Searching for the Higgs

Replication of $H \to \gamma\gamma$ decay signals using data collected by the ATLAS experiment at the Large Hadron Collider (LHC)

Ellie Crossley-Fells

July 16, 2025

Introduction

I've been interested in high energy physics research at CERN since I was a kid. During my high school years I followed the search for the Higgs Boson, and distinctly remember the announcement of its discovery in 2012. 13 years on, going from Registered Nurse to Junior Data Analyst and part-time physics undergraduate, I've gone from being a side-line enjoyer of physics, to someone who is beginning understand, and able to interpret the data.

As my first physics project, I've chosen a topic that initially got me into physics. I will be using proton-proton (di-photon) collision data from the ATLAS experiment at CERN to measure Higgs Boson properties in the di-photon decay channel ($H \to \gamma\gamma$). It is a simplifed version of the research conducted on the Higgs: the data is already curated (skimmed) to only contain di-photon events, and the photon selection criteria and methodology is simplified.

The ATLAS Collaboration releases its data for public use; a statement of belief in the power of open science, and making science accessible. Science has not always been accessible to me throughout my younger years, and I love the idea that this data is available for people to learn with.

I don't claim to have created this code entirely independently. I followed along from the $H \to \gamma\gamma$ notebook, created by ATLAS Open Data for educational purposes. However I have attempted my own bits for the visualisations, and added a section in the data download part to check for existing .root files to improve re-runability (is that a word?). I also created this write-up in my own words to consolidate my learning, which I hope provides an easy to understand breakdown of the research for someone with basic physics knowledge, and also shows the thrilling endevours that are my weekends and evenings.

I'm not an expert and I don't claim to be. This is an informal write-up of a personal project and not an attempt at an academic paper, but I have made sure this document is as scientifically accurate and factually correct to the best of my ability, and appropriate for my level of knowledge as a first year undergraduate.

The Higgs Boson

Quantum Field Theory describes particles as excitations of underlying *quantum fields*, rather than fundamental, indivisible objects like we often think of them. These quantum fields permeate the entire universe, and every particle can be represented as a *wave* permeating these quantum fields. For example, the photon is a wave (or excitation) within the electromagnetic field.

The Higgs field is another quantum field. When fundamental particles (electrons, quarks etc) move through the Higgs field, they interact with it. This *interaction* gives particles mass and the stronger the interaction, the more massive the particle.

The Higgs Boson is the *carrier* of the Higgs field, but it is slightly different in the sense that it does not mediate a force between particles in the usual way (like photons in the electromagnetic field). Instead, the presence of the Higgs field with a non-zero value *everywhere in the universe* is what gives a particle its mass. *Spontaneous symmetry breaking* (a concept I won't go into here) gives a rest mass to all massive elementary particles of the Standard Model.

Higgs Bosons are unstable and difficult to produce: they require extremely high energy particle collisions and decay rapidly (10^{-22}s) into other particles, making direct observation of the Higgs impossible. We therefore look at its *decay channels* (different ways a particle transforms into other particles to become more stable).

Using the standard model, predictions can be made about the decay properties of the Higgs, and the probability of each decay channel can be calculated.

There are 4 main decay channels for the Higgs:

1. $H \to b\bar{b}$ (Higgs to b-quark and anti-b-quark 56.9%)
2. $H \to ZZ \to LLLL$ (Higgs boson to two Z bosons to four leptons 0.0131%)
3. $H \to \tau\bar{\tau}$ (Higgs to a tau lepton and an anti-tau lepton 6.28%)
4. $H \to \gamma\gamma$ (Higgs boson to two photons 0.229%)

The decay of the Higgs Boson into 2 gamma photons ($H \to \gamma\gamma$) is a distinguishable decay channel during proton-proton collisions in the LHC. It is not the most probable decay channel, but it is one of the most detectable.

Data Analysis

Working with ROOT files

I have never worked with ROOT files before, but they are a standard for working with data in particle physics and something I intend to become more familiar with. The primary data structure in a ROOT file is the tree, which contains branches (columns) and leaves (data entries or values). Branches can also contain complex nested collections.

There are 2 distinct parts to the analysis: background photon reduction using data cuts, and invariant mass calculation.

Photon Background Reduction

During proton-proton collisions many photons are detected that do not originate from a Higgs decay, these form the background. (For example, from neutral pion decay and initial-state radiation). The aim is to isolate only collisions that produce 2 gamma photons.

Reducible Backgrounds can be minimized using data cut techniques that involves applying strict selection criteria to events, ensuring detected particles are photons and have not been misidentified.

Irreducible Backgrounds cannot be distinguished from the Higgs decay signal because they involve the same final states or processes as the signal. For example, any particle-antiparticle annihilation also produces 2 gamma photons. To address this, the *total invariant mass* of the photon products is calculated. By conservation of energy and momentum, the invariant mass of the products must equal the Higgs mass, whereas other background processes will have different invariant masses.

The final step of the analysis is to plot the invariant mass of each event to identify the peak around 125 GeV. (The Higgs boson mass range was predicted based on theoretical predictions and experimental constraint to be in the region of 124–126 GeV.)

Invariant Mass Calculation
The data to be plotted is the di-photon invariant mass, which can be calculated using the equation:

$$m_{\gamma\gamma} = \sqrt{\mathbf{E}_{\text{tot}}^2 - \mathbf{p}_{\text{tot}} \cdot \mathbf{p}_{\text{tot}}}$$

Where $c = 1$. $\mathbf{E}_{\text{tot}}$ is the total energy and $\mathbf{p}_{\text{tot}}$ is the total momentum. This calculation is performed using the vector array method .M on the sum of the photon's 4 momenta: (photon_pt,photon_eta,photon_phi,photon_e).

Data Cuts (Applying Photon Selection Criteria)
The photon selection criteria in this project is more simplified than ATLAS orignal analysis.

1. **Transverse Energy Cut**: The leading photon must have E_t (Transverse Energy) > 50 GeV and the sub-leading photon > 30 GeV because background processes frequently generate photons with lower E_t. (As photons are massless, enforcing restriction on E_t also enforces restriction on P_t (Transverse Momentum)).

2. **Calorimeter Isolation**: This cut distinguishes *prompt photons* (photons produced from the Higgs decay rather than background photons produced in jets or hadronic processes). First a "spatial cone" (Figure 1) is mapped around the photon, then the isolation energy the sum of the transverse energies of all calorimeter energy clusters inside this cone is calculated. Additional photon transverse energy relative to the diphoton mass isolation is also required. .For each photon, the extra energy in the cone must be less than 5.5% of the photon's own transverse momentum. Only these events are kept.

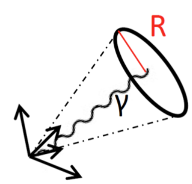

Figure 1: Photon Isolation Cone (Ronflette, 2014)

3. **Barrel-End Cap Transition Region Restriction** The transition region is where the barrel and the end-cap sections inside the calorimeter connect, and it can be a complex area with varying material compositions and a result can introduce uncertainties in measurements of particles. It is resolved by excluding events around the calorimeter barrel/end-cap transition region $1.37 < |n| < 1.52$..

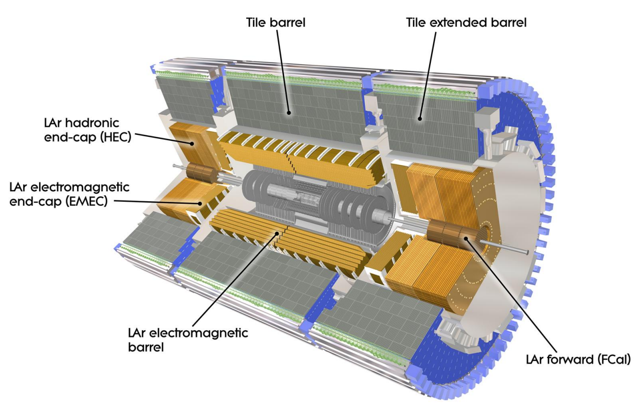

Figure 2: ATLAS Calorimetry System (ATLAS Collaboration, 2025)

4. **Invariant Mass Criteria**: The leading photon must have a transverse momentum greater than 35% of the invariant mass of the photon pair. Only events that match this criteria are kept for analysis.

```python
# Cut on the photon reconstruction quality
def cut_photon_reconstruction(photon_isTightID):
    # Only the events which have True for both photons are kept
    return (photon_isTightID[:,0]==False) | (photon_isTightID[:,1]==False)

# Cut on the transverse momentum
def cut_photon_pt(photon_pt):
# Only the events where photon_pt[0] > 50 GeV and photon_pt[1] > 30 GeV are kept
    return (photon_pt[:,0] < 50) | (photon_pt[:,1] < 30)

# Cut on the energy isolation
def cut_isolation_pt(photon_ptcone20, photon_pt):
# Only the events where the calorimeter isolation is less than 5.5% are kept
    return ((photon_ptcone20[:,0]/photon_pt[:,0]) > 0.055) | ((photon_ptcone20[:,1]/photon_pt[:,1]) > 0.055)

# Cut on the pseudorapidity in barrel/end-cap transition region
def cut_photon_eta_transition(photon_eta):
# Only the events where modulus of photon_eta is outside the range 1.37 to 1.52 are kept
    condition_0 = (np.abs(photon_eta[:, 0]) < 1.52) & (np.abs(photon_eta[:, 0]) > 1.37)
    condition_1 = (np.abs(photon_eta[:, 1]) < 1.52) & (np.abs(photon_eta[:, 1]) > 1.37)
    return condition_0 | condition_1

# This function calculates the invariant mass of the 2-photon state
def calc_mass(photon_pt, photon_eta, photon_phi, photon_e):
    p4 = vector.zip({"pt": photon_pt, "eta": photon_eta, "phi": photon_phi, "e": photon_e})
    invariant_mass = (p4[:, 0] + p4[:, 1]).M # .M calculates the invariant mass
    return invariant_mass

# Cut on null diphoton invariant mass
def cut_mass(invariant_mass):
    return (invariant_mass == 0)

# Cut on diphoton invariant mass based isolation
# Only the events where the invididual photon invariant mass based isolation is larger than 35% are kept
def cut_iso_mass(photon_pt, invariant_mass):
    return ((photon_pt[:,0]/invariant_mass) < 0.35) | ((photon_pt[:,1]/invariant_mass) < 0.35)
```

Figure 3: Python functions used to perform the data cuts

The functions are quite straightforward if you are familiar with the data. Having some metadata would have been useful, and it was only until after I finished the project that I realised this might have been available in the ROOT file.

Plotting

I used all of the available data, which on my laptop took a while to download. I did initially attempt with only 70% of the data, but the peak in invariant mass was shifted slightly too far to the right.

A **4th order polynomial** is used to represents the background photons, because it is flexible; it captures the overall shape of the background well without overfitting, and reduces the impact of random, irrelevant fluctuations that do not correspond to the true signal or background.

The Gaussian model is used to fit the signal due to the nature of the detector's resolution. The true Higgs mass is a fixed value, but the measured mass for each event is smeared out by the detector's finite energy resolution. This smearing is well-approximated by a Gaussian distribution.

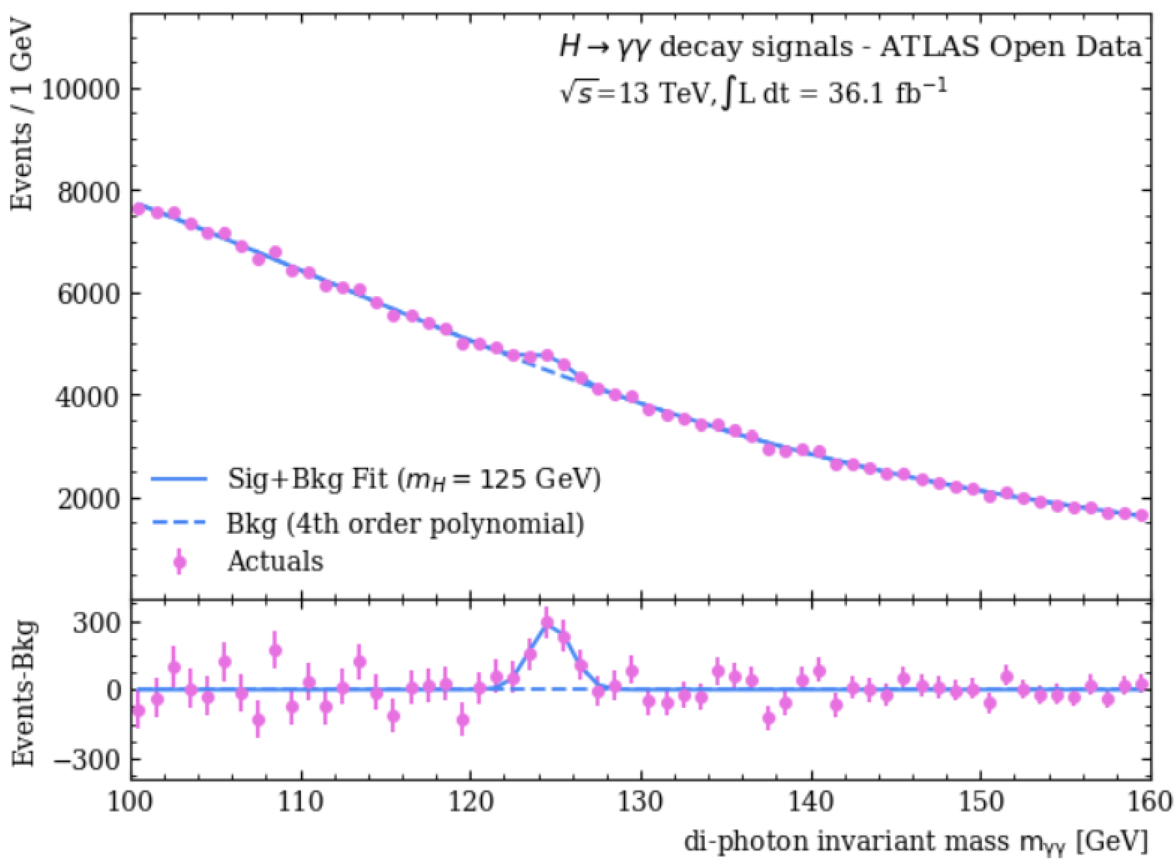

Figure 4: $H \rightarrow \gamma\gamma$ decay signals, showing a clear peak in invariant mass at c.125 GeV.

Above is one of the nicest looking charts I have made in my short data analytics career! Albeit, ATLAS did most of the hardwork, considering I didn't have to spend hours data cleaning.

The chart displays the results of the measurement of the Higgs boson decaying into two photons, using data collected by the ATLAS experiment at the Large Hadron Collider (LHC) in 2015 and 2016. The data corresponds to proton-proton collisions at a center-of-mass energy of 13 TeV, with a luminosity of 36 fb^{-1}.

- **X-axis:** Shows the invariant mass of the photon pairs, measured in GeV. If these two photons originate from the decay of a single particle (like the Higgs boson), their invariant mass will cluster around the mass of that particle.

- **Y-axis:** The number of events per GeV (event rate) for each bin of diphoton invariant mass.

- **Data Points (pink dots):** The actual measured data from the ATLAS detector, showing how many diphoton events were observed at each invariant mass.

- **Background Fit (blue dashed line):** The expected number of diphoton events from all Standard Model processes except the Higgs boson. This background is mostly due to other processes that produce two photons but not from Higgs decay.

- **Signal + Background Fit (solid blue line):** The sum of the background and the expected Higgs boson signal, as predicted by the Standard Model and fit to the data.

The narrow bump at c.125 GeV is a signature of the Higgs boson decaying into two photons. The Standard Model predicts the Higgs boson mass to be about 125 GeV.

I really enjoyed this project and look forward to doing more as my knowledge improves throughout my studies.

References

[1] ATLAS Collaboration (2012) ‘Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC’, *Physics Letters B*, vol. 716, no. 1, pp. 1–29. Available at: https://www.sciencedirect.com/science/article/pii/S037026931200857X (Accessed: 8 July 2025).

[2] ATLAS Collaboration (2018) ‘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’, *Physical Review D*, vol. 98, no. 5, 052005. Available at: https://journals.aps.org/prd/abstract/10.1103/PhysRevD.98.052005 (Accessed: 8 July 2025).

[3] ATLAS Collaboration (2019) ‘ATLAS Open Data Release: Higgs to diphoton analysis (2019 update)’, *ATLAS Public Note*, ANA-OTRC-2019-01-PUB. Available at: https://cds.cern.ch/record/2707171/files/ANA-OTRC-2019-01-PUB-updated.pdf (Accessed: 8 July 2025).

[4] ATLAS Collaboration (no date) ‘EM Calorimeter Image’. Available at: https://atlas.cern/Discover/Detector/Calorimeter (Accessed: 14 July 2025).

[5] ATLAS Collaboration (2018) ‘Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run 2 data collected in 2015 and 2016’, CERN-EP-2018-216. Available at: https://cds.cern.ch/record/2643391/files/PhotonIDPaper.pdf?version=1 (Accessed: 14 July 2025).

[6] Barillari, T., on behalf of the ATLAS Collaboration (2012) ‘Jet energy scale uncertainties in ATLAS’, ATL-PHYS-PROC-2012-137. Available at: https://cds.cern.ch/record/1473875/files/ATL-PHYS-PROC-2012-137.pdf (Accessed: 14 July 2025).

[7] Ronflette, L. (2014) ‘Isolated photon cross section studies in ALICE experiment’, *QGP France 2014*, 17 September 2014. Available at: https://llr.in2p3.fr/sites/qgp2014/Talks/QGP-France2014Ronflette.pdf (Accessed: 14 July 2025).