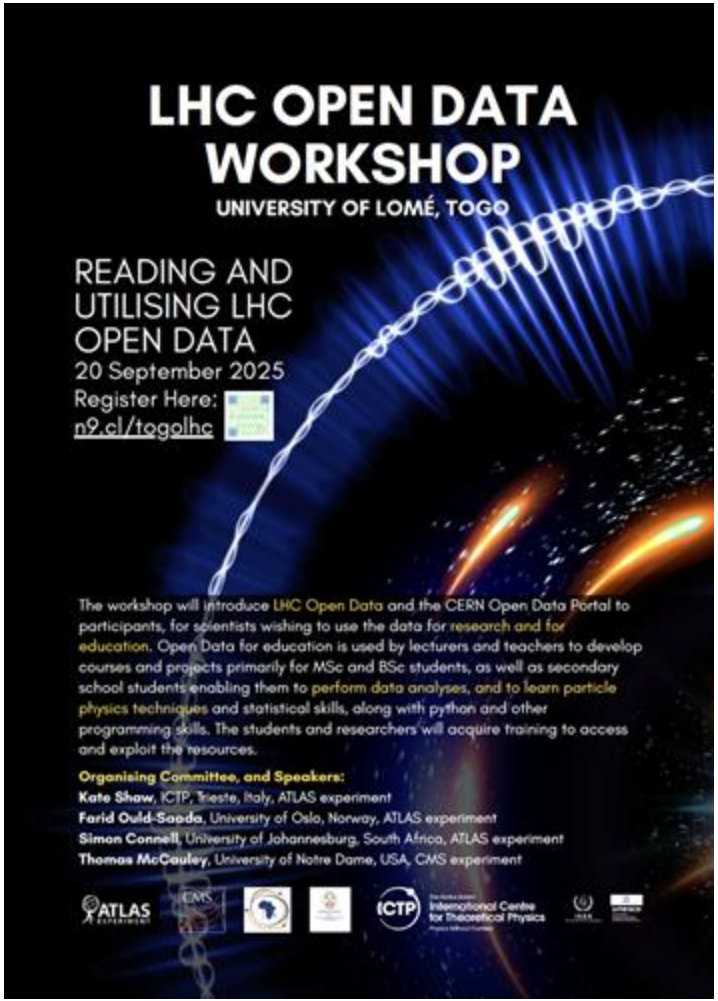
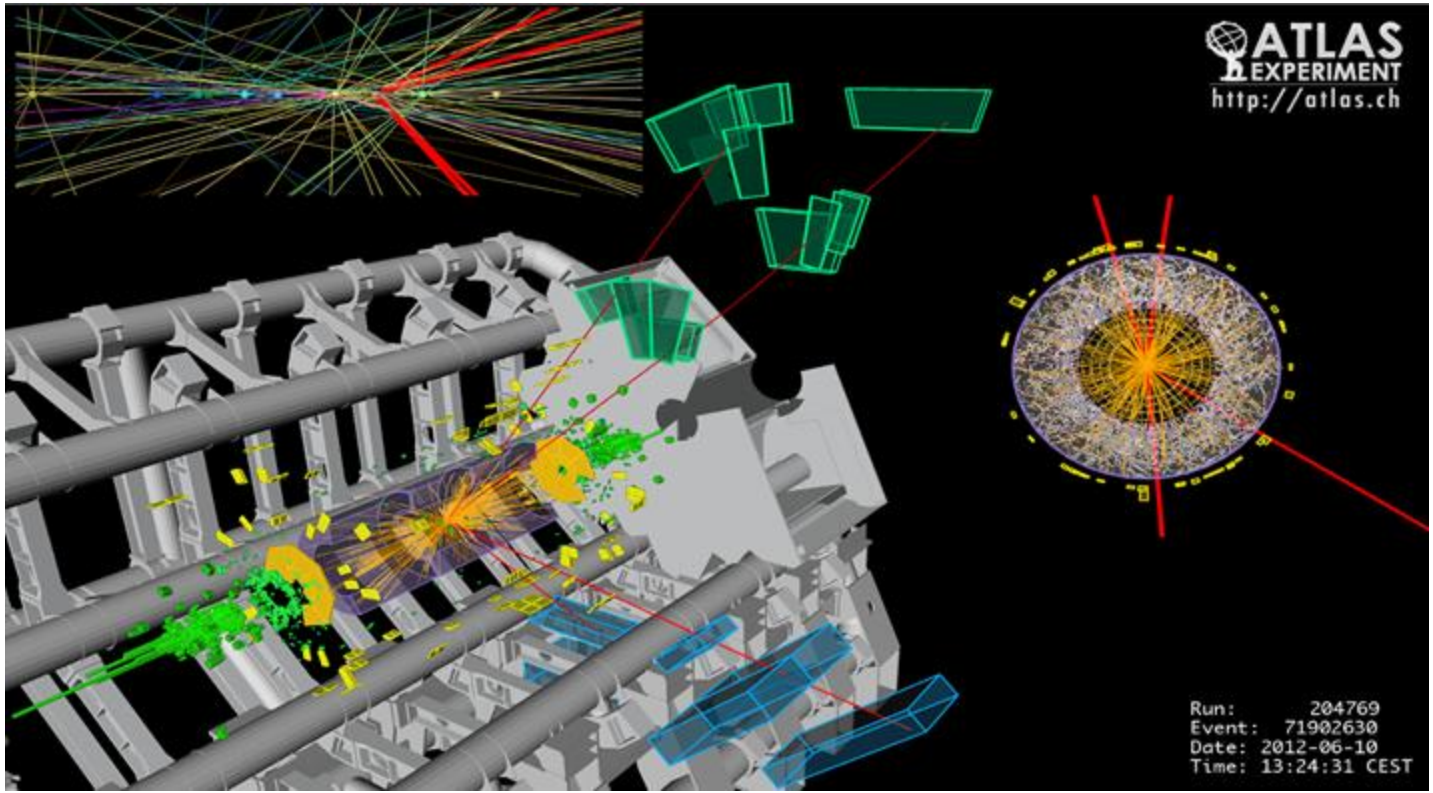
ATLAS Open Data Tutorial
Discover the Higgs Boson

Farid Ould-Saada, University of Oslo

African Conference on Fundamental and Applied Physics, Lomé, Togo 15-20.8.2025

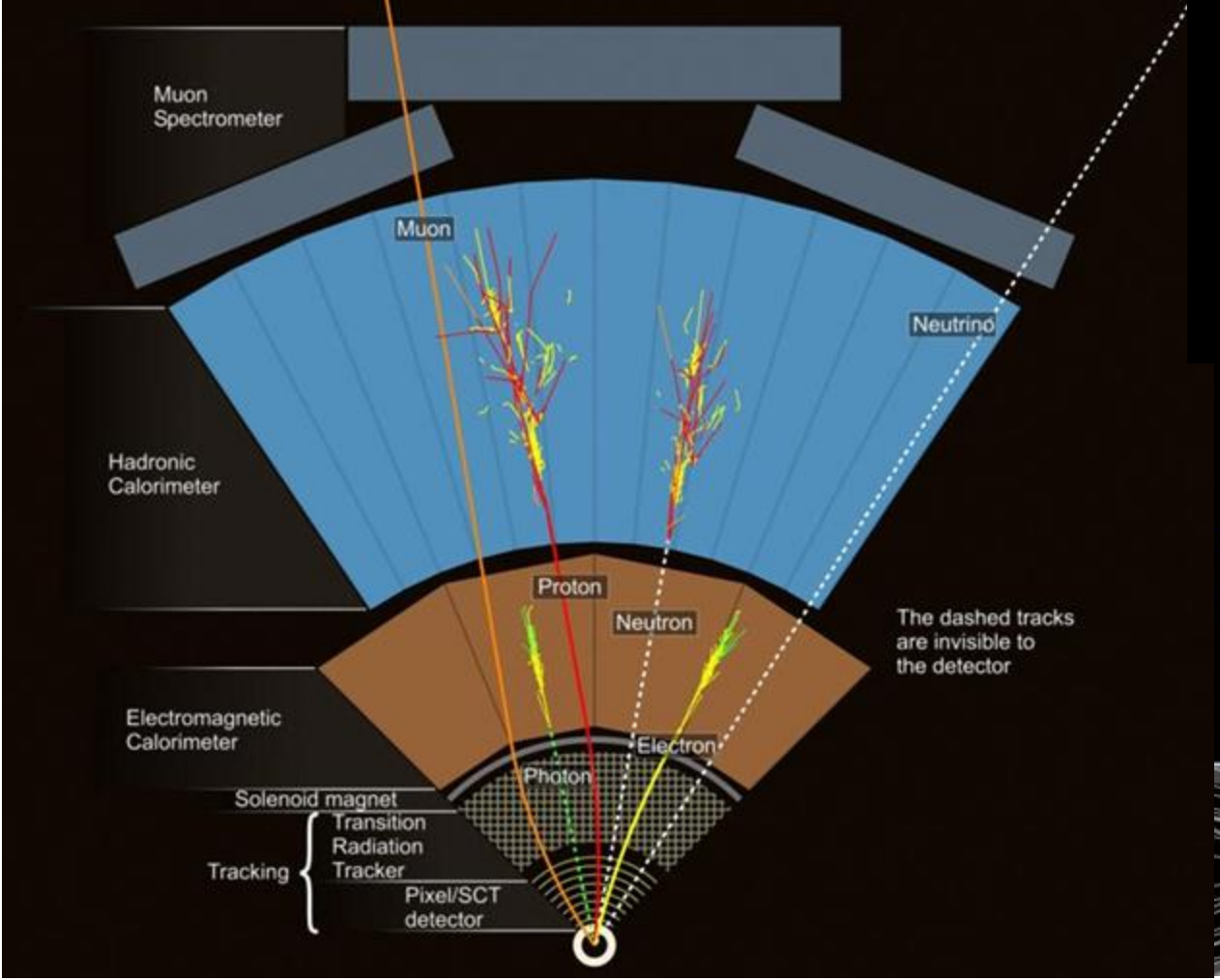


Proton - proton collisions at LHC

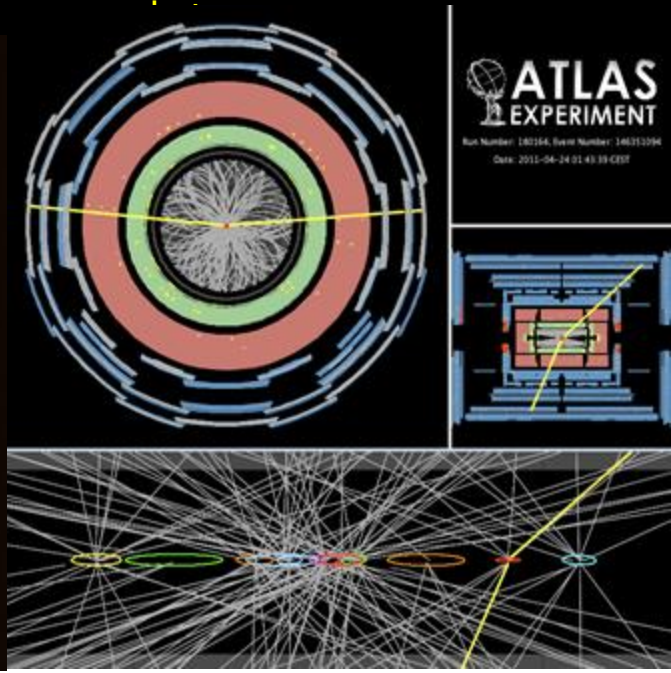


$$N_{\text{Physics}} = \sigma_{\text{Physics}} \left(\int \mathcal{L} dt \right)_{\text{Beams}}$$

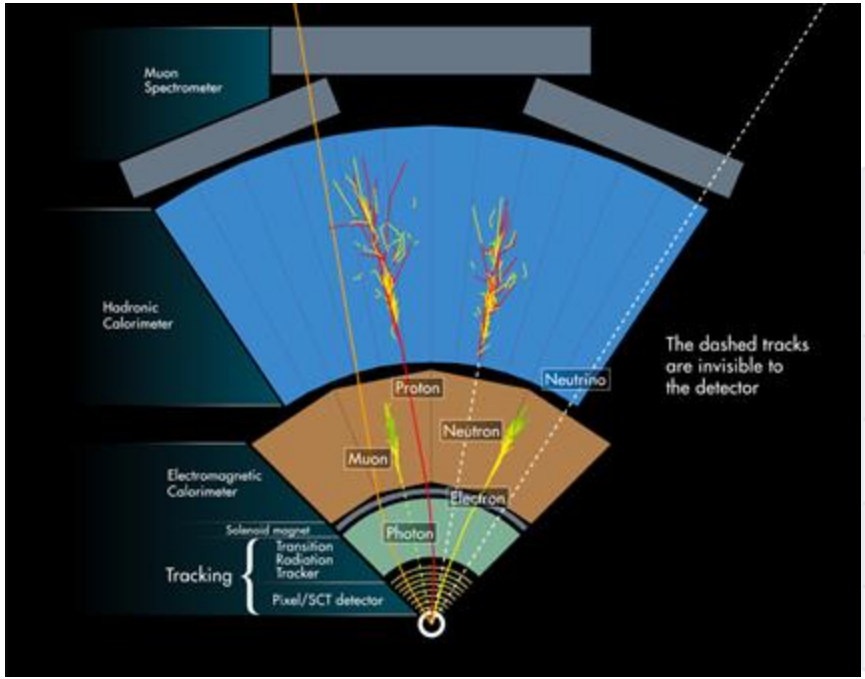
- Sensitivity to rare phenomena
  - with small cross sections
  - depends on the luminosity



- Particles have various signatures in different parts of the detector
- by combining the various signatures,
  - ➢ can reconstruct how particle moved through the detector
- Short-lived particles (Z, Higgs, top) reconstructed through their decay



Reconstructed physics objects and kinematical variables



Electron (e)	Muon (μ)	Photon (γ)
InDet & EMCAL rec.	InDet & MS rec.	InDet & EMCAL rec.
$p_T > 7$ GeV	$p_T > 7$ GeV	$E_T > 25$ GeV (>10)
$\|\eta\| < 2.47$	$\|\eta\| < 2.5$	$\|\eta\| < 2.37$

τ-leptons	Small-R jets	Large-R jets
InDet & EMCAL rec.	EMCAL & HCAL rec.	EMCAL & HCAL rec.
$p_T > 20$ GeV	anti-kt, R = 0.4	anti-kt, R = 1.0
$\|\eta\| < 2.5$	$p_T > 20$ GeV	$p_T > 250$ GeV
	$\|\eta\| < 2.5$ (<4.5)	$\|\eta\| < 2.0$
	b-tagging (DL1dv01)	trimming: $R_{sub} = 0.2$, $f_{cut} = 0.05$

Datasets available for Open Data for Education

- Several final-state collections specifically tailored towards physics analysis example notebooks
- Non-specific final state samples are also available

Selection	Collection Name
At least one lepton with at least 7 GeV of p_T and 30 GeV of missing transverse momentum (i.e. a leptonically-decaying W-boson enhanced selection)	1LMET30
Two to four leptons with at least 7 GeV of p_T each	2to4lep
At least two muons with at least 10 GeV of p_T (i.e. a leptonically-decaying Z-boson enhanced selection)	2muons
At least three jets with at least 20 GeV of p_T, at least one lepton passing tight identification requirements with at least 7 GeV of p_T, and 30 GeV of missing transverse momentum (i.e. a semi-leptonic top-quark enhanced selection)	3J1LMET30
At least two photons with at least 25 GeV of p_T each (i.e. a Higgs boson decaying to two photons enhanced selection)	GamGam
At least two jets with at least 20 GeV of p_T, at least two leptons passing tight identification requirements with at least 7 GeV of p_T, and 30 GeV of missing transverse momentum (i.e. a di-leptonic top-quark enhanced selection)	2J2LMET30
At least two jets with at least 20 GeV of p_T identified as containing at least one heavy flavor hadron using the 70% working point (i.e. a Higgs boson decaying to b-quarks enhanced selection)	2bjets
At least three leptons with at least 7 GeV of p_T each	3lep
Exactly three leptons with at least 7 GeV of p_T (i.e. a leptonically-decaying W+Z boson enhanced selection)	exactly3lep
At least four leptons with at least 7 GeV of p_T each	4lep
Exactly four leptons with at least 7 GeV of p_T (i.e. a leptonically-decaying ZZ boson or Higgs to four leptons enhanced selection)	exactly4lep

How to access ATLAS Open Data

- **`atlasopenmagic`**
  - a Python package to simplify working with ATLAS Open Data
  - providing utilities to manage metadata and URLs for streaming the data
    - Data and Metadata retrieval functions
    - Streaming URLs to CERN Open Data Portal
    - Notebooks and analysis utilities

Installation

```
pip install atlasopenmagic
```

- Alternatively,

```
git clone https://github.com/atlas-outreach-data-tools/atlasopenmagic.git
cd atlasopenmagic
pip install .
```

Quick start

```
import atlasopenmagic as atom
```

- available releases

```
atom.available_releases()
set_release('2024r-pp')
```

- which datasets to retrieve and use the 'Dataset ID'? *Pythia8EvtGen_...*:

```
all_metadata = atom.get_metadata('301204')
```

- a specific variable?

```
xsec = atom.get_metadata('301204', 'cross_section')
```

- URLs to stream the files for that MC dataset?

```
all_mc = atom.get_urls('301204')
```

- some data instead? check the available options:

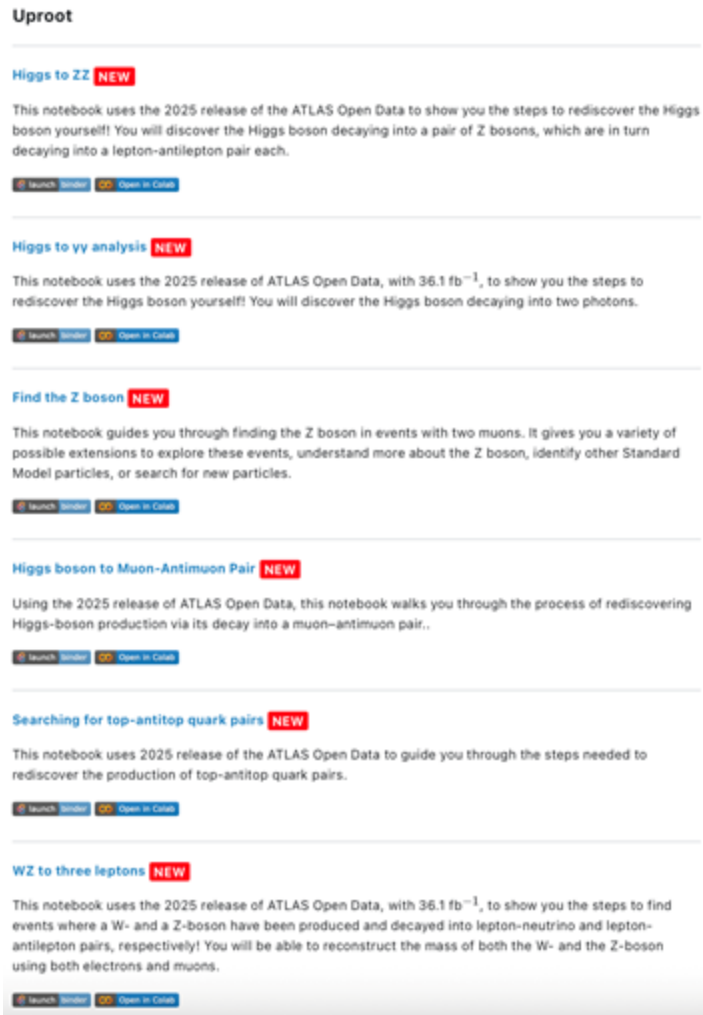
```
atom.available_data()
```

- URLs for the one that's to be used?

```
all_mc = atom.get_urls('data')
```

Open Data for Education

- **ROOT** is a C++ toolkit for data processing, widely used in High-Energy Physics
  - provides file format (ROOT files) and a full suite of analysis, statistical and visualisation tools
- **PyROOT** is a set of Python bindings to ROOT, giving direct Python access to ROOT's C++ features and libraries
- **Uproot** is an independent, pure Python library for reading and writing ROOT files
  - does not require C++ ROOT,
  - focusing on fast I/O and seamless compatibility with the wider Python data science ecosystem
- Jupyter Notebooks
  - Examples available to run on Binder or Colab



Jupyter Notebooks with Uproot - further details

- **Discover the Higgs boson in its decays to ZZ and γγ**

launch binder | Open in Colab

Higgs to ZZ **NEW** HZZAnalysis in Binder | HZZAnalysis.ipyn in Colab

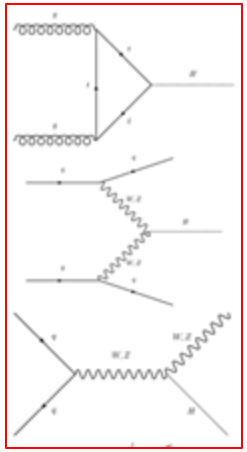
This notebook uses the 2025 release of the ATLAS Open Data to show you the steps to rediscover the Higgs boson yourself! The Higgs boson decays into a pair of Z bosons, which are in turn decaying into a lepton-antilepton pair each.

Higgs to γγ analysis **NEW** HyyAnalysis in Binder | HyyAnalysis.ipynb in Colab

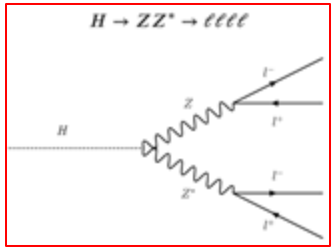
This notebook uses the 2025 release of ATLAS Open Data, with 36.1 fb−1, to show you the steps to rediscover the Higgs boson yourself! The Higgs boson decays into two photons.

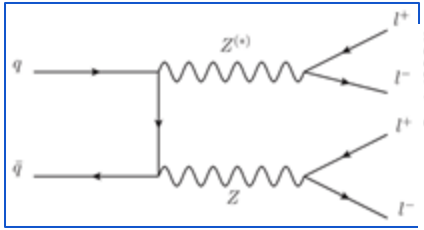
Discover Higgs: H→ZZ→ 4-leptons

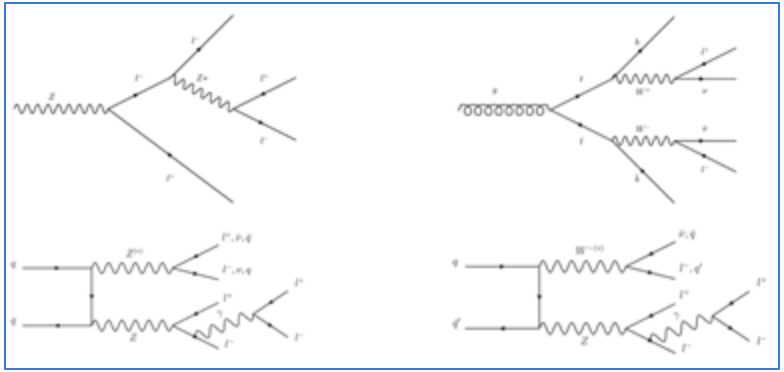
HZZAnalysis in Binder



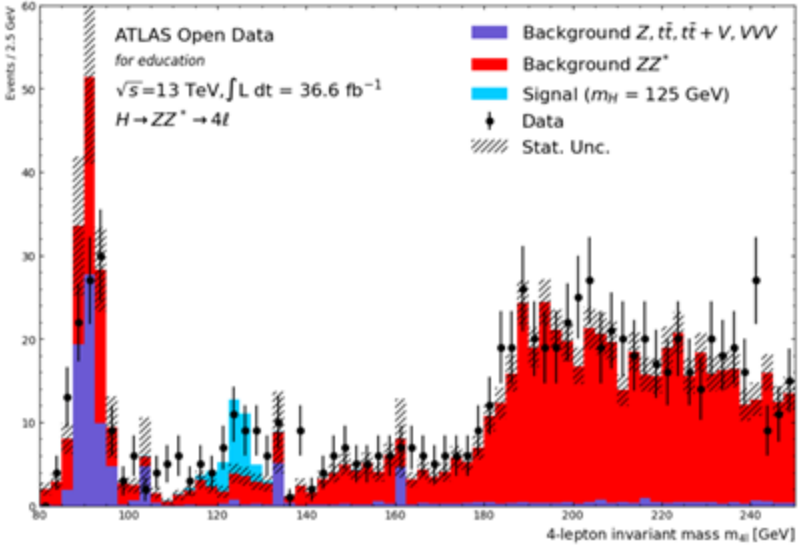
- ← Higgs produced in proton-proton collisions
  - Higgs can decay to 2 Z-bosons (signal) →
  - Other SM processes mimic Higgs signal →





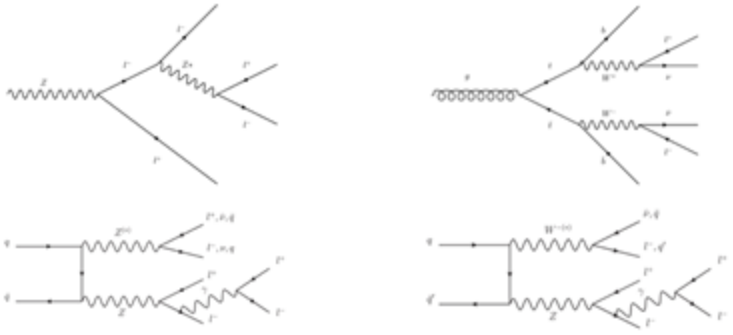


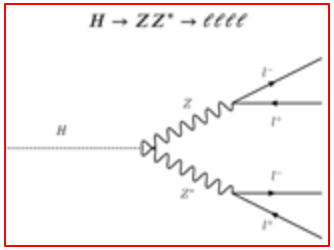
- Try H to ZZ* → 4-lepton analysis
  - available to run on Binder
  - Invariant mass of 4 leptons ⇒ Higgs signal?

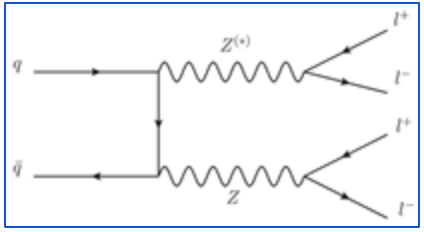


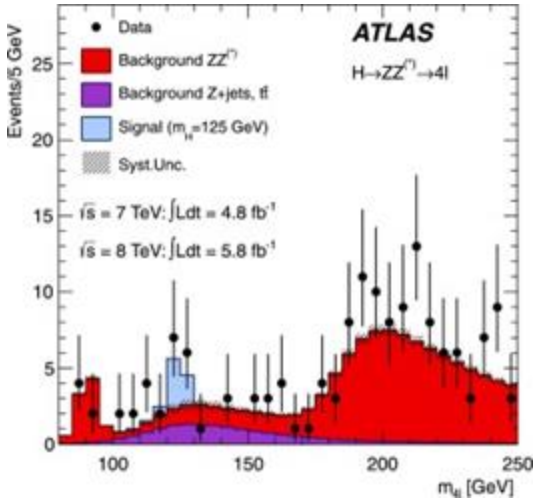
$$m_{4l} = \sqrt{E_{\text{tot}}^2 - \mathbf{p}_{\text{tot}} \cdot \mathbf{p}_{\text{tot}}}$$

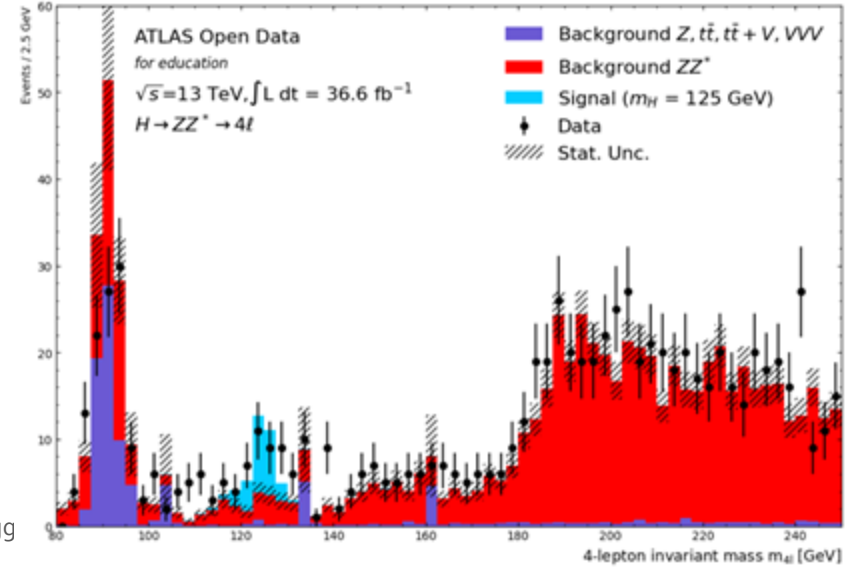
H→ZZ→ 4-leptons











- Clear signal observed in Open Data →
  - Compatible with what ATLAS has published →

$$S = \frac{N_{sig}}{\sqrt{N_{bg}}}$$

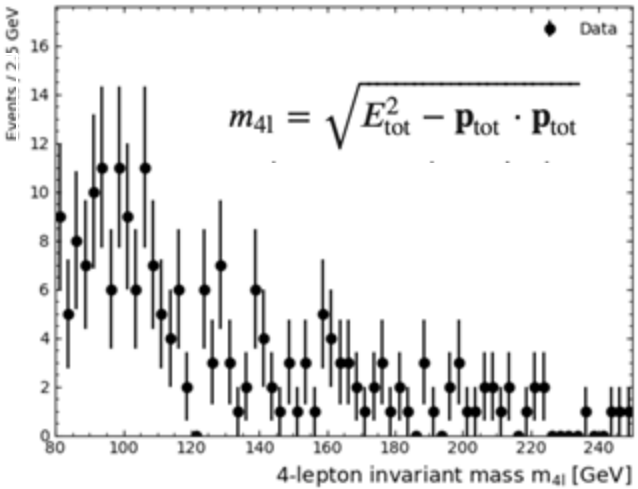
Results:
N_sig = 28.73
N_bg = 10.30
Significance S = 4.426

$H \rightarrow ZZ^* \rightarrow l^+l^-l^+l^-$ analysis steps in Jupyter Notebook

- Reading the data
  - *lumi = 36.2 fb^{-1}* → data size of the full release
  - *fraction = 1.0* → reduce this is if you want the code to run quicker
  - *skim = "exactly4lep"* → Select the skim to use for the analysis, exactly 4 leptons (electrons or muons) in the final state
  - Access `data15_periodD` - a small fraction of the data released (36.1 fb-1)
    - `data` refers to the event information collected from real experiment,
    - `Background` & `Signal` keys refer to Monte-Carlo (MC) simulations of experiment.
  - Both real data and MC data will then be analysed and compared together to discover the Higgs!

- Define important variables and cuts
  - Variables['lep_pt','lep_eta','lep_phi','lep_e','lep_charge','lep_type','trigE','trigM','lep_isTrigMatched','lep_isLooseID','lep_isMediumID','lep_isLooseIso','lep_type']
  - Identify 2 lepton-pairs of same flavour & opposite charge: $e^+e^-e^+e^-$, $\mu^+\mu^-e^+e^-$, $\mu^+\mu^-\mu^+\mu^-$
  - Cut on p_t, m_ll, m_4l … require matched triggers,
  - Plot m_4l

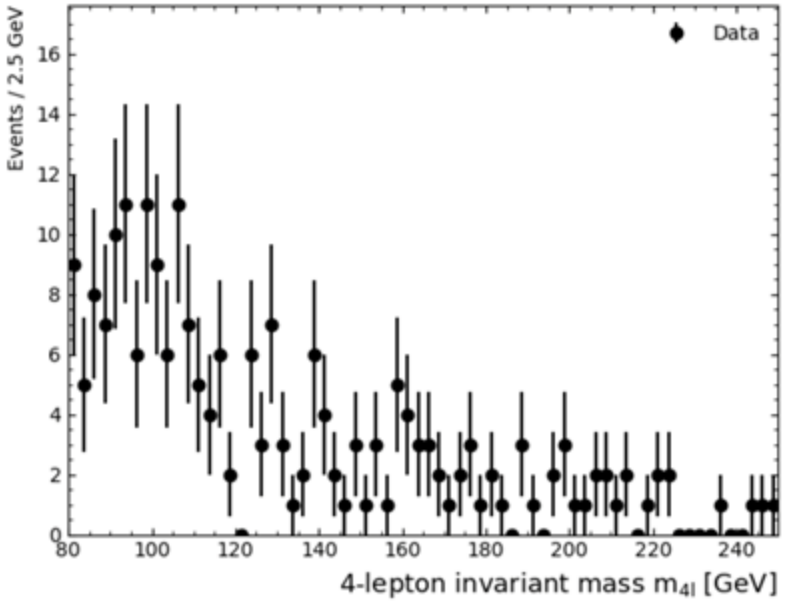
⇒



$H \rightarrow ZZ^* \rightarrow l^+l^-l^+l^-$ analysis

- Look at some period (D) of real data
  - Plot invariant mass distribution distribution $\Rightarrow$

$$m_{4l} = \sqrt{E_{\text{tot}}^2 - \mathbf{p}_{\text{tot}} \cdot \mathbf{p}_{\text{tot}}}$$



- Reading Monte-Carlo simulation data
  - SM randomised simulations to compare to ATLAS data
- Need to apply event *Weight* taking into account real detector inefficiencies, resolutions, …
  - Attribute appropriate *weight* to each data point, thus affecting how it contributes to histogram count for its bin

- Weights which are important to our analysis:
  - scaleFactor_PILEUP - scale factor for pileup reweighting
  - scaleFactor_ELE - scale factor for electron efficiency
  - scaleFactor_MUON- scale factor for muon efficiency
  - scaleFactor_LepTRIGGER - scale factor for lepton triggers

$H \rightarrow ZZ^* \rightarrow l^+l^-l^+l^-$ analysis

- Add a cross-section (σκ) weight xsec_weight associated with each MC file.
  - normalise the entire Monte-Carlo distribution based on # events in data
  - Integrated luminosity: $\int Ldt$
  - Sum of all weights: $\sum_i w_i$
  - Generator filter efficiency: η
  - Correction factor to account for higher-order effects in theoretical calculations of cross section: κ

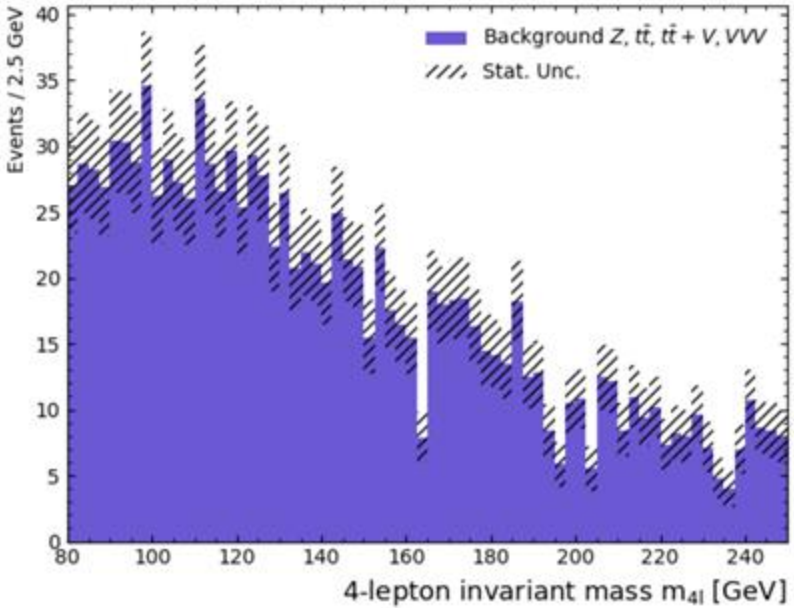
$$w_\sigma = \frac{\int Ldt\, \sigma\, \eta\, k\, w}{\sum_i w_i}$$

- For data_D period, integrated luminosity is 0.105 fb-1

- **Plot invariant mass distribution for weighted MC background**

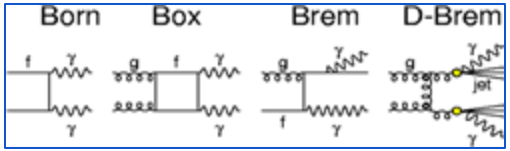
$$m_{4l} = \sqrt{E_{tot}^2 - \mathbf{p}_{tot} \cdot \mathbf{p}_{tot}}$$

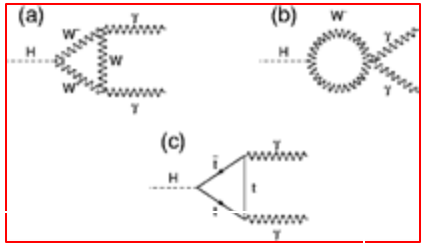
$\Rightarrow$

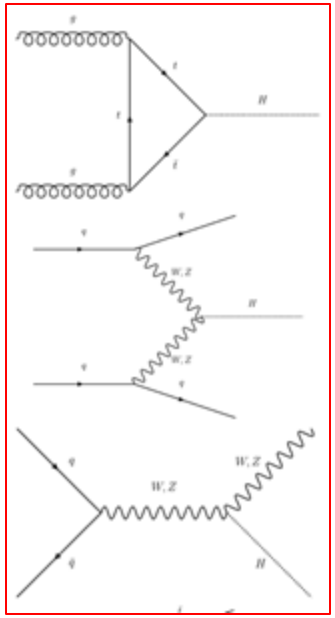


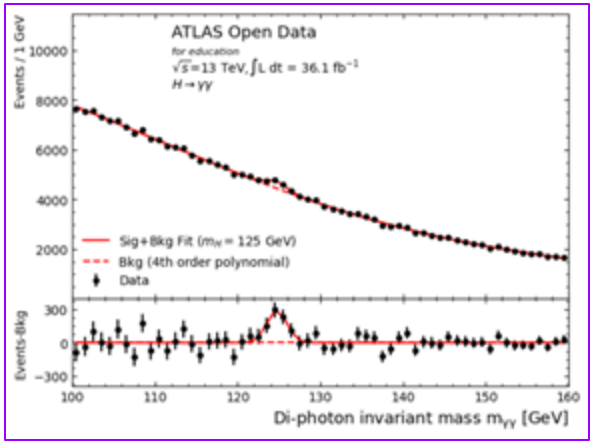
Discover the Higgs boson at the LHC
H→ γγ

- Higgs produced in proton-proton collisions →
  - Higgs can decay to 2 photons (signal) →
  - Other SM processes mimic Higgs signal →

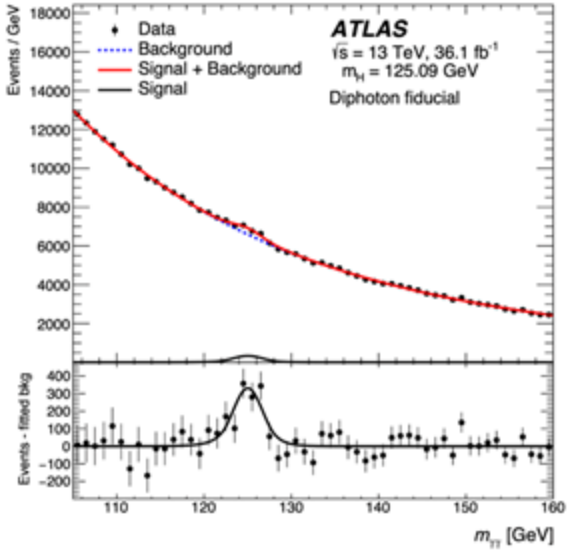








- Try H to γγ analysis
  - available to run on Binder
  - Invariant mass of 2 the photons ⇒ Higgs signal?



- Clear signal observed in Open Data →
  - Compatible with what ATLAS has published →
  - Measurements of Higgs boson properties in the diphoton decay channel (Section 5)

HyyAnalysis in Binder