).} \tag{2.11}$$

The Lagrangian density can then be written as follows:

$$\mathcal{L}_{\text{EW}} = \sum_j \left(i\bar{\psi}_j \gamma^\mu D_\mu \psi_j \right) - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} - \frac{1}{4} W^i_{\mu\nu} W_i^{\mu\nu}. \tag{2.12}$$

Here D_μ is a covariant derivative, and how it acts depends on the field ψ_j. For the right-handed singlets ψ_2 and ψ_3 it depends only on the vector potential B_μ, while for left-handed doublets it also depends on the matrix potential W^i_μ. The structure of B_μ and W^i_μ follows from the $\text{SU}(2)_L$ and $\text{U}(1)_Y$ symmetries and implies four massless gauge bosons: W^1, W^2 (charged), and W^3 and B (neutral). The W^1 and W^2 bosons correspond to the observed states: $W^\pm$ bosons. However, unlike in QCD and QED, the neutral gauage bosons W^3 and B do not directly correspond to the observed states due to the *electroweak symmetry breaking*, which is discussed in the next Section. There is the following relation between the gauge bosons and the observed states:

$$\begin{pmatrix} W^3_\mu \\ B_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix}, \tag{2.13}$$

where θ_W is called the weak mixing angle. The conserved quantity in the electroweak interaction is the weak hypercharge Y. It is related to the electric charge Q in the following way:

$$Y_W = 2(Q - T_3), \tag{2.14}$$

where T_3 is a projection on the third axis of the vector quantity $\vec{T}$ called the weak isospin. T_3 is conserved in weak interactions. Because the $\vec{T}$ itself is not related to any conservation law, the mention of "projection on the third axis" is often omitted and "weak isospin" directly refers to T_3.

Apart from the mixing of gauge bosons, the description given above does not account for the masses of Z, W^+ and W^- bosons, which are in fact very large (around 91 GeV and 80 GeV, respectively). Adding corresponding mass terms into the Lagrangian density would violate the local gauge invariance and lead to the theory being non-renormalizible and thus to infinities in theory predictions. Instead, gauge bosons acquire mass through *spontaneous breaking* of the $\text{SU}(2)_L \times \text{U}(1)_Y$ symmetry by the Higgs mechanism. The Higgs mechanism is also responsible for masses of quarks and leptons in the SM through the Yukawa interaction. The Higgs mechanism leads to an emergence of an additional (Higgs) field and its particle manifestation – the Higgs boson.

2.4 Higgs mechanism

In order to produce mass terms in the Lagragian density corresponding to Z, W^+ and W^- bosons, the minimal Higgs model must have four degrees of freedom (the fourth corresponding to the Higgs boson itself) and therefore consists of two complex scalar fields, one of which is neutral and the other one charged:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \tag{2.15}$$

The corresponding Lagrangian density is

$$\mathcal{L} = (\partial_\mu \phi)^\dagger (\partial^\mu \phi) - V(\phi), \tag{2.16}$$

where † means the Hermitian conjugate, and $V(\phi)$ is the potential of the Higgs field:

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2. \tag{2.17}$$

The parameter λ should be greater than zero for the potential to have a minimum, while the sign of μ^2 defines location of the minimum, with $\mu^2 > 0$ corresponding to the minimum at zero, and $\mu^2 < 0$ to a non-zero minimum, in which case the field is said to have a *non-zero vacuum expectation value* v (shown in Figure 2.4).

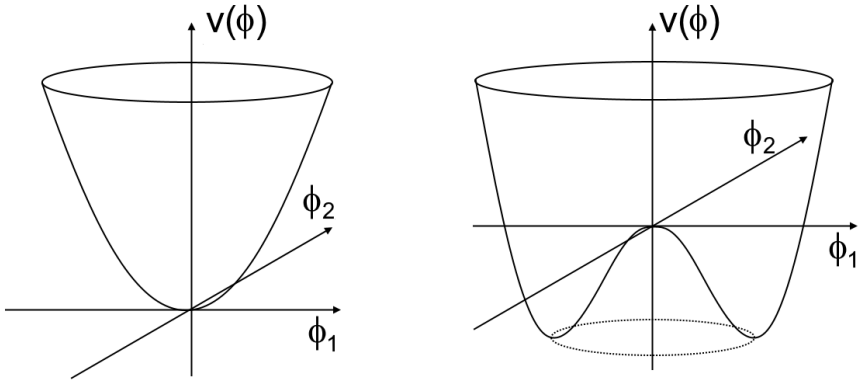


Figure 2.4: Complex scalar field doublet potential $V(\phi)$ for $\mu^2 > 0$ (left) and $\mu^2 < 0$ (right).

Non-zero vacuum expectation value for the Higgs field breaks the $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ local gauge symmetry, but it should be noted that the theory itself still retains local gauge invariance and is therefore renormalizible. It is said that the symmetry is *spontaneously* broken. The symmetry breaking is providing the necessary additional degrees of freedom for boson masses. However, since the neutral photon is massless, the non-zero vacuum expectation value is required only for the neutral scalar field ϕ^0. The fields then can be expanded around the corresponding minima:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1(x) + i\phi_2(x) \\ v + \eta(x) + i\phi_4(x) \end{pmatrix}, \tag{2.18}$$

where $\eta(x)$ is the linear term in the expansion of ϕ^0 around the minimum.

Written as such, the Higgs mechanism will give rise to three additional massless Goldstone bosons. The Goldstone bosons can be eliminated from the theory by rewriting Higgs potential in the *unitary gauge*:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}. \tag{2.19}$$

The resulting Lagrangian density is known as Salam–Weinberg model. Analyzing the Lagrangian

density and extracting the mass terms yields to the following relations for the masses of the bosons:

$$\begin{aligned} m_W &= \tfrac{1}{2} g_W v, \\ m_A &= 0, \\ \tfrac{m_W}{m_Z} &= \cos\theta_W, \end{aligned} \tag{2.20}$$

where m_W, m_Z and m_A are masses of W, Z and photon respectively, g_W is the $\mathrm{SU}(2)_L$ (weak) interaction coupling strength, and θ_W is the weak mixing angle.

The (trilinear) couplings of the Higgs boson to the vector bosons can be expressed as follows:

$$g_{HWW} = g_W m_W, \; g_{HZZ} = g_Z m_Z, \tag{2.21}$$

where HWW is the HW^+W^- interaction vertex, HZZ is the HZ^*Z interaction vertex, and $g_Z = g_W / \cos\theta_W$.

2.4.1 Fermion masses

Assuming that the Higgs field couples to fermions, the corresponding term in the Lagrangian density will look like (taking $\mathrm{SU}(2)_L$ doublet containing the electron e and electron neutrino ν_e as an example):

$$\mathcal{L}_e = -g_e \left[\begin{pmatrix} \bar{\nu}_e & \bar{e} \end{pmatrix}_L \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} e_R + \bar{e_R} \begin{pmatrix} \phi^{+*} & \phi^{0*} \end{pmatrix} \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L \right], \tag{2.22}$$

where g_e is the coupling of the electron to the Higgs field, known as the Yukawa coupling. After spontaneous symmetry breaking, the Higgs potential in unitary gauge is given by Equation 2.19. The Lagrangian density thus becomes:

$$\mathcal{L}_e = -\frac{g_e}{\sqrt{2}} h(\bar{e_L} e_R + \bar{e_R} e_L) - \frac{g_e}{\sqrt{2}} v(\bar{e_L} e_R + \bar{e_R} e_L). \tag{2.23}$$

The first term in this Lagrangian density corresponds to a coupling between the electron and Higgs boson itself. The second term has the exact form for the electron mass, if the Yukawa coupling fulfils the following relation:

$$g_e = \sqrt{2} \frac{m_e}{v}. \tag{2.24}$$

The same logic applies for all other fermions (heavier lepton and quarks), with one caveat that for the upper component of an $\mathrm{SU}(2)_L$ doublet, e.g. for u quark in the $\binom{u}{d}_L$, the conjugate doublet ϕ_c is required. All fermion Yukawa couplings then follow similar relation:

$$g_f = \sqrt{2} \frac{m_f}{v}. \tag{2.25}$$

Since the Higgs field vacuum expectation value v is constrained by the W mass measurement (see Equation 2.20, v is found to be around 246 GeV), precise measurement of Yukawa couplings will settle the question if their masses are generated (only) by the Higgs mechanism.

2.5 Higgs boson

The Higgs mechanism, first invoked to explain W and Z masses and then also fermion masses, predicts the existence of an additional massive scalar particle – the Higgs boson – a quantum of the Higgs field. By construction, the Higgs boson couples to massive vector bosons W and Z and to fermions. The Higgs boson mass is a free parameter of the theory depending on Higgs boson potential and vacuum expectation value:

$$m_H = 2\lambda v^2. \tag{2.26}$$

The Higgs boson can decay to fermions via $H \to f\bar{f}$ for all kinematically allowed decay modes with $m_H > 2m_f$. The Higgs boson can also decay to photons, although only through a fermion or vector boson loop. For the vector bosons, the mass relation is not satisfied: $m_H < 2m_W < 2m_Z$. Nevertheless decays are still possible to vector boson pairs, where one boson is virtual (sometimes said that one boson is *off the mass shell*): $H \to ZZ^*$ and $H \to WW^*$. This thesis focuses on the Higgs boson decay to a low-mass lepton pair and a photon $H \to \ell\ell\gamma$ – a very rare decay (see Table 2.1 for comparison of the branching ratios).

2.5.1 Higgs boson at the LHC

Since particles at the LHC originate from proton-proton collisions, it is important to consider the Higgs boson production mechanism. At the LHC, Higgs bosons are most often produced in gluon-gluon fusion. However, another mechanism – vector-boson fusion – produces a final state with two jets, achieving a better signal over background ratio because the event can be tagged through these additional objects. It also introduces more kinematic constraints on possible values of momentum of the decay products. Other production modes include quark-initiated associated production with a W or Z boson or with two top quarks (ttH). These production modes are shown in Figure 2.5.

A particle consistent with the Higgs boson properties was found in 2012 by the ATLAS and CMS collaborations [7, 8]. In their papers, the ATLAS and CMS collaborations combined the data in many decay channels to reach the significance threshold of 5σ for the discovery. Data in the $H \to \gamma\gamma$, $H \to ZZ^* \to \ell^+\ell^-\ell^+\ell^-$, $H \to WW^* \to e\nu\mu\nu$, $H \to \tau\tau$ decay channels, and $H \to b\bar{b}$ associated production with a Z or W boson were used. The choice of the decay channels was motivated by the predicted SM branching ratios of the Higgs boson, as well as the experimental sensitivity in these channels. The ability to separate signal from background when reconstructing events from the decay products is the reason, for example, for the significant contribution of the very pure $H \to ZZ^* \to \ell^+\ell^-\ell^+\ell^-$ and $H \to \gamma\gamma$ channels into the combination, despite the relatively low branching ratio. The Higgs boson branching ratios in different channels and their experimental status is shown (for the $H \to \ell\ell\gamma$, before the results presented in this thesis) in Table 2.1.

The best measurement to date estimates the Higgs boson mass at 125.38 ± 0.14 [26]. The Higgs boson decay width is measured to be $\Gamma_H = 3.2^{+2.8}_{-2.2}$ MeV [27] (SM prediction for the total width is about 4 MeV). The Higgs boson cross-sections for the main production modes are summarized in Table 2.2.

The Higgs boson couplings to fermions and vector bosons are also constrained by the measurements at the LHC, shown in Table 2.3.

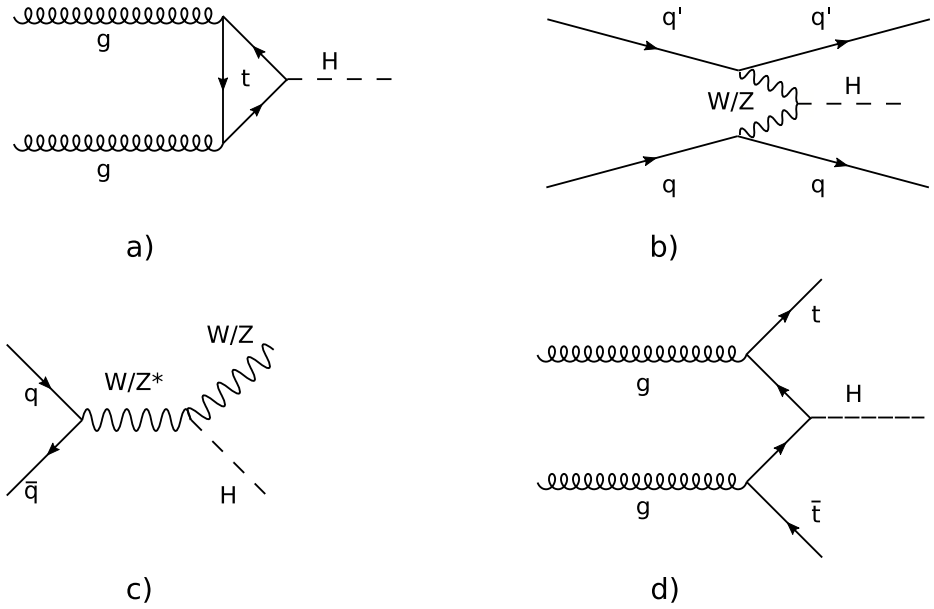


Figure 2.5: Main Feynman diagrams for Higgs boson production in pp collisions: gluon-gluon fusion (a), vector-boson fusion (b), production associated with a vector boson (c), and with two top quarks (d).

Decay channel	Branching ratio (SM prediction)	Signal strength (ratio of the observed signal yield to the SM expectation)
$H \to b\bar{b}$	0.5809	1.04 ± 0.13
$H \to WW^*$	0.2152	1.19 ± 0.12
$H \to \tau\tau$	0.06256	$1.15^{+0.16}_{-0.15}$
$H \to ZZ^*$	0.02641	1.01 ± 0.11 [19]
$H \to \gamma\gamma$	0.002270	$1.11^{+0.10}_{-0.09}$
$H \to Z\gamma$	0.001541	Limit: $3.6 \cdot \sigma_{\text{SM}} \times \mathcal{B}_{\text{SM}}$ [20]
$H \to \mu^+\mu^-$	0.0002171	$1.19^{+0.40\ +0.15}_{-0.39\ -0.14}$ (3.0σ evidence) [21]
$H \to \mu^+\mu^-\gamma$ (low-mass)	0.0000342	Limit: $4.0 \cdot \sigma_{\text{SM}} \times \mathcal{B}_{\text{SM}}$ [22]
$H \to e^+e^-\gamma$ (low-mass)	0.0000720	Limit: $7.62 \cdot \sigma_{\text{SM}} \times \mathcal{B}_{\text{SM}}$ [23]

Table 2.1: Higgs boson branching ratios and signal strength, defined as a ratio of the observed signal yield to the SM expectation, in different decay channels. For all processes, except the last two, the branching ratios are taken from Ref. [24] for $m_H = 125.09$ GeV, corresponding to the central value of the combined ATLAS and CMS measurement [25]. For the $H \to \ell\ell\gamma$ (low-mass) the branching ratios are estimated with `Pythia8` for $m_{\ell\ell} < 30$ GeV (see details in Section 4.3). Where not stated otherwise, the signal strength values are taken from Ref. [9] and the decay is observed with a statistical significance greater than 5σ. For the $H \to \mu^+\mu^-\gamma$ and $H \to e^+e^-\gamma$ channels, the experimental status shown before the results presented in this thesis (see Section 4.9).

H Production mode	σ [pb]	σ_{SM} [pb]
ggF	45 ± 4.0	44.7 ± 2.2
VBF	$4.25^{+0.84}_{-0.77}$	3.515 ± 0.075
WH	$1.57^{+0.48}_{-0.46}$	1.204 ± 0.024
ZH	$0.84^{+0.25}_{-0.23}$	$0.797^{+0.033}_{-0.026}$
ttH	$0.71^{+0.15}_{-0.14}$	$0.586^{+0.034}_{-0.049}$

Table 2.2: Higgs boson measured and predicted cross-sections for different production modes: gluon-gluon fusion (ggF), vector-boson fusion (VBF), production associated with a vector boson (WH, ZH), and with two top quarks (ttH) [28].

Parameter	Measured value
κ_Z	1.10 ± 0.08
κ_W	1.05 ± 0.08
κ_b	$1.06^{+0.19}_{-0.18}$
κ_t	$1.02^{+0.11}_{-0.10}$
κ_τ	1.07 ± 0.15
κ_μ	< 1.57 at 95% CL

Table 2.3: Higgs boson measured couplings to vector bosons and fermions [28], expressed in terms of coupling strength modifiers κ. The κ modifiers are based on the leading-order contributions to each production and decay process and relate the measurement to the SM expectation for a given production mode or decay process j: $\kappa_j = \sqrt{\sigma_j/\sigma_{\text{SM}}}$ or $\kappa_j = \sqrt{\Gamma_j/\Gamma_{\text{SM}}}$, respectively. For κ_μ an upper limit at 95% confidence level is reported.

2.6 Open questions in the Standard Model

The SM is a very successful theory, incorporating all known particle species. In all experiments to date, studying particles and their interactions, no significant deviation from the SM was found.

However the SM is known to be incomplete. For example, it does not include gravity. While gravity is incredibly weak and thus does not affect particle physics results at the achievable sensitivity, at the extreme energy densities, especially in the early Universe, gravity cannot be neglected and therefore adequate description cannot be provided by the SM. The best theory of gravity available today – General Relativity – is not a quantum theory and cannot be easily incorporated into the SM. The vast difference in coupling strengths for gravity and other interactions (e.g. weak force is about 10^{24} times as strong as gravity) is in itself a challenge (known as the *hierarchy problem*) which a potential unified theory must address.

A selection of open questions which could be potentially related to the Higgs mechanism is discussed below. For a more complete overview of problems in the SM, see Ref. [13].

Nature of dark matter

The SM does not readily provide a particle candidate for dark matter, abundance of which in the Universe is supported by multiple sources of evidence based on its gravitational interactions. Due to

its relationship with the mass of particles, it is possible that the Higgs boson couples to dark matter. Models of such scenarios are called Higgs portal models and are probed at the LHC, for example in searches for invisible Higgs boson decays. So far, there is no evidence supporting Higgs portal models.

Extended Higgs sector

The SM does not place a restriction on the exact form of the Higgs mechanism. The complex scalar field doublet given by Equation 2.15 defines the *minimal* Higgs mechanism. It can easily be extended, for example to contain two Higgs doublets, which is known as Two-Higgs-doublet (2HDM) model. It predicts two neutral scalar bosons h, H, two charged H^+, H^- and a pseudoscalar A. The discovered Higgs boson would be either h or H, and other heavier Higgs bosons could still remain undiscovered. To date, there is no experimental evidence supporting any extended Higgs sector models.

Neutrino masses

There is evidence that freely propagating neutrinos are changing flavors (this effect is called neutrino oscillations). This suggests that neutrinos have masses, albeit too low to be directly measured by modern experiments, current constraints put them on the order of an eV or less. If these masses are also provided by the Higgs mechanism, that would mean that neutrinos couple much less to the Higgs field than other fermions. In this case, neutrinos would also be Dirak particles, but since only left-handed neutrinos ν_L participate in weak interactions, the ν_R neutrinos cannot be detected. Alternative mechanism (so-called "seesaw") more readily accommodates small neutrino masses, however in that case neutrinos must be their own antiparticles (Majorana particles). One consequence of that would be a possibility of lepton-number violating decays, for example the so called neutrinoless double-beta decay, shown in Figure 2.6. No such decays have been observed to date.

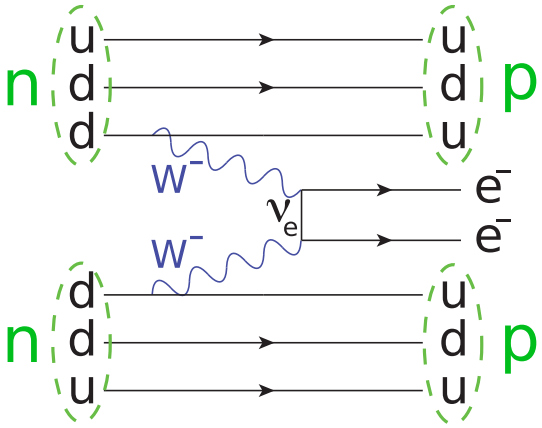


Figure 2.6: Feynman diagram of a neutrinoless double beta decay: simultaneous beta-decay of two neutrons to two protons and two electrons (and **no neutrinos**) in the final state, only possible if neutrinos are Majorana particles. Taken from Ref. [29].

CP violation

It is known that the charge conjugation and parity (CP) symmetry of the SM is violated. In particular, CP violation was observed in weak interactions. It was measured in weak interactions involving quarks, but the amount of CP violation found is not sufficient to account for observed matter density in our Universe (without CP violation, all matter and anti-matter would annihilate into radiation, so the amount of CP violation places constraints on the present state and evolution of the Universe).

Another source of CP violation in the SM could be in neutrino oscillations, where the current sensitivity is not yet enough for a precise measurement.

Yet another source of CP violation could be associated with the Higgs field, therefore it is crucial to study CP properties of the Higgs boson. CP properties can be studied in the Higgs boson production, i.e. in the interaction of quarks and bosons with the Higgs boson, for example in ttH [30] and vector boson fusion [31] production modes. With more data from LHC, it will also be possible to study the CP properties of the Higgs boson in its decays, such as the decay to a photon and two leptons presented in this thesis. More details about this measurement are given in Chapter 4.1.

CHAPTER 3

Experimental setup

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is located at the border between France and Switzerland near Geneva. With a circumference of around 27 km, it is the largest particle collider in the world. It collides proton beams, achieving a center-of-mass energy of $\sqrt{s} = 13$ TeV and instantaneous luminosities up to $2 \cdot 10^{34}\text{cm}^{-2}\text{s}^{-1}$. Such unprecedented conditions allow for abundant production and, consequently, detailed study of elementary particles. This includes very heavy short-lived particles, such as the W and Z bosons, as well as the Higgs boson, which was discovered at the LHC. Additionally short runs are performed with heavy ion collisions (lead-lead, proton-lead, and one run with Xe ions), where nuclear matter states, including the exotic quark-gluon plasma, are studied. The description of the LHC below focuses on its proton-proton operation mode. A more detailed description of the LHC is given in [32].

The LHC has two adjacent parallel vacuum beam pipes for oppositely directed proton beams. Eight arcs and straight sections form a closed loop. In the arcs, the beam pipes are surrounded by superconducting dipole magnets generating magnetic fields up to 7.7 T to direct the protons along the curved path. Straight sections contain radio frequency (RF) cavities, producing RF electric fields which accelerate the protons from 450 GeV to 6.5 TeV. Beam injection, beam dump and four beam interaction points (IPs), where the beams can cross, are also located in the straight sections. Quadrupole magnets are used for beam redirection during injection and dump, as well as for focusing in the IPs. A schematic view of the LHC is presented in Figure 3.1.

The main particle detectors of the LHC are: CMS [33], ATLAS [34], ALICE [35] and LHCb [36]. The LHCb detector is optimized to study heavy quark flavour physics, specifically b and c quarks and composite states containing them. The ALICE detector is optimized to study heavy ion collisions. The CMS and ATLAS are general-purpose detectors and are able to detect hadrons, photons, electrons and muons from the collision, covering almost all possible directions of the collision products (nearly 4π solid angle). This allows for a detailed reconstruction of events where short-lived particles, such as heavy bosons, are produced. The ATLAS and CMS detectors have comparable technical capabilities, but incorporate different design choices, therefore providing independent confirmation of each other's results.

Protons which enter the LHC are already pre-accelerated to 450 GeV. To achieve this, they follow the following acceleration chain:

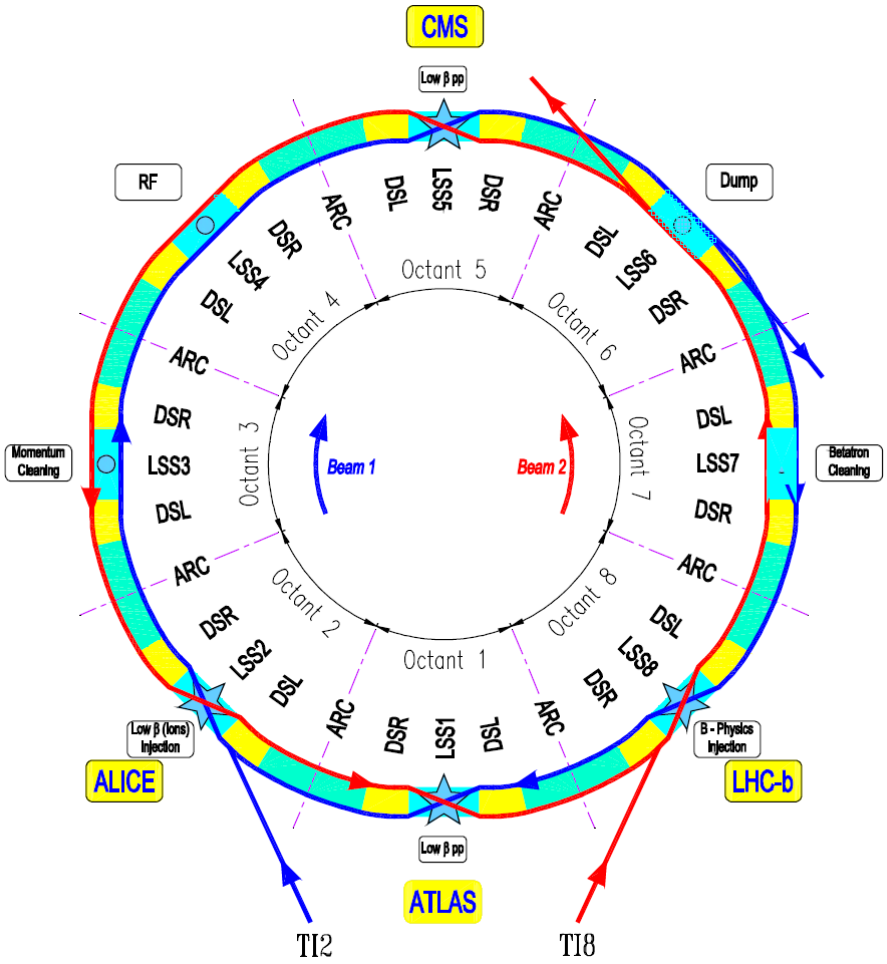


Figure 3.1: Sections of the LHC ring: 8 arcs containing dipole beam bending magnets and 8 straight sections containing RF acceleration cavities, beam injection points and interaction points, where the four main LHC experiments are located. Taken from Ref. [32].

- Hydrogen gas is passed though an electric field, which strips the electrons from H atoms. The resulting protons are accelerated by the linear accelerator Linac2 to an energy of 50 MeV. Protons are then accelerated by progressively larger and more powerful circular accelerators in three stages.
- The Proton Synchrotron Booster (PSB), with a circumference of 157 metres, accelerates the protons to 1.4 GeV.
- The Proton Synchrotron (PS), with a circumference of 628 metres, subsequently accelerates the protons to 26 GeV.
- The Super Proton Synchrotron (SPS), with a circumference of 6.9 km, further accelerates the protons to 450 GeV. At this stage protons are injected into the LHC.

A schematic view of the LHC and its acceleration chain is shown in Figure 3.2.

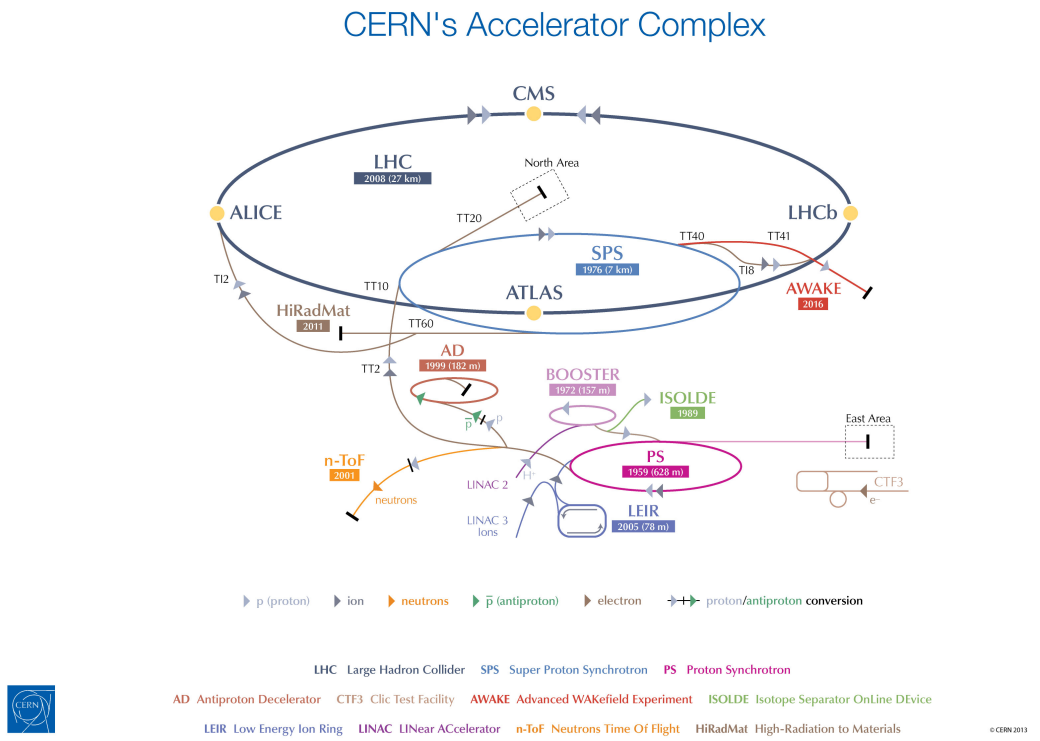


Figure 3.2: CERN's accelerator complex schematic, with the accelerator chain of the LHC (Linac2 → Proton Synchrotron Booster (PSB) → Proton Synchrotron (PS) → Super Proton Synchrotron (SPS) → LHC) and experiments shown. Taken from Ref. [37].

When protons are accelerated to their maximum energy of 6.5 TeV, they circulate in the LHC with the revolution frequency $f_{\text{rev}} = 11.245$ kHz. This corresponds to one rotation every $\sim$ 90 μs. Rather than a continuous stream, the protons are grouped in bunch trains with up to 2808 bunches consisting of approximately $1.15 \cdot 10^{11}$ protons each. The bunches are precisely spaced, such that each bunch

follows the other after 25 ns. Consequently, when beams are focused and directed to the IPs, collisions happen every 25 ns.

Despite the focusing, most bunch-bunch interactions at the IP occur with protons passing each other without an inelastic (or *hard-scatter*) collision, in which a large fraction of the proton's energy can be converted to mass and therefore heavy particles can be produced. To filter out these events, experiments use on-line triggering systems. For the ATLAS detector, it is described in Section 3.2.4.

In the events with hard-scatter, multiple other interactions occur (between different protons in the bunch) that produce detectable particles. Moreover, because readout time of some detectors is larger than 25 ns, particles produced in several preceding and following bunch-crossings can also influence the results. These effects are collectively referred to as *pile-up* and have to be taken into account by the experiments (they are a source of experimental uncertainty). The mean number of interactions per bunch crossing, as recorded by the ATLAS detector, is shown in Figure 3.3. The MC events are reweighted to match the pile-up profile measured in data for each data taking campaign.

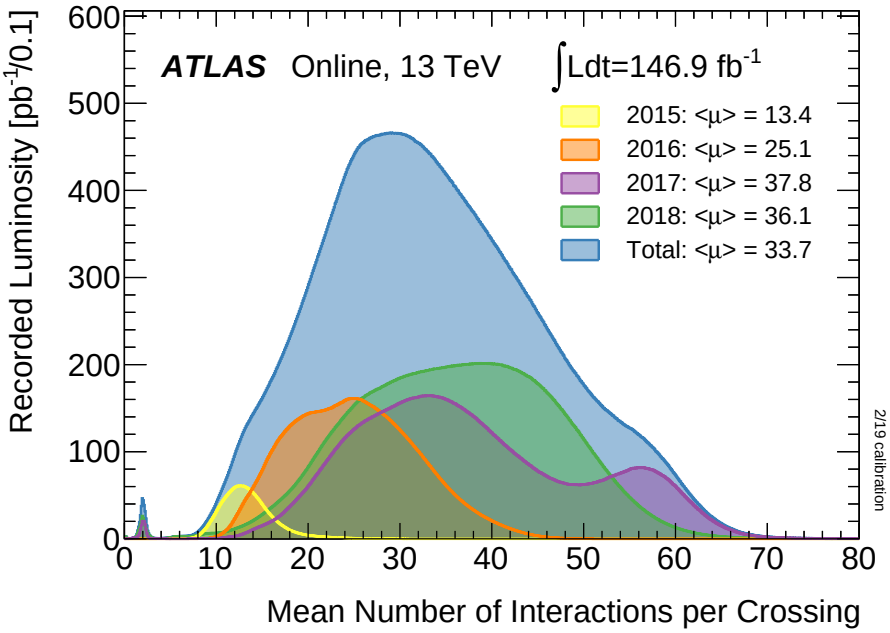


Figure 3.3: Mean number of interactions per bunch crossing recorded by the ATLAS detector in different data taking periods [38].

3.1.1 Luminosity

The number of expected events of interest N_{event} (for example, Higgs boson decays) depends on the process cross-section σ_{event}, as well as on the total (integrated) luminosity L_{int}:

$$N_{\text{event}} = L_{int} \cdot \sigma_{\text{event}}. \tag{3.1}$$

The instantaneous (per unit of time) luminosity L depends on the parameters of the colliding beams:

$$L = \frac{N_{\text{bunch}}^2 n_{\text{beam}} f_{\text{rev}} \gamma}{4\pi \epsilon_{\text{norm}} \beta^*} F, \tag{3.2}$$

where $N_{\text{bunch}} \approx 1.15 \cdot 10^{11}$ is the number of protons per bunch, $n_{\text{beam}} = 2808$ is the number of bunches per beam, $f_{\text{rev}} = 11.245$ kHz is the revolution frequency, γ is the Lorentz-factor given by:

$$\gamma = \frac{E_k + m_p}{m_p}, \tag{3.3}$$

where E_k is the proton's kinetic energy (6.5 TeV for the LHC during its second run) and m_p is the proton's rest mass, resulting in a Lorentz-factor of around 6930. $F = 0.836$ is the geometric luminosity reduction factor related to the crossing angle of the beams at the IP, ϵ_{norm} is the normalized (corrected for relativistic effects) transverse beam emittance and β^* is the beta function β at the collision point.

Together β and ϵ, taken at the IP, determine how particles of the beam are distributed, specifically their displacement from beam centre, x, and their deflection angle relative to the beam line, x'. In these coordinates, particles' distribution forms an ellipse with an area proportional to emittance ϵ and a ratio of short to long axis proportional to $\sqrt{\beta}$, as shown in Figure 3.4. At the LHC, the nominal values are $\epsilon_n = 3.75$ μm and $\beta^* = 0.55$ m.

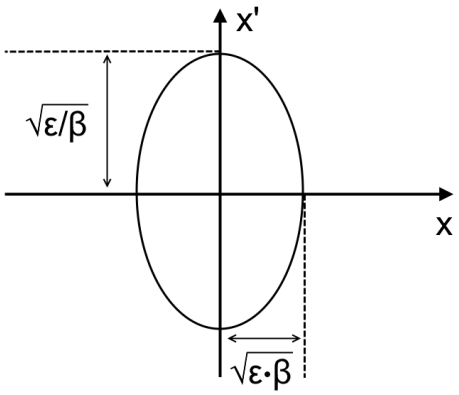


Figure 3.4: Correspondence between the beam parameters at the interaction point: emittance ϵ and beta function β define the ellipse of particles' distribution in the deflection angles x' and the displacement x.

Substituting all variables in Equation 3.2 with their nominal values, the resulting nominal luminosity is close to $L = 1 \cdot 10^{34}\text{cm}^{-2}\text{s}^{-1}$. However conditions vary during the fill, in particular the average number of protons per bunch N_{bunch} is maximal in the beginning of the fill and is reduced over time. In the beginning of some fills, the LHC managed to achieve luminosities up to $L = 2 \cdot 10^{34}\text{cm}^{-2}\text{s}^{-1}$.

Since the instantaneous luminosity is not constant during the run, it is crucial for the experiments to measure it to evaluate the integrated luminosity for the collected data. The luminosity measurement with ATLAS detector is described in Section 3.2.5. The total integrated luminosity delivered by the LHC during Run 2 data taking (2015-2018) at the IP of ATLAS detector amounted to 156 fb^{-1}. However, the luminosity corresponding to the time when the ATLAS detector was recording data was slightly less, at 147 fb^{-1}. Out of that, the 139 fb^{-1} correspond to high-quality data where all detector

systems operated nominally, and therefore this number is the most relevant for majority of the results, including the results presented in this thesis. The integrated luminosity as a function of time during Run 2 is shown in Figure 3.5.

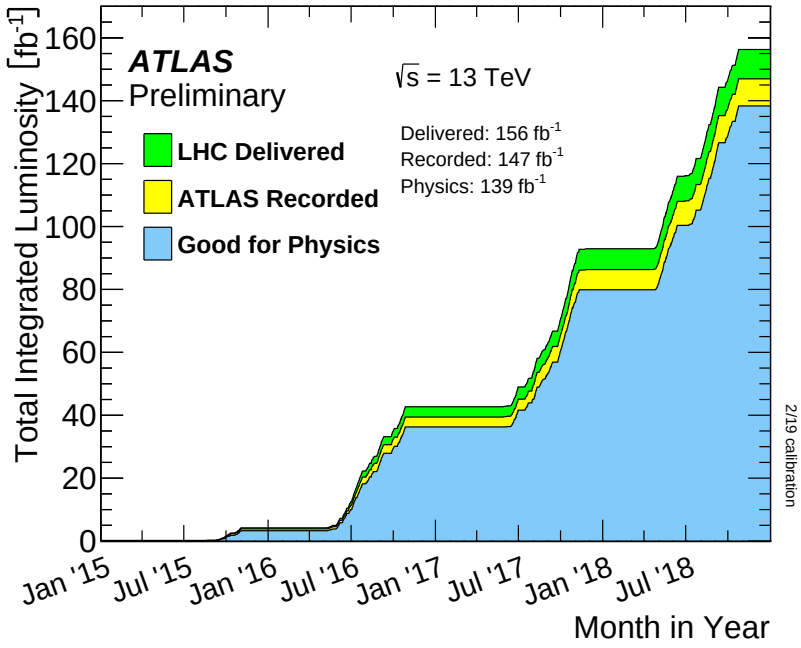


Figure 3.5: The integrated luminosity delivered by the LHC, the amount corresponding to the data recorded by the ATLAS detector, and the amount corresponding to the high-quality data recorded by the ATLAS detector in the Run 2 (2015-2018) data taking [38].

3.2 The ATLAS detector

The ATLAS detector covers nearly the entire solid angle around the IP (i.e. most of the possible directions of collision products) with sensitive detector systems. The presented description is based on Ref. [34], unless explicitly mentioned otherwise. The detector systems register signals from electrons, photons, muons and hadronic jets, which comprises nearly all (known) particle types sufficiently long-lived to reach the detection area. One notable exception are neutrinos; their presence in the event can only be inferred from a kinematic imbalance of the visible decay products in the plane transverse to the beam pipe.

The ATLAS experiment uses a right-handed Cartesian coordinate system centred at the nominal IP. The z-axis points along the beam pipe, the x-axis points from the IP into the LHC ring, and the y-axis points upwards. The azimuthal angle ϕ around the beam axis, and the polar angle θ from the beam axis are also often used. Instead of the polar angle θ, especially when discussing detector coverage, the pseudorapidity η is used. It is defined as $\eta = -\ln\tan(\theta/2)$ and approximates rapidity $y = 1/2 \ln\left[(E + p_z)/(E - p_z)\right]$ for light objects. The notion of pseudorapidity-azimuthal angle space is also often used, defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

The ATLAS detector is forward-backward symmetric and follows a general cylindrical arrangement, with barrel-shaped parts and discs (end-caps) on both ends. Several cylindrical detector systems are nested inside one another, forming a layered structure. The innermost part forms the inner tracking detector, subjected to a magnetic field of 2 T supplied by a solenoid magnet. The magnetic field bends charged particle trajectories, which provides information about their charge and momentum. The inner detector is surrounded by the calorimeter system, where most of the particles are stopped, and their energy is measured. Muons pass through the calorimeters without losing much energy, and are tracked by the surrounding muon tracking system, subjected to a magnetic field supplied by toroid magnets. A schematic view of the ATLAS detector and its subsystems is shown in Figure 3.6.

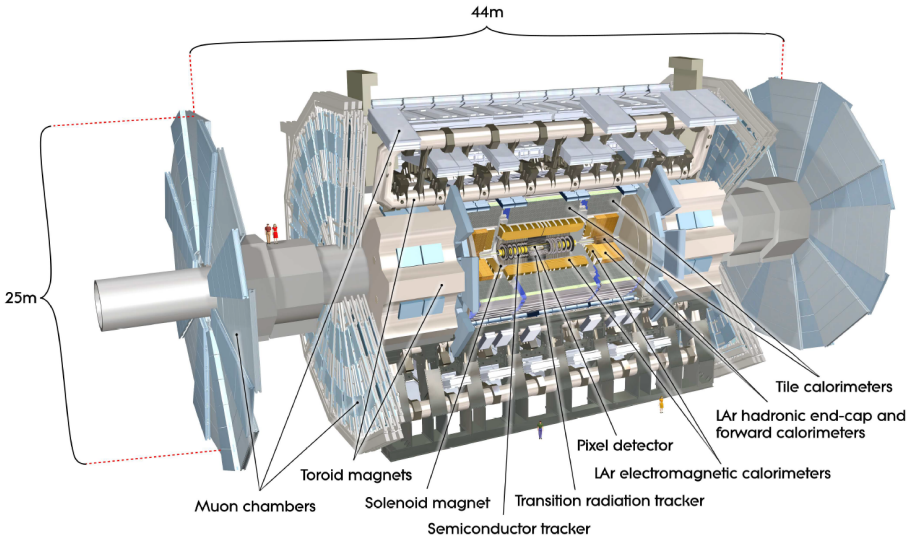


Figure 3.6: The ATLAS detector and its subsystems [34]. The inner tracking detector, composed of the pixel detector, the semiconductor tracker and the transition radiation tracker is nested inside the calorimeter system, composed of the electromagnetic and hadronic calorimeters. Surrounding it are muon chambers of the muon tracking system.

3.2.1 Inner detector

The inner detector provides tracking capabilities for pseudorapidities up to $|\eta| = 2.5$. Sufficient performance for physics results requires reconstructing the tracks of particles with momentum in the transverse plane (transverse momentum) as low as $p_{\mathrm{T}} = 0.5$ GeV and at the same time accurately measuring high-p_{T} tracks. These requirements necessitate high granularity and a large number of readout channels for the tracking detectors.

The inner detector is composed of three nested tracking detectors of different types: the pixel detector, the semiconductor tracker (SCT) and the transition radiation tracker (TRT). A schematic view of the inner detector is shown in Figure 3.7.

The pixel detector is composed of sensors with silicon modules ("pixels"), with a typical sensor measuring $50 \times 400\ \mu\mathrm{m}^2$. With $\sim$ 80.4 million readout channels, the pixel detector provides high-resolution tracking near the IP. The inside of a semi-conducting module is subjected to a high voltage difference, and passing particles cause a current impulse. The modules are organized in layers, with

three 2-layer discs in the end-cap and 3 layers in the barrel, covering 50.5 mm $< R <$ 122.5 mm distances from the beam. From the start of Run 2, the pixel detector was augmented with an insertable B-layer (IBL) at a radius of 33.3 mm [39, 40], adding another $\sim$ 12 million readout channels to the detector. The pixel detector provides up to four measurements (hits) per track in the barrel.

The semiconductor tracker (SCT) utilizes the same technology as the pixel detector, but is composed of daisy-chained silicon sensors (strips) with a strip pitch of 80 μm. The SCT detector has $\sim$ 6.3 million readout channels. The measurement of space points is provided by having two sets of strips: one running parallel to the beam (z-axis), providing measurements of the R and ϕ coordinates, and another one (stereo strips), rotated at an angle of $\pm$20 mrad with respect to z-axis, providing measurement of the z coordinate. Eight strip layers, covering 255 mm $< R <$ 549 mm in the barrel, provide up to four space points per track.

The transition radiation tracker (TRT) is composed of 4 mm diameter straw tubes filled with Ar-based gas mixture [41] (a Xe-based mixture was used in Run 1, and some of the straw tubes were still filled with it in Run 2). The TRT detector has 351 000 readout channels. Inside the tubes are anode wires, and a voltage difference is maintained between the cathode walls and the wires. Traversing particles ionize the gas, leading to a collection of charge on the anode and an induced current in the straw. In addition to that, charged particles with high Lorentz factors (most notably electrons) traversing the radiating foil between the tubes emit x-ray transition radiation which is also capable of ionizing the gas[1]. By having high and low signal thresholds in the front-end electronics, it is then possible to separate particle types based on the probability to exceed the high threshold. As a result, in addition to being a tracking detector, the TRT has electron identification capabilities. In the barrel it covers 563 mm $< R <$ 1066 mm and provides up to 36 measurements per track.

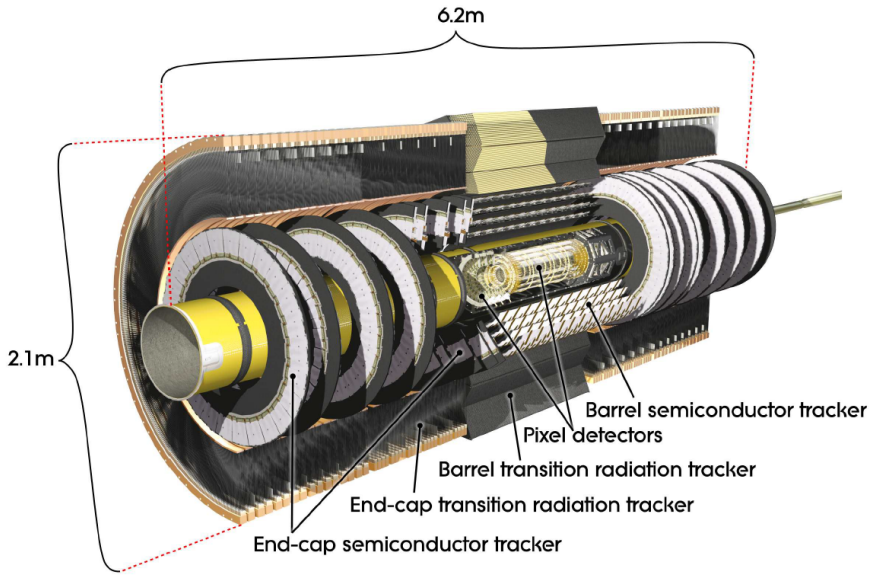


Figure 3.7: The inner detector of the ATLAS experiment [34].

[1] It should be noted that the ability to absorb the transition radiation photons is greatly reduced with Ar-based mixture compared to Xe-based.

Track and primary vertex reconstruction

Tracks are usually characterized by their curvature expressed as q/p, azimuthal angle ϕ and polar angle θ of the momentum vector $\vec{p}$, as well as impact parameters at the point of closest approach to the primary vertex (defined below): longitudinal z_0 and transverse d_0. Often also derivative quantities are used: $|\Delta z_0| \sin\theta$, where $\Delta z_0 = z_0 - z_v$ is the distance in z to the vertex, and significance of d_0, defined as $|d_0|/\sigma_{d_0}$, where σ_{d_0} is the uncertainty on the d_0 measurement.

Tracks are reconstructed from the hits in the inner detector. Track reconstruction starts with the formation of clusters, from which candidate tracks are formed using filtering and fitting algorithms.

Clusters are formed from hits in the pixel and SCT detectors. The SCT clusters are transformed into space-points by analyzing which crossed strips have hits. The TRT timing information is used to evaluate at which radius in the tube the hit occurred, producing so-called calibrated drift circles.

Track seeds are formed from a combination of space-points in the pixel layers and the first SCT layer. A pattern-recognition algorithm attempts to extend the track seeds to full tracks, containing at least seven hits in the pixel and SCT combined. In case of remaining tracks not satisfying these criteria, the pattern-recognition algorithm is run again under a larger energy loss hypothesis (targeted mainly at electrons due to their high bremsstrahlung energy loss [42]). Afterwards, possible ambiguities in cluster-to-track associations are removed by applying quality cuts, e.g. on the number of clusters associated to a track and the number of holes (i.e. where the track crosses a sensitive layer without leaving a hit).

Track candidates are then extended into the TRT, adding drift circles to them. Finally, the extended tracks are refitted and the fit quality is compared with silicon-only track candidates, and outlier hits (those reducing the fit quality) are marked. Additionally, backtracking is performed, starting instead with the TRT information and extending track candidates into the silicon detector. This can recover tracks from photon conversions (discussed in more detail below) and long-lived particles.

The vertex finding algorithm reconstructs the position of the primary vertex, where the main hard-scatter event originates. The vertex with the largest $\sum p_{\mathrm{T}}^2$ of the tracks is considered as the primary vertex.

3.2.2 Calorimeters

The inner detector is surrounded by the calorimeter system where most of the particles are stopped, i.e. lose nearly all of their energy traversing through the material. They produce cascades of secondary particles (showers) and their energy is measured, thereby measuring the energy of the original particles.

Calorimeters cover the pseudorapidity range $|\eta| < 4.9$. The calorimeter system consists of a fine-grained electromagnetic calorimeter for reconstructing and identifying electrons and photons, and a more coarsely instrumented hadronic calorimeter for reconstructing jets. Wide coverage in $|\eta|$ is important for measuring the missing transverse momentum $E_{\mathrm{T}}^{\mathrm{miss}}$ (kinematic imbalance[2]) in the event, which is necessary to identify events containing neutrinos and potential new weakly interacting particles (for example dark matter candidates). A schematic view of the ATLAS calorimeter system is shown in Figure 3.8.

The liquid-argon-based (LAr) EM calorimeter consists of layers of lead absorbers and sensitive volumes filled with liquid argon, arranged in an accordion shape such as to cover the full range in

[2] Vector sum of the momenta in the transverse (x–y) plane of all objects in the event, taken with a negative sign.

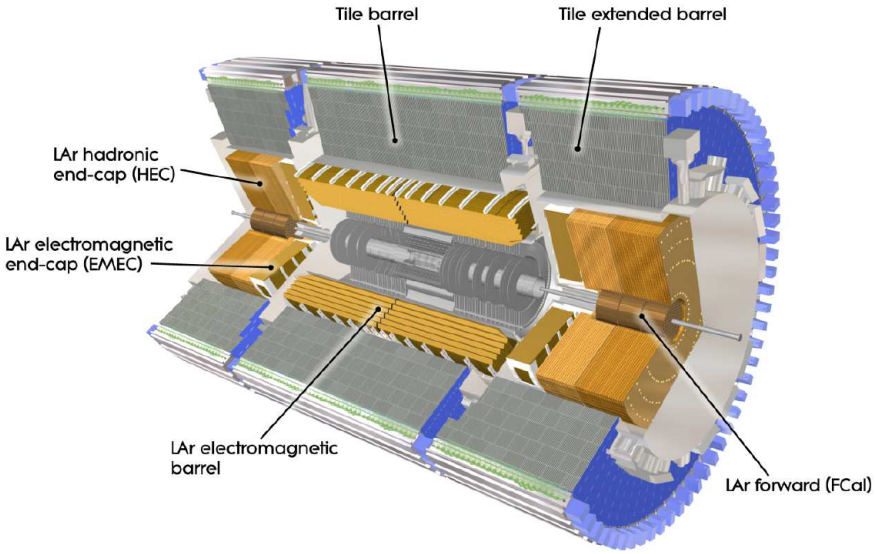


Figure 3.8: Calorimeter system of the ATLAS experiment [34].

ϕ. The particles hit lead plates and produce cascades of secondary particles, which enter the LAr volumes and ionize the argon. This ionization is subsequently detected by electronics. The energy deposited in the EM calorimeter is measured per layer and per cell in the η–ϕ plane; the cell structure is shown in Figure 3.9.

Hadronic calorimeters, surrounding the EM calorimeters, are mostly targeted at hadronic jets which tend to propagate further than electrons and photons. In the endcap ($|\eta| > 1.5$) and forward ($|\eta| > 3.1$) regions, they are based on the same LAr technology, except that copper and tungsten is used for absorber plates instead of lead. In the range $0.8 < |\eta| < 1.7$, the calorimeters are based on tile technology (the slight overlap with the LAr calorimeter in $1.5 < |\eta| < 1.7$ is by design). Tile calorimeters use steel as an absorber and scintillator as a sensitive medium, with photomultipliers to detect the optical signal.

Photon and electron reconstruction

In order to reconstruct photons and electrons (described in more detail in Ref. [43]), the first step is reconstructing their energy in the calorimeter. The signals from different calorimeter cells are combined to form topologically connected clusters (topoclusters). An algorithm is employed which combines neighbouring cells that have a high signal-to-noise ratio. After the initial clusters are formed they are analyzed for local maxima of deposited energy, and in case of multiple maxima they can be split (that would reflect the fact that they likely correspond to different particles). Photon and electron energy reconstruction uses only the EM energy of the topoclusters, i.e. energy from cells in the EM calorimeter.

While traversing the inner detector material, photons sometimes convert to e^+e^- pairs (in principle conversions to $\mu^+\mu^-$ are also possible but extremely rare due to the larger muon mass). Such photons are referred to as *converted* photons, and it is important to locate the conversion vertex because

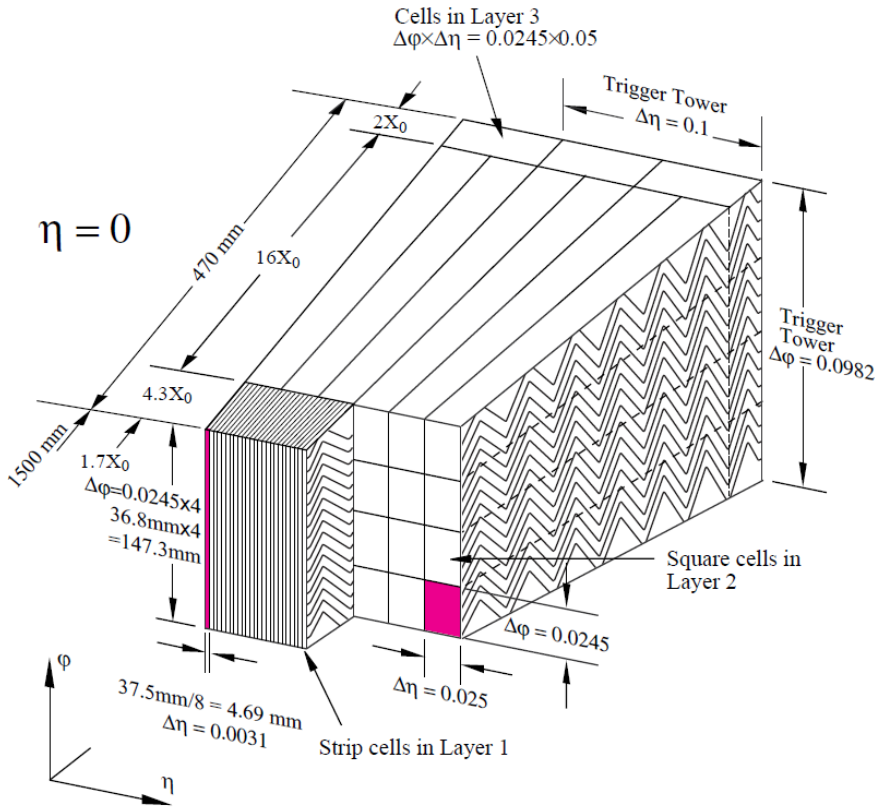


Figure 3.9: A sketch of the LAr calorimeter module illustrating calorimeter cells and how they are combined for triggering into trigger towers [34].

a converted photon will have a different shower profile in the EM calorimeter compared to an unconverted photon, which will affect both calibration and identification. The conversion vertex also provides additional directional information.

To reconstruct the conversion vertex, fixed-size (3×5 cells in the η–ϕ plane) EM calorimeter clusters and tracks loosely matched to them are used (more details are given in Ref. [43]). The conversion vertex is found by a combinatorial algorithm scanning through the tracks to find candidates consistent with a decay of a massless particle (a photon) inside the inner detector. Single tracks without hits in the innermost sensitive layers can also be attributed to a conversion vertex. To improve the efficiency of the attribution, the tracks must have a high probability to be electron tracks as determined by the TRT. In case of several candidate vertices, the ones with tracks in the following order are preferred: pairs of tracks with hits in the pixel or in both the SCT and the TRT; pairs of tracks reconstructed only in the TRT; single tracks. Within each category the vertex with smallest conversion radius (distance from the IP) is preferred.

In the context of this thesis, with highly collimated electrons (with typical separations $\Delta R(ee) < 0.1$), converted photons serve an important role of a model process. In particular, the close-by electrons are calibrated using the energy calibration derived for converted photons. At the same time, in the event selection, converted photons must be distinguished from the true $\ell^+\ell^-$ candidate (i.e. the one

originated from the IP), in particular to suppress the $H \to \gamma\gamma$ background where one photon is converted. More details are given in Section 4.4.

Electron tracks and photon conversion vertices are then matched to the topoclusters. In case of multiple matching tracks, tracks with pixel detector hits are preferred, and subsequently a track which is closer to the deposit in ΔR is preferred. After the tracks are refitted allowing for bremsstrahlung as described in Section 3.2.1, neighboring topoclusters can be combined to form *superclusters* [43], taking into account that these topoclusters can be attributed to an associated bremsstrahlung photon, or one leg of a conversion vertex, and therefore belong to the same final-state particle. Superclusters that are not associated to a good track serve as candidates for the unconverted photons, which do not leave tracks in the ID.

In the case of very close-by electrons originating from the interaction point, their respective energy deposits in the calorimeter may become merged into a single supercluster. This is the case for many events with close-by electrons relevant for this thesis. Close-by electron candidates are initially selected among objects where there are two (or more) tracks matching a single supercluster. Then a tighter selection is applied, described in Section 4.4.

The information from the tracking system, the calorimeter system, and quantities that combine both tracking and calorimeter information (e.g. characterizing track-to-calorimeter matching) is used to identify electrons and photons. For electrons, a likelihood-based identification algorithm [43] is used. The algorithm provides several working points depending on the desired quality of candidates for the analysis, e.g. it may be possible to select a smaller number of high likelihood candidates (*tight* working point) or a higher number of lower-likelihood candidates (*loose* working point). For the purposes of this thesis, *medium* electron candidates are used. Photon identification is based on sets of rectangular cuts (also defining a working point) on the variables characterizing the properties of the showers of secondary particles (shower shapes) in the calorimeter [44]. For the purposes of this thesis, *tight* photon candidates are used.

It is important that objects are well-isolated to suppress backgrounds from jets. This is achieved by defining *isolation* criteria. The track isolation is based on the sum of the transverse momenta of the tracks found within a cone of a set radius ΔR of the track or the photon cluster direction. Tracks matched to the candidate object itself (if it is an electron or a converted photon) are excluded from the calculation. The calorimeter isolation is based on the sum of the transverse energy within a cone of a set radius ΔR (typically larger than for the track isolation). From this energy, the candidate's own energy is subtracted. In the context of this thesis, the ΔR separation between electrons or muons is often so small that they contribute to each other's isolation cone. Therefore the isolation calculation must be corrected to subtract this contribution (see Section 4.5.2).

Finally, the energy of the electron and photon candidates must be calibrated. The calibration procedure is performed in several steps, described in more detail in Ref. [43]. First, the energy reconstruction is optimized using a multivariate regression algorithm trained on samples of simulated events based on the properties of the shower development in the EM calorimeter. Second, universal (for photons and electrons) correction factors for the energy scale are derived from the comparison of $Z \to ee$ decays in data and simulation. Finally, the universality of these corrections is verified based on comparisons of data and simulation for radiative Z decays (for photons), or $J/\Psi \to ee$ decays (for electrons).

Jet reconstruction

Jet reconstruction starts from topoclusters in calorimeter (described earlier). Jets are usually composed of both charged particles (such as pions) and neutral particles (such as photons). The charged particles leave tracks in the ID, and the *particle flow* algorithm (described in Ref. [45]) takes advantage of this fact. It starts with tracks satisfying quality requirements, and attempts to match each track to a single topocluster. It then computes the expected energy deposited by the particle in calorimeter based on the topocluster position and the track momentum. More topoclusters can be added at this stage if the probability that the particle deposited energy in more than one topocluster is high. Then the energy is subtracted from the matched topoclusters on cell-by-cell basis until all the expected energy is removed. The remnant cells are then removed from the topoclusters. The result of the procedure are modified topoclusters (matched to charged particle tracks) as well as unmodified topoclusters corresponding to neutral particles. This procedure significantly improves the jet energy resolution, in particular by reducing the pile-up contribution.

The topoclusters that are likely to have originated from the same initial parton are subsequently grouped together to form a jet. This is achieved using the anti-k_t algorithm [46] which treats each topocluster as corresponding to an energy deposit from a massless particle directed towards its barycenter, forming a four-vector. Starting from highest p_{T} topocluster, the algorithm combines topoclusters using a measure of distance between the four-vectors.

As with electron and photon candidates, the jet energy must be calibrated. The calibration procedure (described in Ref. [45]) involves correction for pile-up contamination, MC-based jet energy scale and resolution correction, and validation of corrections by comparing $Z \to \mu\mu$ events with a recoiling jet in data and MC.

After performing the reconstruction procedure described above, there could still be some jets originating from pile-up interactions. A multivariate discriminant called the jet-vertex-tagger (JVT) [47] is constructed based on variables related to tracks associated with the jets, in particular defining how likely these tracks are to originate from the primary vertex, in order to suppress pile-up jets.

3.2.3 Muon system

Muon trajectories are bent by the toroid magnetic field supplied by the large superconducting toroid magnets, and they are tracked by the muon system, providing coverage for $|\eta| < 2.7$. Several detector technologies are used with some of them focused on triggering muons, and others on tracking. A schematic view of ATLAS muon system is shown in Figure 3.10.

Over most of the η range, precision measurement of muon tracks is provided by monitored drift tube chambers (MDT). Drift tubes, 29.970 mm in diameter, are filled with an Ar/CO_2 gas mixture, with a voltage difference applied between the tube and an anode wire in the center. Muons ionize the gas, producing current in the readout. At large pseudorapidities, cathode strip chambers (CSC) are used instead due to their higher rate capability and better time resolution. They feature multiple anode wires in the single chamber and multiple cathode strips.

The muon triggering system is required to have good time resolution for its operation, and therefore proportional gas chambers with small gaps are used. The detectors of triggering system also compliment track measurement from the MDT in the track bending direction. The resistive plate chambers (RPC) each contain two parallel resistive plates 2 mm apart (without anode wires) subjected to high voltage. The gap is filled with a special gas mixture, which is ionized by passing muons. The thin gap

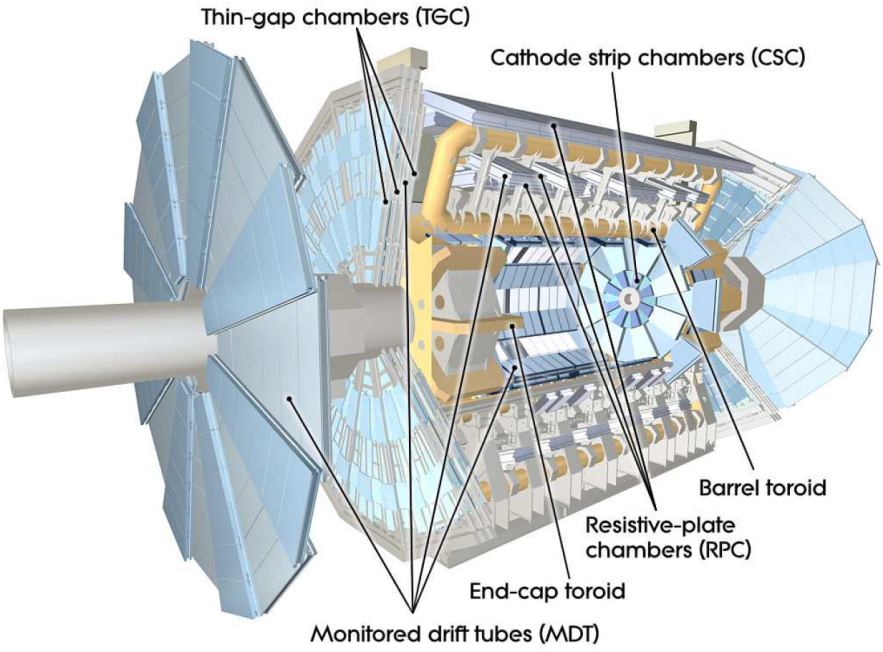


Figure 3.10: Muon system of the ATLAS experiment [34].

chambers (TGC) are multi-wire chambers with a wire-to-cathode distance (1.4 mm) smaller than the wire-to-wire distance (1.8 mm), which enables very good time resolution.

Muon reconstruction

Muon identification and reconstruction is described in detail in Ref. [48]. Muons may originate both from processes of interest (e.g. a Higgs boson decay), or from background processes such as decay of hadrons or heavy quarks (b or c). Muons from these background processes tend to feature a distinctive "kink" in the reconstructed track, and can usually be rejected by the goodness of fit and isolation criteria. Additional criteria are imposed on the number of hits on the track, depending on the desired quality of muon candidates for the analysis (as with photons and electrons, it is possible to select a smaller number of high-probability muon candidates, or a higher number of lower-probability candidates). For the purposes of this thesis, the *medium*-criteria muons are used, and they can be of the following two types:

- Combined: tracks are initially reconstructed independently in the muon system and the inner detector and matched. Then a fit is performed simultaneously for all hits of the matched tracks, and some of the hits may be rejected. It may happen at this stage that only the inner detector part of the track remains.

- Extrapolated: track is reconstructed based only on the hits in the muon system and the requirement to originate from the IP. These criteria are mainly used to extend muon reconstruction into the region $2.5 < |\eta| < 2.7$ not covered by the inner detector.

The muon momentum scale and resolution is then corrected based on a study of $Z \to \mu\mu$ and $J/\Psi \to \mu^+\mu^-$ decays in data and simulation [48]. As with electron and photon candidates, the isolation is calculated for the muon candidates to suppress backgrounds, most notably from b and c jets. As mentioned above, in the context of this thesis the isolation calculation must be corrected to subtract the contribution of close-by particles from each other's isolation cone (see Section 4.5.2).

Finally, the reconstructed muons that have poor momentum resolution are flagged, defined such that relative uncertainties on the combined track q/p measurement are larger (with some tolerance) than the corresponding ones from the inner detector and the muon system track fits. In principle such muons can be used for some studies, but in the present analysis events containing them are rejected.

3.2.4 Triggers

With collisions happening every 25 ns, or with a frequency of 40 MHz, it is not possible to reconstruct and record all events. Therefore events are filtered with the goal of keeping only the most interesting events from the perspective of the ATLAS physics programme, including events with W, Z or Higgs bosons, hadronic events with high momentum transfer, or events with large missing transverse momentum $E_{\mathrm{T}}^{\mathrm{miss}}$, to name a few examples.

The trigger system [49] applies event filtering on-line, reducing the event rate to a level which can be recorded. It uses fast algorithms to reconstruct the most prominent features of the event on the fly. The triggering system is composed of two levels: the first level (L1) reduces the rate to about 100 kHz. Among these events, the high-level trigger (HLT) performs a more detailed reconstruction and further reduces the event rate to an acceptable rate of up to $\sim$ 2 kHz.

The L1 trigger system is purely hardware-based due to strict latency requirements. It is using only muon tracking detectors and coarser calorimeter elements, combining multiple cells (trigger towers, shown in Figure 3.9). In the calorimeter (L1Calo), it aims to identify objects (electrons, photons, jets) with high E_{T}. In the muon system (L1Muon), it is searching for patterns of hits indicating high-p_{T} muon tracks. In addition, for Run 2 a new L1 topological trigger was added [49] that performs selections based on geometric or kinematic association between trigger objects from L1Calo and L1Muon. It also performs calculations of global event quantities such as missing transverse momentum $E_{\mathrm{T}}^{\mathrm{miss}}$. The L1 marks the areas in η, ϕ coordinates where such objects were identified, producing regions of interest (RoIs) which are then used by the HLT.

The software-based HLT performs fast reconstruction of tracks in the inner detector and muon tracker, as well as energies in the calorimeter. It is using information about RoIs from the L1 to optimize the reconstruction. The HLT triggers are targeted at specific signatures, e.g. single high-p_{T} objects, pairs of objects above certain p_{T} threshold, or a large $E_{\mathrm{T}}^{\mathrm{miss}}$ in the event. The rates of different triggers are combined into streams, related to a physics process. An example of HLT rates for a fill that reached the peak luminosity of $L = 2 \cdot 10^{34}\mathrm{cm}^{-2}\mathrm{s}^{-1}$ is shown in Figure 3.11.

In 2017 and 2018 data taking, a new HLT trigger was introduced, specifically targeted at events with close-by electrons. It increased the number of selected events for the presented analysis. More details are given in Section 4.5.1.

3.2.5 Luminosity measurement

In order to measure luminosity, elastic collisions are used, where deflections of protons are relatively small, and therefore they propagate at high η. Luminosity is measured in ATLAS using the

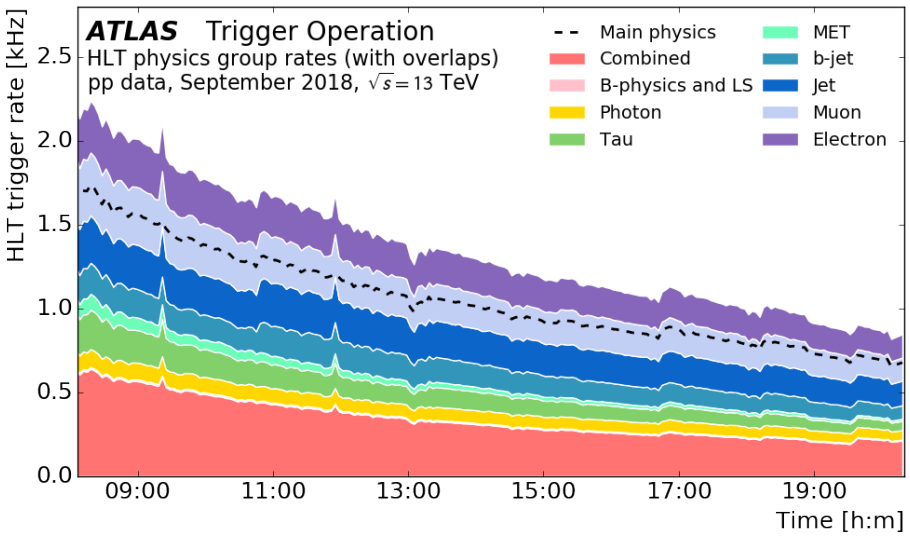


Figure 3.11: HLT rates for a fill that reached peak luminosity of $L = 2 \cdot 10^{34}\text{cm}^{-2}\text{s}^{-1}$ in September 2018, as a function of time [50]. Fractional rates in different streams are also shown.

LUCID-2 [51] detector, which is placed 17 m away from the IP at $|\eta| \sim 5.6$.

LUCID-2 is based on photomultipliers that detect Cherenkov radiation produced by particles passing through their quartz windows and in the quartz optical fibers. The number of interactions in the bunch-crossing is proportional to the number of particles detected by LUCID-2 and therefore LUCID-2 can only measure *relative* luminosity. The absolute instantaneous luminosity L can be expressed in terms of events per bunch-crossing μ and inelastic cross-section σ_{inel} as follows:

$$L = \frac{\mu f_{\text{rev}} N_{\text{bunch}}}{\sigma_{\text{inel}}}, \tag{3.4}$$

where μ is measured by LUCID-2 during running in 1-minute intervals, while σ_{inel} is measured in a separate calibration procedure using a method developed by van der Meer (referred to as "van der Meer scan"). During the van der Meer scan, the beams are swept across each other, and μ is measured, obtaining a distribution of μ values as a result (shown in Figure 3.12). The method uses the fact that from the standard deviation Σ and median μ_{max} of this distribution, the σ_{inel} can be extracted:

$$\sigma_{\text{inel}} = \mu_{\text{max}} \frac{2\pi\Sigma_x, \Sigma_y}{n_1 n_2}, \tag{3.5}$$

where Σ_x and Σ_y are extracted from the scans along orthogonal directions, and the beam current product $n_1 n_2$ is determined by the LHC.

The μ measurement from LUCID-2 is supplemented by measurements from different sources, such as counting tracks in the ID and measuring currents in the calorimeters. The consistency between these is used to determine the systematic uncertainty on the luminosity measurement.

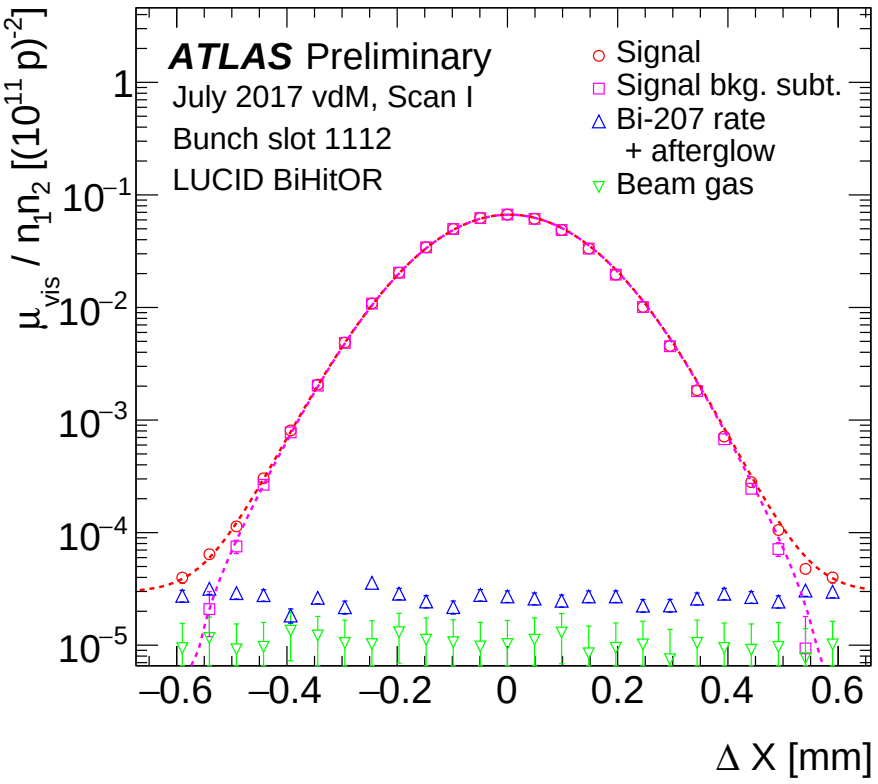


Figure 3.12: Van der Meer scan performed in 2017. Distribution of μ_{vis} ($\mu_{\text{vis}} = \varepsilon\mu$, where ε is detector efficiency) with respect to the relative displacement of the colliding bunches ΔX [52].

3.2.6 Detector simulation

The techniques of object identification and reconstruction from the detector data are discussed above in relation to particular detector systems. However it is not enough to have the reconstructed events. It is essential to create simulated events to obtain the expected result against which data can be compared, be it the expectation based on the Standard Model or based on a theory extending the Standard Model. Moreover, in some specific cases the simulation is also used to correct data, notably for energy calibration of objects in the event, based on comparisons of benchmark processes in data and simulation. Simulated events are also invaluable for detector development, in particular for studying new potential detector designs.

The theory prediction for pp collision resulting in the production of particles under study (for example, Higgs bosons) and their subsequent decay is independent from the detector due to very short time scales (and, consequently, lengths) involved, because there is no interaction with the detector material at this stage. This prediction is obtained using Monte-Carlo methods and is referred to as Monte-Carlo *event generation*. It is described in Section 2.2.2.

However, a prediction in the form of generated events is not directly comparable with the reconstructed data, because detector effects, such as limited detection efficiencies in different regions of the phase space, are not taken into account. To overcome this, the entire ATLAS detector is simulated.

The simulation is done in two main steps: simulation of particle propagation through the detector material, and simulation of the detector response.

In the ATLAS experiment, the simulation is done with a custom open source software called `Athena` [53]. It features a highly modular structure, including using external (developed outside of ATLAS) packages as modules. In addition to the full detector simulation, event generation (using external packages – event generators) and reconstruction (of both simulated samples and data) is also performed with `Athena`. It allows for seamless transition between outputs and inputs for different steps in the chain (employing custom file formats), and preservation of the original event information (*truth* information) in the simulated samples. Detector geometry and conditions are stored in a unified on-line database, which is accessed by simulation jobs in `Athena` at run-time. Detailed description of the ATLAS Simulation infrastructure is given in Ref. [54].

The simulation of particle propagation through the detector geometry is done with the `GEANT4` [55] package (used as an external module in `Athena`). `GEANT4` is a general purpose software using Monte-Carlo methods to simulate particle trajectories and energy loss in the form of energy deposits along the trajectories (referred to as G4 hits). `GEANT4` has the following features:

- It simulates many different types of particles, from elementary particles such as electrons and photons to heavy ions.
- It can operate in a wide particle energy range (the exact ranges depend on particle type), switching between different models approximating physics of the interaction for each range.
- It supports a wide range of materials through which particles can propagate, including the ability to construct user-defined materials with specific properties.

`GEANT4` operates on step-by-step basis. The particle direction and energy can change per step based on the encountered material (and decay rate for unstable particles). The step size itself is determined by the material properties (e.g., density) and particle properties, such as its energy. For each step, the energy loss is registered in the form of G4 hits. For the ATLAS detector, the G4 hits in the sensitive detector areas are saved for subsequent simulation of the detector response. Effects of pile-up are taken into account by adding G4 hits from additional simulated events without a hard-scatter component.

Simulation of the detector response (referred to as *digitization*) is highly dependent on detector type and therefore is not possible with a general-purpose software. There are separate modules in `Athena` corresponding to each detector system. These modules simulate specific detection technologies (e.g. silicon-based modules for the pixel and SCT detectors; gas tubes in the TRT, muon chambers) as well as performance of the electronics of the corresponding detectors. After digitization, the output has the same format (signals from detector readout) as the data, and is sent through the same reconstruction chain.

CHAPTER 4

Higgs boson decays to a low mass lepton pair and a photon

4.1 Overview and motivation

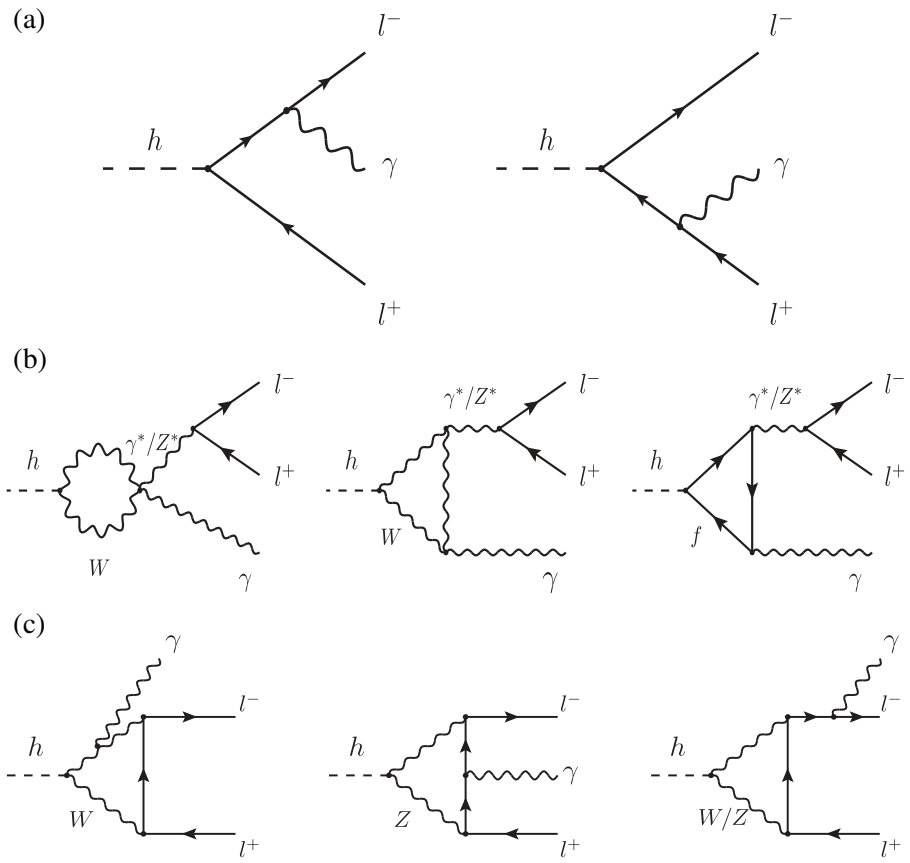


Figure 4.1: Dominant Feynman diagrams for the $\ell\ell\gamma$ final state of the Higgs boson decay [10]: tree-level (a), γ^* and Z pole one-loop (b), and type II (c).

Following the Higgs boson discovery in 2012, during Run 2 of the LHC (2015-2018), the focus has shifted to studying its properties, including the observation of its decays independently in individual channels. Since then, the Higgs boson was observed with a significance of more than 5σ in the $H \to b\bar{b}$, $H \to WW^* \to e\nu\mu\nu$, $H \to \tau\tau$, $H \to ZZ^* \to \ell^+\ell^-\ell^+\ell^-$ and $H \to \gamma\gamma$ channels. Properties of the Higgs boson have been measured with increasing precision, such as its mass, cross-section and its couplings to vector bosons and heavy fermions (discussed in Section 2.5). All measurements so far are compatible with the Standard Model (SM) expectations, however many of them are still dominated by the experimental and statistical uncertainties. In addition to gaining a better understanding of the SM Higgs boson, studying its properties is a way of testing for possible deviations from the SM. For example, limits have been set on potential anomalous Higgs boson couplings to gauge bosons [56], and a model-independent interpretation of the combined measurement of Higgs boson production and decay based on an effective field theory was performed [57].

The full Run 2 data set, with just under 140 fb^{-1} of high-quality data recorded by the ATLAS and CMS detectors each, also makes it possible to probe more rare Higgs boson decays. For example, the evidence threshold was reached for the $H \to \mu^+\mu^-$ decays [21] by the CMS detector. Another

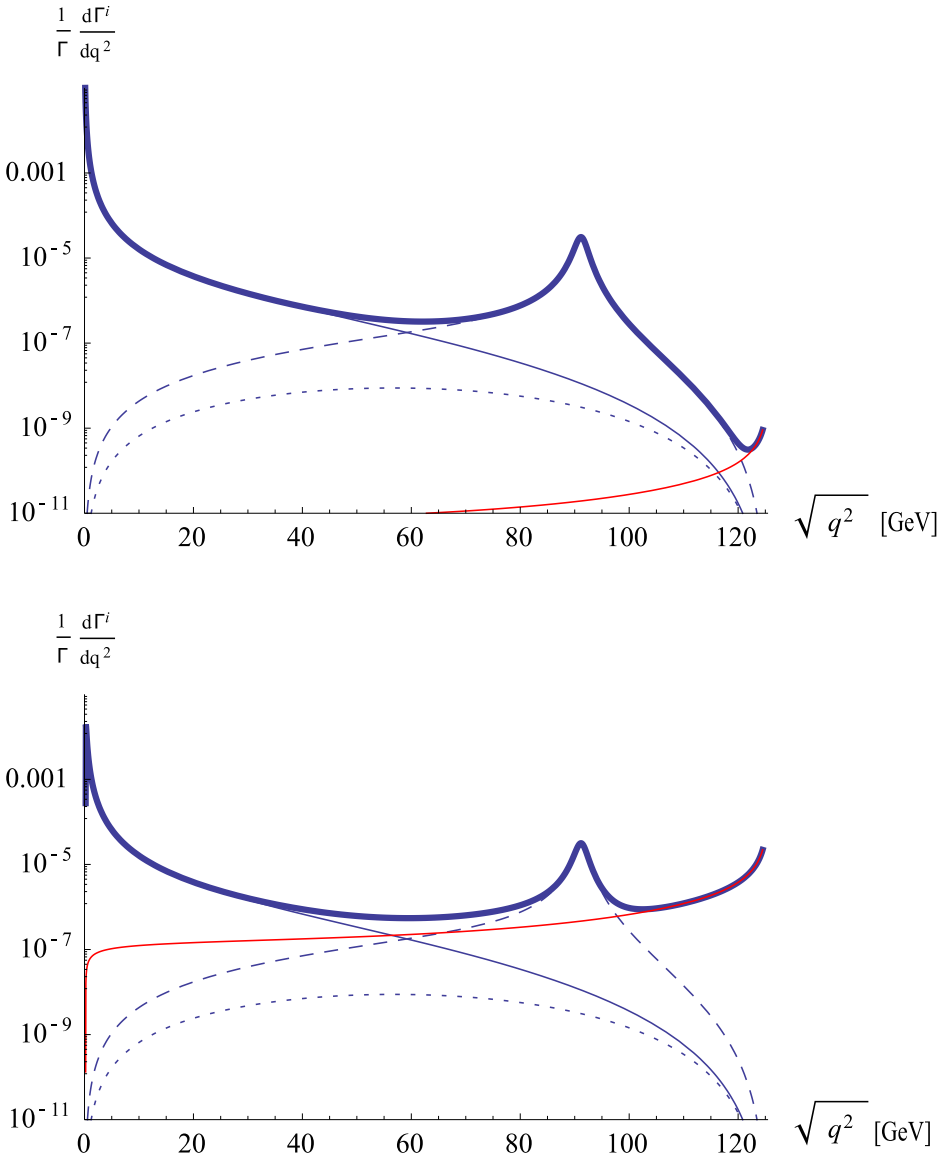


Figure 4.2: The dilepton invariant mass distributions of $H \to e^+e^-\gamma$ (top) and $H \to \mu^+\mu^-\gamma$ (bottom) normalized to $\Gamma(H \to \gamma\gamma)$ [10]. The thin red solid line denotes the contribution of the tree level diagrams (including $H \to \ell^+\ell^-$ process). The thin blue solid line denotes the contribution from the γ^* pole diagrams. The dashed line denotes the contribution from the Z pole diagrams. The thick line gives the total contributions. The dotted line denotes the contribution from type-II diagrams. The diagram types are defined in Figure 4.1.

example is the three-body decay of the Higgs boson to the $\ell\ell\gamma$ final state. With three particles, two of which are oppositely charged, a reference frame can be defined in which it is possible to assess relative frequencies of the decays in different directions. In the dilepton rest frame, decays in the forward and backward direction can be used to measure a property called forward-backward asymmetry. Defined like this, the forward-backward asymmetry directly relates to the CP of the decay and can be used to probe the magnitude of CP symmetry violation in the SM [11].

There are several processes that contribute to the $\ell\ell\gamma$ final state. At tree-level, the $H \to \ell^+\ell^-$ decay with a photon radiated from a final-state lepton (final state radiation) contributes. Decays through a Z boson ($H \to Z\gamma \to \ell^+\ell^-\gamma$) or a virtual photon ($H \to \gamma^*\gamma \to \ell^+\ell^-\gamma$) arise at the one-loop level. One-loop level Feynman diagrams with bosons and fermions in the loop (type II) also have a small contribution. The dominant Feynman diagrams contributing to the $\ell\ell\gamma$ final state are shown in Figure 4.1. The experimental status of searches for the above mentioned decays is presented in Table 2.1.

All of these decays contribute to the $H \to \ell^+\ell^-\gamma$ branching ratio calculation, but a good separation can be achieved by selecting events only in a specific range of the dilepton invariant mass. $H \to \gamma^*\gamma \to \ell^+\ell^-\gamma$ decays tend to have events with a low dilepton mass, while $H \to Z\gamma \to \ell^+\ell^-\gamma$ decays are concentrated around the Z peak of 91 GeV. $H \to \ell^+\ell^-$ decays with a final-state photon tend to have a higher dilepton mass above the Z peak. The contribution from type-II diagrams is small everywhere, and especially in the low- and high-m_{ll} regions. The contribution of the different diagrams to the decay rate as a function of the dilepton invariant mass is shown in Figure 4.2.

In this thesis, the decay of the Higgs boson to a low-mass lepton pair ($m_{\ell\ell} < 30$ GeV) and a photon is studied with the full Run-2 data set from the ATLAS detector. In this regime, the decay proceeds almost exclusively through a virtual photon $H \to \gamma^*\gamma \to \ell^+\ell^-\gamma$, which provides complementarity to the $H \to Z\gamma \to \ell^+\ell^-\gamma$ search. However, the low dilepton mass poses a unique challenge for the analysis, because in this regime the leptons tend to be highly collimated which leads to overlapping energy deposits in the calorimeter for the $ee\gamma$ channel. Among previously published studies by the CMS Collaboration [22, 23], the latest result reported an exclusion limit between 4 and 5 times the SM cross-section times branching ratio using only the $\mu\mu\gamma$ channel and a portion of the Run-2 data set (35.9 fb^{-1}). In the present study, which is the first such publication by the ATLAS Collaboration, both the $\mu\mu\gamma$ and $ee\gamma$ channels are used. A dedicated off-line identification algorithm was developed to identify dielectron pairs with overlapping calorimeter clusters. A special trigger was also employed during data taking to increase the number of candidate events.

4.2 Analysis strategy

The analysis is performed with a data set corresponding to 139 fb^{-1} of high-quality pp data at a center-of-mass energy $\sqrt{s} = 13$ TeV recorded by the ATLAS detector in Run 2. The samples of events simulated with Monte-Carlo methods (MC samples, listed in Section 4.3) are used to model the SM prediction for the signal $H \to \gamma^*\gamma \to \ell^+\ell^-\gamma$ process, to aid in estimation of the background sources and to develop a custom algorithm to identify overlapping dielectron signatures in the calorimeter (merged electron ID).

Candidate events and objects satisfying quality criteria are selected, with additional selections applied to increase the signal-to-background ratio and, consequently, the sensitivity. Candidate signal events are separated into three channels depending on the lepton content: muon, resolved electron and

merged electron channels. The latter two both constitute the same $ee\gamma$ physical final state, but when electron energy deposits merge in the calorimeter due to the close proximity of the electron trajectories, only a single electron object is reconstructed by the standard reconstruction algorithms. Therefore, in the merged electron channel, events are initially selected with one electron object and multiple associated tracks, and are then identified using the merged electron ID, described in Section 4.4.

Candidate events are further separated into three categories (with three channels, a total of nine categories): VBF-like, targeting vector boson fusion mode of Higgs boson production, which has a higher signal-to-background ratio; high-$p_{\text{TThrust}}^{\ell\ell\gamma}$ [1], targeting the high transverse momentum of the $\gamma^*\gamma$ system; and low-$p_{\text{TThrust}}^{\ell\ell\gamma}$, with the rest of the events. More details are provided in Section 4.5.

The analysis of the selected events is carried out in the $m_{\ell\ell\gamma}$ distribution, where the dominant non-resonant background processes are expected to follow a smoothly falling function, while signal forms a characteristic peak around 125 GeV. The signal model, described in Section 4.6, is constructed using $H \to \gamma^*\gamma \to \ell^+\ell^-\gamma$ events from MC samples and fitting their $m_{\ell\ell\gamma}$ distribution with a double-sided Crystal Ball function.

The background model (also in Section 4.6) is constructed from selected events in data by fitting their $m_{\ell\ell\gamma}$ distribution with a smoothly falling function chosen from a list of simple, low-degree-of-freedom functions. The signal peak range in data between 120 and 130 GeV is kept blinded for the expected results. After unblinding of the data, the full spectrum is fitted using a signal-plus-background model to obtain the observed results. In the electron channels, a small resonant component is added to the background model representing the $H \to \gamma\gamma$ process where a photon undergoes conversion into an electron pair. Its estimation is based on the resonance shape from the signal model, scaled to the expected $H \to \gamma\gamma$ yield estimated from MC samples.

To control for a possible bias associated with the fit function choice for the non-resonant background, a background-only template is constructed using MC samples of events in the $\ell\ell\gamma$ final state (without the H resonance), combined with a data-driven estimate of backgrounds where one or more objects are mis-identified or mis-reconstructed. This template is then fitted with the previously constructed signal-plus-background model. Any events classified as signal by this fit are *spurious*, and hence a measure of the potential bias in reporting signal yields, since no signal was injected into the template. Spurious signal criteria are used to determine the best fitting function for the background in each category, and any residual bias is taken into account as a systematic uncertainty.

The analysis is dominated by statistical uncertainty. Among systematic uncertainties, the uncertainty associated with spurious signal is dominant. Many other sources of systematic uncertainties with smaller contribution are taken into account, the details of which are given in Section 4.7.

Finally, the log-likelihood ratio method is used to characterize the excess of observed events over the background-only hypothesis. In addition, the Higgs boson cross-section times branching ratio in the fiducial region with $m_{\ell\ell} < 30$ GeV is measured. The statistical analysis procedure is described in Section 4.8, and the results and their interpretation are presented in Section 4.9.

[1] $p_{\text{TThrust}}^{\ell\ell\gamma} = |\vec{p}_{\text{T}}^{\,\ell\ell\gamma} \times \hat{t}|$, where $\hat{t} = (\vec{p}_{\text{T}}^{\,\ell\ell} - \vec{p}_{\text{T}}^{\,\gamma})/|\vec{p}_{\text{T}}^{\,\ell\ell} - \vec{p}_{\text{T}}^{\,\gamma}|$. This quantity is highly correlated with $\vec{p}_{T}^{\,\gamma^*\gamma}$ but has better experimental resolution [58].

4.3 Simulated samples

The analysis relies on MC simulation to model the expected signal $H \to \gamma^*\gamma \to \ell^+\ell^-\gamma$ process and some background processes. The MC generated events (see Section 2.2.2) are taken through the full ATLAS detector simulation chain (described in Section 3.2.6), unless stated otherwise.

Estimation of a small resonant $H \to \gamma\gamma$ background is based purely on MC samples, while the main, non-resonant background is modelled as a simple functional form and estimated using a simultaneous signal-plus-background fit to the data. A background template is needed to choose the best fitting function and gauge a systematic bias in the estimation (see Section 4.6.2). It is constructed primarily based on the MC simulation of the non-resonant SM processes in the $\ell\ell\gamma$ final state.

MC simulation is also used for certain other studies and cross-checks, most notably in relation to the merged electron ID. Corresponding MC samples are described in their respective sections. Finally, some of the standard reconstruction techniques in ATLAS rely at least in part on MC simulation. For example, the energy calibration of photons and electrons is based on a measurement of $Z \to ee$ decays in data and subsequent comparison to MC simulation; details can be found in the corresponding references in Section 3.2.

4.3.1 Signal samples

In the event generation of the $H \to \gamma^*\gamma \to \ell^+\ell^-\gamma$ process, the Higgs boson mass was set to 125 GeV and its width to $\Gamma_H = 4.07$ MeV, corresponding to the SM prediction for this mass as given in Ref. [59]. The mass value is corrected in the final fit model to 125.09 GeV, corresponding to the central value from the combined ATLAS and CMS measurement [25]. The best-available theoretical calculations of the corresponding SM production cross-sections for $m_H = 125.09$ GeV were taken from Ref. [24]. The branching ratios $\mathcal{B}(H \to \mu^+\mu^-\gamma)$ and $\mathcal{B}(H \to e^+e^-\gamma)$ (given in Table 2.1) were estimated with `Pythia 8` [60] for $m_{\ell\ell} < 30$ GeV, which corresponds to the analysis phase space. When extrapolated to a common phase space, this estimation agrees with theory predictions in Ref. [61, 62] within 3%. As there are no calculated systematic uncertainties for these predictions, same uncertainty as for $\mathcal{B}(H \to Z\gamma)$ was assumed, which corresponds to 5.8%. This conservative value was used because it is the largest theory uncertainty on branching ratios among similar processes, namely $H \to Z\gamma$ and $H \to \gamma\gamma$, according to Ref. [24].

The Higgs boson production was simulated with `Powheg-Box v2` [63] at next-to-next-to-leading-order (NNLO) accuracy in QCD for the ggF production mode and NLO accuracy in QCD for VBF, WH and $qqZH$ production modes. The samples were subsequently normalized to the next-to-next-to-next-to-leading order (N3LO) cross section in QCD with NLO electroweak (EW) corrections for ggF, and NNLO in QCD with NLO EW corrections for WH, $qqZH$ and VBF (for VBF, the NNLO level in QCD is approximate). No dedicated samples were generated for the ttH, bbH and $ggZH$ production modes due to their small contribution, and these production modes were not considered for optimization of event selection and categorization. However, their contribution to the total expected yield was taken into account, as described below.

For the bbH production mode, acceptance is assumed to be similar to the ggF production mode. Therefore, to obtain bbH yields, ggF yields are scaled by a factor $\sigma_{bbH}/\sigma_{\text{ggF}}$ (σ_{bbH} is available at NNLO in QCD). Similarly, $ggZH$ is assumed to be similar to $qqZH$ and hence $qqZH$ yields are scaled by $\sigma_{ggZH}/\sigma_{\text{qqZH}}$ (σ_{ggZH} is available at NLO in QCD with next-to-leading-logarithm corrections). For the ttH, the procedure is somewhat more nuanced and involves an MC sample

simulated with `Powheg-Box v2` [63] at NLO accuracy in QCD, containing $H \to \gamma^*\gamma \to \ell^+\ell^-\gamma$ decays for lepton pair masses up to $m_{\ell\ell} < 10$ GeV. A correction of $\sigma_{ttH} \times \mathcal{B}$ in the the full $m_{\ell\ell}$ spectrum to this phase space region was calculated with `Pythia 8`. Since the analysis is performed for lepton pair masses up to $m_{\ell\ell} < 30$ GeV, the ttH yields are further scaled up by a factor $N_{\text{SR}}(m_{\ell\ell} < 30\text{ GeV})/N_{\text{SR}}(m_{\ell\ell} < 10\text{ GeV})$, where N_{SR} are combined yields from ggF, VBF, WH and $qqZH$ production modes.

The combined yields for the ttH, bbH and $ggZH$ modes obtained with such scaling amount to a few percent of the total yields (1-4% depending on the analysis category) and therefore do not have a decisive role in the selection optimization and creation of the signal model in the statistical analysis. A conservative 50% uncertainty was assumed for these yields to cover the assumptions about similarity in acceptance with respect to the production modes and kinematic regions on which the estimates are based.

`Pythia 8` was used for simulation of the Higgs boson decay as well as parton showering, final state radiation, hadronisation and underlying event. The decay only via $H \to \gamma^*\gamma \to \ell^+\ell^-\gamma$ was simulated, while any interference with other processes in the final state, such as $H \to Z\gamma \to \ell^+\ell^-\gamma$ or $H \to \ell^+\ell^-$, was ignored. This is justified by the fact that these processes have a negligible contribution in the $m_{\ell\ell} < 30$ GeV region used for the analysis, as shown in Figure 4.2. The contribution from pile-up was also simulated with `Pythia 8` and added to all simulated events at the detector simulation stage, as described in Section 3.2.6.

To take into account systematic uncertainties in parton showering, the Higgs boson decay simulation was also performed with `Herwig 7` [64]. Since no MC sample of the $H \to \gamma^*\gamma \to \ell^+\ell^-\gamma$ process with `Herwig 7` was available, the estimation was performed with a $H \to \gamma\gamma$ sample instead. Several kinematic variables were compared: number of jets in the event, p_{T}^{H} and $p_{\text{T}}^{\text{lead jet}}$ (highest p_{T} among jets). Based on this comparison, binned weights were calculated and applied to `Pythia 8` simulated events such that they match kinematics of `Herwig 7` events. These reweighted events were then used to calculate the uncertainty, described in Section 4.7.

4.3.2 Background samples

For the resonant $H \to \gamma\gamma$ process, where a photon undergoes conversion into an electron pair, only ggF and VBF production modes were considered, because the contribution of other production modes in the analysis phase space is negligible. The samples were simulated with the same software and at the same accuracies as the signal process. The same mass and width of Higgs boson were used as well. The cross sections for $m_H = 125.09$ GeV were also taken from Ref. [24]. In the background model, only yields from $H \to \gamma\gamma$ samples were used, while shape of the $m_{\ell\ell\gamma}$ distribution was assumed to be the same as for signal; more details are given in Section 4.6.1.

The non-resonant SM processes in the $\ell\ell\gamma$ final state were simulated with `Sherpa` [65] program with LO accuracy in QCD for number of jets in the event from 0 up to 3. To produce a reliable estimate of these processes, a sample with high statistical power (i.e. with many simulated events) is needed. In order to reduce computational inefficiency, a phase space similar to the one used in analysis was required in event generation. Over 30 million events in the electron channel and over 20 million events in the muon channel were produced. Taking such a large number of events through the full detector simulation chain would come at a prohibitive computational expense. A hybrid approach was used, where $\sim$ 1.5 million of the events were fully simulated in each channel, and differences in several key kinematic distributions (excluding $m_{\ell\ell\gamma}$) with respect to generator-level distributions were

extracted. Then, generator-level events were reweighted to match the fully-simulated distributions. The validity of this procedure is demonstrated by the fact that, although no reweighting was applied to the $m_{\ell\ell\gamma}$ distribution, a good agreement between the fully-simulated and the reweighted generator-level distributions is observed.

4.4 Close-by electrons

In the low-dilepton mass regime ($m_{\ell\ell} < 30$ GeV) probed in this analysis, the leptons tend to be highly collimated. For the electrons, this often leads to merging of the calorimeter energy deposits of the respective electrons, and their reconstruction as a single electron associated with multiple tracks using standard reconstruction methods. With standard reconstruction having failed to identify the two electrons, these events would be lost for the analysis. The ability to recover such electrons (referred to as merged electrons) can significantly increase the number of selected Higgs boson events in the $ee\gamma$ channel, therefore a special identification algorithm was developed for this purpose. This improvement more than doubles the number of selected $ee\gamma$ events (event yields are given in Table 4.10).

The identification of merged electrons starts with picking a suitable candidate among the electrons (as identified by the standard reconstruction) in the event with multiple associated tracks. The electrons are required to have $p_{\mathrm{T}} > 20$ GeV, $|\eta| < 2.37$ and have a ratio of E_{T} in the hadronic calorimeter to the E_{T} of the EM cluster less than 0.1. The electrons are rejected if they are classified by the standard reconstruction algorithm as highly likely to originate from converted photons. Tracks are required to have $p_{\mathrm{T}} > 0.5$ GeV, $|\eta| < 2.50$ and at least 7 hits in the silicon detectors. To further suppress background from converted photons, they also must have at least 1 hit in the innermost pixel layer. To form a candidate track pair, tracks must be oppositely charged. To suppress SM backgrounds from one real electron and an additional track, a p_{T}-dependent requirement on maximum separation between the two tracks at the IP, $\Delta\phi^{\mathrm{IP}}_{\mathrm{trks}}$, is imposed. For these backgrounds, the tracks are typically well-separated near the IP, while the lower-p_{T} track is subsequently bent by the magnetic field back toward the higher-p_{T} track. Conversely, tracks from real γ^* objects typically have a small separation in $\Delta\phi^{\mathrm{IP}}_{\mathrm{trks}}$ because of the small γ^* mass. Finally, if there are more than two tracks satisfying all above requirements, the ranking as in the standard reconstruction (see Section 3.2.1) is used to select the best candidates.

Merged electrons are identified among these candidates using a multivariate algorithm (merged electron ID) described below. The efficiency of the merged electron ID is measured in data using early[2] converted photons from from $Z \to \ell\ell\gamma$ decays, and energy of the merged electron object is calibrated using the early converted photon energy calibration, due to the similarity of an early converted photon to a merged electron object from the calorimeter perspective.

4.4.1 Identification

The merged electron identification algorithm exploits the difference between the merged $\ell^+\ell^-$ pair in $H \to \gamma^*\gamma \to \ell^+\ell^-\gamma$ signal events and misidentified jets in background events as well as single electrons originating from decays of heavy quarks (the background containing events with single electrons originating from the IP is reduced by preselection requirements discussed above). In particular, background objects typically feature a higher proportion of transverse energy in the hadronic calorimeter relative to the EM calorimeter, R_{Had}.

Another difference exists in the ratio of the energy difference between the maximum energy deposit and the energy deposit in a secondary maximum in the cluster to the sum of these energies, E_{Ratio}, calculated in the first layer of EM calorimeter. For the single electrons identified by the standard identification algorithm, this difference is more prominent. Single electrons typically feature E_{Ratio}

[2] An early converted photon is defined as having radius R<160 mm for the efficiency measurement, and R=30 mm for calibrating the merged objects.

close to 1, while background objects tend to have a large spread in E_{Ratio} distribution, with low E_{Ratio} values more likely. For the merged electrons, due to a small but non-zero $\Delta\eta$ separation between the two electrons from low-mass γ^* decay, E_{Ratio} has extra tail structure. However, when considered only for the merged electrons with $\Delta\eta$ of less than approximately one strip size ($\Delta\eta < 0.003$), E_{Ratio} still has sufficient discriminating power.

Similarly, larger spread is typically observed for background objects in R_η and R_ϕ distributions. R_η (R_ϕ) is the ratio of $\sum E_{\text{T}}$ in a rectangle containing 3×7 (3×3) calorimeter cells in $\eta - \phi$, centred at the cluster position in the second layer of EM calorimeter, to the $\sum E_{\text{T}}$ in a rectangle containing 7×7 (3×7) calorimeter cells, respectively. The second layer of EM calorimeter, where R_η and R_ϕ, are calculated, has a coarser structure than the first layer. Consequently, the small $\Delta\eta$ separation between the two electrons does not play such a significant role as for the E_{Ratio} distribution, and therefore no extra $\Delta\eta$ requirement is applied.

A multivariate discriminator is trained using signal MC samples (see Section 4.3) and background samples containing jets and single electrons from decays of heavy quarks simulated with `Pythia 8` at LO accuracy in QCD. It is known that distributions of some calorimeter shower shape-associated variables are not well-modelled in the MC simulation. To correct for that, the so-called *fudge factors* are used, which correct shower shape variables to bring their distributions closer to what is seen in data. For the merged electron ID to be effective in data, these corrections are taken into account.

The multivariate discriminant is trained using the following variables, with selection requirements (cuts) applied to each variable separately:

- Shower-shape related: the above mentioned R_{Had}, E_{Ratio}, R_η, R_ϕ, as well as some others: f_3, $w_{\eta 2}$, $w_{\text{tots}1}$ (definitions are given in Ref [42]).
- Track-to-calorimeter pointing variables: angular distances between the extrapolated dielectron vertex candidate and the cluster center, $\Delta\phi_{\text{vtx}}$ and $\Delta\eta_{\text{vtx}}$; separation in $|\eta|$ between the leading track and the cluster center as measured in the first calorimeter layer, $\Delta\eta_1^{\text{lead track}}$; ratio of cluster energy to the momentum reconstructed from tracks and vertex position E/p_{vtx}. Electrons in signal events typically have tracks pointing to the calorimeter cluster, i.e. feature lower values of $\Delta\phi_{\text{vtx}}$, $\Delta\eta_{\text{vtx}}$, $\Delta\eta_1^{\text{lead track}}$ and E/p_{vtx}.
- Difference in the longitudinal impact parameter of the two tracks, $|z_0^{\text{trk1}} - z_0^{\text{trk2}}|$, with signal events typically featuring lower values.
- TRT identification information: probability for each of the tracks to be an electron track, based on a likelihood estimator using high threshold hit information.

The constraints in the training were set such as to achieve maximum possible background rejection while retaining at least 50% signal efficiency, i.e. at least 50% of preselected candidate merged electron events must pass the ID requirement. The optimal cuts on the variables listed above depend on the candidate merged electron p_{T} and η , and therefore different cuts are used in different p_{T} and η ranges. For η, it is sufficient to divide the barrel ($|\eta| < 1.37$) and endcap ($|\eta| > 1.52$) regions into two ranges (bins) each. (Objects in the instrumented transition region between the barrel and endcap, $1.37 < |\eta| < 1.52$, are rejected.) The background p_{T} distribution is steeply falling, while the signal distribution peaks at around $p_{\text{T}} = 55$–60 GeV, and therefore finer granularity is required for $p_{\text{T}} < 60$ GeV, while one bin is sufficient above this value. As a result, the following granularity (binning) is used:

- for p_{T}: 20–30 GeV, 30–40 GeV, 40–50 GeV, 50–60 GeV, 60+ GeV
- for $|\eta|$: 0.0–0.8, 0.8–1.37, 1.52–2.01, 2.01–2.37

After the multivariate discriminator was trained, the cut values were further optimized to provide a smooth efficiency profile in the electron E_{T} and $|\eta|$ distributions. In its final version, the merged electron ID has 61% overall signal efficiency (defined $N_{\mathrm{ID}}^{\mathrm{presel}}/N^{\mathrm{presel}} \times 100\%$, where N^{presel} is the number of preselected candidate merged electrons, and $N_{\mathrm{ID}}^{\mathrm{presel}}$ is the number of candidates passing the ID among them) and 0.13% background efficiency in the MC simulation, i.e. using the signal $H \to \gamma^* \gamma \to \ell^+ \ell^- \gamma$ and background MC samples discussed above. The signal efficiency with respect to the separation between electrons in ΔR is shown in Figure 4.3. It can be seen that the efficiency curve is relatively flat and high efficiency is retained even for the lowest ΔR values. In the standard reconstruction, there is a pronounced efficiency loss for the low ΔR values. The combined merged electron ID and isolation efficiency is subsequently measured in data (described below), because measuring both at the same time is technically simpler. In the MC simulation, the merged electron ID together with isolation, i.e. defined as $N_{\mathrm{ID + iso}}^{\mathrm{presel}}/N^{\mathrm{presel}} \times 100\%$, where $N_{\mathrm{ID + iso}}^{\mathrm{presel}}$ is the number of objects passing ID and isolation, has 52% signal efficiency and 0.06% background efficiency.

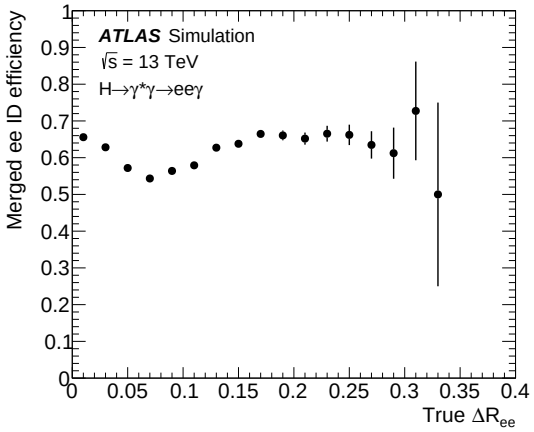


Figure 4.3: Merged-*ee* identification efficiency as a function of the truth ΔR (i.e. at the level of event generation) between the electrons for simulated $H \to \gamma^* \gamma \to \ell^+ \ell^- \gamma$ events [66]. The efficiency is defined as $N_{\mathrm{ID}}^{\mathrm{presel}}/N^{\mathrm{presel}} \times 100\%$, where N^{presel} is the number of preselected candidate merged electron objects, and $N_{\mathrm{ID}}^{\mathrm{presel}}$ is the number of objects passing ID among them.

4.4.2 Efficiency measurement

The combined efficiency of the merged electron ID and isolation requirement is estimated in data with $Z \to \ell\ell\gamma$ events, where a photon is radiated off one of the leptons, based on the method for estimating the photon identification efficiency described in Ref. [44]. From the reconstruction perspective, merged electrons are are very similar to converted photons (with one difference that a dielectron pair from the decay of γ^* has mass, which can lead to the electrons being separated in the non-bending plane η instead of just in the bending plane ϕ). Because there is no clean, abundant source of γ^*

available in the data for an efficiency measurement, the merged electron ID efficiency is estimated with this source of converted photons.

The $Z \to \ell\ell\gamma$ events are required to contain two same-flavour, opposite-sign leptons and one photon. The leptons are selected to ensure that they are well-reconstructed, using similar requirements as described in Section 4.5.2. The photon, on the other hand, is only required to have $p_T > 20$ GeV and satisfy the standard set of object quality and photon cleaning requirements used in photon reconstruction. Only converted photons with conversion radius $R < 160$ mm are considered, i.e. the ones which are kinematically the most similar to a γ^*. Crucially, no identification requirements are applied to the photon. Instead, a requirement 85 GeV $< m_{\ell\ell\gamma} <$ 97 GeV is made to select events with a three-body mass very close to m_Z. This ensures that the event is very likely to be a $Z \to \ell\ell\gamma$ decay, and therefore the third object in the event besides the two leptons is very likely to be a photon. Then the efficiency of the (photon) identification algorithm can be measured in data, by comparing $Z \to \ell\ell\gamma$ event yields before and after the identification requirement. The ratio of efficiencies in data and MC simulation (scale factor, SF $= \epsilon_{\text{data}}/\epsilon_{\text{MC}}$) is then used to correct the simulated events. But first backgrounds for the $Z \to \ell\ell\gamma$ process have to be taken into account.

A requirement on dilepton mass, $m_{\ell\ell} < 83$ GeV, is used to remove a background with a jet or a photon in initial state radiation, so that a phase space is isolated that contains predominantly Z boson with a final state radiation photon, plus possibly a background with a Z boson associated with jets. Crucially, this is a non-resonant background and can therefore be differentiated from the resonant $\ell\ell\gamma$ peak. This background is estimated in data by performing a template fit to the $m_{\ell\ell\gamma}$ variable with a requirement that the photon fails both tight identification and loose isolation requirements.

The signal template is constructed with MC simulation using `Sherpa`. The fraction of signal and background is then extracted from the signal-plus-background fit to data. The fits are performed in several p_{T} and η bins for the photon. To reduce statistical fluctuations of background templates, p_{T} bins are merged in the background template construction. The fits are performed once with photons required to pass the merged electron ID, and the other time without this requirement. The ratio of signal yields from the two fits gives the efficiency of the merged electron ID in data.

The scale factor is then calculated for the photon in p_{T} and η bins. All scale factor values are between 0.9 and 1.1, which demonstrates that the variables used in the merged electron ID (and also the isolation variable) are reasonably well modelled in MC simulation. The central value for the efficiency is defined as a weighted average of the measurements in the $\mu\mu\gamma$ and $ee\gamma$ channels:

$$\epsilon_{comb} = \frac{\epsilon_e/\sigma_e^2 + \epsilon_\mu/\sigma_\mu^2}{1/\sigma_e^2 + 1/\sigma_\mu^2}\,, \tag{4.1}$$

where σ denotes statistical uncertainty. Averaged over all bins, the combined efficiency of the merged electron ID and isolation requirement as measured in data is $\sim 50\%$.

The total statistical uncertainty is given by:

$$\sigma_{stat,tot} = \frac{1}{\sqrt{1/\sigma_e^2 + 1/\sigma_\mu^2}}\,. \tag{4.2}$$

There is a also a systematic uncertainty associated with the background estimation, which is measured by comparing background yields with η (instead of p_{T}) bins merged in the background template construction.

The statistical and systematic uncertainties for the merged electron ID efficiency measurement are estimated in p_{T} and η bins and are typically around 5% for the statistical and 3%-5% for the systematic uncertainty. Both statistical and systematic uncertainties in the merged electron ID efficiency measurement are considered *systematic* uncertainties for the purposes of this analysis. They are applied on an event-by-event basis depending on the merged electron object p_{T} and η. In this sense they are treated in the same way as all other experimental systematic uncertainties described in Section 4.7.

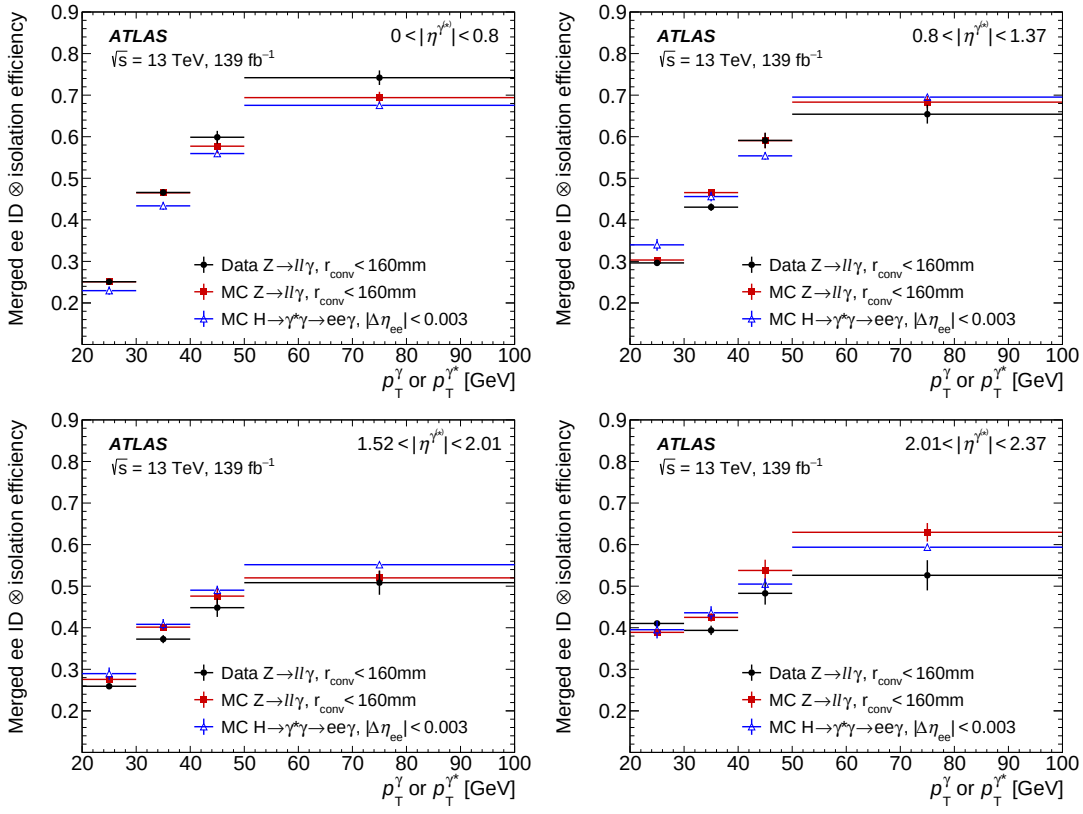


Figure 4.4: Merged electron ID identification and isolation efficiency as a function of merged-ee p_{T} in the $0 < |\eta| < 0.8$ (top left), $0.8 < |\eta| < 1.37$ (top right), $1.52 < |\eta| < 2.01$ (bottom left) and $2.01 < |\eta| < 2.37$ (bottom right) range [66]. The efficiencies for early converted photons in simulated $Z \to \ell\ell\gamma$ events (red squares) compared to the efficiencies of merged objects (γ^*) in simulated $H \to \gamma^*\gamma \to \ell^+\ell^-\gamma$ events (blue triangles). For the $H \to \gamma^*\gamma \to \ell^+\ell^-\gamma$ simulated events an additional generator-level requirement of $|\Delta\eta_{ee}| < 0.003$ is used to better match the converted photon signature. The efficiencies in $Z \to \ell\ell\gamma$ events for early converted photons measured in data (black circles) are also shown.

The efficiency measurement described above and its uncertainties are based on the assumption that the γ^* object and early converted photons behave in the same way for the purposes of their identification. To ensure that this is the case, the efficiency of the merged electron ID (together with isolation) for the early converted photons in $Z \to \ell\ell\gamma$ MC-simulated events is compared with the efficiency for the γ^* objects in $H \to \gamma^*\gamma \to \ell^+\ell^-\gamma$ MC-simulated events, shown in Figure 4.4. For the $H \to \gamma^*\gamma \to \ell^+\ell^-\gamma$ MC simulation an additional generator-level requirement of $|\Delta\eta_{ee}| < 0.003$ is

used to better match the converted photon signature, since converted photons are separated only in the bending plane ϕ. Because good agreement is observed, no additional systematic uncertainty due to differences between converted photons and merged electrons is considered.

4.4.3 Energy calibration

As with other objects in the event, the energy of the merged electron has to be calibrated. As already discussed, merged electron objects are more similar to early converted photons than to single electrons from the reconstruction perspective. For these merged electrons, a larger fraction of their energy falls outside the calibration window compared to single electrons. Additionally, a larger fraction of the energy is absorbed in material upstream of the calorimeter in case of the merged electron. A comparison is demonstrated in Figure 4.5 where a bias in reconstructed energy with respect to true energy is shown when applying electron energy calibration and converted photon energy calibration to the merged electron object. It can be seen that the converted photon energy calibration with minimal radius (R = 30 mm) exhibits the least bias. Therefore, the converted photon calibration at R = 30 mm was used to calibrate the merged electron energy.

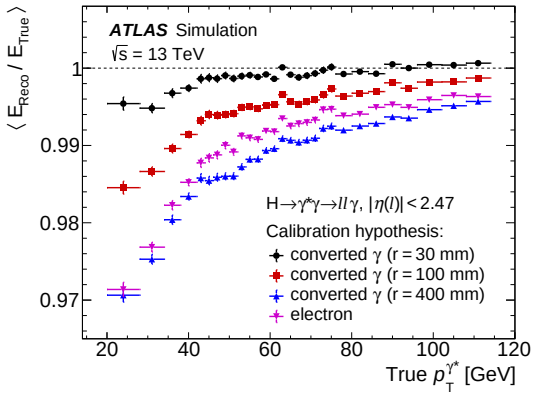


Figure 4.5: Ratio of reconstructed to true merged electron energy in simulated $H \to \gamma^*\gamma \to \ell^+\ell^-\gamma$ events as a function of the true merged electron p_{T} for several energy calibration techniques [66]: calibrated as a photon with a conversion radius of 30 mm (black circles), 100 mm (red squares), and 400 mm (blue upward triangles) or as an electron (purple downward triangles).

A systematic uncertainty associated with the converted photon energy calibration is estimated with the standard methods described in Section 3.2.2. However, energy resolution for the merged electrons is slightly different in comparison to the converted photons. The conversions reconstructed with two tracks in silicon detectors (SiSi) are conceptually the closest object to the merged electron. Therefore an additional uncertainty term is defined using the quadratic difference between the converted photon energy resolution, σ_{SiSi}, and merged electron energy resolution, σ_{ME}:

$$\Delta\sigma = \sqrt{\sigma^2_{\mathrm{ME}} - \sigma^2_{\mathrm{SiSi}}}\ . \tag{4.3}$$

The energy resolution uncertainty up (+) and down (−) variations are then defined as the following:

$$E^{\pm}_{\text{reco}} = E_{\text{true}} + \Delta E(\text{true,reco}) \cdot \frac{\sqrt{\sigma^2_{\text{ME}} \pm (\Delta\sigma)^2}}{\sigma_{\text{ME}}}, \tag{4.4}$$

where ΔE(true,reco) is the difference between true and reconstructed energy:

$$\Delta E(\text{true,reco}) = E^{\text{nom}}_{\text{reco}} - E_{\text{true}}. \tag{4.5}$$

As can be seen in Equation 4.4, the resolution effectively increases (decreases) quadratically by $\Delta\sigma$ for up and down uncertainty variations. These uncertainty variations are then propagated to the analysis results in the same way as all other experimental systematic uncertainties described in Section 4.7.

4.5 Event and objects selection

4.5.1 Trigger

The events in data are selected on-line by the trigger system described in Section 3.2.4. The analysis uses a combination of of single-lepton, dilepton, lepton+photon, and diphoton HLT triggers. These triggers match the signature of the signal $H \to \gamma^* \gamma \to \ell^+\ell^-\gamma$ events, with leptons and photons in the final state. Diphoton triggers, in particular, are useful to collect merged electron plus photon topologies, because of looser identification requirements and no tracking or track-matching requirements, allowing more collimated electrons to pass. Trigger algorithms do only fast on-line reconstruction (e.g. single lepton triggers target high p_{T} leptons and provide only the initial preselection of candidate events). Besides, triggers run in parallel, and a single selected event can match several triggers.

In 2017 and 2018 data taking, a new trigger `e25_mergedtight_g35_medium_Heg`[3] was introduced to target events with collimated electrons. In comparison with other electron triggers, it has stricter requirements on isolation in the calorimeter, while allowing for wider showers, which are a typical signature for merged deposits from two collimated electrons. Additionally, there is a requirement on the invariant mass of the photon and merged electron (i.e. likely an electron pair reconstructed as a single electron), $m_{e\gamma} > 90$ GeV, to reduce a background from $Z \to ee$ events. Studies were performed to confirm that the behavior of this trigger is well-modelled in MC simulation, therefore no additional systematic uncertainty is associated to it.

Triggers are further evaluated by their efficiency, defined as ratio of the number of selected signal events (with full selection requirements described further) that fired a given trigger to the total number of selected signal events in the absence of any trigger requirement. Trigger efficiency defined like this is estimated in MC simulation using $H \to \gamma^* \gamma \to \ell^+\ell^-\gamma$ MC samples. Among the available triggers, on the one hand, it is better to use more of them to increase the cumulative signal selection efficiency. On the other hand, it is preferable to avoid redundant triggers (i.e. ones which have nearly the same efficiency profile) to simplify the shape of signal distribution (different triggers have different turn-on curves in efficiency). To determine how useful each trigger is, they were studied on the following metrics:

- Selected events **inclusive fraction**: defined as $N_{\mathrm{SR}}^{\mathrm{trig}_1}/N_{\mathrm{SR}}$, where $N_{\mathrm{SR}}^{\mathrm{trig}_1}$ is the number of selected events in the signal region that fired a given trigger trig_1, and N_{SR} is the total number of events in the signal region N_{SR} (without a trigger requirement). Since events counted as $N_{\mathrm{SR}}^{\mathrm{trig}_1}$ can *also* fire any other trigger trig_i, these are *inclusive* fractions and may add up to more than 1.

- Selected events **unique fraction**: defined as $N_{\mathrm{SR}}^{only\ \mathrm{trig}_1}/N_{\mathrm{SR}}$, where $N_{\mathrm{SR}}^{only\ \mathrm{trig}_1}$ is the number of selected events in the signal region that fired *only* a given trigger trig_1, and N_{SR} is the total number of events in the signal region N_{SR} (without a trigger requirement). These are *unique* fractions and add up to much less than 1, since there is a significant overlap.

A high unique fraction, as defined above, indicates an important contribution to the signal yield that would otherwise have been lost. The triggers scoring low on unique fraction can therefore be

[3] ATLAS naming convention for triggers is to specify an object (`e` for electron, `mu` for muon and `g` for photon) and p_{T} threshold in GeV following it. Other parts of the name refer to reconstruction working points (e.g. `mergedtight`), some additional features, such as topological requirements, identifying a specific trigger. If a trigger targets multiple objects with the same p_{T} threshold, their number is specified first, e.g. `2mu`.

Trigger (group)	Inclusive fraction		Unique fraction	
	Resolved-*ee*	Merged-*ee*	Resolved-*ee*	Merged-*ee*
`HLT_e24_lhmedium_L1EM20VH`, `HLT_e60_lhmedium`, `HLT_e120_lhloose`	83.7	73.4	2.1	1.5
`HLT_2e12_lhloose_L12EM10VH`	36.3	6.6	2.1	0.1
`HLT_e20_lhmedium_g35_loose`	82.5	68.4	2.0	0.5
`HLT_g35_loose_g25_loose`	89.9	96.9	9.1	24.6

Table 4.1: Fraction of events in percent (inclusive, $N_{\mathrm{SR}}^{\mathrm{trig_1}}/N_{\mathrm{SR}} \times 100\%$, and unique, $N_{\mathrm{SR}}^{\mathit{only}\ \mathrm{trig_1}}/N_{\mathrm{SR}} \times 100\%$) selected with electron triggers in 2015. All electron triggers suitable for the analysis are considered and shown here; the single electron triggers are bundled together in one group.

removed. However after any one trigger is removed, unique fractions of other triggers change, and should be recalculated before removing any more triggers. Some triggers can also be bundled up together for the unique fraction analysis, such as all single lepton triggers. This makes the results more easily interpretable for that group as a whole. Care should also be taken to consider some expected difference in the trigger efficiency, e.g. between resolved and merged channels due to specifics in the kinematics of a typical event.

An example of unique fraction analysis for electron triggers in 2015 is shown in Table 4.1. As can be seen, some triggers are complimentary, e.g. with low unique fractions in the merged electron channel, but significant in the resolved electron channel. Lower inclusive fraction for single lepton triggers in the merged electron channel is expected because there is a significant fraction of events with merged electrons that don't pass the standard medium electron identification; these events are instead collected with a diphoton trigger (`HLT_g35_loose_g25_loose`). For exactly this reason the apparent unique fraction for single lepton triggers in the merged electron channel is very low, while for the `HLT_g35_loose_g25_loose` is considerably higher than in the resolved electron channel. Both inclusive and unique fractions for the dielectron trigger (`HLT_2e12_lhloose_L12EM10VH`) are lower in the merged electron channel, because only one EM cluster (a merged cluster) is expected there. The final list of triggers chosen in the analysis is given in Table 4.2.

After the list of triggers is finalized, the cumulative trigger efficiency is calculated. For that, also trigger *matching* is taken into account. Trigger matching ensures that the *selected* reconstructed objects were indeed the objects that triggered the event. This correction reduces the trigger efficiency by a small amount. The trigger efficiency is therefore measured as $N_{\mathrm{SR}}^{\mathit{any}\ \mathrm{trig\ (matched)}}/N_{\mathrm{SR}} \times 100\%$, where $N_{\mathrm{SR}}^{\mathit{any}\ \mathrm{trig\ (matched)}}$ is the number of selected events in the signal region that fired any trigger from the final list of triggers, *provided* the objects that fired the trigger are matched to the selected objects, and N_{SR} is the number of selected events in the signal region without a trigger requirement. The trigger efficiency is determined to be 97.2% overall, and efficiencies per data taking and in different channels are shown in Table 4.3.

As a standard procedure in ATLAS, a trigger scale factor is often used, defined as a ratio of trigger efficiency in data to efficiency in MC simulation, $\mathrm{SF} = \epsilon_{\mathrm{data}}/\epsilon_{\mathrm{MC}}$, and applied as a correction to MC. Typically there is also a systematic uncertainty associated with the scale factor. However, due to the very high trigger efficiency, a trigger scale factor is not used in this analysis, as any effect would be

Year	(Di)Muon	(Di)Electron	(Di)Lepton-photon, Diphoton
2015	mu40 mu20_iloose_L1MU15 2mu10 mu18_mu8noL1	e60_lhmedium e24_lhmedium_L1EM20VH e120_lhloose 2e12_lhloose_L12EM10VH	e20_lhmedium_g35_loose g25_medium_mu24 g15_loose_2mu10_msonly g35_loose_g25_loose
2016	mu50 mu26_ivarmedium 2mu14 mu22_mu8noL1	e60_lhmedium_nod0 e26_lhtight_nod0_ivarloose e140_lhloose_nod0 2e17_lhvloose_nod0	e20_lhmedium_nod0_g35_loose g25_medium_mu24 g15_loose_2mu10_msonly g35_loose_L1EM22VHI_mu18noL1 g35_loose_L1EM22VHI_mu15noL1_mu2noL1 g35_loose_g25_loose
2017-18	mu50 mu26_ivarmedium 2mu14 mu22_mu8noL1	e60_lhmedium_nod0 e26_lhtight_nod0_ivarloose e140_lhloose_nod0 2e24_lhvloose_nod0	**e25_mergedtight_g35_medium_Heg** g25_medium_mu24 g15_loose_2mu10_msonly g35_loose_L1EM24VHI_mu18 g35_loose_L1EM24VHI_mu15_mu2noL1 g35_tight_icalotight_L1EM24VHI_mu18noL1 g35_tight_icalotight_L1EM24VHI_mu15noL1_mu2noL1 g35_medium_g25_medium_L12EM20VH

Table 4.2: Triggers used in the analysis per data taking period. Highlighted is the new trigger aimed at collimated electrons used in 2017 and 2018 data taking.

Year	Channel	Trigger efficiency
All (2015-2018)	All	97.2%
2015	All	88.6%
2016	All	98.3%
2017	All	97.1%
2018	All	97.1%
All (2015-2018)	Muons	96.2%
All (2015-2018)	All electrons	98.4%
All (2015-2018)	Resolved-*ee*	96.5%
All (2015-2018)	Merged-*ee*	99.8%

Table 4.3: Trigger efficiency in different data taking periods and different channels.

quite small, and therefore no additional systematic uncertainty is assigned.

4.5.2 Event and objects preselection

Events passing the trigger requirement are required to contain a primary vertex that has at least two tracks with $p_{\mathrm{T}} > 0.5$ GeV each. Track reconstruction is described in Section 3.2.1. For events in data there are additional data quality requirements [67] to ensure that they are recorded when all detector systems performed nominally.

Electron and photon, jet, muon reconstruction is described in more detail in Sections 3.2.2, 3.2.2 and 3.2.3, respectively. The reconstruction of merged electron candidates is a novel aspect of this analysis and is described in Section 4.4.

Initially the objects in the event are preselected using somewhat relaxed criteria, which is necessary for overlap removal procedure (described below) to work properly. Then, the selection criteria are tightened in the final selection. At the preselection step, photon and electron candidates are required to satisfy only the *loose* and *very loose* criteria respectively, while muon candidates are required to satisfy *medium* identification criteria (same as in the final selection). To ensure that objects are well-reconstructed and to reduce backgrounds, additional kinematic requirements are applied to candidates already at preselection step: $p_{\mathrm{T}} > 4.5$ GeV, $|\eta| < 2.47$ for electrons, $p_{\mathrm{T}} > 3$ GeV, $|\eta| < 2.7$ for muons, $p_{\mathrm{T}} > 20$ GeV, $|\eta| < 2.37$ for photons and $p_{\mathrm{T}} > 20$ GeV, $|\eta| < 4.4$ for jets. Additionally, jets originating from pile-up collisions in the range $|\eta| < 2.4$ and $p_{\mathrm{T}} < 60$ GeV are suppressed using the JVT discriminant (see Section 3.2.2).

At this stage, the objects are preselected and a candidate γ^* is formed. The leading (the one with higher p_{T}) muon must satisfy $p_{\mathrm{T}} > 11$ GeV (leading resolved electron $p_{\mathrm{T}} > 13$ GeV) in accordance with minimum p_{T} requirements of the dilepton triggers (the imposed requirements are slightly higher than the minimum trigger p_{T} to avoid turn-on effects in the trigger efficiency). Among lepton pairs satisfying these requirements and merged electron candidates, muon pairs are prioritized. If there are multiple γ^* candidates of the same lepton flavor, the one with highest $p_{\mathrm{T}}^{\gamma^*}$ is preferred (the p_{T} of the tracks is used in the $p_{\mathrm{T}}^{\gamma^*}$ calculation for electrons, both merged and resolved). In the rare case where two tracks match both a resolved electron pair and a merged electron candidate, the resolved pair is preferred.

After the γ^* candidate is selected, the following objects are kept: leptons belonging to the γ^* candidate, the highest p_{T} photon, and possibly jets (if there are any in the event). In the case where the the highest-p_{T} photon is within $\Delta R < 0.02$ of the γ^* candidate, that photon is discarded and the next highest-p_{T} preselected photon candidate is considered, with the same motivation as for the subsequent overlap removal procedure.

An overlap removal procedure is performed to avoid double-counting objects. Specifically, if two or more objects are too close to each other in ΔR, they are more likely to originate from the same deposits in the detector which are reconstructed twice (or more) as different objects. Another reason is to remove background events where the photon is too close to the electron, e.g. photon bremsstrahlung events. The cone sizes used are related to the nature of the object's typical detector response, while prioritization (which objects to keep) is determined by the specifics of the analysis. For example, in this analysis a muon is prioritized if it is too close to a jet, while some other analyses, e.g. with final states containing jets, may prefer to keep a jet in this case. The resulting procedure is the following:

- If an electron (merged or resolved) is within $\Delta R < 0.4$ from the photon, the event is discarded.

- If a jet is within $\Delta R < 0.4$ from the photon, the jet is discarded.
- If a remaining jet is within $\Delta R < 0.2$ from an electron (merged or resolved), the jet is discarded.
- If a remaining jet is within $\Delta R < 0.4$ from an electron (merged or resolved), the electron is discarded (and, since this effectively removes the γ^* candidate, the event is discarded).
- If a muon is within $\Delta R < 0.4$ from the photon, the muon (and the event) is discarded.
- If a remaining jet is within $\Delta R < 0.4$ from the muon, the jet is discarded.

After overlap removal, events containing muons with poor momentum resolution (see Section 3.2.3) are rejected. Then, trigger matching is performed and the event is discarded if the objects that triggered the event do not match the selected objects, for any of the triggers used in the analysis. The identification requirements for resolved electrons and photons are tightened to *medium* and *tight* respectively. Merged electrons are required to pass a merged electron ID described in Section 4.4. The impact parameters of tracks associated to the signal objects must also satisfy the following requirements: $|z_0 \sin\theta| < 0.5$ mm for all tracks and $|d_0|/\sigma_{d_0} < 5$ for resolved and merged electrons ($|d_0|/\sigma_{d_0} < 3$ for muons).

Electrons, muons and photons are required to be well-isolated. For muons, the $\sum p_{\mathrm{T}}$ of tracks in a small cone around it[4] must not be greater than 6% of the muon p_{T} (15% of the p_{T} for resolved electrons and 5% of the p_{T} for photons). In the calorimeter, for resolved electrons the $\sum E_{\mathrm{T}}$ around respective clusters must not exceed 20% of electron p_{T} (6.5% of the p_{T} for photons). For merged electrons, the isolation calculation is performed in the same way as for resolved electrons, but the requirements are somewhat more strict: the $\sum p_{\mathrm{T}}$ of tracks and $\sum E_{\mathrm{T}}$ in the calorimeter around it must both be less than 6% of the merged electron p_{T}.

In the events with highly collimated leptons, isolation estimates for resolved electrons and muons need to be modified to ensure that the track p_{T} or calorimeter E_{T} of the second object is not counted against the isolation of the first, and vice versa. An example from an early study of this isolation correction for resolved electrons is shown in Figure 4.6.

Another consequence of predominantly collimated objects is that the isolation requirement on the subleading lepton (the one with lower p_{T}) can be dropped with a slight *gain* in signal efficiency. This is possible because of two effects: isolation is typically less efficient for objects with lower p_{T}, and objects contributing to the isolation of the lower-p_{T} subleading lepton are also likely to contribute to the isolation of the leading lepton. As a consequence, the leading lepton isolation can be exploited as a highly efficient isolation for both collimated objects. Therefore, in this analysis, the isolation requirement is applied only to the leading lepton.

4.5.3 Signal region and categories

At this point, the event contains a single well-defined H candidate. Further selection is performed mainly to boost the sensitivity of the search. This gives more freedom in choosing the exact requirements, while at the same time it can be non-trivial to pick a set of requirements giving the best possible sensitivity. The event selection for this analysis is based on fixed requirements in different kinematic

[4] See references in Section 3.2 for ΔR sizes for track isolation of corresponding objects, and how transverse energy is evaluated for calorimeter isolation.

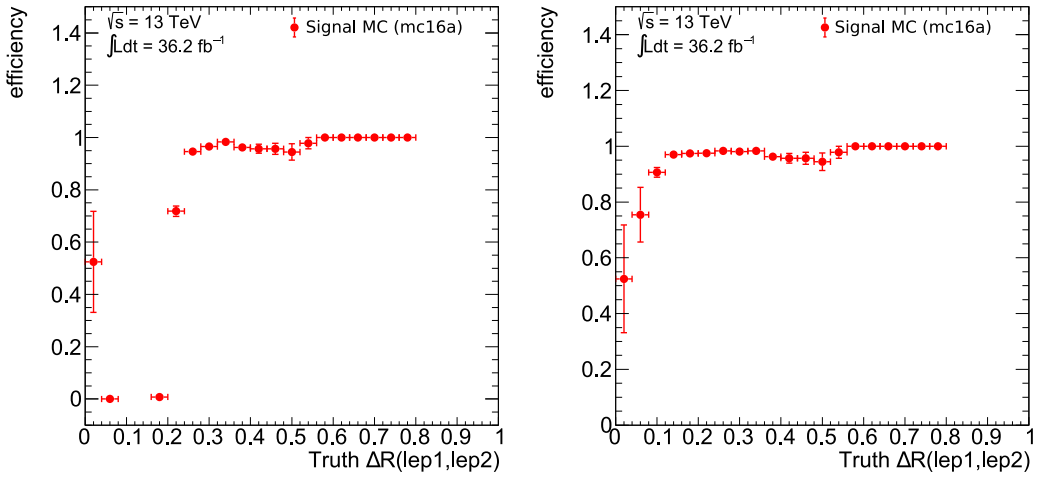


Figure 4.6: Isolation requirement efficiency as a function of truth ΔR (i.e. at the level of event generation) between resolved electrons without (left) and with (right) isolation correction for close-by objects applied, obtained with the signal $H \to \gamma^* \gamma \to \ell^+ \ell^- \gamma$ MC sample corresponding to 2015-2016 data taking conditions (mc16a campaign). The efficiency is defined as the ratio of events passing preselection and the isolation requirement to events passing only preselection. It can be seen that the isolation requirement without correction has significantly lower efficiency for low ΔR (highly collimated electrons).

variables (a *cut-based* approach). The events were also categorized in several mutually exclusive categories, which results in a greater cumulative sensitivity. This is possible mostly because of the fact that among reconstructed events with a Higgs boson, those produced via VBF production mode are more rare, but allow for a greater background suppression due to the additional requirement of two jets in the event with characteristic kinematics. Another highly sensitive category with good background suppression can be constructed targeting the high-$p^{\ell\ell\gamma}_{\text{TThrust}}$ events. Therefore, three types of categories were defined (ordered by priority): VBF, high-$p^{\ell\ell\gamma}_{\text{TThrust}}$ and low-$p^{\ell\ell\gamma}_{\text{TThrust}}$. The events failing VBF requirements are considered for high-$p^{\ell\ell\gamma}_{\text{TThrust}}$ category, while low-$p^{\ell\ell\gamma}_{\text{TThrust}}$ category contains the remaining events. Together with the three channels (muons, resolved electrons and merged electrons), this gives a total of nine analysis categories.

To evaluate different possible selection sets, an approximate calculation of the significance of the signal excess over background was used, defined as $Z_{\text{approx}} = \sqrt{2((s+b)\ln(1+s/b)-s)}$ [68], where s is the signal and b is the background yield. To calculate the expected significance, the signal contribution is estimated with signal Monte-Carlo samples, while the background contribution is estimated in data. While the MC signal yields in the 122 GeV $< m_{\ell\ell\gamma} <$ 128 GeV region (where the signal peak is located) are simply added up, the background in the same region is estimated from the fit of the $m_{\ell\ell\gamma}$ distribution in data in the range 105 GeV $< m_{\ell\ell\gamma} <$ 160 GeV with a falling exponential function, with the signal region 120 GeV $< m_{\ell\ell\gamma} <$ 130 GeV excluded (referred to as data sidebands); the function is extrapolated to the excluded region. This model is much more simplified than the full signal background models described in Section 4.6, and the extraction of s and b yields is not done with the more advanced statistical methods described in Section 4.8. Nevertheless, Z_{approx} was found to closely approximate the true significance Z as calculated with full signal and background models and advanced statistical methods. The advantage of using Z_{approx} is that it can be evaluated quickly

for many different possible selection requirements under consideration.

A selection optimization procedure was developed based on Z_{approx}. The procedure starts with choosing a set of kinematic variables, initial values and ranges. The following kinematic variables are considered:

- Some variables already used in preselection (this means that these requirements should not be relaxed, but can be tightened): $p_{\text{T}}^{\ell\ \text{lead}}$, $p_{\text{T}}^{\ell\ \text{sublead}}$, p_{T}^{γ}.
- The dilepton system (or merged electron) mass, $m_{\ell\ell}$.
- The $p_{\text{T}}^{\ell\ell}$ and p_{T}^{γ} relative to the $m_{\ell\ell\gamma}$: $p_{\text{T}}^{\ell\ell}/m_{\ell\ell\gamma}$, $p_{\text{T}}^{\gamma}/m_{\ell\ell\gamma}$.
- Variables targeting VBF production mode signature (in case of more than two jets, the leading and subleading are considered): momentum $p_{\text{T}}^{\text{jet}}$ of both jets, separation $\Delta\eta_{jj}$ between them, minimum distance between a jet and other objects in the event (a photon, a muon, resolved or merged electron), $\min(\Delta R_{j-\gamma/\ell})$, angular separation between dijet and $\ell\ell\gamma$ systems, $\Delta\phi_{\ell\ell\gamma,jj}$, mass of the dijet system, m_{jj}, momentum of the H candidate, $p_{\text{T}}^{\ell\ell jj\gamma}$, and $|\eta_{\text{Zeppenfeld}}^{\ell\ell\gamma}| = |\eta_{\ell\ell\gamma} - 0.5(\eta_{j1} + \eta_{j2})|$, where $\eta_{\ell\ell\gamma}$ is the pseudorapidity of the $\ell\ell\gamma$ system and η_{j1} and η_{j2} is the pseudorapidity of the leading and subleading jet respectively. The latter quantity was demonstrated to be a powerful discriminant for the VBF events in Ref. [69].

The optimization procedure in general can be outlined as follows:

- For the first kinematic variable v_1 on the list, evaluate Z_{approx} for many different values of cuts in a given range, e.g. $v_1 > c_1$, $v_1 > c_2$, and so on (note that the sign depends on the nature of the variable, can also be "<"). No cuts are initially applied on any other variables $v_2, ..v_n$ under consideration for this calculation, unless a requirement was already imposed on preselection step (such as $p_{\text{T}} > 11$ GeV for leading muon).
- Fix a cut on v_1 at c_{best}, which has the highest Z_{approx}. Repeat with the rest of variables on the list.
- Use the final cut values for all variables as initial values and repeat the procedure several times (this time applying the other cut values) until it converges to a set of cuts that does not change, or until the improvement in Z_{approx} becomes negligible.
- Start the procedure with different initial values. If the resulting set of selections is close to the one obtained initially, it is considered robust: this makes sure to avoid narrow local maxima in Z_{approx}.

The optimization procedure is performed separately for the VBF production mode (using VBF signal MC samples and all variables listed above), and for ggF production mode (using ggF signal MC samples and all variables except the ones targeting VBF signature). Other production modes are not specifically targeted in the optimization because of their small contribution. The procedure is also slightly modified to account for low signal yields, especially for VBF production mode. For example, the yields in all channels are combined to optimize the VBF category, while for the ggF production mode the procedure is performed separately in all three channels (the optimal values turned out to be very similar, and in the end a harmonized selection set was used for ggF categories). To cope with

still relatively low VBF yields, significance profiles were plotted in ROOT program [70] and smoothed with TGraphSmooth to avoid picking random noise as the optimal value. Additionally, the Z_{approx} was multiplied by a factor $\text{Erf}(s/2.16)$, which was tuned to avoid going into a too-low signal yields regime (less than 2 events), where Z_{approx} becomes an unreliable estimation of the true Z, mostly because statistical and systematic uncertainties are not taken into account.

As a result of optimization, the following values of kinematic requirements were chosen: $p_{\text{T}}^{\ell\ell}/m_{\ell\ell\gamma} > 0.3$, $p_{\text{T}}^{\gamma}/m_{\ell\ell\gamma} > 0.3$, $m_{\ell\ell} < 30$ GeV. The remaining variables tested were not tightened, and therefore the cut values remain the same as in preselection. Distribution shapes after selection and significance profiles for the ggF production mode are shown in Figure 4.7. These requirements are optimal for both ggF and VBF production modes, and are therefore applied in all categories.

The choice of $m_{\ell\ell}$ requirement, while being a quite good discriminator, is mostly influenced by restricting the phase space to the region where $H \to \gamma^*\gamma \to \ell^+\ell^-\gamma$ decays dominate and any interference with $H \to Z\gamma$ can be ignored; it also serves to simplify the background composition.

Additionally, to remove events with a J/Ψ, the $m_{\ell\ell}$ region corresponding to J/Ψ is excluded: $2.9 < m_{\mu\mu} < 3.3$ GeV in the muon channel [71] and $2.5 < m_{ee} < 3.5$ GeV [72] in the electron channel (the difference is due to different mass resolution in these channels). To remove events with an Υ, the region $9.1 < m_{\mu\mu} < 10.6$ GeV in muon channel [71] and $8.0 < m_{ee} < 11.0$ GeV in electron channel[5] is excluded.

Finally, only events in the $m_{\ell\ell\gamma}$ window of $105 < m_{\ell\ell\gamma} < 160$ GeV are used for the construction of the signal and background models; this region contains the signal as well as background-dominated regions on either side of the signal peak, which is necessary for the signal-plus-background model fit.

All kinematic requirements mentioned so far are optimal for both ggF and VBF production modes. The VBF-specific requirements, defining the VBF analysis category, are applied on top of them: $|\eta^{\ell\ell\gamma}_{\text{Zeppenfeld}}| < 2.0$, $\Delta\eta_{jj} > 2.7$, $\min(\Delta R_{j-\gamma/\ell}) > 1.5$, $\Delta\phi_{\ell\ell\gamma,jj} > 2.8$, $m_{jj} > 500$ GeV; for leading and subleading jets: $p_{\text{T}}^{\text{jet}} > 25$ GeV ($p_{\text{T}}^{\text{jet}} > 30$ GeV for forward jets, with $|\eta| > 2.5$). Distribution shapes after selection and significance profiles are shown in Figure 4.8.

The $p^{\ell\ell\gamma}_{\text{TThrust}}$ requirement defining the high-$p^{\ell\ell\gamma}_{\text{TThrust}}$ category is optimized using a different approach, because there is a large interdependence between low-$p^{\ell\ell\gamma}_{\text{TThrust}}$ and high-$p^{\ell\ell\gamma}_{\text{TThrust}}$ categories (the $p^{\ell\ell\gamma}_{\text{TThrust}}$ requirement impacts significance in the low-$p^{\ell\ell\gamma}_{\text{TThrust}}$ category as well). The $p^{\ell\ell\gamma}_{\text{TThrust}}$ requirement is optimized by varying the value from 30 to 120 GeV and studying the impact of the resulting categorization on the global expected limit with the statistical methods described in Section 4.8, using a falling exponential function for the estimation of background from the $m_{\ell\ell\gamma}$ fit in the data sidebands. The value of 100 GeV was found to yield the maximum sensitivity.

The event selection requirements are summarized in Table 4.4.

[5] The Υ window was determined by plotting the $m_{\ell\ell}$ distribution in the resolved electron channel in data, with only preselection applied in order to increase the sample size.

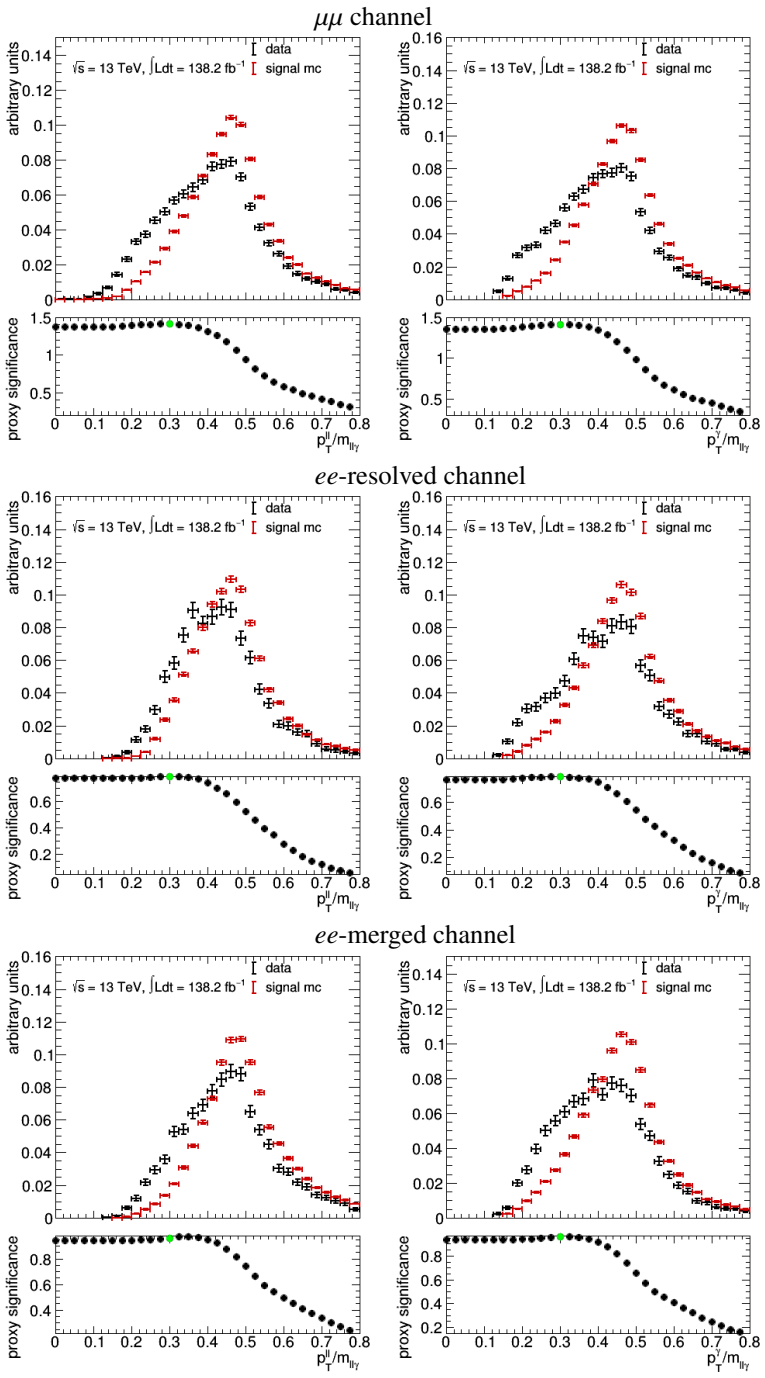


Figure 4.7: Kinematic distributions of $p_{\mathrm{T}}^{\ell\ell}/m_{\ell\ell\gamma}$ and $p_{\mathrm{T}}^{\gamma}/m_{\ell\ell\gamma}$ in the low-$p_{\mathrm{TThrust}}^{\ell\ell\gamma}$ category. The red histogram is obtained with a ggF signal MC sample, the black histogram is obtained with data in sidebands. The muon channel is shown in the top, resolved electron in the middle and merged electron in the bottom. All low-$p_{\mathrm{TThrust}}^{\ell\ell\gamma}$ category cuts are applied except for the variable being plotted. The Z_{approx} profile is shown in the bottom part of each plot. The green point represents the selected cut value. Note that the optimization procedure works sequentially to determine the best *set* of cuts, therefore the final values may not necessarily maximize Z_{approx} for each of the distributions individually, and cut values at $p_{\mathrm{T}}^{\ell\ell}/m_{\ell\ell\gamma} > 0.3$, $p_{\mathrm{T}}^{\gamma}/m_{\ell\ell\gamma} > 0.3$ were harmonized across channels.

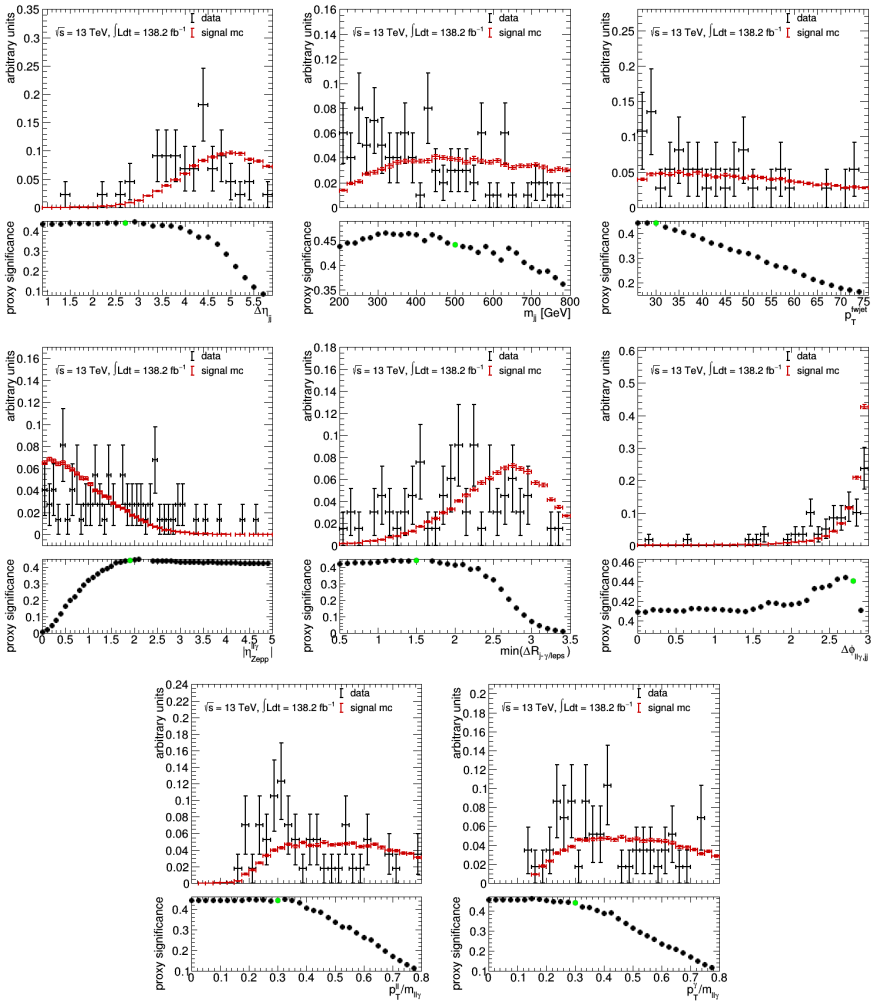


Figure 4.8: Kinematic distributions of variables in the VBF category. The red histogram is obtained with a VBF signal MC sample, the black histogram obtained with data in sidebands. All channels (electrons resolved, merged, and muons) are combined to increase yields. All VBF category cuts are applied except for the variable being plotted. The Z_{approx} profile is shown in the bottom part of each plot. The green point represents the selected cut value. Note that the optimization procedure works sequentially to determine the best *set* of cuts, therefore the final values may not necessarily maximize Z_{approx} for each of the distributions individually. Note also that the $p_{\text{T}}^{\ell\ell}/m_{\ell\ell\gamma} > 0.3$, $p_{\text{T}}^{\gamma}/m_{\ell\ell\gamma} > 0.3$ cuts do not have a significant effect, but also do not reduce the Z_{approx}. These cuts are much more useful in the other categories, but are kept here too for simplifying selection procedure.

Object preselection and Higgs candidate selection		
Passing trigger requirements.		
In data: passing data quality requirements (all systems operate nominally).		
Has a *very loose* γ.		
Has a *medium* $\mu\mu$ pair with $p_{\mathrm{T}}^{\mu\ \mathrm{lead}} > 11$ GeV?		
Yes:	No. Has a resolved ee pair with $p_{\mathrm{T}}^{e\ \mathrm{lead}} > 13$ GeV or a merged e candidate with $p_{\mathrm{T}} > 20$ GeV?	
↓	Yes: pick the one with highest $\sum p_{\mathrm{T}}^{\mathrm{trk}}$.	
$\mu\mu$ channel	ee-resolved channel	ee-merged channel
A H candidate remains after overlaps are removed.		
Final Object Selection		
Reject events with "bad" μ	e pass *medium* ID	Merged e passes merged ID
Tracks pass $\lvert z_0 \sin\theta\rvert < 0.5$, $\lvert d_0\rvert/\sigma_{d_0} < 3$	$\lvert z_0 \sin\theta\rvert < 0.5$, $\lvert d_0\rvert/\sigma_{d_0} < 5$	$\lvert z_0 \sin\theta\rvert < 0.5$, $\lvert d_0\rvert/\sigma_{d_0} < 5$
Leading μ is isolated	Leading e is isolated	Merged e is isolated
The γ is well-isolated and passes *tight* ID.		
Event-level selection		
$m_{\ell\ell} < 30$ GeV		
$105 < m_{\ell\ell\gamma} < 160$ GeV		
Passes J/Ψ and Υ veto.		
$p_{\mathrm{T}}^{\ell\ell}/m_{\ell\ell\gamma} > 0.3$, $p_{\mathrm{T}}^{\gamma}/m_{\ell\ell\gamma} > 0.3$		
Categorization		
Passes VBF requirements?		
Yes:	No. Passes $p_{\mathrm{TThrust}}^{\ell\ell\gamma} > 100$ GeV?	
↓	Yes:	No:
VBF category	high-$p_{\mathrm{TThrust}}^{\ell\ell\gamma}$ category	low-$p_{\mathrm{TThrust}}^{\ell\ell\gamma}$ category.

Table 4.4: Summary of the event selection requirements.

4.6 Signal and background modelling

The signal and background $m_{\ell\ell\gamma}$ distribution shapes are modelled using parametric functions. These functions are incorporated into the extended likelihood model described in Section 4.8. Subsequently the signal and background contributions are extracted from the simultaneous fit to data in all analysis categories in the range 110 GeV $< m_{\ell\ell\gamma} <$ 160 GeV [6] and further statistical analysis is performed.

4.6.1 Signal and resonant background

The signal is modelled based on MC simulation of the $H \to \gamma^*\gamma \to \ell^+\ell^-\gamma$ process. The MC samples used and Higgs boson production modes considered are described in Section 4.3.1. To determine the signal shape, the $m_{\ell\ell\gamma}$ distribution of selected signal MC events is fitted with a double-sided Crystal Ball function independently in each analysis category. The double-sided Crystal Ball function[7] features a Gaussian core and asymmetric power-law tails. The following parameters are floating in the fit: Gaussian core mean, μ^{CB}, initially set to Higgs boson mass peak, which in MC simulation corresponds to 125 GeV; Gaussian core width, σ^{CB}, corresponding to the mass resolution in the category; $\alpha^{\text{CB}}_{\text{low}}$ and $\alpha^{\text{CB}}_{\text{high}}$, which correspond to points where the Gaussian core transforms into a power law, measured in standard deviations from μ^{CB}; and $n^{\text{CB}}_{\text{low}}$ and $n^{\text{CB}}_{\text{high}}$, which are the exponents in the power law of the low- and high-mass tails. The double-sided Crystal Ball function can be expressed as the following:

$$\mathrm{S}(m_{ll\gamma}) = N \times \begin{cases} e^{-t^2/2}, & \text{if } -\alpha^{\text{CB}}_{\text{low}} \le t \le \alpha^{\text{CB}}_{\text{high}} \\ e^{-\frac{1}{2}\left(\alpha^{\text{CB}}_{\text{low}}\right)^2}\left[\frac{n^{\text{CB}}_{\text{low}}}{\alpha^{\text{CB}}_{\text{low}}}\left(\frac{\alpha^{\text{CB}}_{\text{low}}}{n^{\text{CB}}_{\text{low}}} - \alpha^{\text{CB}}_{\text{low}} - t\right)\right]^{-n^{\text{CB}}_{\text{low}}}, & \text{if } t < -\alpha^{\text{CB}}_{\text{low}} \\ e^{-\frac{1}{2}\left(\alpha^{\text{CB}}_{\text{high}}\right)^2}\left[\frac{n^{\text{CB}}_{\text{high}}}{\alpha^{\text{CB}}_{\text{high}}}\left(\frac{\alpha^{\text{CB}}_{\text{high}}}{n^{\text{CB}}_{\text{high}}} - \alpha^{\text{CB}}_{\text{high}} - t\right)\right]^{-n^{\text{CB}}_{\text{high}}}, & \text{if } t > \alpha^{\text{CB}}_{\text{high}} \end{cases} \tag{4.6}$$

where t is the distance from μ^{CB} in standard deviations, i.e. $t = (m_{ll\gamma} - \mu^{\text{CB}})/\sigma^{\text{CB}}$, and N is a normalization constant.

After the fitting is performed, the resulting μ^{CB} values are shifted +90 MeV to correct the $m_H =$ 125 GeV set in simulation to the central value from the combined ATLAS and CMS measurement $m_H = 125.09$ GeV [25]. The resulting mass resolution from the fit is shown in Table 4.5. For high-$p^{\ell\ell\gamma}_{\text{TThrust}}$ and VBF categories, the electron channels have better resolution than the muon channels, which is believed to be due to these categories being enriched in leptons with higher p_{T}, where the muon resolution degrades. Examples of signal fit in high-$p^{\ell\ell\gamma}_{\text{TThrust}}$ merged electron category and low-$p^{\ell\ell\gamma}_{\text{TThrust}}$ resolved electron category are shown in Figure 4.9. All signal fit functions along with their parameters are available in Appendix A.

In this analysis, there is a small resonant $H \to \gamma\gamma$ background (yields are given in Table 4.10) in the electron channels, where one of the decay photons converts and is identified as a merged electron,

[6] This slightly narrower mass window than in the event selection further reduces the uncertainty associated with a possible systematic bias in the signal yields extracted from the statistical model, described in Section 4.6.3. The non-resonant background template described in Section 4.6.2 is constructed and corresponding background contributions are estimated using selected events in the full 105 GeV $< m_{\ell\ell\gamma} <$ 160 GeV mass window.

[7] First described in the Crystal Ball experiment [73], this function is now widely used (e.g. in Ref. [74]) to model resonances, because it models tails due to radiative losses.

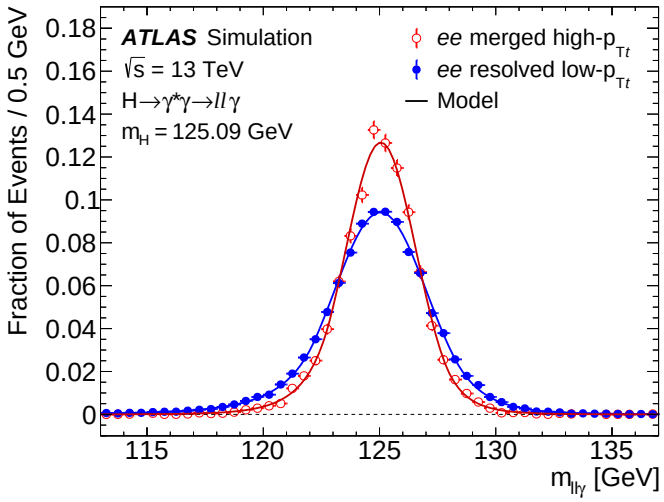


Figure 4.9: A comparison of the signal fits in the high-$p^{\ell\ell\gamma}_{\text{TThrust}}$ merged electron (red) and low-$p^{\ell\ell\gamma}_{\text{TThrust}}$ resolved electron (blue) categories [66]. Histograms of MC yields are shown with circles, while double-sided Crystal Ball functions from the fit are shown with lines. The mass distribution is shifted up by 90 MeV in both to match the measured Higgs boson mass [25], $m_H = 125.09$ GeV.

Category	σ_{68} [GeV]	σ_{90} [GeV]
Low-$p^{\ell\ell\gamma}_{\text{TThrust}}$ Dimuon	2.03	3.73
Low-$p^{\ell\ell\gamma}_{\text{TThrust}}$ Resolved Electron	2.17	3.99
Low-$p^{\ell\ell\gamma}_{\text{TThrust}}$ Merged Electron	2.13	3.76
VBF Dimuon	2.01	3.83
VBF Resolved Electron	1.87	3.38
VBF Merged Electron	1.77	3.12
High-$p^{\ell\ell\gamma}_{\text{TThrust}}$ Dimuon	2.09	3.90
High-$p^{\ell\ell\gamma}_{\text{TThrust}}$ Resolved Electron	1.69	3.06
High-$p^{\ell\ell\gamma}_{\text{TThrust}}$ Merged Electron	1.60	2.82

Table 4.5: Signal mass resolutions σ_{68} and σ_{90} in each analysis category, where σ_{68} (σ_{90}) is defined as half of the smallest interval expected to contain 68% (90%) of the signal around the resonance mass peak.

Category	μ_{CB}^{sig} [GeV]	$\mu_{CB}^{H\to\gamma\gamma}$ [GeV]	σ_{CB}^{sig} [GeV]	$\sigma_{CB}^{H\to\gamma\gamma}$ [GeV]
Low-$p_{\text{TThrust}}^{\ell\ell\gamma}$ Resolved Electron	125.04	124.86	1.99	1.97
Low-$p_{\text{TThrust}}^{\ell\ell\gamma}$ Merged Electron	124.90	124.61	2.00	2.26

Table 4.6: Comparison of the main (σ_{CB}, μ_{CB}) fit parameters from the $H \to \gamma^*\gamma \to \ell^+\ell^-\gamma$ signal fit and from $H \to \gamma\gamma$ background fit with a double-sided Crystal Ball in low-$p_{\text{TThrust}}^{\ell\ell\gamma}$ categories, where there are enough events for such comparison. The μ^{CB} values in this table are shifted +90 MeV to match the the central value from the combined ATLAS and CMS measurement, m_H = 125.09 GeV [25].

or as two resolved electrons. Its contribution is estimated from corresponding MC samples (see Section 4.3.2) and modelled using the same shape as signal in each category. This means that all parameters of the double-sided Crystal Ball function are fixed to those from the signal fit, and only the normalization is adjusted to $H \to \gamma\gamma$ MC yields. This is done because there are insufficient $H \to \gamma\gamma$ yields in MC in low-populated categories (high-$p_{\text{TThrust}}^{\ell\ell\gamma}$ and VBF) to perform a reliable fit. At the same time, the Higgs boson peak from the $H \to \gamma\gamma$ process is expected to look very similar to the $H \to \gamma^*\gamma \to \ell^+\ell^-\gamma$ signal peak, specifically in terms of resonance mass and resolution. Since there are enough $H \to \gamma\gamma$ events in low-$p_{\text{TThrust}}^{\ell\ell\gamma}$ categories, a comparison was performed there between the double-sided Crystal Ball fit with all parameters floating and the renormalized one with parameters from the signal fit. The fits were found to be in good agreement, further justifying the use of the signal shape for $H \to \gamma\gamma$ background modelling in all categories. The main parameters (σ_{CB} and μ_{CB}) of the signal fit function and the $H \to \gamma\gamma$ fit function in the low-$p_{\text{TThrust}}^{\ell\ell\gamma}$ categories are compared in Table 4.6. $H \to \gamma\gamma$ fit functions along with their parameters are shown in Appendix A.

4.6.2 Non-resonant background

The dominant background in this analysis comes from non-resonant SM processes in the $\ell\ell\gamma$ final state. There are small contributions from events with misidentified photons, electrons, or muons. The expected $m_{\ell\ell\gamma}$ distribution for the combined non-resonant background is smoothly falling, and can be reliably fitted with a function of just a few parameters. This makes it possible to perform the fit in data, combining the signal function (described earlier) with a chosen smoothly falling function in each category (when the data is blinded, the fit is performed in sidebands and is extrapolated into the 120 GeV $< m_{\ell\ell} <$ 130 GeV blinded region). This fit is performed at the statistical analysis step, where the parameters of background functions in each category serve as additional unconstrained nuisance parameters in the statistical model, described in Section 4.8.

However the choice to use a specific function for the background fit introduces the possibility of a systematic bias in the signal yields extracted from the statistical model. The evaluation of this bias relies on a background template constructed using MC simulation and data-driven techniques, and serves two purposes: choosing a function with a low bias, and estimating the systematic uncertainty associated with any residual bias.

To construct the background template, the dominant SM non-resonant $\ell\ell\gamma$ contribution is estimated with MC simulation (details are given in Section 4.3.2), while smaller contributions from events with a misidentified photon ($\ell\ell$ + jet) and events with misidentified leptons (γ + jets) are estimated in data. First, fractions of $\ell\ell$ + jet and γ + jets events are estimated in each category. Then, shapes of the corresponding $m_{\ell\ell\gamma}$ distributions are extracted from data, and the $\ell\ell$ + jet and γ + jets background

templates are constructed. To complete the combined background template, the templates for non-resonant SM $\ell\ell\gamma$, $\ell\ell$ + jet, and γ + jets backgrounds are stacked in each category, according to the estimated fractions of the corresponding background. The combined background template is normalized to the sidebands of the $m_{\ell\ell\gamma}$ distribution in data.

Estimating fractions of background events

Events with misidentified photon typically come from $\ell\ell$ + jet events, with a jet misidentified as a photon. This background is greatly reduced by photon isolation and identification requirements, because it typically features a poorly isolated calorimeter deposit (sometimes with two energy maxima resolved in the calorimeter strip layer) and some leakage into the hadronic calorimeter. Therefore a sample enriched in $\ell\ell$ + jet events can be created by inverting photon isolation and identification requirements.

The fraction of events with a misidentified photon in the signal region is estimated in the following way: first, events are selected in data using the full analysis selection but without the photon isolation requirement. Then, calorimeter isolation (defined in Section 3.2.2) is employed to separate events into several different samples. Calorimeter isolation is calculated inside a cone with $\Delta R = 0.2$, referred to as topoetcone20.

Events with photon candidates satisfying topoetcone20$/E_{\mathrm{T}} < 0.06$, where E_{T} is the transverse energy corresponding to the photon, and passing the tight identification criteria are labelled as "sample A" (see Figure 4.10). The quantity that is needed to be estimated is the number of events with misidentified photons $N_A^{\text{fake }\gamma}$ in sample A.

Events with photon candidates with topoetcone20$/E_{\mathrm{T}} > 0.16$, on the other hand, represent a sample ("sample B") of almost exclusively $\ell\ell$ + jet events. Therefore, $\ell\ell$ + jet event yields are nearly equal to the total yields in this sample, $N_B \simeq N_B^{\text{fake }\gamma}$.

Two more samples of events are obtained by selecting events with photon candidates that *failed* the tight identification criteria used in the analysis. Sample C corresponds to events failing photon identification with topoetcone20$/E_{\mathrm{T}} < 0.06$, while sample D contains events failing photon identification with topoetcone20$/E_{\mathrm{T}} > 0.16$. Assuming no correlation between isolation and identification criteria[8], the number of events with misidentified photon, $N_A^{\text{fake }\gamma}$, can then be estimated as follows (the so-called $ABCD$ method):

$$N_A^{\text{fake }\gamma} = N_C \times \frac{N_B}{N_D}. \tag{4.7}$$

The method is illustrated schematically in Figure 4.10. There is a correction to remove the contribution from events with real photons from the SM $\ell\ell\gamma$ background among events failing the photon identification requirement, i.e. in samples C and D. The MC simulated $\ell\ell\gamma$ events failing photon identification are used to estimate this contribution, which is then subtracted from N_C and N_D. As a final result, the contribution of events with a misidentified photon is estimated at around 6-11% depending on the category; precise fractions are given in Table 4.7.

Another background arises from events with a misidentified electron or muon that is in fact a hadronic jet, a lepton from heavy-flavour decays or an electron from a photon conversion (γ+jets). This background is estimated in data by performing a template fit to the isolation variable topoetcone20$/E_{\mathrm{T}}$

[8] In fact there is a small correlation, but it is negligible comparing to statistical uncertainty associated with spurious signal (described in Section 4.6.3).

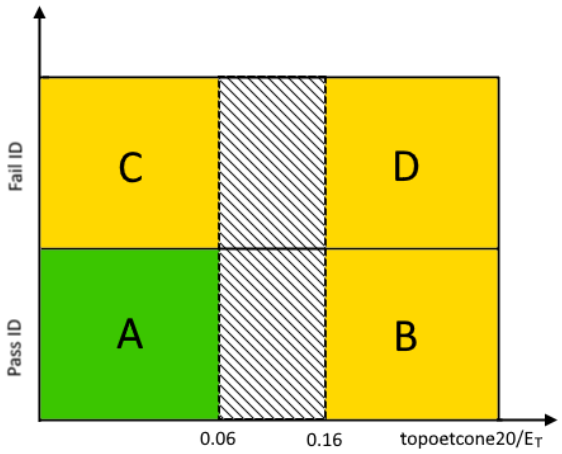


Figure 4.10: An illustration of the *ABCD* method used for $\ell\ell$ + jet background estimation. Events in data are separated into one signal region, *A*, and three control regions *B*, *C* and *D* using isolation and identification criteria. The number of events with a misidentified photon, $N_A^{\text{fake }\gamma}$, can be estimated from the control regions as $N_A^{\text{fake }\gamma} = N_C \times \frac{N_B}{N_D}$ (assuming no correlation between isolation and identification criteria).

of the subleading lepton.

The template for the SM $\ell\ell\gamma$ background is built using $\ell\ell\gamma$ MC samples. The template for the background from misidentified leptons (γ + jets) is built from events in data failing subleading lepton identification requirements: medium ID criteria in the resolved electron channel, merged electron ID in the merged electron channel and impact parameter requirements for the muon channel. This template is then corrected by subtracting the contribution from SM $\ell\ell\gamma$ events failing the identification requirements, estimated using $\ell\ell\gamma$ MC samples. The two templates are then used in a fit to the topoetcone20/E_{T} distribution in the data sidebands (with all signal selection but the region 120 GeV $< m_H <$ 130 GeV excluded). In this fit, the shapes of the topoetcone20/E_{T} distributions for the two templates remain unchanged, while fractions corresponding to the SM $\ell\ell\gamma$ background, f_{SM}, and to γ + jets, f_{fake}, are floating. The best-fit fraction f_{fake} is then extracted.

The estimated fractions of γ + jets background vary between categories (primarily because the rates at which objects are misidentified are different for different objects), with 4% in the muon low-$p_{\text{TThrust}}^{\ell\ell\gamma}$ category, 2% in the merged electron low-$p_{\text{TThrust}}^{\ell\ell\gamma}$ category, and 30% in the resolved electron low-$p_{\text{TThrust}}^{\ell\ell\gamma}$ category. In the high-$p_{\text{TThrust}}^{\ell\ell\gamma}$ and VBF categories the estimate has large statistical uncertainties. All fractions are summarized in Table 4.7.

Modelling the $m_{\ell\ell\gamma}$ shape of background sources

To complete the combined background template, besides the contribution of γ + jets and $\ell\ell$ + jet backgrounds in the signal region in terms of yields fraction as described above, also the corresponding $m_{\ell\ell\gamma}$ shape is needed. It is extracted in data from similar control regions as used for the yield fraction estimates.

For the $\ell\ell$ + jet background, the photon candidate is required to fail either the final identification or the isolation requirements. For the γ + jets background, a subleading lepton candidate is required to fail medium ID criteria in the resolved electron channel, merged electron ID in the merged electron

Category	$f_{\ell\ell+\text{jet}}$	$f_{\gamma+\text{jets}}$
Low-$p_{\text{TThrust}}^{\ell\ell\gamma}$ muons	0.08 ± 0.00	0.040 ± 0.005
Low-$p_{\text{TThrust}}^{\ell\ell\gamma}$ resolved electrons	0.08 ± 0.00	0.27 ± 0.02 (0.21 for $\Delta R_{\ell\ell} < 0.2$; 0.06 for $\Delta R_{\ell\ell} > 0.2$)
Low-$p_{\text{TThrust}}^{\ell\ell\gamma}$ merged electrons	0.11 ± 0.00	0.016 ± 0.001
VBF muons	0.09 ± 0.09	0.25 ± 0.10
VBF resolved electrons	0.06 ± 0.04	0.67 ± 0.43
VBF merged electrons	—	0.046 ± 0.026
High-$p_{\text{TThrust}}^{\ell\ell\gamma}$ muons	0.06 ± 0.02	0.15 ± 0.05
High-$p_{\text{TThrust}}^{\ell\ell\gamma}$ resolved electrons	0.08 ± 0.03	0.49 ± 0.12
High-$p_{\text{TThrust}}^{\ell\ell\gamma}$ merged electrons	0.06 ± 0.02	0.017 ± 0.006

Table 4.7: Extracted $\ell\ell$ + jet and γ + jets background fractions in data for analysis categories together with statistical uncertainties. For the $\ell\ell$ + jet background in the VBF merged electrons category there was an insufficient number of events to perform a fit.

channel, and impact parameter requirements for the muon channel. For both backgrounds, the contribution from SM $\ell\ell\gamma$ is removed from the data template using $\ell\ell\gamma$ MC samples. The procedure is done in this way for each category. However in the VBF categories, due to large statistical uncertainties, the selection requirements are somewhat relaxed to increase the number of selected events, requiring only $m_{jj} > 400$ GeV and $\Delta\eta_{jj} > 2.5$. Also in the resolved electron low-$p_{\text{TThrust}}^{\ell\ell\gamma}$ category, the estimate for γ + jets is performed separately for two groups of events, with $\Delta R > 0.2$ and $\Delta R < 0.2$, as two populations with distinct $m_{\ell\ell\gamma}$ distributions and rates of misidentified objects were found.

The non-resonant SM $\ell\ell\gamma$, $\ell\ell$ + jet and γ + jets background templates in each category are then stacked, according to the estimated fractions of the corresponding background (the SM $\ell\ell\gamma$ is assumed to take the remaining fraction, i.e. one minus sum of the $\ell\ell$ + jet and γ + jets fractions). In the VBF and resolved electron high-$p_{\text{TThrust}}^{\ell\ell\gamma}$ categories, due to large statistical uncertainties, the fractions are taken from the corresponding low-$p_{\text{TThrust}}^{\ell\ell\gamma}$ categories. Finally, the combined background template is normalized to the sidebands of the $m_{\ell\ell\gamma}$ distribution in data. The resulting templates in the low-$p_{\text{TThrust}}^{\ell\ell\gamma}$ categories are shown in Figure 4.11. For comparison, the data in the sidebands is also shown, and good agreement of the shapes can be seen between the templates and data sidebands.

4.6.3 Background functions and spurious signal

In each category, the non-resonant background contribution is fitted in data with a smoothly falling function (combined with the signal function discussed before). The fit functions considered are: exponential of a polynomial of orders 1 to 3, Bernstein polynomials [75] of order 4 and 5 and a power-law function. The exponential of a polynomial is defined as

$$\mathcal{B}\left(m_{\ell\ell\gamma}\right) = N \cdot \exp\left(-\sum_{i=1}^{n} \alpha_i m_{\ell\ell\gamma}^i\right) \tag{4.8}$$

where n is degree of the polynomial, α_i are free parameters, and N is the normalization factor. Bernstein polynominals are defined as

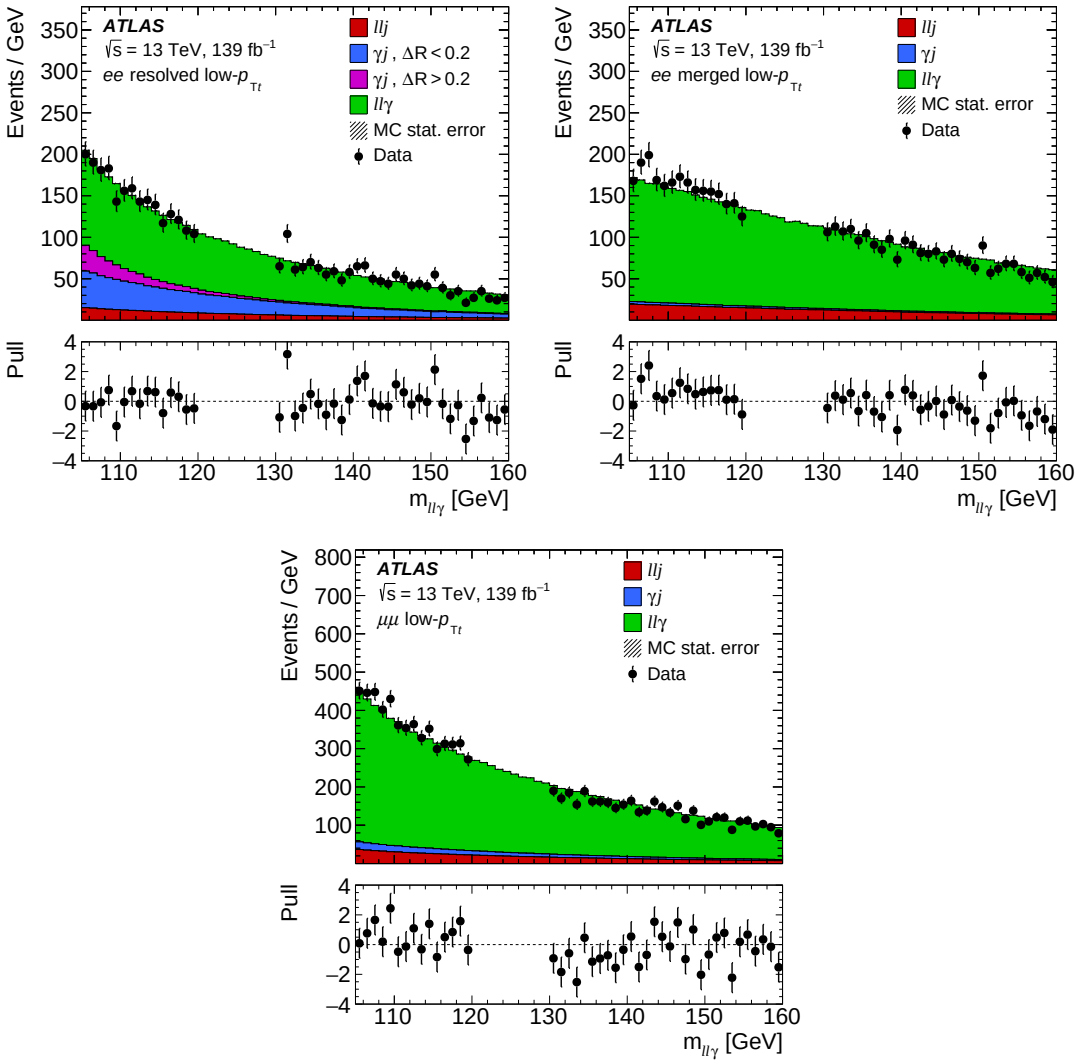


Figure 4.11: Non-resonant background templates in the low-$p^{\ell\ell\gamma}_{\text{TThrust}}$ categories [66]. Contributions from SM $\ell\ell\gamma$, γ + jets and $\ell\ell$ + jet (including two ΔR populations for resolved electrons) are shown. Data in the $m_{\ell\ell\gamma}$ sidebands is also shown. The bottom panels show the pull distribution of the data with respect to the sum of background components.

$$\mathcal{B}\left(m_{\ell\ell\gamma}\right) = N \cdot \sum_{i=0}^{n} \left(\alpha_i m_{\ell\ell\gamma}^{i} \cdot (1 - m_{\ell\ell\gamma})^{n-i} \cdot \frac{n!}{i! \cdot (n-i)!} \right), \tag{4.9}$$

where all α_i are free parameters as well, except α_n which is fixed to 1 (because Bernstein polynomials of degree n require $n + 1$ parameters).

The power law function is defined as

$$\mathcal{B}\left(m_{\ell\ell\gamma}\right) = N \cdot m_{\ell\ell\gamma}^{\alpha_1}. \tag{4.10}$$

Each of the considered functions is then used in a signal-plus-background fit to the background template (discussed in the previous section) instead of data. Since the background template contains no signal, any signal yield resulting from this fit is *spurious* and represents a bias associated with the background function choice. To keep this bias small, requirements are placed on the maximum allowed spurious signal, and functions that do not satisfy these requirements are removed from consideration. To measure the spurious signal, fits are performed in the mass range [110, 160] GeV; the fits are repeated with the Higgs boson mass hypothesis scanned from 121 GeV to 129 GeV (roughly 2σ below and above the $m_H = 125$ GeV value) in steps of 1 GeV, and the maximum value of spurious signal $|N_s|$ over this range is considered:

$$N_{sp} = \max_{121\ \text{GeV}<m_H<129\ \text{GeV}} |N_s(m_H)|. \tag{4.11}$$

For a function to pass the spurious signal test, it should satisfy at least one of the following criteria:

- $N_{sp} < 10\%\ N_{s,exp}$
- $N_{sp} < 20\%\ \sigma_{bkg}$

where $N_{s,exp}$ is the expected number of signal events and σ_{bkg} is the statistical uncertainty on the number of signal events from the fit (obtained from the fit of the signal-plus-background model to the background-only template).

These criteria are slightly modified to accommodate the categories where the background template has large statistical uncertainties compared to the expected number of signal events. An additional metric ζ_s for each function is evaluated, defined as

$$\zeta_s(m_H) = \begin{cases} N_s + \Delta_\text{T}, & N_s + \Delta_\text{T} < 0 \\ N_s - \Delta_\text{T}, & N_s - \Delta_\text{T} > 0 \\ 0, & \text{otherwise} \end{cases} \tag{4.12}$$

where Δ_T is the local (for the particular value of m_H) statistical uncertainty of the background template, i.e. the statistical uncertainties associated with the limited number of MC events. As previously, the maximum ζ_s in the range 121 GeV $< m_H <$ 129 GeV is taken as the spurious signal, ζ_{sp}. Then ζ_{sp} is used instead of N_{sp} to check whether a function passes. In other words, the criteria are slightly relaxed for categories with large MC template statistical uncertainties. It should be noted that this only allows more functions to pass for the statistically limited categories, but the systematic uncertainty associated with the bias is still assigned as N_{sp}, so there is no underestimation of uncertainty.

Finally, a goodness of fit criteria must be satisfied (p-value(χ^2) $> 1\%$) for a background-only function fit to the template, and among passing functions the function with the fewest degrees of freedom is chosen. The latter minimizes the statistical uncertainty of the resulting fit, and therefore improves the sensitivity. The resulting selected functions in each category and residual spurious signal yields are summarized in Table 4.8. These residual spurious signal yields are taken into account as an additional source of systematic uncertainty, discussed in Section 4.7.

Category	Function	N_{sp}	$N_{sp}/N_{s,exp}$ [%]
Low-$p_{\mathrm{TThrust}}^{\ell\ell\gamma}$ muons	$\exp\left(-(\alpha_1 m_{\ell\ell\gamma} + \alpha_2 m_{\ell\ell\gamma}^2)\right)$	-12.1	-17.7
Low-$p_{\mathrm{TThrust}}^{\ell\ell\gamma}$ resolved electrons	$m_{\ell\ell\gamma}^{\alpha_1}$	-8.39	-34.5
Low-$p_{\mathrm{TThrust}}^{\ell\ell\gamma}$ merged electrons	$\exp\left(-(\alpha_1 m_{\ell\ell\gamma} + \alpha_2 m_{\ell\ell\gamma}^2)\right)$	4.47	13.7
VBF muons	$m_{\ell\ell\gamma}^{\alpha_1}$	-0.467	-32.9
VBF resolved electrons	$\exp\left(-\alpha_1 m_{\ell\ell\gamma}\right)$	-0.137	-29.7
VBF merged electrons	$m_{\ell\ell\gamma}^{\alpha_1}$	-0.254	-29.9
High-$p_{\mathrm{TThrust}}^{\ell\ell\gamma}$ muons	$m_{\ell\ell\gamma}^{\alpha_1}$	0.315	7.34
High-$p_{\mathrm{TThrust}}^{\ell\ell\gamma}$ resolved electrons	$m_{\ell\ell\gamma}^{\alpha_1}$	-0.737	-61.6
High-$p_{\mathrm{TThrust}}^{\ell\ell\gamma}$ merged electrons	$m_{\ell\ell\gamma}^{\alpha_1}$	0.231	8.75

Table 4.8: Functions for the non-resonant background fit selected with the spurious signal procedure. The residual spurious signal N_{sp} is also shown, as well as its size relative to the expected signal yields, $N_{sp}/N_{s,exp}$.

4.7 Systematic uncertainties

The sources of systematic uncertainty can be broadly divided into two main categories: *theory*, associated with the theory predictions, and *experimental*, associated with the measurement. Some *statistical* uncertainties associated with the employed experimental methods (e.g. in background templates, reconstruction algorithms, etc[9]) are also treated as experimental systematic uncertainties, to identify them as uncertainties that could be reduced by improving the experimental methods, and not by collecting more data.

Systematic uncertainties affecting the expected yield and shape of the signal and the resonant $H \to \gamma\gamma$ background $m_{\ell\ell\gamma}$ distributions are taken into account by first measuring the impact of the $\pm 1\sigma$ uncertainty variations on the fitted signal yield. For the non-resonant background, a bias associated with the fitting function choice is treated as an additional uncertainty on the signal yields, based on the spurious signal procedure described in Section 4.6.3.

The impact of the $\pm 1\sigma$ uncertainty variations is gauged for the signal and resonant background using the respective MC samples. The $m_{\ell\ell\gamma}$ distribution in the signal region is obtained using the modified event weights (or modified object's properties) and fitted with the double-sided Crystal Ball function described in Section 4.6.1. The relevant parameter X of the signal function is extracted from this fit, depending on the uncertainty source. For sources affecting signal yields, $X = N$ (the normalization); for uncertainties affecting the energy scale, $X = \mu^{\mathrm{CB}}$ (the mean of the gaussian component); for the energy resolution uncertainties, $X = \sigma^{\mathrm{CB}}$ (the width of the gaussian component). The relative uncertainty for the parameter X is then obtained using the following relations:

$$\Delta(X)^{+1\sigma} = \frac{X_{\mathrm{syst}}^{+1\sigma} - X_{\mathrm{nom}}}{X_{\mathrm{nom}}}, \Delta(X)^{-1\sigma} = \frac{X_{\mathrm{syst}}^{-1\sigma} - X_{\mathrm{nom}}}{X_{\mathrm{nom}}}, \tag{4.13}$$

where X_{nom} is the value of the parameter X from the baseline (nominal) signal fit and $X_{\mathrm{syst}}^{\pm 1\sigma}$ is the value of the parameter X from the fit to the $m_{\ell\ell\gamma}$ distribution corresponding to the $\pm 1\sigma$ uncertainty variations. For the special case of spurious signal uncertainty, the absolute values of spurious signal yields for each category are used, given in Table 4.8.

The obtained $\Delta(X)^{\pm 1\sigma}$ values are used to define nuisance parameters added into the likelihood model, and the impact of systematic uncertainties on the ratio of the observed event yield classified as signal to the expected SM signal yield, μ, is determined as described in Section 4.8.2. Here the impact is reported in percent of the observed μ value.

The combined observed impact of systematic uncertainties on μ is 11%, which corresponds to only 35% of the size of the statistical uncertainty. Therefore the result is dominated by the statistical uncertainty and will significantly improve with more data. The observed impact of individual sources (groups of sources) of systematic uncertainty is discussed in more detail below and is shown in Table 4.9. The impact is either quoted with respect to the measured μ value, or otherwise (where explicitly mentioned) with respect to the reported cross-section times branching ratio, $\sigma \times \mathcal{B}$.

4.7.1 Experimental uncertainties

The dominating experimental uncertainty is associated with the non-resonant background fit function choice (spurious signal, detailed in Section 4.6.3), with an impact of 6.1%.

[9] One example is statistical uncertainty in the merged electron ID efficiency measurement, discussed in Section 4.4.2.

Uncertainty source	μ	$\sigma \times \mathcal{B}$
Spurious Signal	6.1	
$\mathcal{B}(H \to \ell\ell\gamma)$	5.8	–
QCD scale	4.4	1.1
ℓ, γ, jets	4.0	
PDF	2.3	0.9
Luminosity	1.7	
Pile-up	1.7	
α_S	1.6	0.3
Minor prod. modes	0.8	
$H \to \gamma\gamma$ background	0.7	
Parton Shower	0.3	
Total systematic	11	7.9
Statistical	31	
Total	33	32

Table 4.9: Relative systematic uncertainties (in percent of the observed μ value) in the measured signal strength and the measured cross-section times branching ratio. The uncertainties are symmetrized and ordered by the observed impact. The statistical and total uncertainties are shown for comparison.

Other significant experimental uncertainties (with a combined impact of 4.0%) are associated with objects in the event. They arise from the measurement of the object's identification and isolation efficiencies, as well as the object's energy/momentum calibration. Notably, that includes uncertainties associated with the merged electron efficiency measurement and energy calibration, described in Section 4.4. For the photons, resolved electrons, muons and jets, the uncertainties are estimated centrally (see references in Section 3.2).

The uncertainty in the combined 2015–2018 integrated luminosity is 1.7 % [52], obtained using the LUCID-2 detector [51] for the primary luminosity measurements. Another uncertainty with 1.7% impact is associated with the modelling of pile-up in MC simulation. An uncertainty on the Higgs boson mass measurement, estimated at ±240 MeV [25], also contributes albeit with a very small (less than 0.2%) impact.

Two uncertainties are considered for the $H \to \gamma\gamma$ background, with a combined impact of 0.7%. The $H \to \gamma\gamma$ cross-section measurement uncertainty is estimated at ±18% of the $H \to \gamma\gamma$ cross-section [56]. The uncertainty on the rate with which a photon is misidentified or misreconstructed (as a dilepton pair) is estimated at ±50%. In Ref. [44], this rate is measured as a function of η, and is 50% at its maximum; therefore ±50% is used as a conservative estimate for this uncertainty. Additionally, uncertainties associated with luminosity and Higgs boson mass measurements are also taken into account for the $H \to \gamma\gamma$, fully correlated with the corresponding signal uncertainties, and are included in the reported impacts.

4.7.2 Theory uncertainties

A significant theory uncertainty is associated with the branching ratio $\mathcal{B}(H \to \ell\ell\gamma)$ prediction. An estimate of 5.8% from Ref. [24] is used; more details are given in Section 4.3.1.

Other sources of theory uncertainty are associated with predictions and assumptions in the MC generation of the signal $H \to \gamma^*\gamma \to \ell^+\ell^-\gamma$ samples (see Section 2.2.1). One of these uncertainties is associated with the choice of QCD factorization and renormalization scales, with a total impact of 4.4% (estimation of this uncertainty is explained in more detail below).

Another uncertainty arises from the measurement of parton density functions (PDFs), evaluated using the eigenvectors of the `PDF4LHC15` PDF set, with an impact of 2.3%. The assumed value of strong interaction coupling strength, α_s, used in the PDF set also introduces a source of uncertainty with an impact of 1.6%. The PDF and α_s variations are provided in Ref. [15].

Finally, there is an uncertainty associated with the modelling of the underlying event and the QCD radiation (parton showers) in `Pythia 8` with an impact of 0.3%. Its estimation is based on a comparison with `Herwig 7`, described in Section 4.3.1.

For the $\sigma \times \mathcal{B}$ measurement, most of the uncertainties have the same relative impact as for the signal strength measurement. However the impact of the QCD scale, PDF and α_s uncertainties is reduced, to 1.1% for the QCD scale, 0.9% for the PDF uncertainties, and 0.3% for the α_s uncertainty, respectively. This is caused by the fact that for the $\sigma \times \mathcal{B}$ measurement, only the contribution of the uncertainties associated with acceptance and bin-to-bin migration remains. This also means that there is no uncertainty on branching ratio associated with this measurement.

QCD scale uncertainties

For the MC generation, the assumed values for the QCD factorization and renormalization scales (μ_F and μ_R, respectively) introduce a systematic uncertainty (more details are given in Section 2.2.1). Traditionally these uncertainties are estimated by varying the corresponding scales, setting them to 2μ and $\mu/2$, and obtaining modified event weights, which are then stored in the MC samples. However, this method may underestimate the true uncertainty in cases where events are categorized into exclusive regions of phase space, especially when it is done using cuts on jet-related observables. An alternative approach, originally developed for the Simplified Template Cross Sections (STXS) [76], is based on categorizing the events in several mutually exclusive regions of truth-particle phase space[10] (bins), depending on p_{T}^{H}, jet multiplicity and other variables, and correctly propagating QCD scale uncertainties into the categories. Another advantage of the STXS approach is that it serves to unify event categorization in different Higgs boson measurements (most notably measurements of cross sections in different production modes and kinematic regions) and thereby provides a common framework for evaluation of the QCD scale uncertainties.

The STXS scheme features different categorization (binning) depending on the Higgs boson production mode, shown in Figure 4.12. The propagation of uncertainties is based on the extension of the Stewart-Tackmann (ST) method. The original method for binning in jet multiplicity is described in Ref [77, 78] and is briefly illustrated here.

The main idea of the ST method is the following:

[10] A phase space of the original event, not modified by the detector effects during reconstruction. More information is provided in Section 3.2.6.

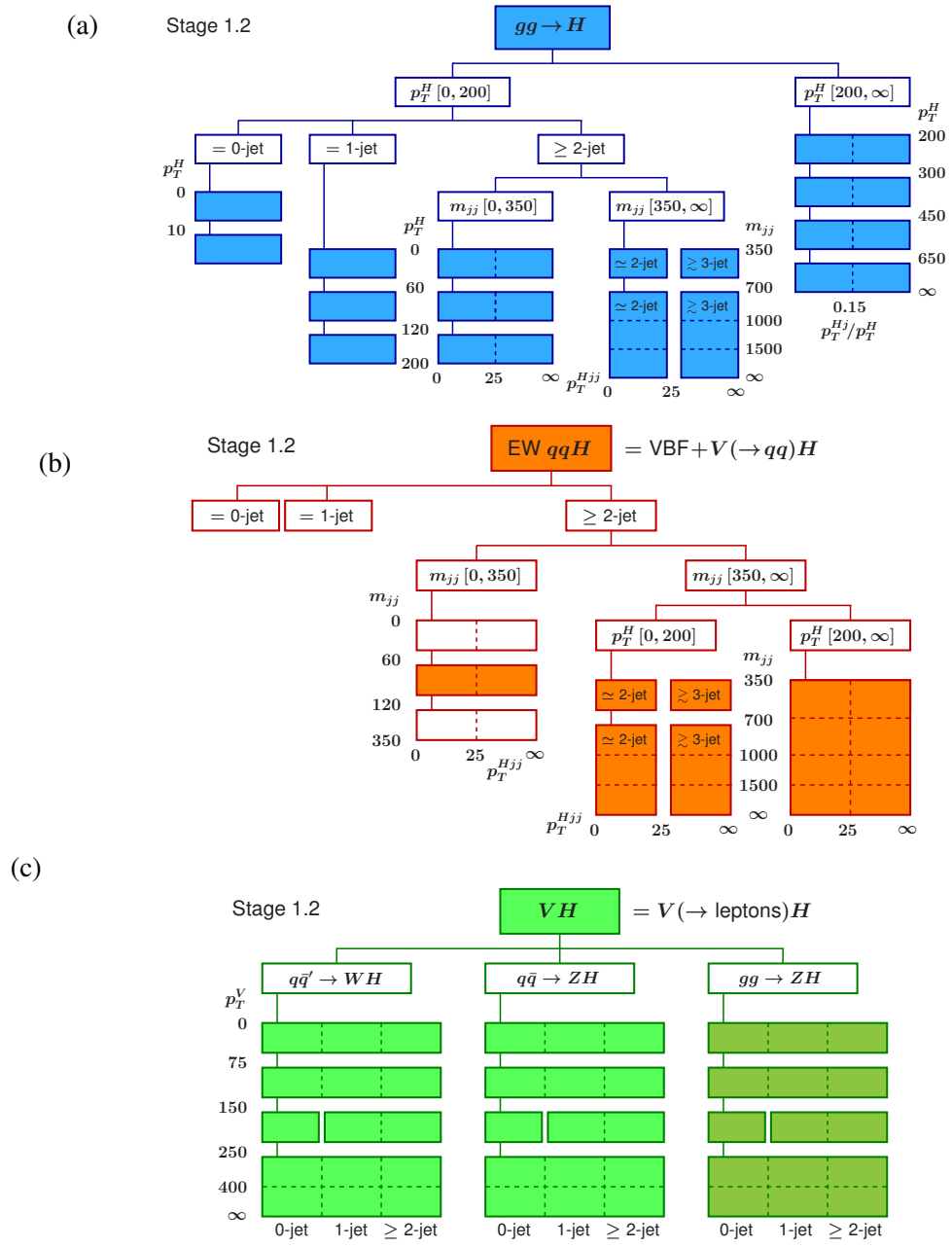


Figure 4.12: Categorization of events in Simplified template cross sections (STXS) [76] for the ggH (a), VBF (b) and VH (c) Higgs boson production modes [79].

- The uncertainties for the *exclusive* categories, with exactly N jets, can be calculated using the uncertainties for *inclusive* categories, with $\geq N+1$ jets, and the uncertainty for the combined measurement (with $\geq N$ jets).
- The uncertainties for the inclusive categories, with $\geq N+1$ jets, and the combined measurement,

with $\geq N$ jets, can be reliably estimated from traditional scale variations.

For the case of two categories with N jets and $\geq N+1$ jets, separated by a certain p_{T}^{j} requirement, it is evident that the uncertainty for the total cross-section, $\sigma_{\geq N}$, should not depend on the p_{T}^{j} requirement. Therefore the uncertainty can be split into two components: migration uncertainty, Δ_M, 100% anti-correlated between categories, and a 100% correlated yields uncertainty Δ^Y. The correlations between uncertainties are expressed with a 2×2 covariance matrix:

$$\begin{pmatrix} (\Delta_N^Y)^2 & \Delta_N^Y \Delta_{\geq N+1}^Y \\ \Delta_N^Y \Delta_{\geq N+1}^Y & (\Delta_{\geq N+1}^Y)^2 \end{pmatrix} + \begin{pmatrix} \Delta_M^2 & -\Delta_M^2 \\ -\Delta_M^2 & \Delta_M^2 \end{pmatrix}. \tag{4.14}$$

The uncertainty for the exclusive and inclusive categories, Δ_N and $\Delta_{\geq N+1}$, as well as for the combined measurement, $\Delta_{\geq N}$, are expressed as the following:

$$\begin{aligned} \Delta_N^2 &= (\Delta_N^Y)^2 + \Delta_M^2, \\ \Delta_{\geq N+1}^2 &= (\Delta_{\geq N+1}^Y)^2 + \Delta_M^2, \\ \Delta_{\geq N} &= \Delta_{\geq N}^Y = \Delta_N^Y + \Delta_{\geq N+1}^Y. \end{aligned} \tag{4.15}$$

The $\Delta_{\geq N}$, $\Delta_{\geq N+1}$ are estimated from QCD scale variations, $\Delta_{\geq N} = \Delta_{\geq N}^{\mu}$, $\Delta_{\geq N+1} = \Delta_{\geq N+1}^{\mu}$. Given that they also satisfy the relation

$$\Delta_{\geq N}^{\mu} = \Delta_N^{\mu} + \Delta_{\geq N+1}^{\mu}, \tag{4.16}$$

only two uncertainties are independent. Therefore one of the values, either Δ_N^Y, $\Delta_{\geq N+1}^Y$, or Δ_M, or a certain combination of them must be set explicitly. In Ref. [78] it is shown that setting $\Delta_{\geq N+1}^Y = 0$ achieves the best result, by correctly reflecting the correlations between uncertainties for different jet multiplicities. With this, all three components can be explicitly defined:

$$\begin{aligned} \Delta_{\geq N+1}^Y &= 0, \\ \Delta_N^Y &= \Delta_{\geq N}^{\mu}, \\ \Delta_M &= \Delta_{\geq N+1}^{\mu}. \end{aligned} \tag{4.17}$$

In this analysis, QCD scale uncertainties are calculated with the ST method for the STXS categorization. The analysis categories (described in Section 4.5.3) are defined with different variables than the STXS categories shown in Figure 4.12. However the QCD uncertainties are still applicable, as long as the analysis categorization is highly correlated with the STXS categories, or in other words if the category boundaries in the analysis correspond to the binning assumed in the STXS scheme. For the VBF and VH Higgs boson production modes, for which the event yields are comparatively low, this is simply assumed to be true (because a deviation there will not have a significant effect on the total uncertainty estimate). However, if the category boundaries for the ggF production mode do not correspond to the STXS categories, it could introduce a large discrepancy. In particular there are two requirements used in the analysis categorization that are expected to be correlated with the STXS categorization:

- The $p_{\mathrm{TThrust}}^{\ell\ell\gamma}$ cut, defining high-$p_{\mathrm{TThrust}}^{\ell\ell\gamma}$ categories, is highly correlated with p_{T}^{H}.
- The angular separation between dijet and $\ell\ell\gamma$ systems, $\Delta\phi_{\ell\ell\gamma,jj}$ (one of the requirements defining VBF analysis category), serves as an implicit third jet veto, whereas p_{T}^{Hjj} is used for the same

purpose in the STXS.

To test whether the $p_{\mathrm{TThrust}}^{\ell\ell\gamma}$ cut has a similar effect as any of the p_{T}^{H} boundaries in the STXS scheme, two p_{T}^{H} distributions of ggF MC signal events satisfying all selection requirements are plotted, one with and the other without the $p_{\mathrm{TThrust}}^{\ell\ell\gamma} > 100$ GeV requirement, shown in Figure 4.13. If the effect of the $p_{\mathrm{TThrust}}^{\ell\ell\gamma}$ cut is similar to any of the p_{T}^{H} boundaries, a change in the shape of the two distributions would occur close to one of the STXS boundaries in p_{T}^{H}. This appears to be the case; specifically the shapes diverge at around $p_{\mathrm{T}}^{H} = 120$ GeV.

The same test is performed with the events in the ggF MC signal samples that satisfy the VBF category requirements, by plotting one p_{T}^{Hjj} distribution with all VBF category requirements, and another with the same selection but with the $\Delta\phi_{\ell\ell\gamma,jj}$ requirement removed. The corresponding p_{T}^{Hjj} distributions are shown in Figure 4.13. As can be seen, the two shapes do not start to diverge around the STXS boundary of $p_{\mathrm{T}}^{Hjj} = 25$ GeV, but rather at a much smaller p_{T}^{Hjj} value, and there is no apparent strong correlation between these two variables. This indicates that the QCD uncertainties for the VBF-like events in the ggF category should be evaluated separately, not using the pre-defined STXS bins.

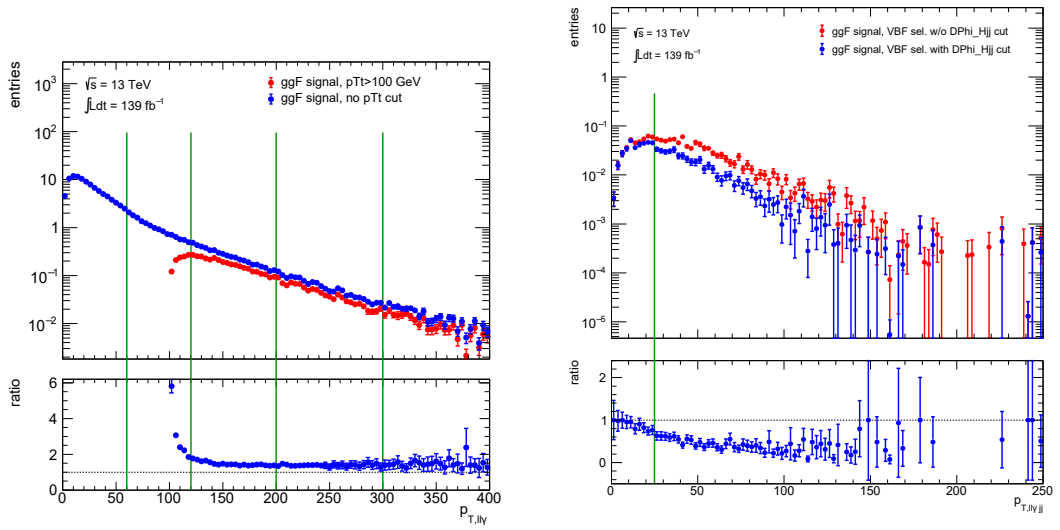


Figure 4.13: Left: p_{T}^{H} distribution of selected events with and without a $p_{\mathrm{TThrust}}^{\ell\ell\gamma} > 100$ GeV requirement. Right: p_{T}^{Hjj} distribution of selected events passing VBF category requirements, with and without the $\Delta\phi_{\ell\ell\gamma,jj}$ requirement. Both distributions are obtained using signal MC ggF samples. The STXS category boundaries are superimposed in green. The dashed line in the lower panel indicates unity ratio.

The $\Delta\phi(H, jj)$ requirement serves as an implicit third jet veto, and the ST method is used in the same way as described above for two categories in jet multiplicity: exactly 2 jets, and $\geq$ 2 jets. The evaluation is based on the result published in Ref. [80]. This result contains the Higgs boson differential cross-section measured in bins of $\pi - \Delta\phi(H, jj)$ along with uncertainties for all bins and the corresponding correlation matrix. These uncertainties are translated into relative uncertainties in each bin, which are assumed to be the same for the $H \to \gamma^{*}\gamma \to \ell^{+}\ell^{-}\gamma$ events from signal MC samples, provided they are selected with the same (VBF) selection requirements and one of the values

of the $\Delta\phi(H, jj)$ cut for which the measurement is performed in Ref. [80]. Therefore signal MC events passing VBF selection requirements are used, and uncertainties for the following scenarios are obtained: *no* $\Delta\phi(H, jj)$ requirement corresponds to the inclusive $\sigma_{\geq 2}$ cross-section, the *same* cut as in the VBF category $\Delta\phi_{\ell\ell\gamma,jj} > 2.8$ corresponds to the exclusive $\sigma_{=2}$ cross-section, and the *inverted* cut $\Delta\phi_{\ell\ell\gamma,jj} < 2.8$ corresponds to the inclusive $\sigma_{\geq 3}$ cross-section. The total uncertainties for the sum of all bins, $\Delta^{\mu}_{\geq 2}$, $\Delta^{\mu}_{=2}$, $\Delta^{\mu}_{\geq 3}$ are obtained using the correlation matrix. The yields and migration uncertainties can be subsequently estimated from Equation 4.17:

$$\begin{aligned} \Delta^{Y}_{\geq N+1} &= 0, \\ \Delta^{Y}_{N} &= \Delta^{\mu}_{\geq 2}, \\ \Delta_{M} &= \Delta^{\mu}_{\geq 3}. \end{aligned} \tag{4.18}$$

It should be noted that this estimation relies on the VBF selection being the same in this analysis as in Ref. [80] (this ensures validity of the relative uncertainties per bin and their correlations). This is not strictly the case, however the differences are minor, and therefore the uncertainties are still considered valid. The VBF selection used in this analysis is detailed in Section 4.5.3, while the following VBF selection was used for Ref. [80]: 2 jets with $p_{\mathrm{T}} > 25$ (30) GeV and $|\eta| < 2.5$ ($2.5 < |\eta| < 4.5$), $m_{jj} > 400$ GeV, $\Delta\eta_{jj} > 2.8$.

The uncertainties amount to $\Delta^{Y}_{N} = 20.8\%$ and $\Delta_{M} = 71.3\%$ compared to $\Delta^{Y}_{N} = 20\%$ and $\Delta_{M} = 23.5\%$ in the STXS scheme for the $\approx$ 2-jet ($p_{\mathrm{T}}^{Hjj} < 25$ GeV) and $\geq$ 3-jet ($p_{\mathrm{T}}^{Hjj} > 25$ GeV) categories. It should be emphasized that the 71.3% uncertainty is applied to a very small subset of events: ggF events passing the VBF selection *except* for the $\Delta\phi_{\ell\ell\gamma,jj}$ cut, and therefore the cumulative impact of this uncertainty on the results is very small.

4.8 Statistical analysis

A statistical analysis is employed to achieve the following goals:

- To quantify to our best knowledge which fraction of the observed event yield is the **signal** and which is the **background**;
- To measure the ratio between the observed event yield classified as signal and the expected SM signal yield (**signal strength**, μ), and evaluate statistical and systematic uncertainties in this measurement;
- To characterize the level of confidence in observing the signal, expressed by the ***p*-value** (and the corresponding **significance**, Z);
- To set an **upper limit** on the signal strength that is excluded at 95% confidence level;
- To measure the Higgs boson production **cross-section times $H \to \ell\ell\gamma$ branching ratio** in the fiducial region for $m_{\ell\ell} < 30$ GeV.

To achieve the goals outlined above, the extended likelihood formalism is used, in particular the profile likelihood ratio method [68]. The signal parametrization and the background functional forms obtained previously are used to construct the background-only and signal-plus-background likelihood models. The best-fit signal strength is found by maximizing the signal plus-background likelihood taking into account the statistical and systematic uncertainties, which are represented in the likelihood.

The p-value[11] – the probability of finding data with equal or greater incompatibility with the background-only hypothesis – is calculated from the test statistic employed by the profile likelihood ratio method using the asymptotic formulas detailed in Section 4.8.1. For the presented search, this result was additionally checked with a large number of pseudo-experiments (see Section 4.8.4 for details). The p-value is then converted into an equivalent significance Z, defined such that for a Gaussian-distributed random variable the probability to find it more than Z standard deviations above its mean is equal to the p-value. Consequently, the lower the p-value, the higher the significance Z. The significance is *observed* if the likelihood is constrained using data, and *expected* if the signal-plus-background Asimov dataset [68] is used, constructed from the expected signal-plus-background model, which is based on the SM expectation of the signal and background function parameters taken from fits to the data sidebands. The Asimov dataset is an artificially constructed dataset, and the calculated maximum likelihood estimators of each parameter should exactly match the Asimov dataset input parameter values [12] (see Section 4.8.1 for its role in the calculation).

The 95% confidence level upper limit on the signal strength is set by constraining parameters in the likelihood model using data, while the expected limit is set with a signal-plus-background Asimov dataset or with a background-only Asimov dataset. The two results have different interpretation: the result based on the background-only Asimov dataset represents an upper limit in the absence of the Higgs boson, while the result based on the signal-plus-background Asimov dataset represents an upper limit assuming the SM Higgs boson. The upper limits are typically only reported if there is

[11] A low p-value means that observed result is highly unlikely to be a random fluctuation and very likely to be a genuine signal.

[12] For an example with event counts in bins, that would mean that counts in the Asimov dataset are equal to their expectation values.

no significant ($Z > 3\sigma$) excess over the background-only prediction. In the case of this analysis, the observed excess surpasses the 3σ *evidence* threshold and therefore the upper limits are not reported in publications of the results, but are presented in this thesis.

The Higgs boson production cross-section times $H \to \ell\ell\gamma$ branching ratio in the $m_{\ell\ell} < 30$ GeV region is measured by multiplying the expected SM value with the measured signal strength. To correctly propagate the uncertainty for this measurement, some theory uncertainties are recalculated, because only the contribution of the uncertainties associated with the signal acceptance and bin-to-bin migration remains (more details are given in Section 4.7).

Because the sensitivity of the presented search with the current data set was not enough to improve the Higgs boson mass measurement, its mass was fixed at the value $m_H = 125.09$ GeV, corresponding to the central value of the combined ATLAS and CMS measurement [25].

Several cross-checks of the statistical procedure, both from the technical implementation side and for the consistency of the results themselves, are described in Section 4.9.3. The extended likelihood model formalism used is explained below.

4.8.1 Extended likelihood formalism

Likelihood definition

To construct the likelihood function, the probability density function $f_{\text{tot}}(m^i_{\ell\ell\gamma})$ of the invariant mass distribution for each candidate event i is evaluated from the signal and background functional forms obtained previously (see Section 4.6). The event counts are assumed to be Poisson-distributed and the likelihood definition is extended by multiplying it with the Poisson term.

The likelihood definition is further modified by multiplying it with constraint terms $G(\boldsymbol{\theta})$, which quantify the impact of the nuisance parameters $\boldsymbol{\theta}$ representing systematic uncertainties. Most of the constraint terms are Gaussian-distributed, with standard deviation set to $\pm 1\sigma$ uncertainty variations. Where the uncertainties are not exactly symmetric (the most common case), two Gaussian distributions are used, with standard deviation set to the $+1\sigma$ and to -1σ uncertainty variations, respectively. In the vicinity of the nominal value of the parameter a 6-degree polynomial interpolation is used to avoid discontinuity. In some cases, constraint terms are log-normal-distributed, particularly for certain parameters with nominal values in the vicinity of zero (such as energy resolution in GeV), to prevent the corresponding nuisance parameters from acquiring unphysical negative values. The background function parameters are also treated as nuisance parameters, but they are not constrained in the likelihood.

The resulting extended likelihood can be written as the following:

$$L\left(\mu, \boldsymbol{\theta} \middle| \{m^i_{\ell\ell\gamma}\}_{i=1..n}\right) = \frac{\mathrm{e}^{-N(\mu,\boldsymbol{\theta})} N^n(\mu, \boldsymbol{\theta})}{n!} \left(\prod_{i=1}^{n} f_{\text{tot}}(m^i_{\ell\ell\gamma}; \mu, \boldsymbol{\theta}) \right) \times G(\boldsymbol{\theta}). \tag{4.19}$$

Here, μ is the signal strength, $\boldsymbol{\theta}$ are the nuisance parameters, and n is the observed number of events.

The expected event yield N is the total number of expected events, including the number of signal events N_{sig}, $H \to \gamma\gamma$ (resonant background) events $N_{H\to\gamma\gamma}$, non-resonant background events N_{bkg}, and the spurious signal yield $N_{\text{spur}}(\theta_{\text{spur}})$. The number of signal events is defined as:

$$N_{sig} = L_{\text{int}} \times \mu \times \sigma_{SM}(pp \to H) \times BR_{SM}(H \to ll\gamma) \times \varepsilon, \tag{4.20}$$

where L_{int} is the integrated luminosity and ε is the signal acceptance. The spurious signal $N_{\text{spur}}(\theta_{\text{spur}})$ is constrained within the spurious signal uncertainty σ_{spur} by the corresponding nuisance parameter, θ_{spur}, centred at zero. The probability density function $f_{\text{tot}}(m^i_{\ell\ell\gamma};\mu,\boldsymbol{\theta})$ is built from the component probability density functions of f_{sig}, $f_{H\gamma\gamma}$ and f_{bkg}, and from the expected event yields $N^{(k)}_{\text{sig}}$, $N^{(k)}_{\text{spur}}$, $N^{(k)}_{H\gamma\gamma}$ and $N^{(k)}_{\text{bkg}}$ in the different analysis categories k (listed in Section 4.5.3):

$$f_{\text{tot}}(m^i_{\ell\ell\gamma};\mu,\boldsymbol{\theta}) = \frac{1}{N}\sum_k \Big\{\Big[N^{(k)}_{\text{sig}}(\mu,\boldsymbol{\theta}_{\text{sig}},\boldsymbol{\theta}_{\text{comm}}) + N^{(k)}_{\text{spur}}(\theta^{(k)}_{\text{spur}})\Big] \times f^{(k)}_{\text{sig}}(m^i_{\ell\ell\gamma};\boldsymbol{\theta}_{\text{sig}},\boldsymbol{\theta}_{\text{comm}}) + N^{(k)}_{H\gamma\gamma}(\boldsymbol{\theta}_{H\gamma\gamma},\boldsymbol{\theta}_{\text{comm}}) \times f^{(k)}_{H\gamma\gamma}(m^i_{\ell\ell\gamma};\boldsymbol{\theta}_{\text{comm}}) + N^{(k)}_{\text{bkg}} \times f^{(k)}_{\text{bkg}}(m^i_{\ell\ell\gamma};\boldsymbol{\theta}^{(k)}_{\text{bkg}})\Big\}, \quad (4.21)$$

where $N^{(k)}_{\text{sig}}$ are summarized yields from all considered Higgs boson production modes (ggF, VBF, WH, $qqZH$, $ggZH$, bbH, ttH) in each category. The nuisance parameters associated with the signal yields and shape, $\boldsymbol{\theta}_{\text{sig}}$, and with the resonant background yields, $\boldsymbol{\theta}_{H\gamma\gamma}$, and common nuisance parameters, $\boldsymbol{\theta}_{\text{comm}}$, quantify the impact of the systematic uncertainties and act across categories (i.e. their response for the same uncertainty in different categories is fully correlated), while nuisance parameters quantifying the spurious signal, $\theta^{(k)}_{\text{spur}}$, and parameters of the background functions, $\boldsymbol{\theta}^{(k)}_{\text{bkg}}$, are unique for each category, i.e. assumed to be independent. The dependence of the $f^{(k)}_{H\gamma\gamma}(m^i_{\ell\ell\gamma};\boldsymbol{\theta}_{\text{comm}})$ term only on the common nuisance parameters $\boldsymbol{\theta}_{\text{comm}}$ is motivated by specific choices for this analysis in estimation of systematic uncertainties for the $H \to \gamma\gamma$ background, described in more detail in Section 4.7. As constructed, it depends only on the nuisance parameters representing the common Higgs boson properties, notably corresponding to the uncertainty associated with Higgs boson mass measurement.

Significance calculation and CL_s limits

To quantify the level of agreement of a tested signal strength μ with respect to the signal strength $\hat{\mu}$ maximizing the likelihood, the following test statistic [68] can be used:

$$t_\mu = -2\ln\lambda(\mu) = -2\ln\frac{L(\mu,\hat{\hat{\boldsymbol{\theta}}}(\mu))}{L(\hat{\mu},\hat{\boldsymbol{\theta}})}, \quad (4.22)$$

where L is the likelihood function described by Equation 4.19; $\hat{\boldsymbol{\theta}}$ are the values of the nuisance parameters maximizing the likelihood, and $\hat{\hat{\boldsymbol{\theta}}}$ are the values that maximize the likelihood for the tested value μ. The likelihood ratio defined thus can take values ranging from close to 0 (for poor levels of agreement for a given μ with $\hat{\mu}$) to 1 (for $\mu = \hat{\mu}$). Taking the negative logarithm of the ratio translates it into *low* values of the test statistic for good levels of agreement and *large* values for poor levels of agreement. Quantitatively, the level of agreement is assessed by calculating the p-value:

$$p_\mu = \int_{t_{\mu,\text{obs}}}^{\text{inf}} f(t_\mu|\mu)dt_\mu, \quad (4.23)$$

where f is the probability density distribution of the test statistic.

To calculate the significance of the excess, a slightly different test statistic is employed:

$$q_0 = \begin{cases} -2\ln\frac{L(0,\hat{\hat{\boldsymbol{\theta}}}(0))}{L(\hat{\mu},\hat{\boldsymbol{\theta}})} & \hat{\mu} \geq 0, \\ 0 & \hat{\mu} < 0, \end{cases} \tag{4.24}$$

where $\hat{\hat{\boldsymbol{\theta}}}(0)$ is the value of the parameters $\boldsymbol{\theta}$ maximizing the likelihood assuming the background-only (fixing μ to zero) hypothesis. This test statistic considers only upward fluctuations of the signal ($\hat{\mu} \geq 0$), i.e. any *downward* fluctuation of the data cannot be used to reject the background-only hypothesis. Subsequently, the p-value is calculated thus:

$$p_0 = \int_{q_{0,\mathrm{obs}}}^{\mathrm{inf}} f(q_0|0)dq_0. \tag{4.25}$$

In this scheme, small p_0 values occur only for $q_0 > 0$, which happens only for $\hat{\mu} \geq 0$. The difficulty in this calculation is that f is not available in the analytical form. It can be assessed using a large number of pseudo-experiments. Or, alternatively, it is possible to use Wald's approximation, derived in Ref. [81]:

$$-2\ln\frac{L(0,\hat{\hat{\boldsymbol{\theta}}}(0))}{L(\hat{\mu},\hat{\boldsymbol{\theta}})} = \frac{\hat{\mu}^2}{\sigma^2} + O(1/\sqrt{N}). \tag{4.26}$$

Here, $\hat{\mu}$ is Gaussian-distributed with zero mean (corresponding to zero signal strength for background-only hypothesis) and standard deviation σ, and N represents the data sample size. Hence for small sample sizes such approximation may not hold. Using Wald's approximation, the p-value can be calculated simply as (see Ref. [68] for derivation):

$$p_0 = 1 - \Phi(\sqrt{q_0}), \tag{4.27}$$

where Φ is the cumulative distribution of the standard (zero mean, unit variance) Gaussian.

Once p_0 is calculated, it is translated into a significance Z, using the following relation:

$$Z = \Phi^{-1}(1-p), \tag{4.28}$$

where Φ is is the quantile (inverse of the cumulative distribution) of the standard Gaussian.

To set an upper limit, a different test statistic is employed:

$$\tilde{q}_\mu = \begin{cases} -2\ln\tilde{\lambda}(\mu) & \text{for } \mu \geq \hat{\mu} \\ 0 & \text{for } \mu < \hat{\mu} \end{cases} \qquad \text{with} \qquad \tilde{\lambda}(\mu) = \begin{cases} \ln\frac{L(\mu,\hat{\hat{\theta}}(\mu))}{L(\hat{\mu},\hat{\theta})} & \text{for } \hat{\mu} \geq 0, \\ \ln\frac{L(\mu,\hat{\hat{\theta}}(\mu))}{L(0,\hat{\theta}(0))} & \text{for } \hat{\mu} < 0\,. \end{cases} \tag{4.29}$$

As can be seen above, the test statistic is set to zero for $\mu < \hat{\mu}$, i.e. if a larger signal $\hat{\mu}$ than tested signal μ maximizes the likelihood, it is not considered as evidence against the presence of a signal. Furthermore, the signal is assumed to be non-negative, which is reflected by the fact that $\hat{\mu}$ (the best-fit μ) is effectively never allowed below zero in the denominator of the likelihood ratio. Based on this test statistic, the p-value is calculated as

$$p_\mu = \int_{\tilde{q}_{\mu,\mathrm{obs}}}^{\infty} f(\tilde{q}'_\mu|\mu)d\tilde{q}'_\mu = CL_{s+b}\,. \tag{4.30}$$

In addition to this, the CL_s procedure [82] is applied to modify the p-value, which is more robust in low signal-to-background ratio scenarios[13]:

$$CL_s = \frac{CL_{s+b}}{CL_b}, \tag{4.31}$$

where

$$CL_b = \int_{\tilde{q}_{\mu,\text{obs}}}^{\infty} f(\tilde{q}'_\mu|\mu = 0) d\tilde{q}'_\mu \tag{4.32}$$

is the p-value for the background-only case. As above, using Wald's approximation, one can arrive at the following result for calculating CL_s (derived in Ref. [68] for p_μ and adjusted here for CL_s):

$$CL_s = \frac{1 - F(\tilde{q}_\mu|\mu)}{1 - F(\tilde{q}_\mu|0)}. \tag{4.33}$$

Here,

$$F = \begin{cases} \Phi(\sqrt{\tilde{q}_\mu}) & \text{for } 0 < \tilde{q}_\mu \leq \mu^2/\sigma^2, \\ \Phi(\frac{\tilde{q}_\mu + \mu^2/\sigma^2}{2\mu/\sigma}) & \text{for } \tilde{q}_\mu > \mu^2/\sigma^2. \end{cases} \tag{4.34}$$

As seen above, this result now depends on knowing σ – the standard deviation of $\hat{\mu}$. It can be found by using pseudo-experiments, which is however computationally expensive. An alternative approach is used, exploiting properties of the Asimov dataset. Specifically, the Asimov data set is used to evaluate the "Asimov likelihood" L_A, and the $\tilde{q}_{\mu,A}$ test statistic is constructed in the same way as in Equation 4.29. As shown in Ref. [68], the standard deviation σ can be obtained from the following relation:

$$\sigma^2 = \frac{\mu^2}{\tilde{q}_{\mu,A}}. \tag{4.35}$$

A hypothesized signal strength μ is excluded at the 95% confidence level when $CL_s < 1 - 0.95 = 0.05$. Consequently, different values of μ are tested, each time constructing an Asimov dataset for the corresponding μ, in order to find the largest signal strength that cannot be rejected given the experimental data.

4.8.2 Calculation of post-fit systematic uncertainties

Systematic uncertainties are taken into account by the means of nuisance parameters in the likelihood model (as explained above). The $\pm 1\sigma$ variations of systematic uncertainties are first estimated as described in Section 4.7. Subsequently a fit of the likelihood model to data is performed, where nuisance parameters are allowed to float. From this fit, new (*post-fit*) $\pm 1\sigma$ variations of systematic uncertainties are obtained. To determine the impact of a given uncertainty source on the reported μ and $\sigma \times \mathcal{B}$, the corresponding nuisance parameter is fixed to its best-fit value, while all other parameters in the likelihood are allowed to float. The modified total uncertainty is then used to determine the

[13] As can be seen from Equation 4.31, in case of low values of CL_b (which implies good compatibility of data with the background-only hypothesis), the value of CL_s increases relative to CL_{s+b}, thereby decreasing maximum possible signal strength μ that can be excluded and therefore reducing the reported sensitivity of the search, providing a more realistic estimate than when relying purely on CL_{s+b}.

impact of the uncertainty source:

$$\Delta^{\pm 1\sigma}_{\text{syst src}} = \sqrt{(\Delta^{\pm 1\sigma}_{\text{total}})^2 - (\Delta^{\pm 1\sigma}_{\text{syst NP fixed at best-fit}})^2} \tag{4.36}$$

The same technique is used to estimate the impact of a group of systematic uncertainty sources, by fixing several nuisance parameters at the same time. A special case of this is when all nuisance parameters corresponding to systematic uncertainties are fixed, obtaining the total systematic uncertainty as a result.

4.8.3 Profiling the negative log-likelihood ratio

The best-fit signal strength value μ corresponds to the minimum of the negative log-likelihood ratio $-2\ln\lambda(\mu)$ (a point where it approaches zero), and $\pm 1\sigma$ uncertainties correspond to the two points where it crosses the $-2\ln\lambda(\mu) = 1$ line. Typically this minimum is found by minimizing the negative log-likelihood ratio as a function of all its parameters, with the help of minimization algorithms such as gradient descent [83]. These algorithms also provide estimates of $\pm 1\sigma$ uncertainties for all parameters, including the signal strength μ.

However, in case of this analysis, the $\pm 1\sigma$ uncertainty bands obtained with `Minos` algorithm of the `Minuit2` minimizer [84] are not precise enough. This is especially relevant for additional results such as fits with separate parameters of interest in different categories (shown in Figures 4.17 and 4.18), where statistical uncertainties in some individual categories are very large, and for studying the impact of individual systematic uncertainties (shown in Table 4.9), where the systematic uncertainty in question is very small compared to the total uncertainty, and the numerical precision of the calculation is insufficient to resolve its impact.

Therefore, instead of relying on minimization algorithms for the $\pm 1\sigma$ uncertainties, all results in this analysis are obtained by profiling the negative log-likelihood ratio as a function of μ using the algorithm described in Appendix B.

4.8.4 Pseudo-experiments

The main results of this analysis are obtained using the approximation in Equation 4.26. As mentioned, for small data sample sizes that approximation may not hold. In this analysis the data samples are relatively small in some categories (e.g. in the VBF categories, see yields in Table 4.10), and therefore it is instructive to check the validity of the asymptotic approximation by confirming the results with pseudo-experiments. The reported best-fit μ and significance Z are cross-checked using pseudo-experiments. The results and implications of this cross-check are discussed in Section 4.9.3. A general description of pseudo-experiments is given below.

Pseudo-experiments are performed using the signal-plus-background model from which a dataset (so-called toy dataset) containing realizations of $m_{\ell\ell\gamma}$ values is generated using Poisson sampling. A large number of such datasets, each with independent realizations of $m_{\ell\ell\gamma}$ values, is obtained. To verify the results of this analysis, two different setups of pseudo-experiments are used.

In the first setup, the signal-plus-background model with parameters taken from the fit to data (including the best-fit μ) is used to generate toys. Then, the signal-plus-background model is fitted to each toy dataset, and the value of the signal strength μ maximizing the likelihood is found. A distribution of μ values is obtained, from which the median, corresponding to the best-fit μ, is extracted,

as well as the $\pm 1\sigma$ uncertainty bands. An example of the μ distribution from pseudo-experiments is shown in Figure 4.14. The test of results is to verify that the median μ value and the $\pm 1\sigma$ uncertainty bands from the distribution of μ values for toys is close to the best-fit μ and $\pm 1\sigma$ uncertainties from the fit to data of the signal-plus-background model.

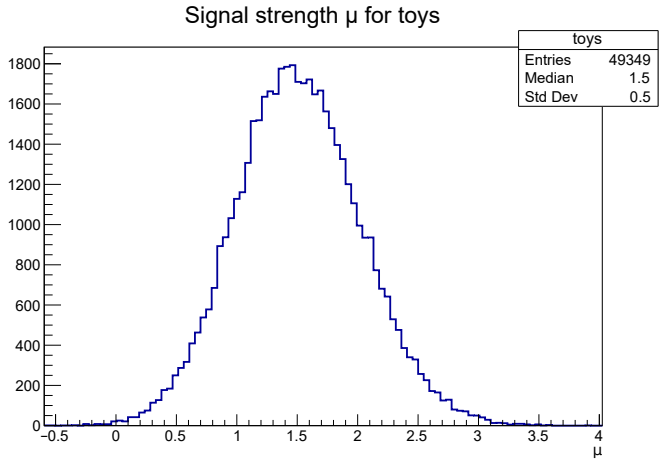


Figure 4.14: Distribution of the signal strength μ in pseudo-experiments drawn from the post-fit signal-plus-background model.

In the second setup, the background-only model is used (all parameters are the same as from the signal-plus-background fit to data, but with the signal strength value μ set to zero), and toys are generated from this model. Then two fits are performed to each toy: the signal-plus-background fit, from which $L(\hat{\mu}, \hat{\boldsymbol{\theta}})$ is obtained, and the background-only fit, from which $L(0, \hat{\hat{\boldsymbol{\theta}}}(0))$ is obtained. The value of the q_0 test statistic is subsequently calculated using Equation 4.24. A distribution of q_0 is obtained with a large number of toys, shown in Figure 4.19. The p_0 value is calculated as given by Equation 4.25, i.e. integrating the q_0 distribution from the q_{obs} value seen in data to the largest value (effectively to $+\infty$). The p_0 is then translated into a significance Z, which can be compared to the significance value obtained from the asymptotic approximation.

4.9 Results and interpretation

The presented results are based on signal-plus-background model described in Section 4.6 and obtained using statistical methods described in Section 4.8.

4.9.1 Expected results

The expected results are based on data in the range 110 GeV $< m_{\ell\ell\gamma} <$ 160 GeV, with the signal region 120 GeV $< m_{\ell\ell\gamma} <$ 130 GeV excluded (data sidebands). The expected event yields for signal and background in each category, as well as the estimated sensitivity, is shown in Table 4.10.

The expected 95% CL_s limit on the Higgs boson production cross-section times the $H \to \ell\ell\gamma$ branching ratio in the $m_{\ell\ell} < 30$ GeV region is 1.86 times the SM prediction. Assuming no SM Higgs boson, the limit is 0.95 times the SM prediction. The best-fit value of the expected signal strength is $\mu_{\text{exp}} = 1.0 \pm 0.5 = 1.0 \pm 0.5\ (\text{stat.})\ ^{+0.2}_{-0.1}\ (\text{syst.})$. The expected significance amounts to $Z_{\text{exp}} = 2.1\sigma$.

4.9.2 Observed results

The observed results are based on data in the full range 110 GeV $< m_{\ell\ell\gamma} <$ 160 GeV. An excess over the background-only expectation is found, consistent with the Standard Model Higgs boson. The number of selected events in data in each analysis category is shown in Table 4.10.

The observed 95% CL_s limit is 2.26 times the SM prediction. The best-fit value of the observed signal strength is $\mu_{\text{obs}} = 1.5 \pm 0.5 = 1.5 \pm 0.5\ (\text{stat.})\ ^{+0.2}_{-0.1}\ (\text{syst.})$, which corresponds to an excess with a significance of $Z_{\text{obs}} = 3.2\sigma$, therefore surpassing the *evidence* threshold of 3.0σ. The fiducial cross-section times branching ratio in the $m_{\ell\ell} < 30$ GeV region measured in data amounts to $\sigma(H) \times \mathcal{B}(H \to \ell\ell\gamma) = 8.7^{+2.8}_{-2.7}$ fb $= 8.7 \pm 2.7\ (\text{stat.})\ ^{+0.7}_{-0.6}\ (\text{syst.})$ fb.

The $m_{\ell\ell\gamma}$ distribution in data in each analysis category, together with the signal-plus-background model based on the full combined fit to all categories, is shown in Figure 4.15. The $m_{\ell\ell\gamma}$ distribution for all categories combined, with data events weighted by a category-dependent weight $\ln(1+S_{90}/B_{90})$, where S_{90} is the number of signal events in the smallest window containing 90% of the expected signal and B_{90} is the expected number of background events in the same window, is presented in Figure 4.16. The data events are weighted merely to visually enhance the signal peak shown in this figure, and no data weights are applied to any results outside of this figure.

Systematic uncertainties are discussed in more detail in Section 4.7. The impact of systematic uncertainties on the reported signal strength μ per group of sources is measured, and reported in percent of the observed μ value in Table 4.9.

Category	Events	S_{90}	B_{90}^{N}	$B_{H\to\gamma\gamma}$	f_{90} [%]	Z_{90}
VBF Resolved Electron	10	0.4	1.6	0.009	20	0.3
VBF Merged Electron	15	0.8	2.0	0.07	27	0.5
VBF Dimuon	33	1.3	5.9	–	18	0.5
High-$p_{\mathrm{TThrust}}^{\ell\ell\gamma}$ Resolved Electron	86	1.1	12	0.02	9	0.3
High-$p_{\mathrm{TThrust}}^{\ell\ell\gamma}$ Merged Electron	162	2.5	18	0.2	12	0.6
High-$p_{\mathrm{TThrust}}^{\ell\ell\gamma}$ Dimuon	210	4.0	34	–	11	0.7
Low-$p_{\mathrm{TThrust}}^{\ell\ell\gamma}$ Resolved Electron	3713	22	729	0.5	2.9	0.8
Low-$p_{\mathrm{TThrust}}^{\ell\ell\gamma}$ Merged Electron	5103	29	942	2	3.0	1.0
Low-$p_{\mathrm{TThrust}}^{\ell\ell\gamma}$ Dimuon	9813	61	1750	–	3.4	1.4

Table 4.10: Number of selected data events in each analysis category in the range 105 GeV $< m_{\ell\ell\gamma} <$ 160 GeV, as well as expected signal (S_{90}) in the minimal $m_{\ell\ell\gamma}$ interval containing 90% of the signal, the resonant $H \to \gamma\gamma$ background ($B_{H\to\gamma\gamma}$), and the non-resonant background in the same interval (B_{90}^{N}) estimated using the expected signal-plus-background model based on data sidebands. $B_{H\to\gamma\gamma}$ only appears in the electron categories and is marked as "–" otherwise. In addition, the expected signal purity ($f_{90} = S_{90}/(S_{90} + B_{90})$) and the expected signal significance ($Z_{90} = \sqrt{2((S_{90} + B_{90})\ \ln(1 + S_{90}/B_{90}) - S_{90})}$) are given.

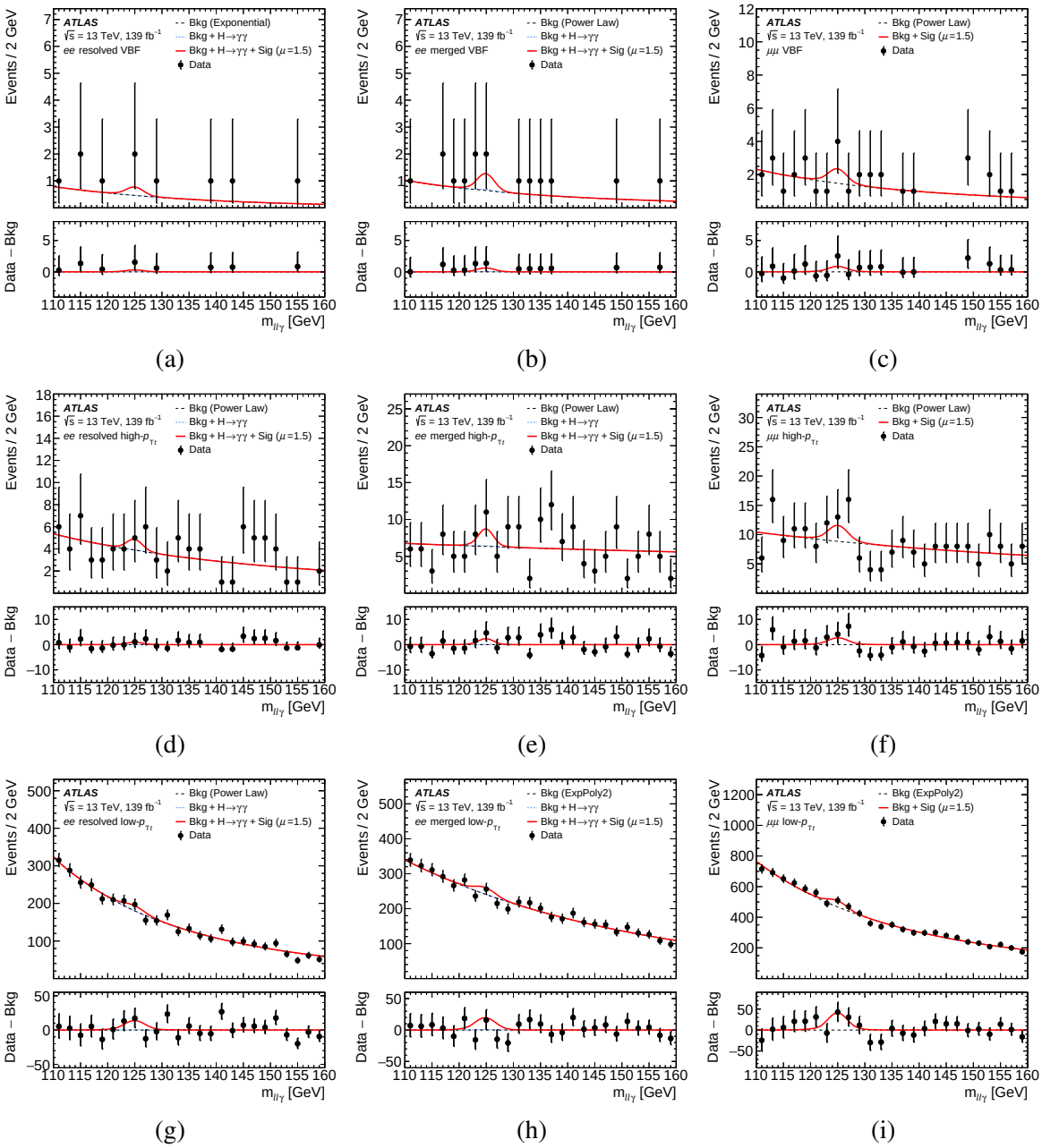


Figure 4.15: $m_{\ell\ell\gamma}$ distributions of the selected data events (black circles) and the signal-plus-background model after the combined fit (red curve) for the VBF categories (a, b, c), the high-$p^{\ell\ell\gamma}_{\text{TThrust}}$ categories (d, e, f), and the low-$p^{\ell\ell\gamma}_{\text{TThrust}}$ categories (g, h, i) [66]. Additionally, the non-resonant background component is shown with a dashed black line and the $H \to \gamma\gamma$ background component is shown with a dotted blue line. The bottom panels show the residuals of the data with respect to the non-resonant background component of the signal-plus-background model. The *ee*-resolved categories are shown in the left column, the *ee*-merged categories in the middle and the $\mu\mu$ categories in the right column.

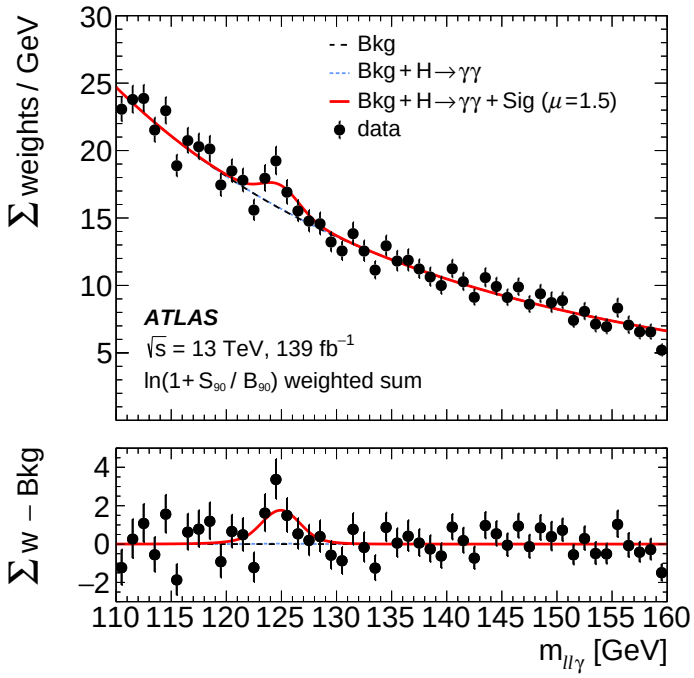


Figure 4.16: $m_{\ell\ell\gamma}$ distribution of the selected data events (black circles), with every data event reweighted by a category-dependent weight, $\ln(1 + S_{90}/B_{90})$, where S_{90} is the number of signal events in the smallest window containing 90% of the expected signal, and B_{90} is the expected number of background events in the same window. The signal-plus-background model after the combined fit is shown with a red curve, its non-resonant background component is shown with a dashed black line and the $H \to \gamma\gamma$ background component is shown with a dotted blue line. The bottom panels show the residuals of the data with respect to the non-resonant background component of the signal-plus-background model.

4.9.3 Cross-checks of consistency of the results

Multiple checks were performed to ensure consistency of the reported results. A first check of consistency is ensuring that the likelihood function profiled with respect to signal strength, μ, behaves smoothly and exhibits a clear minimum. The details of likelihood profiling are described in Section 4.8.3.

A cross-check fit is also performed by specifying independent parameters of interest per category μ_{cat} in the likelihood model. The resulting best-fit μ_{cat} values in individual categories are shown in Figure 4.17. It is also illustrative to group categories (e.g. by flavor or by kinematic channel), specifying one common parameter of interest for each group, and re-running the fit. The results are shown in Figure 4.18. The $\mu\mu$-channels tend to have a slightly higher signal strength than ee-channels, and the same is true for VBF and high-$p^{\ell\ell\gamma}_{\text{TThrust}}$ categories compared to the low-$p^{\ell\ell\gamma}_{\text{TThrust}}$ ones. However it should be noted that given the uncertainties, it is not possible to claim that there is any systematic effect, or in other words, all results agree with each other and with the global signal strength μ within the uncertainties. The likelihood function profiles for the global μ, as well as in different analysis categories individually, are shown in Appendix B.

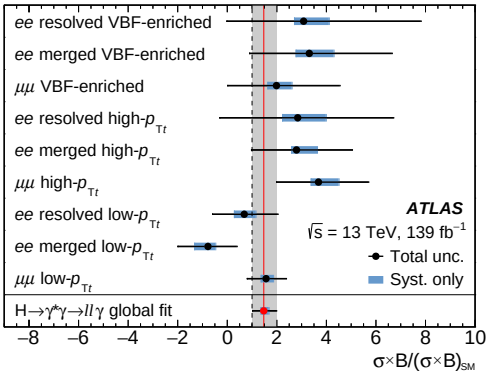


Figure 4.17: Best-fit values of the signal strength parameters μ_{cat} when they are allowed to float independently in the fit for each analysis category (black circles), compared with the result of the global μ fit (red circle and line) including its total uncertainty (grey band) [66].

The results were also cross-checked by running many pseudo-experiments instead of relying on the asymptotic approximation to determine the p-value, see details in Section 4.8.4. The q_0 test statistic for the $\mu = 0$ scenario, with over 500,000 pseudo-experiments, is shown in Figure 4.19. The p-value is found by integrating this distribution from the q_0 value observed in data to the highest value. The resulting p-value corresponds to a signal significance of 3.1σ. For the observed μ scenario, the result from pseudo-experiments is the same as the observed signal strength obtained using the asymptotic approximation, $\mu_{\text{obs}} = 1.5 \pm 0.5$.

The slightly lower significance obtained with pseudo-experiments is likely a consequence of a slight positive bias in the reported μ values from fits to the toy datasets in the $\mu = 0$ scenario. Specifically, when the toy datasets are generated, the μ is set to zero, and hence with a large enough distribution of toys, the median μ from the fit to toys should be also zero. However it was found that the median μ value obtained from the toy ensemble is systematically above zero. This could be a consequence of

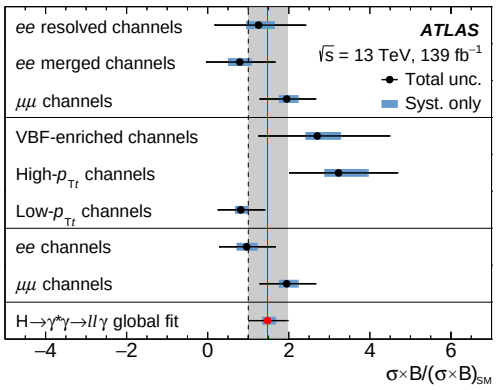


Figure 4.18: Best-fit values of the signal strength parameters μ_{group} when they are allowed to float independently in the fit for different groupings of analysis categories, compared with the result of the global μ fit (red circle and line) including its total uncertainty (grey band) [66].

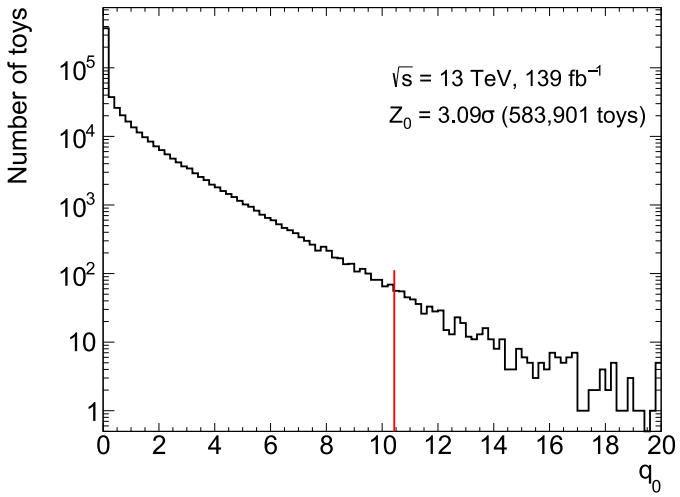


Figure 4.19: Distribution of the q_0 test statistic in pseudo-experiments drawn from the post-fit background-only model, and subsequently performing fits with $\mu = 0$ and floating μ. The red line indicates the q_0 observed in data.

generating purely random toys from the signal-plus-background model, which does not fully reproduce the analysis strategy. In the analysis, the signal regions were constructed with the knowledge of the distribution of data in the sidebands (in the range 110 GeV $< m_{\ell\ell\gamma} <$ 160 GeV with the signal region 120 GeV $< m_{\ell\ell\gamma} <$ 130 GeV excluded). Notably, it was known that even in the very statistically limited VBF categories, the fit with a smoothly falling function is well-behaved. In the case of pseudo-experiments, for these statistically limited categories it is entirely possible to have some toy datasets resulting in "pathological" fits, and indeed this was found to be the case. For example, some fits resulted in a *rising* exponential function instead of a falling one, which completely misrepresents the background expectation. Other toys were completely devoid of events in the region below or above

the 120 GeV $< m_{\ell\ell\gamma} <$ 130 GeV signal region window, resulting in a poorly-constrained background shape. Therefore the results with the asymptotic approximation (and particularly $Z = 3.2\sigma$) are used as the central result. Nevertheless, the results from pseudo-experiments confirm the measured signal strength and the fact that the evidence threshold of 3.0σ is surpassed.

Another test of consistency of the results is a comparison of the expected and the observed impact of systematic uncertainties on signal strength μ. Because systematic uncertainties are represented as nuisance parameters in the likelihood, which are allowed to float in the fit, the impacts of systematic uncertainties of the signal-plus-background model after the fit to data may differ from the impacts obtained with the expected signal-plus-background model. Small differences are expected, but any large deviation may indicate that systematic uncertainties were incorrectly estimated, or suggest some other problems with the likelihood model. For this analysis, there were no large differences observed for any systematic uncertainties. Among small differences, it was found that the impact of the e/γ energy resolution and energy scale uncertainties on the signal strength is slightly different between the expected and the observed result. This effect was studied in more detail and is described in Appendix C.

Finally, even though the analysis does not have enough sensitivity to improve upon the Higgs boson mass measurement, it is instructive to extract the mass value from the fit to ensure that is consistent with the best measurement within the uncertainties. This is technically achieved by introducing another parameter of interest representing the Higgs boson mass, which is allowed to float in the fit. The resulting value $m_H = 124.8 \pm 0.5$ GeV is consistent with the value from the most precise measurement to date, 125.38 ± 0.14 GeV [26]. Some other configurations of the floating mass fit are explored in Appendix C.

4.9.4 Implications and outlook

The presented result, with a signal significance of 3.2σ, constitutes the first evidence for the $H \to \ell\ell\gamma$ decay. With a branching ratio in the analysis fiducial region of only $\sim 10^{-4}$ (see Table 2.1), this result contributes to the general trend[14] of probing rarer and rarer Higgs boson decays at the LHC, thanks to the wealth of data collected in Run 2.

The limitation in the result is largely statistical, with systematic uncertainty constituting only 35% of the statistical uncertainty. Therefore, the result should significantly improve with more data. With more data it will also be possible to explore Higgs boson charge-parity (CP) properties with the $H \to \ell\ell\gamma$ decay [11], which may shed more light on the problem of CP violation in the Standard Model. It will also be possible to constrain Higgs boson coupling modifications introduced by some proposed extensions of the Standard Model [85].

It is informative to make a simple projection to determine, for example, how much more data is required to surpass the discovery threshold of 5σ. This is achieved by scaling up the expected signal-plus-background model, since both signal and background yields are expected to scale up with the increase in the integrated luminosity. Specifically, the signal and background functions are scaled, such that they correspond to an increase in signal and background yields respectively by a fixed factor C. The shape of the signal and background $m_{\ell\ell\gamma}$ distribution remains unchanged. The systematic uncertainties in the signal-plus-background model (see Section 4.7) are already implemented relative to the signal yields and therefore also remain unchanged. One exception is the spurious signal uncertainty, which is based on the absolute value of spurious signal yields, and is therefore scaled up by the same factor C. An Asimov dataset is then constructed, corresponding to the scaled-up expected signal-plus-background model, and a fit is performed to extract expected significance corresponding to the luminosity increase by a factor C, using the same methods as described in Section 4.8. The fit is repeated using different factors C until a factor corresponding to the expected significance $Z = 5\sigma$ is found. Besides the significance Z, the expected signal strength μ is extracted from the fit, together with statistical and systematic uncertainties on it.

As a result of the projection, significance of $Z = 5\sigma$ could be achieved with 950 $\text{fb}^{-1}$of data, which includes already collected 139 $\text{fb}^{-1}$of data. That is certainly within the LHC program goals, with the High Luminosity LHC (HL-LHC) projected to collect $\sim$ 3000 fb^{-1} of data. The corresponding expected signal strength is $\mu_{\text{exp}}^{\text{proj}} = 0.99^{+0.23}_{-0.22} = 0.99 \pm 0.18$ (stat.) $^{+0.15}_{-0.12}$ (syst.). As expected, the statistical uncertainty on μ is largely reduced, and the systematic uncertainty plays a more important role in the result.

This simplistic projection, however, can only be used to broadly highlight future possibilities, and should not be interpreted as an actual target. There is a potential for more improvement beyond simple scaling with luminosity. One possible area of improvement is in the analysis techniques. For example, machine learning, and, specifically, boosted decision trees, could be used to define analysis categories, as explored in Ref. [20] for the $H \to Z\gamma$ analysis. Another possibility is to use neural networks for merged electron identification. These approaches do not necessarily guarantee better performance for this analysis specifically, but they are certainly worth exploring.

Another likely area of improvement is in the systematic uncertainties. There is a concentrated ongoing effort to reduce the uncertainties in many experimental measurements on which the current result relies, as well as in the theoretical calculations. The result has already triggered discussions

[14] For example, another impressive achievement is the evidence for $H \to \mu^+\mu^-$ decay recently reported by the CMS Collaboration [21].

among theorists about a more precise calculation of the $\mathcal{B}(H \to \ell\ell\gamma)$ in the fiducial region of the analysis (and estimating the uncertainty in this calculation), as well as calculating a dedicated fiducial cross-section.

CHAPTER 5

Conclusion

The Higgs boson, discovered in 2012 at the Large Hadron Collider (LHC), is a cornerstone of the Standard Model of particle physics. The Higgs interaction provides a mechanism explaining the observed masses of the elementary particles. Due to its central role, studying the properties of the Higgs boson may shed some light on the issues presently not explained by the Standard Model, such as nature of the dark matter, origin of the neutrino masses, and mechanism of the charge-parity (CP) symmetry violation. With the large amount of data from the LHC collected during Run 2 it becomes possible to measure Higgs boson properties more precisely, including its rare decays. This thesis presents a search for a very rare decay of the Higgs boson to a photon and a low-mass lepton pair (e^+e^- or $\mu^+\mu^-$). The phase space region with $m_{\ell\ell} < 30$ GeV is explored, where the expected SM branching ratio is only $\mathcal{B}(H \to \ell\ell\gamma) \approx 10^{-4}$. This decay has a large contribution from Feynman diagrams with loops and three particles in the final state, which opens up a possibility for more stringent tests of the Standard Model.

The search is performed with the ATLAS detector at the LHC, using 139 fb^{-1}of pp collision data at a center-of-mass energy $\sqrt{s} = 13$ TeV collected during Run 2. An excess is observed over the background-only hypothesis with 3.2σ significance, establishing evidence for this decay. This constitutes the first evidence for the Higgs boson decaying to a photon and a lepton pair. The ratio between the observed event yield classified as signal and the expected SM signal yield is measured at $\mu = 1.5 \pm 0.5$. Additionally the fiducial cross-section times branching ratio in $m_{\ell\ell} < 30$ GeV region is measured at $\sigma(H) \times \mathcal{B}(H \to \ell\ell\gamma) = 8.7^{+2.8}_{-2.7}$ fb.

The main experimental challenge of the analysis is that in the explored dilepton mass regime, $m_{\ell\ell} < 30$ GeV, the leptons tend to be highly collimated, which leads to overlapping of the energy deposits in the calorimeter in the $ee\gamma$ channel. Due to this, such events with close-by electrons would not be correctly reconstructed and therefore would be lost for the analysis. This challenge is overcome by developing a dedicated identification algorithm for close-by electrons based on a multivariate discriminator. The efficiency of the new algorithm is measured in data with converted electrons, and energy calibration for the close-by electrons is also obtained with converted electrons, which from reconstruction perspective look very similar to highly collimated electrons from the γ^* decay.

Another experimental challenge arises from the low branching ratio, which results in a very low number of expected signal events with a large number of expected background events. This challenge, typical for a search for a small signal, is addressed by rigorous statistical analysis and cross-checking of the results.

This analysis paves the way for future measurements of the rare $H \to \ell\ell\gamma$ decay. In particular, the methods developed for present analysis, most notably close-by electron identification algorithm, will find their use in the future analysis with more data. Given the current statistical limitation, the result with more data is expected to improve considerably. Finally, in future measurements, Higgs boson CP properties can be studied with this loop-induced three-body decay, which will have profound implications for understanding the mechanism of CP violation in the Standard Model.

Bibliography

[1] Stella, Bruno R. and Meyer, Hans-Jürgen,
Υ(9.46 GeV) and the gluon discovery (a critical recollection of PLUTO results),
Eur. Phys. J. H **36** (2011) 203 (cit. on p. 1).

[2] G. Arnison et al., *Experimental observation of isolated large transverse energy electrons with associated missing energy at s=540 GeV*, Phys. Lett. B **122** (1983) 103 (cit. on p. 1).

[3] G. Arnison et al., *Experimental observation of lepton pairs of invariant mass around 95 GeV/c^2 at the CERN SPS collider*, Phys. Lett. B **126** (1983) 398 (cit. on p. 1).

[4] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*,
Phys. Rev. Lett. **13** (9 1964) 321 (cit. on p. 1).

[5] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*,
Phys. Rev. Lett. **13** (16 1964) 508 (cit. on p. 1).

[6] G. S. Guralnik, C. R. Hagen and T. W. B. Kibble,
Global Conservation Laws and Massless Particles, Phys. Rev. Lett. **13** (20 1964) 585
(cit. on p. 1).

[7] 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. B **716** (2012) 1
(cit. on pp. 1, 13).

[8] CMS Collaboration,
Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,
Phys. Lett. B **716** (2012) 30 (cit. on pp. 1, 13).

[9] Particle Data Group, *Review of Particle Physics*, PTEP **2020** (2020), 083C01, ISSN: 2050-3911,
eprint: https://academic.oup.com/ptep/article-pdf/2020/8/083C01/34673722/ptaa104.pdf (cit. on pp. 1, 14).

[10] Y. Sun, H. Chang and D. Gao, *Higgs decays to $\gamma\,\ell^+\ell^-$ in the standard model*,
J. High Energy Phys. **61** (2013 2013) (cit. on pp. 2, 38, 39).

[11] A. Korchin and V. Kovalchuk,
Polarization effects in the Higgs boson decay to Z and test of CP and CPT symmetries,
Phys. Rev. D **88** (3 2013) 036009 (cit. on pp. 2, 40, 93).

[12] *Standard Model of Elementary Particles (figure)*,
https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg, Accessed: 2021-09-03. (cit. on p. 4).

[13] M. Thomson, *Modern Particle Physics*, Cambridge University Press, 2013 (cit. on pp. 3, 15).

[14] *Sherpa event generator (online documentation)*, https://sherpa-team.gitlab.io/monte-carlo.html, Accessed: 2021-04-26. (cit. on p. 7).

[15] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, JPhysG **43** (2016) 023001 (cit. on pp. 6, 74).

[16] S. Dulat et al., *New parton distribution functions from a global analysis of quantum chromodynamics*, Physical Review D **93** (2016), ISSN: 2470-0029 (cit. on p. 6).

[17] L. A. Harland-Lang et al., *Parton distributions in the LHC era: MMHT 2014 PDFs*, EPJ C **75** (2015), ISSN: 1434-6052 (cit. on p. 6).

[18] R. D. Ball et al., *Parton distributions for the LHC run II*, JHEP **2015** (2015), ISSN: 1029-8479 (cit. on p. 6).

[19] ATLAS Collaboration, *Higgs boson production cross-section measurements and their EFT interpretation in the* 4ℓ *decay channel at* $\sqrt{s}$ *=13 TeV with the ATLAS detector*, Eur. Phys. J. C **80** (2020) 957, [Erratum: Eur.Phys.J.C 81, 29 (2021), Erratum: Eur.Phys.J.C 81, 398 (2021)], arXiv: 2004.03447 [hep-ex] (cit. on p. 14).

[20] ATLAS Collaboration, *A search for the* $Z\gamma$ *decay mode of the Higgs boson in* pp *collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Phys. Lett. B **809** (2020) 135754, ISSN: 0370-2693 (cit. on pp. 14, 93).

[21] CMS Collaboration, *Evidence for Higgs boson decay to a pair of muons*, JHEP **01** (2021) 148, arXiv: 2009.04363 [hep-ex] (cit. on pp. 14, 38, 93).

[22] CMS Collaboration, *Search for the decay of a Higgs boson in the* $\ell\ell\gamma$ *channel in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, JHEP **11** (2018) 152, arXiv: 1806.05996 [hep-ex] (cit. on pp. 14, 40).

[23] CMS Collaboration, *Search for a Higgs boson decaying into* $\gamma^*\gamma \to \ell\ell\gamma$ *with low dilepton mass in* pp *collisions at* $\sqrt{s} = 8$ *TeV*, Phys. Lett. B **753** (2016) 341, arXiv: 1507.03031 [hep-ex] (cit. on pp. 14, 40).

[24] D. de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, CERN Yellow Reports: Monographs, 2016, URL: https://cds.cern.ch/record/2227475 (cit. on pp. 14, 42, 43, 74).

[25] ATLAS and CMS Collaborations, *Combined Measurement of the Higgs Boson Mass in* pp *Collisions at* $\sqrt{s} = 7$ *and* 8 *TeV with the ATLAS and CMS Experiments*, Phys. Rev. Lett. **114** (2015) 191803, arXiv: 1503.07589 [hep-ex] (cit. on pp. 14, 42, 63–65, 73, 80).

[26] CMS Collaboration, *A measurement of the Higgs boson mass in the diphoton decay channel*, Phys. Lett. B **805** (2020) 135425, ISSN: 0370-2693 (cit. on pp. 13, 92).

[27] CMS Collaboration, *Measurements of the Higgs boson width and anomalous HVV couplings from on-shell and off-shell production in the four-lepton final state*, Phys. Rev. D **99** (11 2019) 112003 (cit. on p. 13).

[28] ATLAS Collaboration,
Combined measurements of Higgs boson production and decay using up to 80 fb^{-1} *of proton-proton collision data at* $\sqrt{s} = 13$ TeV *collected with the ATLAS experiment*,
Phys. Rev. D **101** (1 2020) 012002 (cit. on p. 15).

[29] *Double beta decay feynman (figure)*,
https://commons.wikimedia.org/wiki/File:Double_beta_decay_feynman.svg,
Accessed: 2021-09-03. (cit. on p. 16).

[30] ATLAS Collaboration, *CP Properties of Higgs Boson Interactions with Top Quarks in the $t\bar{t}H$ and tH Processes Using $H \to \gamma\gamma$ with the ATLAS Detector*,
Phys. Rev. Lett. **125** (6 2020) 061802 (cit. on p. 17).

[31] ATLAS Collaboration,
Test of CP invariance in vector-boson fusion production of the Higgs boson in the $H \to \tau\tau$ channel in proton-proton collisions at $\sqrt{s} = 13$ *TeV with the ATLAS detector*,
Phys. Lett. B **805** (2020) 135426, ISSN: 0370-2693 (cit. on p. 17).

[32] O. S. Bruening et al., *LHC Design Report*, CERN Yellow Reports: Monographs, CERN, 2004,
URL: http://cds.cern.ch/record/782076 (cit. on pp. 19, 20).

[33] CMS Collaboration, *The CMS experiment at the CERN LHC*, JINST **3** (2008) S08004
(cit. on p. 19).

[34] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*,
JINST **3** (2008) S08003 (cit. on pp. 19, 24–26, 28, 29, 32).

[35] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002
(cit. on p. 19).

[36] LHCb Collaboration, *The LHCb Detector at the LHC*, JINST **3** (2008) S08005 (cit. on p. 19).

[37] F. Marcastel, *CERN's Accelerator Complex. La chaîne des accélérateurs du CERN*,
(2013), General Photo, URL: http://cds.cern.ch/record/1621583 (cit. on p. 21).

[38] *ATLAS luminosity public results*, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2,
Accessed: 2021-06-18 (cit. on pp. 22, 24).

[39] ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report*,
ATLAS-TDR-19 (2010), URL: https://cds.cern.ch/record/1291633,
ATLAS Insertable B-Layer Technical Design Report Addendum, ATLAS-TDR-19-ADD-1,
2012, URL: https://cds.cern.ch/record/1451888 (cit. on p. 26).

[40] ATLAS Collaboration, *Production and integration of the ATLAS Insertable B-Layer*,
JINST **13** (2018) T05008, arXiv: 1803.00844 [physics.ins-det] (cit. on p. 26).

[41] B. Mindur, *ATLAS Transition Radiation Tracker (TRT): Straw tubes for tracking and particle identification at the Large Hadron Collider*, Nucl. Instrum. Methods Phys. Res. A **845**
(2017) 257, Proceedings of the Vienna Conference on Instrumentation 2016, ISSN: 0168-9002
(cit. on p. 26).

[42] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at* $\sqrt{s} = 13$ *TeV*,
Eur. Phys. J. C **79** (2019) 639, arXiv: 1902.04655 [hep-ex] (cit. on pp. 27, 46).

[43] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data*, JINST **14** (2019) P12006, arXiv: 1908.00005 [hep-ex] (cit. on pp. 28–30).

[44] ATLAS Collaboration, *Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run 2 data collected in 2015 and 2016*, Eur. Phys. J. C **79** (2019) 205, arXiv: 1810.05087 [hep-ex] (cit. on pp. 30, 47, 73).

[45] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, Eur. Phys. J. C **77** (2017) 466, arXiv: 1703.10485 [hep-ex] (cit. on p. 31).

[46] M. Cacciari, G. P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, J. High Energy Phys. **2008** (2008) 063 (cit. on p. 31).

[47] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*, ATLAS-CONF-2014-018 (2014), URL: https://cds.cern.ch/record/1700870 (cit. on p. 31).

[48] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **76** (2016) 292, arXiv: 1603.05598 [hep-ex] (cit. on pp. 32, 33).

[49] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, Eur. Phys. J. C **77** (2017) 317, arXiv: 1611.09661 [hep-ex] (cit. on p. 33).

[50] *ATLAS trigger operation public results*, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/TriggerOperationPublicResults, Accessed: 2021-06-18 (cit. on p. 34).

[51] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, JINST **13** (2018) P07017 (cit. on pp. 34, 73).

[52] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*, tech. rep., CERN, 2019, URL: http://cds.cern.ch/record/2677054 (cit. on pp. 35, 73).

[53] *ATLAS software – Athena*, https://gitlab.cern.ch/atlas/athena, Accessed: 2021-04-23 (cit. on p. 36).

[54] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823, arXiv: 1005.4568 [hep-ex] (cit. on p. 36).

[55] S. Agostinelli et al., *Geant4–a simulation toolkit*, Nucl. Instrum. Methods Phys. Res. A **506** (2003) 250 (cit. on p. 36).

[56] ATLAS Collaboration, *Measurements of Higgs boson properties in the diphoton decay channel with 36 fb^{-1} of pp collision data at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Rev. D **98** (5 2018) 052005 (cit. on pp. 38, 73).

[57] ATLAS Collaboration, *Interpretations of the combined measurement of Higgs boson production and decay*, tech. rep., CERN, 2020, URL: https://cds.cern.ch/record/2743067 (cit. on p. 38).

[58] M. Vesterinen and T. Wyatt, *A novel technique for studying the Z boson transverse momentum distribution at hadron colliders*, Nucl. Instrum. Meth. A **602** (2009) 432, arXiv: 0807.4956 [hep-ph] (cit. on p. 41).

[59] S. Heinemeyer et al., *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of the LHC Higgs Cross Section Working Group*, ed. by S. Heinemeyer, CERN Yellow Reports: Monographs, CERN, 2013, URL: https://cds.cern.ch/record/1559921 (cit. on p. 42).

[60] T. Sjöstrand, S. Mrenna and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv: 0710.3820 [hep-ph] (cit. on p. 42).

[61] A. Firan and R. Stroynowski, *Internal conversions in Higgs decays to two photons*, Phys. Rev. D **76** (2007) 057301, arXiv: 0704.3987 [hep-ph] (cit. on p. 42).

[62] D. A. Dicus and W. W. Repko, *Dalitz decay $H \to f\bar{f}\gamma$ as a background for $H \to \gamma\gamma$*, Phys. Rev. D **89** (2014) 093013, arXiv: 1402.5317 [hep-ph] (cit. on p. 42).

[63] S. Alioli et al., *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043, (using ATLAS svn revisions r2856 for v1, r3080 for v2 ggF, r3052 for v2 VBF, and r3133 for v2 *VH*), arXiv: 1002.2581 [hep-ph] (cit. on pp. 42, 43).

[64] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, Eur. Phys. J. C **76** (2016) 196, arXiv: 1512.01178 [hep-ph] (cit. on p. 43).

[65] E. Bothmann et al., *Event generation with Sherpa 2.2*, SciPost Phys. **7** (2019) 034, arXiv: 1905.09127 [hep-ph] (cit. on p. 43).

[66] ATLAS Collaboration, *Evidence for Higgs boson decays to a low-mass dilepton system and a photon in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Lett. B **819** (2021) 136412, Aux. Mat.: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/HIGG-2018-43/, ISSN: 0370-2693 (cit. on pp. 47, 49, 50, 64, 69, 88, 90, 91, 117).

[67] ATLAS Collaboration, *ATLAS data quality operations and performance for 2015–2018 data-taking*, JINST **15** (2020) P04003, arXiv: 1911.04632 [physics.ins-det] (cit. on p. 55).

[68] G. Cowan et al., *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. C **71** (2011) 1554, [Erratum: Eur. Phys. J. C **73** (2013) 2501], arXiv: 1007.1727 [physics.data-an] (cit. on pp. 57, 79, 81–83).

[69] D. L. Rainwater, R. Szalapski and D. Zeppenfeld, *Probing color singlet exchange in Z + 2-jet events at the CERN LHC*, Phys. Rev. D **54** (1996) 6680, arXiv: hep-ph/9605444 (cit. on p. 58).

[70] R. Brun and F. Rademakers, *ROOT – An object oriented data analysis framework*, Nucl. Instrum. Methods Phys. Res. A **389** (1997) 81, New Computing Techniques in Physics Research V, ISSN: 0168-9002 (cit. on p. 59).

[71] ATLAS Collaboration, *Searches for exclusive Higgs and Z boson decays into $J/\Psi\gamma$, $\Psi(2S)\gamma$, and $\Upsilon(nS)\gamma$ at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Lett. B **786** (2018) 134, ISSN: 0370-2693 (cit. on p. 59).

[72] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using 2012 LHC proton–proton collision data*, Eur. Phys. J. C **77** (2017) 195, arXiv: 1612.01456 [hep-ex] (cit. on p. 59).

[73] M. Oreglia, *A Study of the Reactions $\psi' \to \gamma\gamma\psi$*, 1980, URL: https://www.slac.stanford.edu/cgi-wrap/getdoc/slac-r-236.pdf (cit. on p. 63).

[74] ATLAS Collaboration, *Search for Scalar Diphoton Resonances in the Mass Range 65-600 GeV with the ATLAS Detector in pp Collision Data at $\sqrt{s} = 8$ TeV*, Phys. Rev. Lett. **113** (2014) 171801, arXiv: 1407.6583 [hep-ex] (cit. on p. 63).

[75] S. Bernstein, *Démonstration du théorème de Weierstrass fondée sur la calcul des probabilités*, Comm. Soc. Math. Charkow Sér. 2 **13** (1912) 1, English translation: https://www.mn.uio.no/math/english/people/aca/michaelf/translations/bernstein_english.pdf (cit. on p. 68).

[76] N. Berger et al., *Simplified Template Cross Sections - Stage 1.1*, (2019), arXiv: 1906.02754 [hep-ph] (cit. on pp. 74, 75).

[77] I. W. Stewart and F. J. Tackmann, *Theory uncertainties for Higgs mass and other searches using jet bins*, Phys. Rev. D **85** (3 2012) 034011 (cit. on p. 74).

[78] S. Gangal and F. J. Tackmann, *Next-to-leading-order uncertainties in Higgs+2 jets from gluon fusion*, Phys. Rev. D **87** (9 2013) 093008 (cit. on pp. 74, 76).

[79] *STXS Twiki*, https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHWGFiducialAndSTXS, Accessed: 2021-07-24 (cit. on p. 75).

[80] S. Gangal and F. J. Tackmann, *Next-to-leading-order uncertainties in Higgs+2 jets from gluon fusion*, Phys. Rev. D **87** (9 2013) 093008 (cit. on pp. 77, 78).

[81] A. Wald, *Tests of statistical hypotheses concerning several parameters when the number of observations is large*, Trans. Am. Math. Soc. **54** (3 1943) 426 (cit. on p. 82).

[82] A. Read, *Presentation of search results: the CL_s technique*, JPhysG **28** (2002) 2693 (cit. on p. 83).

[83] H. B. Curry, *The method of steepest descent for non-linear minimization problems*, Quart. Appl. Math. **2** (1944) 258 (cit. on p. 84).

[84] F. James and M. Winkler, *Minuit2*, https://root.cern.ch/root/htmldoc/guides/minuit2/Minuit2.html, Accessed: 2021-07-25 (cit. on pp. 84, 111).

[85] A. Kachanovich, U. Nierste and I. Nisandzic, *Higgs boson decay into a lepton pair and a photon revisited*, Phys. Rev. D **101** (7 2020) 073003 (cit. on p. 93).

[86] *HllyStatAnalysis – statistical analysis software*, https://gitlab.cern.ch/atlas-hgam-sw/h-to-ystary-software/HllyStatAnalysis,
Accessed: 2021-09-01 (access restricted to the ATLAS Collaboration). (cit. on p. 117).

[87] ATLAS Collaboration,
ATLAS finds evidence of a rare Higgs boson Dalitz decay to two leptons and a photon,
https://atlas.cern/updates/briefing/evidence-rare-Higgs-decay, 2021
(cit. on p. 117).

APPENDIX A

Signal and resonant background fits

The selected events $m_{\ell\ell\gamma}$ distribution obtained with MC simulation is fitted with double-sided Crystal Ball function independently in each analysis category. Fit functions and their parameters for the signal are shown in Figures A.1, A.2, A.3 and for the $H \to \gamma\gamma$ background in Figure A.4. Note that μ^{CB} values shown are without the +90 MeV m_H correction applied.

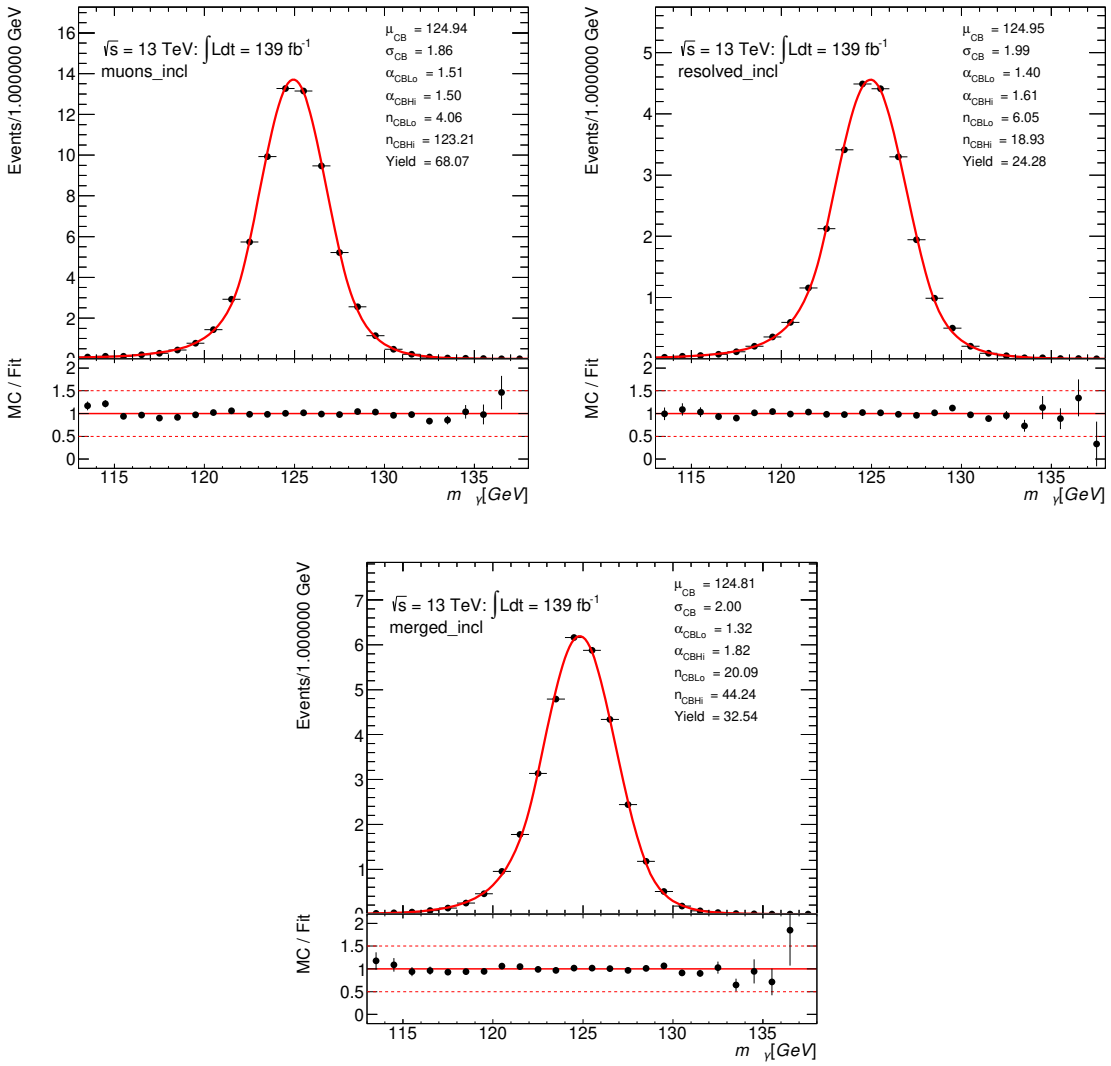


Figure A.1: Signal fit in low-$p_{\mathrm{TThrust}}^{\ell\ell\gamma}$ analysis categories using a double-sided Crystal Ball function.

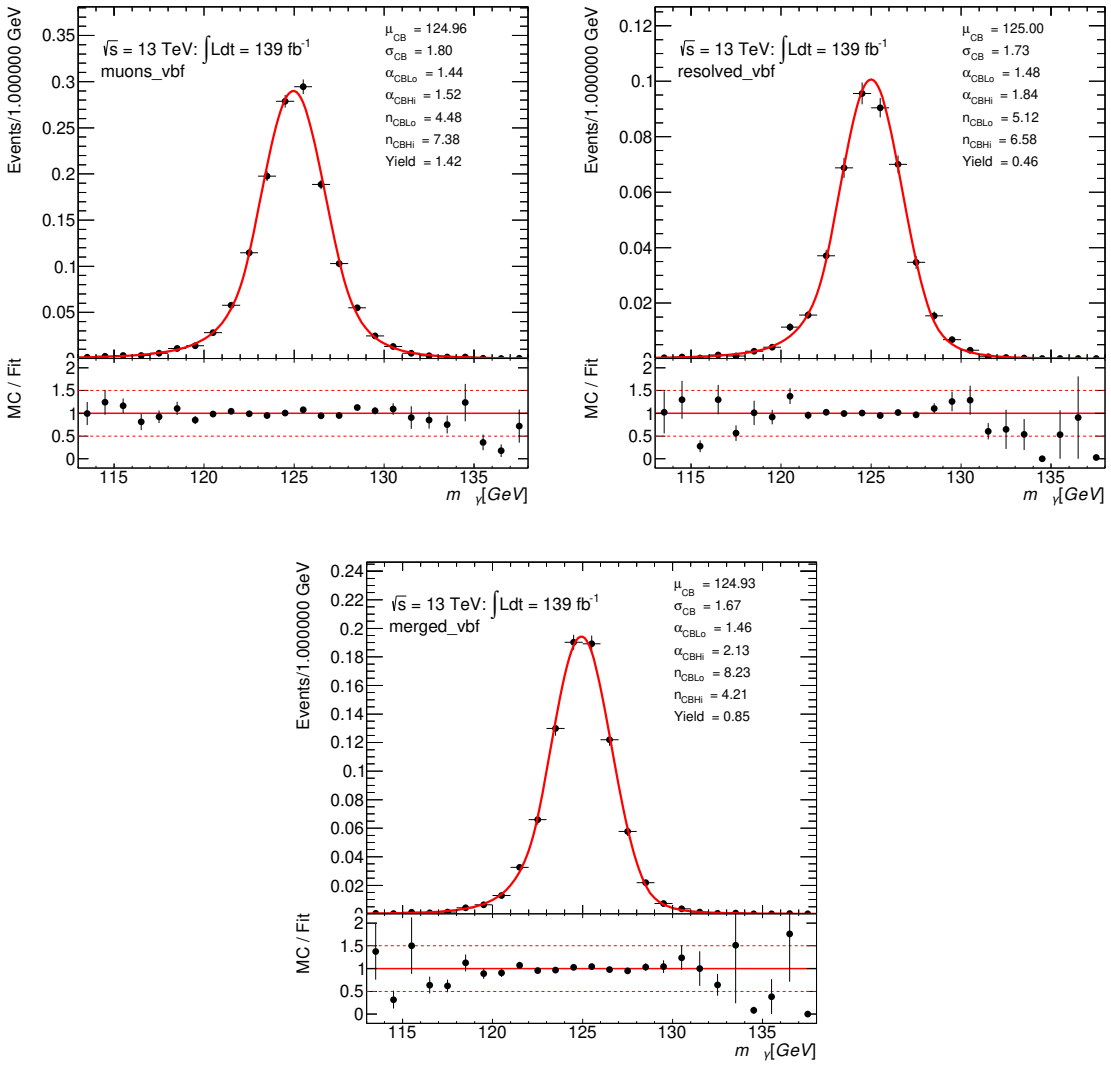


Figure A.2: Signal fit in VBF analysis categories using a double-sided Crystal Ball function.

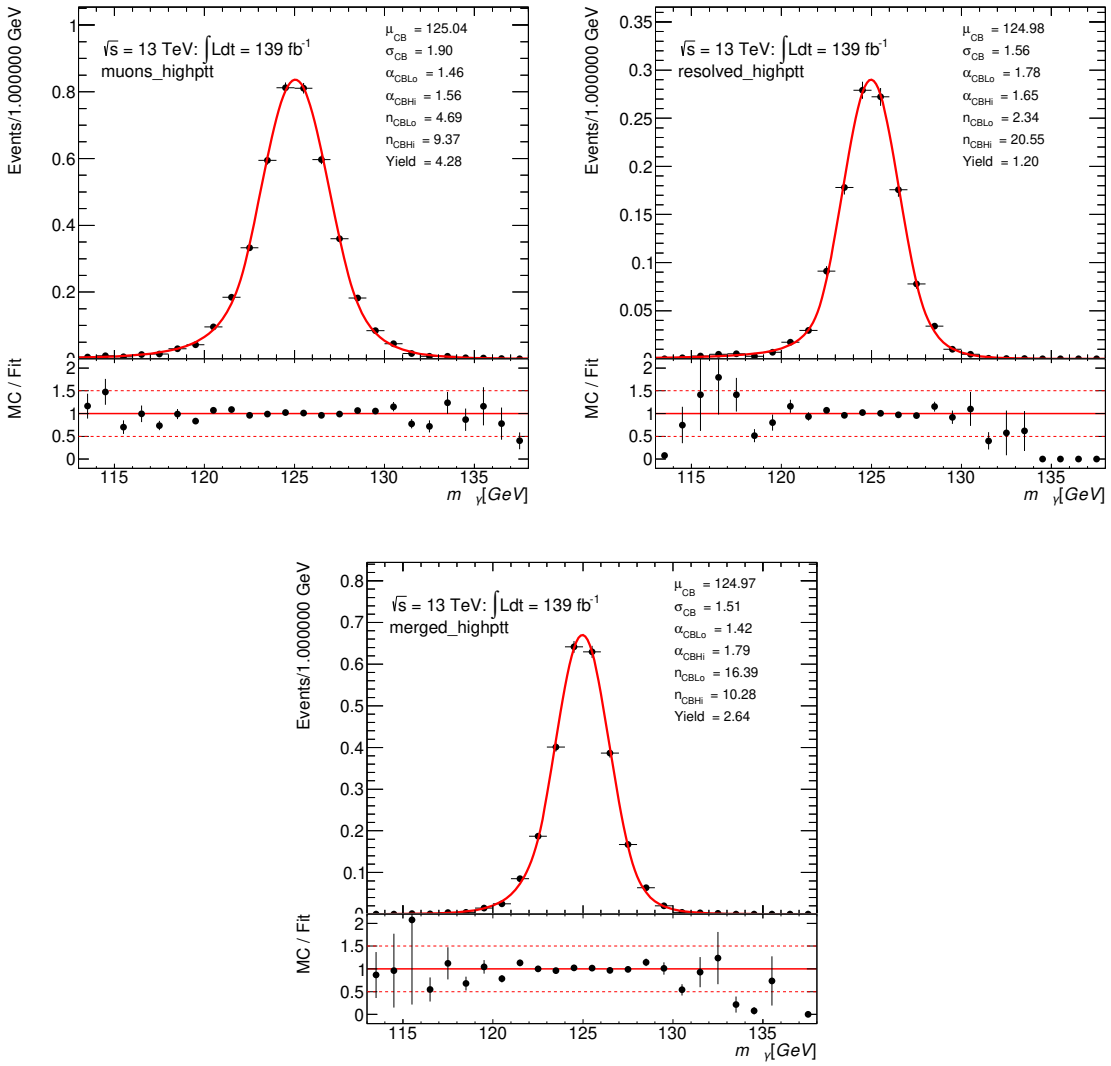


Figure A.3: Signal fit in high-$p_{\mathrm{TThrust}}^{\ell\ell\gamma}$ analysis categories using a double-sided Crystal Ball function.

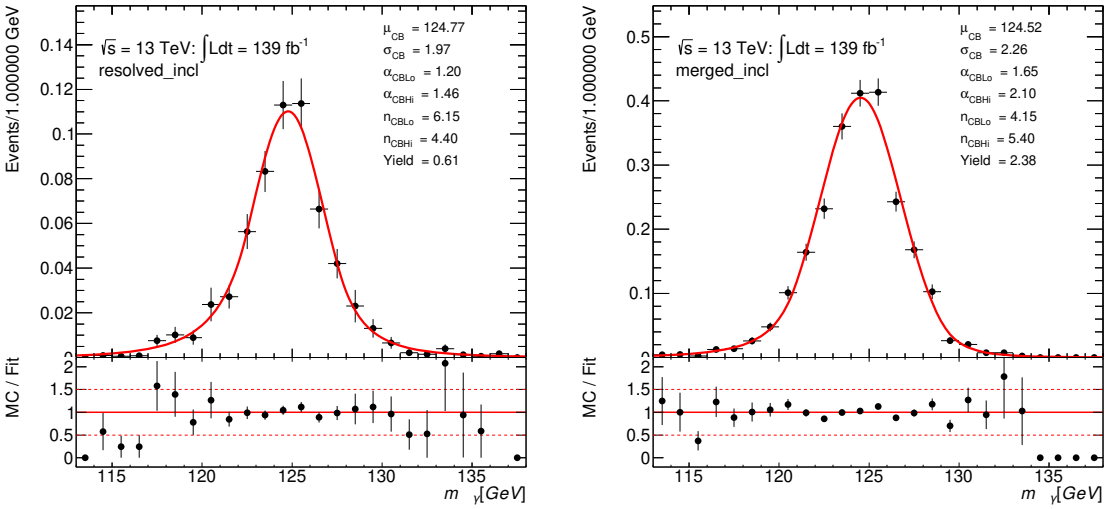


Figure A.4: $H \to \gamma\gamma$ resonant background fit in low-$p^{\ell\ell\gamma}_{\text{TThrust}}$ merged electron and resolved electron categories using a double-sided Crystal Ball function.

APPENDIX B

Negative log-likelihood ratio profiles

The $\pm 1\sigma$ uncertainty bands for the signal strength value μ in different scenarios studied in this analysis are obtained by profiling the negative log-likelihood ratio using the following algorithm:

- Set a sufficiently wide initial range for the μ, for example $-3 < \mu < 7$ (adjusted on a case-by-case basis).
- Uniformly divide the μ range into 500 intervals, and for each μ value minimize the negative log-likelihood ratio function (fixing μ) using `Minuit2` [84] software, which combines several minimization algorithms, to get the $-2\ln\lambda(\mu)$ value.
- Find the values of $-2\ln\lambda(\mu)$ closest to 1 and 0, which correspond to approximate points of interest, $\mu_{-1\sigma}$, $\mu_{+1\sigma}$ and $\mu_{\text{best fit}}$.
- To reach an adequate precision, 250 additional scan points are added in the vicinity of each of the approximate points of interest (roughly ± 0.025 in the μ range around each point of interest), for which $-2\ln\lambda(\mu)$ is calculated.
- Using the above procedure, 1250 points for $-2\ln\lambda(\mu)$ are obtained, with most points concentrated around $\mu_{\pm 1\sigma}$ and $\mu_{\text{best fit}}$, where the highest precision is required.
- Interpolate the $-2\ln\lambda(\mu)$ function with cubic splines. Using the bisection algorithm (set up to reach a precision of 10^{-8} or 999 iterations, whichever comes first), find precise values for $\mu_{\pm 1\sigma}$ and $\mu_{\text{best fit}}$.
- Plot $-2\ln\lambda(\mu)$ in the full μ range.

The $-2\ln\lambda(\mu)$ plots are analyzed to ensure that all negative log-likelihood ratio functions behave smoothly and to verify that the two points $-2\ln\lambda(\mu) \approx 1$ and one point $-2\ln\lambda(\mu) \approx 0$ are within the initial range.

Profiles of $-2\ln\lambda(\mu)$ as a function of signal strength are obtained for several different scenarios. In the main result, there is a single global parameter of interest, μ, in the fit, shown in Figure B.1. In another scenario, μ parameter is floating independently in each category in a simultaneous fit, shown in Figure B.2. Finally a simultaneous fit is performed with μ floating independently in different groupings of categories, shown in Figures B.3, B.4, and B.5. The results of these profiles for different parameter-of-interest scenarios are used to create the results shown in Figure 4.17 and 4.18.

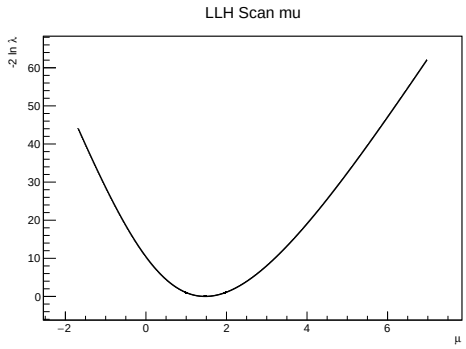


Figure B.1: Profile of the negative log-likelihood ratio as a function of signal strength, with a single global parameter of interest μ (main result).

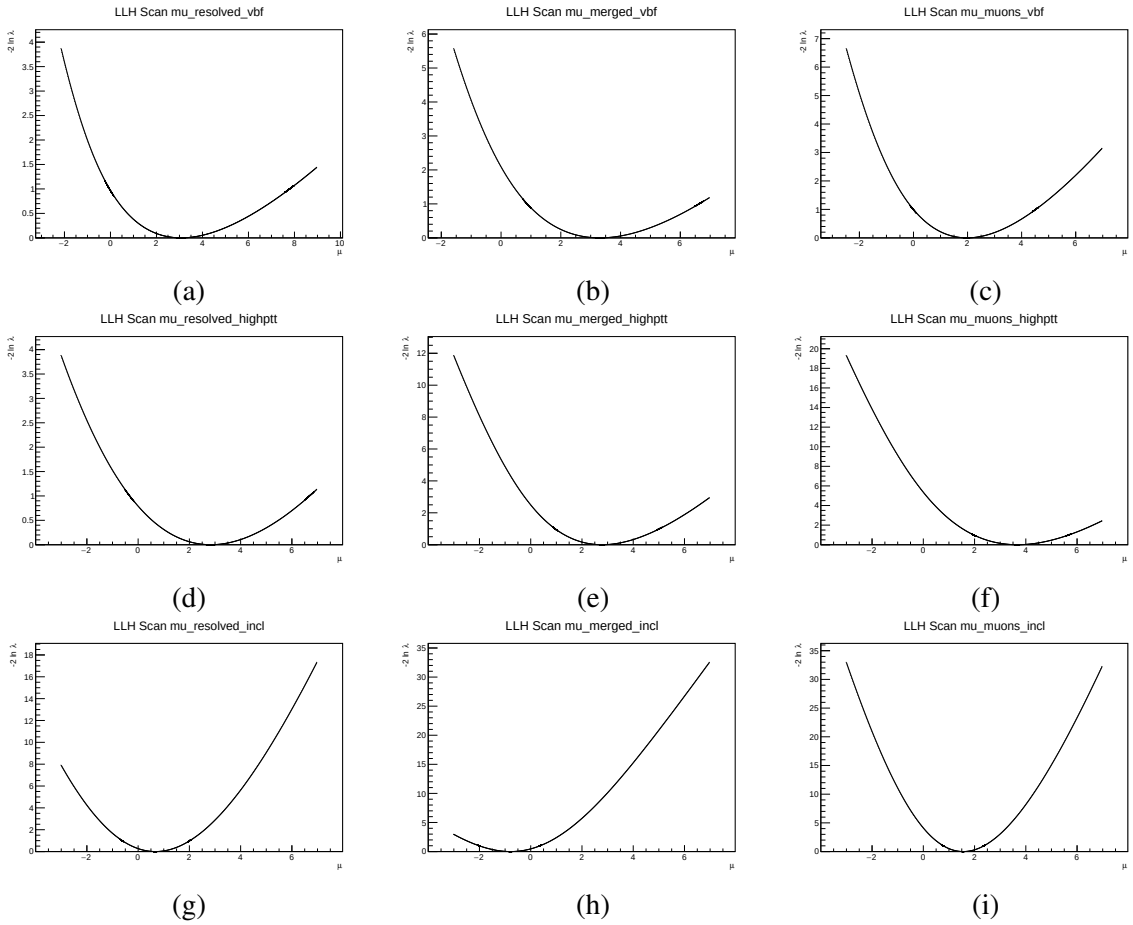
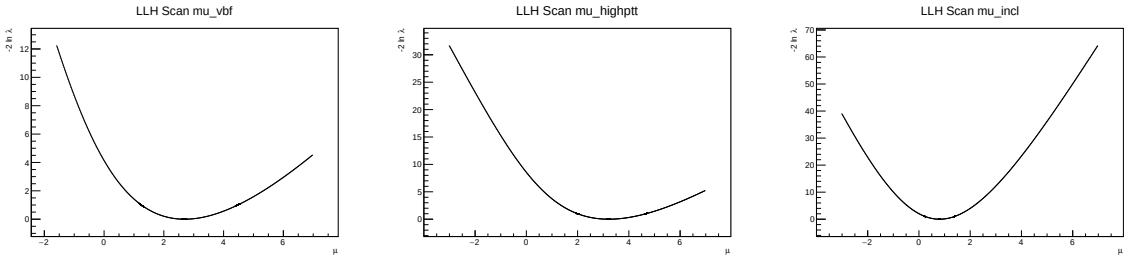
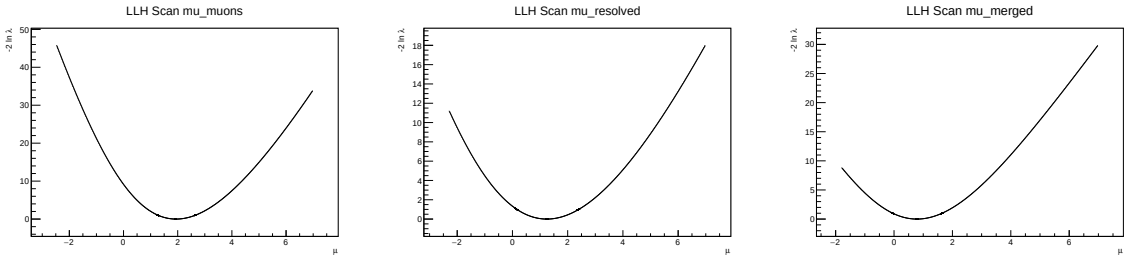


Figure B.2: Profiles of the negative log-likelihood ratio as a function of signal strength μ in the scenario with 9 μ parameters floating independently in each category in a simultaneous fit: the resolved-ee, merged-ee and $\mu\mu$ VBF categories (a, b, c), the resolved-ee, merged-ee and $\mu\mu$ high-$p_{\mathrm{TThrust}}^{\ell\ell\gamma}$ categories (d, e, f), and the resolved-ee, merged-ee and $\mu\mu$ low-$p_{\mathrm{TThrust}}^{\ell\ell\gamma}$ categories (g, h, i).



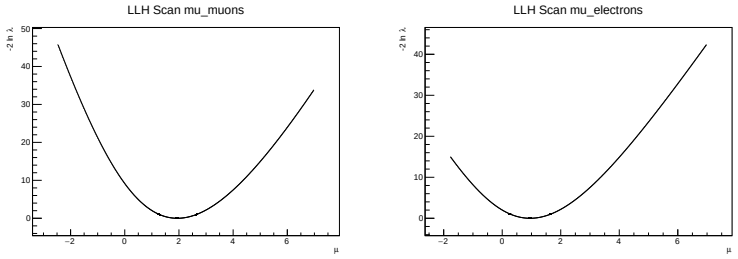
(a) (b) (c)

Figure B.3: Profiles of the negative log-likelihood ratio as a function of signal strength μ in the scenario with 3 μ parameters floating independently in a simultaneous fit: one μ parameter for all VBF categories (a), one for all high-$p_{\mathrm{TThrust}}^{\ell\ell\gamma}$ categories (b), and one for all low-$p_{\mathrm{TThrust}}^{\ell\ell\gamma}$ categories (c).



(a) (b) (c)

Figure B.4: Profiles of the negative log-likelihood ratio as a function of signal strength μ in the scenario with 3 μ parameters floating independently in a simultaneous fit: one μ parameter for all $\mu\mu$ categories (a), one for all resolved-*ee* categories (b), and one for all merged-*ee* categories (c).



(g) (b)

Figure B.5: Profiles of the negative log-likelihood ratio as a function of signal strength μ in the scenario with 2 μ parameters floating independently in a simultaneous fit: one μ parameter for all $\mu\mu$ categories (a), and one for all *ee* categories (b).

APPENDIX C

Cross-checks of systematic uncertainty ranking

The expected impact of systematic uncertainties on the signal strength μ was compared to the observed impact after the fit to data. Only small differences were found, which are within expectation. In particular, small differences between the expected and the observed impact were found for the e/γ energy resolution and e/γ energy scale uncertainties. To study this effect in more detail, the impact was measured separately in the muon and electron channels.

For the e/γ energy resolution uncertainty, in the muon and electron channels an opposite effect is seen between the expected and the observed impact, shown in Table C.1. In the electron channel, the direction of change corresponding to the $+1\sigma$ variations (reflected by sign of the impact value) is opposite in the expected and the observed result. In the muon channel, the direction of change is the same. When both channels are combined, the impact partially cancels out, resulting in reduced total observed impact compared to the expected.

For the e/γ energy scale uncertainty, there is no similar clear trend in comparison of the impact in the electron and muon channels, as also shown in Table C.1. Alternatively, the observed effect

Setup	e/γ resolution impact $\Delta_\mu/\hat{\mu}$			e/γ scale impact $\Delta_\mu/\hat{\mu}$		
	$+1\sigma$	-1σ	symm	$+1\sigma$	-1σ	symm
Expected $S+B$ model (ee-only)	+0.039	-0.046	0.043	-0.018	-0.006	0.012
Expected $S+B$ model ($\mu\mu$-only)	+0.023	-0.026	0.025	-0.007	+0.000	0.003
Expected $S+B$ model (all channels)	+0.026	-0.038	0.032	-0.016	-0.005	0.011
Observed $S+B$ model (ee-only)	-0.044	+0.059	0.052	-0.162	+0.071	0.117
Observed $S+B$ model ($\mu\mu$-only)	+0.011	-0.020	0.016	+0.014	-0.028	0.021
Observed $S+B$ model (all channels)	+0.005	-0.004	0.005	-0.027	+0.006	0.017

Table C.1: Impact of $\pm1\sigma$ uncertainty variations on the signal strength, $\Delta_\mu/\hat{\mu}$, for e/γ energy scale and resolution uncertainty. The impact is reported relative to the expected μ value for the expected $S+B$ model, and relative to the observed μ value for the observed $S+B$ model, respectively. The "symm" column shows the symmetrized impact, $(|\Delta_\mu^{+1\sigma}/\hat{\mu}| + |\Delta_\mu^{-1\sigma}/\hat{\mu}|)/2$. The $ee/\mu\mu$-only results are obtained only with $ee/\mu\mu$ categories. The e/γ energy resolution uncertainty ranks higher than the e/γ energy scale uncertainty in the expected result, while the reverse is true for the observed result.

Setup	Best-fit m_H	e/γ scale pull	e/γ scale constraint		e/γ scale impact $\Delta_\mu/\hat{\mu}$	
			$+1\sigma$	-1σ	$+1\sigma$	-1σ
Fixed $m_H = 125.09$ GeV	–	0.25	+0.83	-0.84	-0.027	0.06
Single POI m_H	$m_H = 124.8 \pm 0.5$	-0.00019	+1.00005	-1.00005	$< 1 \times 10^5$	
Two POIs m_H^{ee}, $m_H^{\mu\mu}$	$m_H^{ee} = 124.2 \pm 0.7$, $m_H^{\mu\mu} = 125.5 \pm 0.7$	-0.00017	+1.00004	-1.00004	$< 2 \times 10^5$	

Table C.2: Constraints for the nuisance parameter, pull and impact on the signal strength, $\Delta_\mu/\hat{\mu}$, for $\pm 1\sigma$ variations of the e/γ energy scale uncertainty. Several scenarios of the fit are considered: fixed $m_H = 125.09$ GeV (as in the main results), a single global floating parameter of interest (POI), m_H, and two independent POIs in the muon and electron channels, m_H^{ee}, $m_H^{\mu\mu}$.

may arise due to the signal peak being at the slightly different mass value in data than in the MC-based expectation due to reasons of a purely statistical nature, which is plausible given the statistical uncertainties of this analysis. To partially test this idea, a fit was performed with the Higgs boson mass allowed to float (also, as an additional cross-check, it was done separately in electron and muon channels), shown in Table C.2. It can be seen that with m_H allowed to float, the corresponding nuisance parameter assumes slightly lower value than $m_H = 125.09$ GeV (i.e. the value which was input into the expected signal-plus-background model) both in the global fit and in the ee-channels separately. At the same time, the pull on the e/γ energy scale uncertainty is removed, and its impact is reduced, which supports the idea of a slightly different e/γ energy scale seen in data. As already mentioned, this difference does not indicate any significant deviation from the prediction, and can be entirely explained by a statistical fluctuation in data.

Author's Contribution

The ATLAS experiment is a collaborative effort, sharing contributions of thousands of scientists, students and engineers in all areas ranging from detector design and operation to the data analysis and review of the results. Without numerous contributions from the collaborators, the results presented in this thesis would be impossible to achieve. The most relevant ones for this thesis are explicitly mentioned in the Acknowledgments. This chapter highlights the areas where the author personally contributed.

Search for Higgs boson decaying to a low mass lepton pair and a photon

The author defined the criteria for the event selection maximizing the expected sensitivity. An iterative procedure for the selection optimization, based on the approximate expected significance was developed. The author performed studies to find the most efficient combination of triggers, lepton identification and isolation criteria. More details are provided in Section 4.5.

The author constructed likelihood-based statistical model for the search (described in Section 4.8). All relevant experimental and theory systematic uncertainties were incorporated into it (described in Section 4.7). The theory uncertainties on the QCD scale were estimated by the author using a custom approach, since techniques commonly used in the collaboration were not applicable for the presented search (see Section 4.7.2 for details).

Using the statistical model, the author calculated all final results presented in Section 4.9 and published in Ref. [66]. The author combined several different tools developed by the collaboration for that purpose, and developed a steering software targeted for the presented search based on them. The steering software significantly stream-lined evaluation of the results by making it possible to reproduce all interpretation results quickly after minor changes to the procedure, such as taking into account additional sources of systematic uncertainties. The software can be used for this or similar searches in the future, and is shared within collaboration, together with documentation [86].

The author performed multiple studies to prove the consistency of the reported results, described in Section 4.9.3. The ability to perform such consistency checks was implemented into the steering software as well.

The author presented the analysis and results in ATLAS internal approval meetings, served as an editor of the internal documentation of the search and contributed to the CERN Physics Briefing article [87] devoted to the search results.

ATLAS Operation Tasks

During the data taking, the ATLAS detector operation is continually supervised in shifts from the ATLAS Control Room at CERN. The author did Run Control shifts, configuring, starting and stopping data taking and calibration runs, and monitored the overall status of all subsystems.

Simulation of the ATLAS detector is indispensable tool for obtaining virtually all the results published by the collaboration (described in Section 3.2.6). Simulation is being constantly improved, both in performance, to meet the growing demand, and in precision with better understanding of the detector effects. Slightly different conditions during experimental runs often require new simulated samples for different years of data taking.

The effort for quality control of the development of simulation software requires continuous shifts for checking both software quality of the daily ATLAS software builds, and the validity of the physics results, achieved by running several "benchmark" simulation tests every day. During his thesis project, the author coordinated this effort, which included direct supervision of the shifters, serving as a liaison between simulation and software coordinators and the shifters, organizing training of new shifters, and managing shift bookings.

List of Figures

List of Tables