Reconstruction and identification of close-by e+e- in the search for Higgs boson decays into a low-mass dilepton system and a photon final state.

Tesi di Laurea Triennale

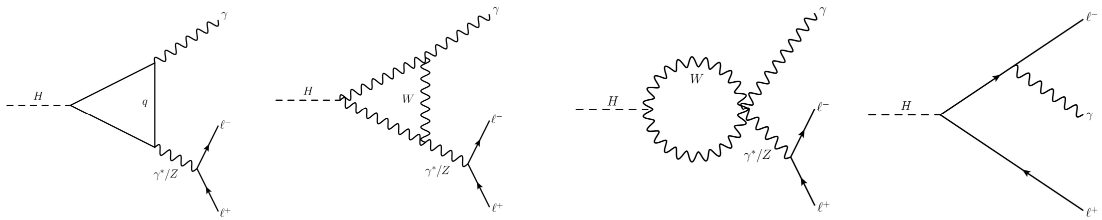
Università degli Studi di Milano

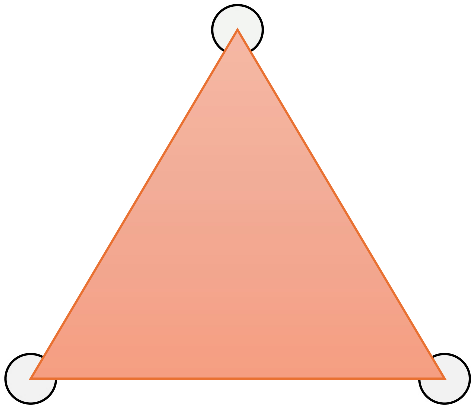
A.A. 2024-2025

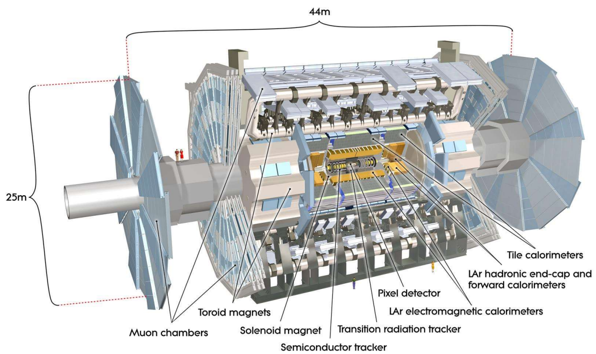
Giuseppe Morinelli (Matricola 966566)

Relatore: Prof. Leonardo Carminati
Correlatori: Dott.ssa Laura Nasella, Dott. Ruggero Turra

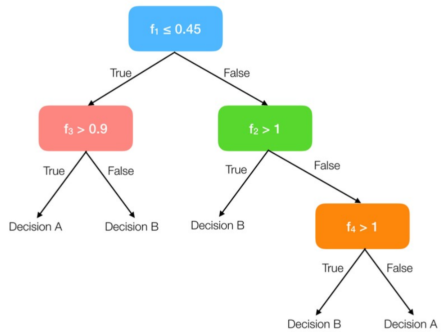
PHYSICS PHENOMENON







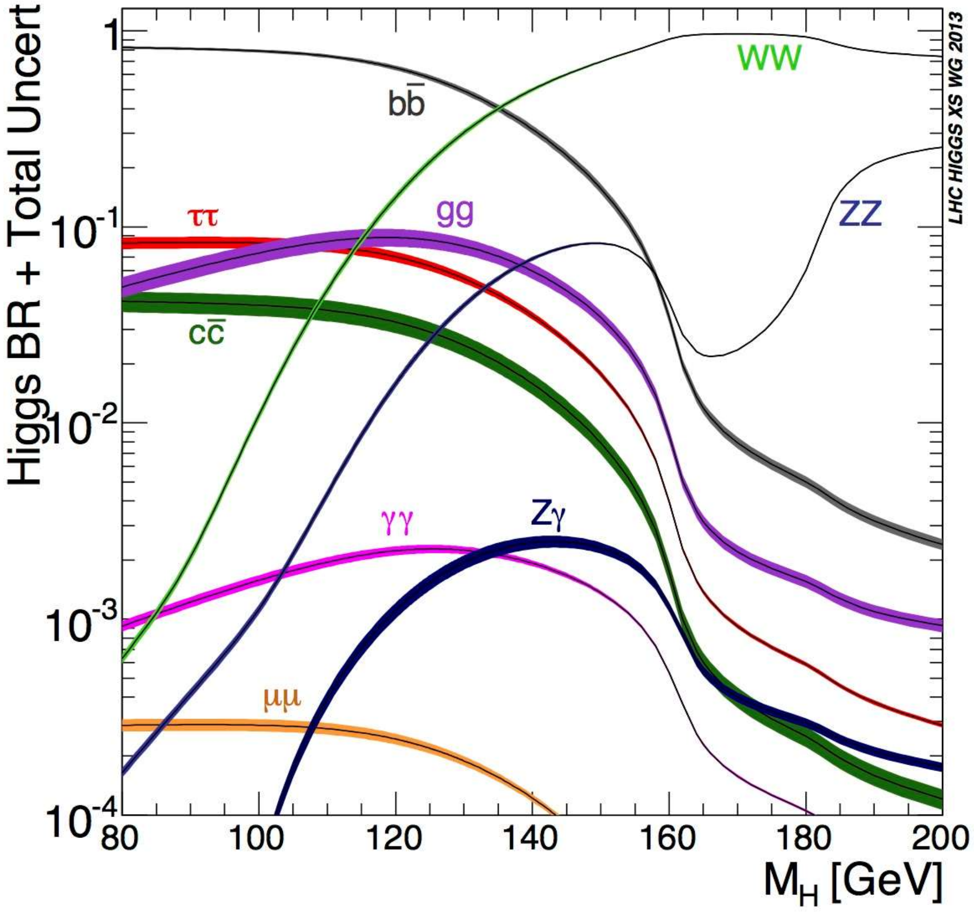
ATLAS EXPERIMENT



BDT ALGORITHM

Higgs boson physics

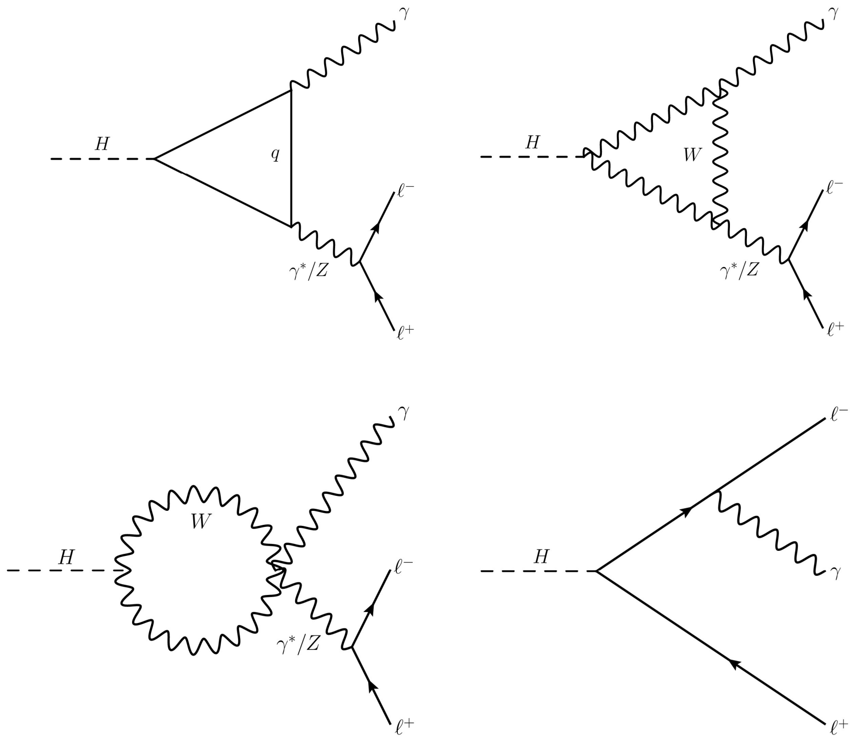
- ATLAS and CMS collaborations proved the existence of the Higgs boson (H) in 2012
- More than ten years later, some aspects of the Higgs sector of the Standard Model (SM) need to be cleared out yet
- H bosons decay almost instantaneously (~10^{-22} s) in a large variety of channels, most of which have been discovered
- Searching for rare decays of H helps testing the SM and could bring evidence of possible extensions of the SM theory



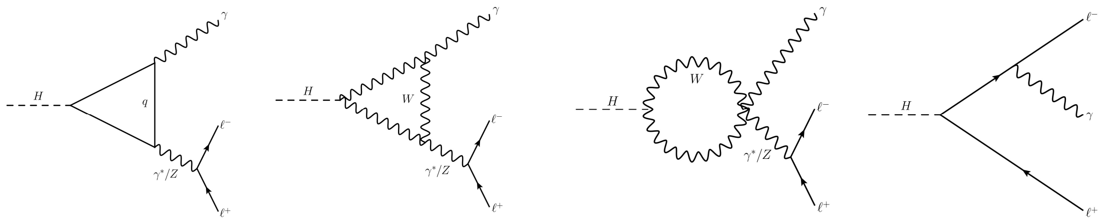
$H \rightarrow \gamma\gamma^* \rightarrow \ell\ell\gamma$ *BR estimate at* $M_H = 125\ GeV$: **~$\mathbf{10^{-5}}$**

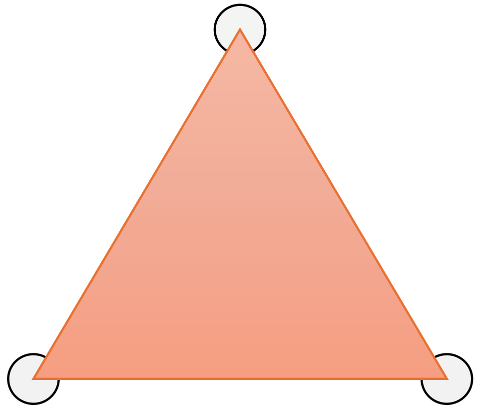
The rare decay $H \to \gamma\gamma^* \to \ell\ell\gamma$

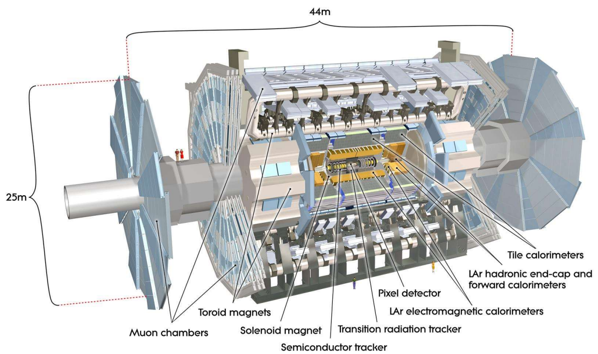
- The rare channel of interest consists in an H decay in photon pair, where one photon is off-shell and decays in a lepton pair
- The lepton pair could be a muon-antimuon or electron-positron
- This decay has been measured with a significance of 3.2σ in the previous analysis during Run 2 (2015-2018)

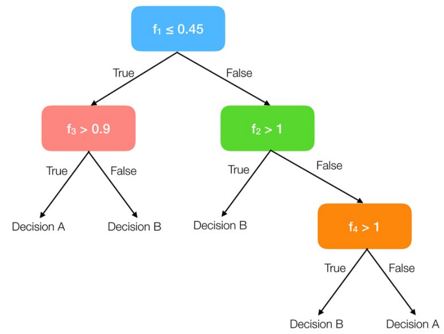


PHYSICS PHENOMENON









ATLAS EXPERIMENT

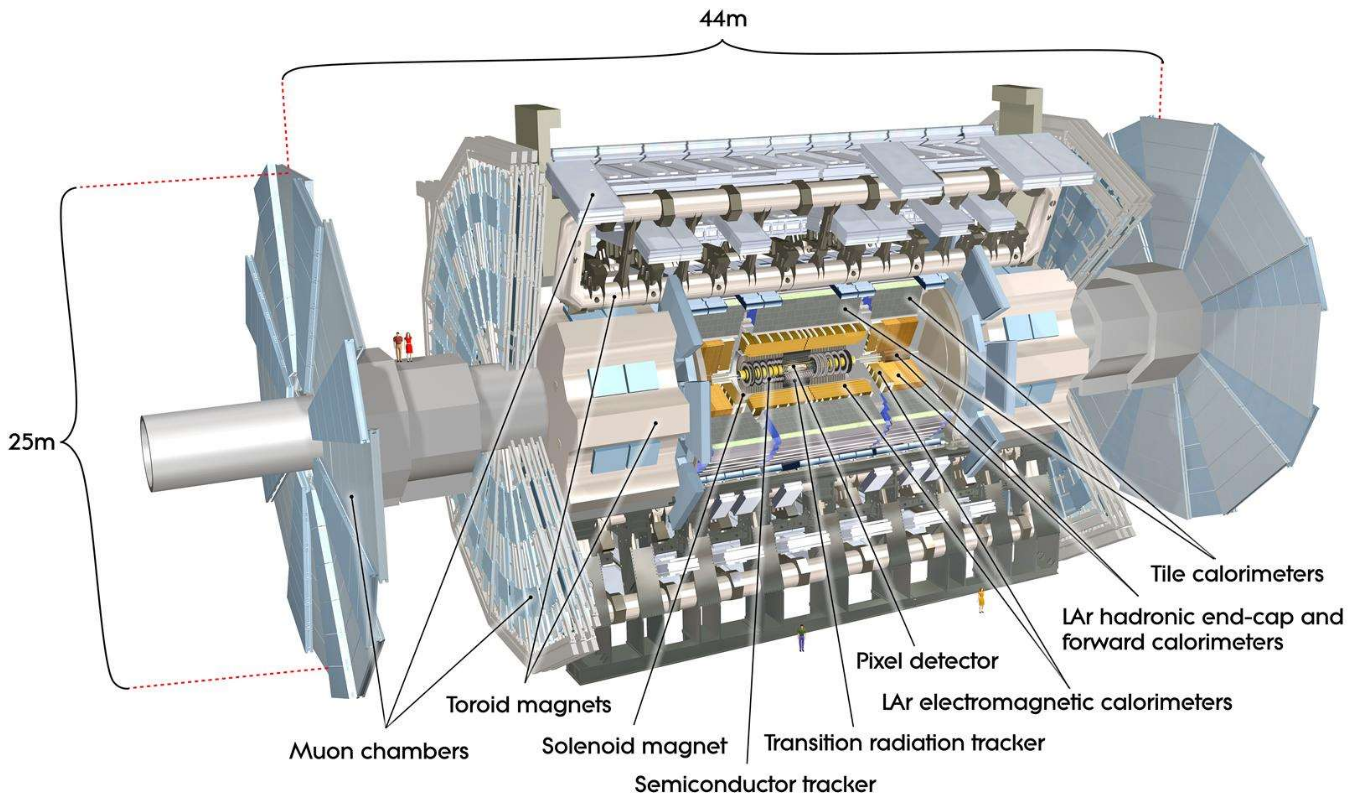
BDT ALGORITHM

The ATLAS experiment

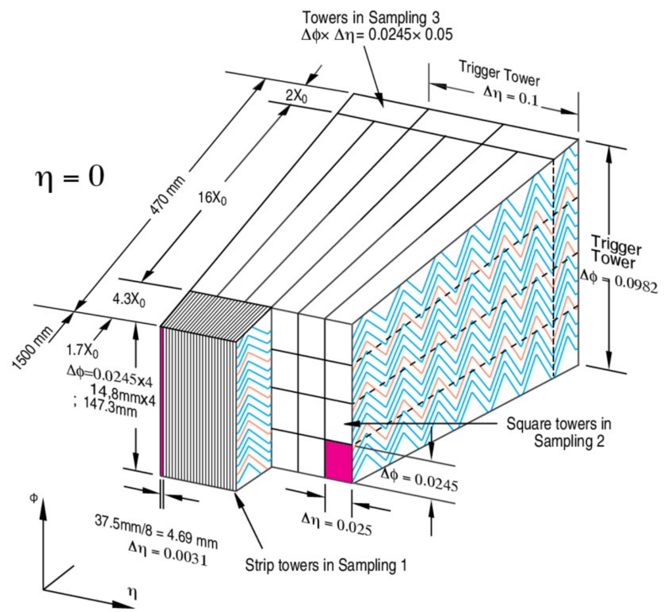
Interaction point:
13.6 TeV pp collision

It is composed of many layers of detectors around an interaction point, in cylindrical shape:

1. Inner detector
2. Electromagnetic and hadronic calorimeters (EMC and HC)
3. Magnet system (solenoid and barrel/endcaps toroids)
4. Muon spectrometer



Reconstruction of electrons and photons

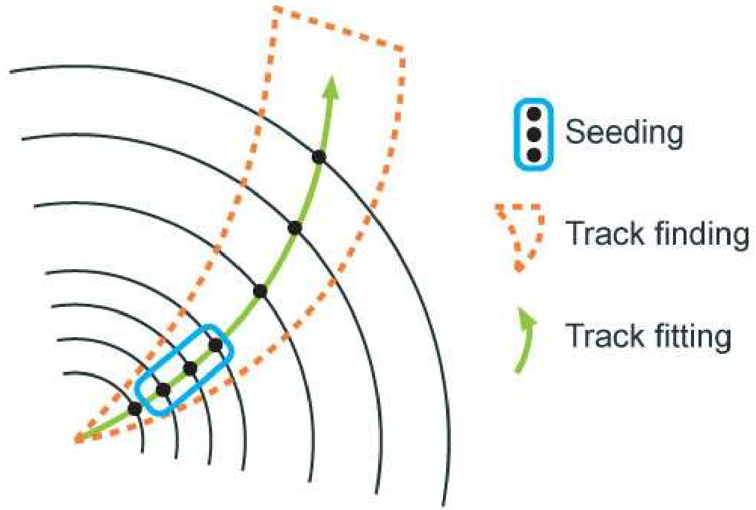


Photons and electrons are reconstructed from

- The Inner Detector: measure hits by particles coming from the collision point and reconstruct tracks
- The EMC: designed to absorb photons and electrons produced by collision events (destructive measure). Energy is deposited in

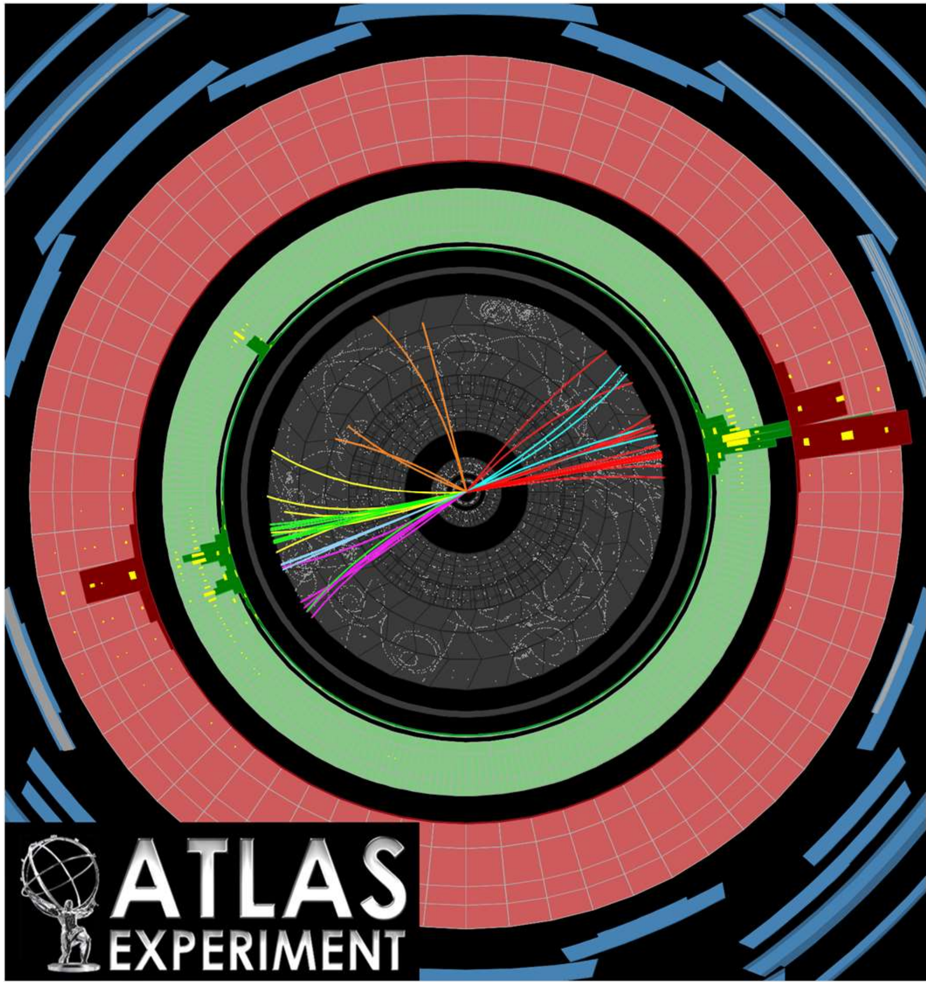
Reconstruction algorithms are used to distinguish between electrons and photons candidates:

- Electron: cluster matched to a track
- Converted photon: cluster matched to a secondary vertex
- Unconverted photon: no tracks pointing to the cluster



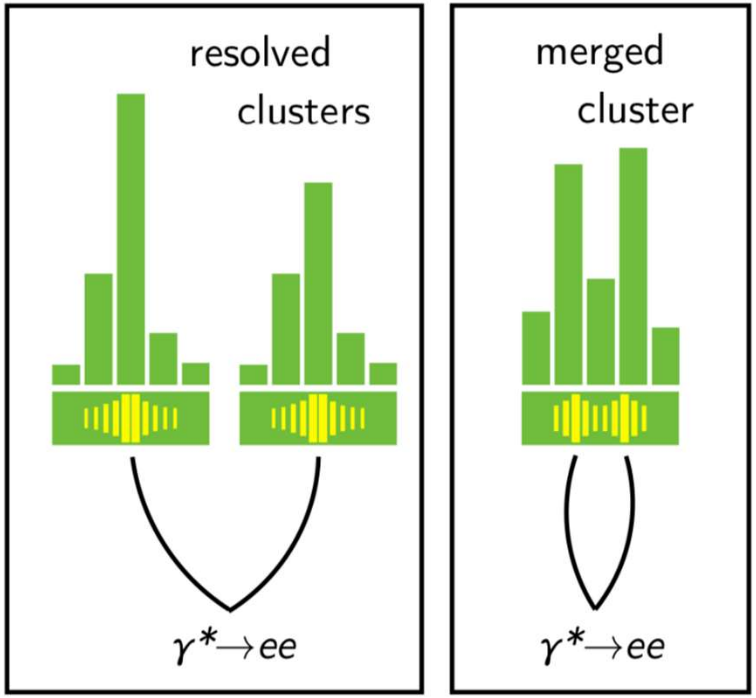
Electron identification

- Exploit information from track and energy deposits variables reconstructed by Inner Detector and calorimeters measures
- Three standard ATLAS identification menus are defined:
  Loose, Medium, Tight
- Distinguish electron candidates from hadronic jets and non-prompt electrons
- ID tuned for isolated electrons

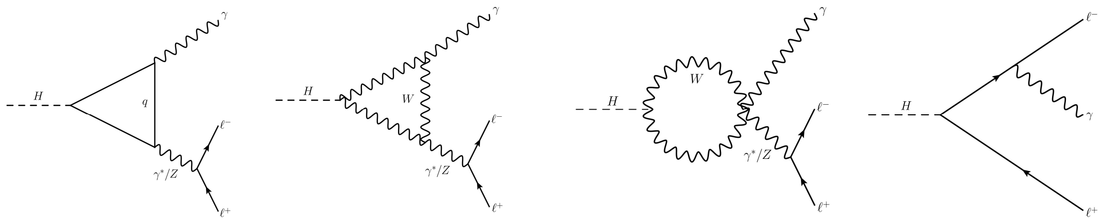


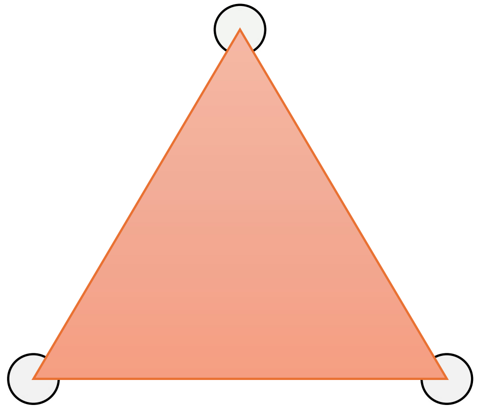
Reconstruction and ID of merged electrons

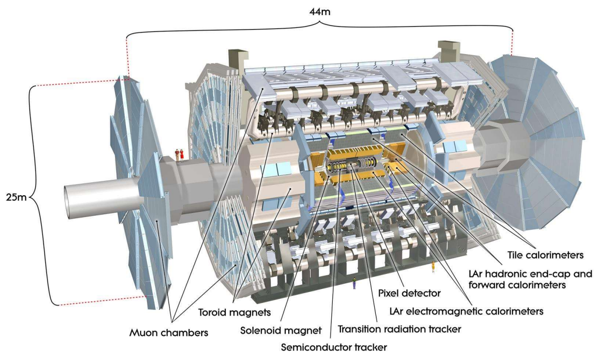
- In case of merged electrons, two clusters of deposited energy overlap
- Merged electrons are reconstructed as particles with a single energy cluster (in the EMC) and two separated tracks (in the inner detector)
- Around 60% of signal electrons are merged
- The standard ATLAS electron ID is not optimal for merged electrons
- A new ID could increase performance and help collecting more $H \to \gamma\gamma^* \to \ell\ell\gamma$ events



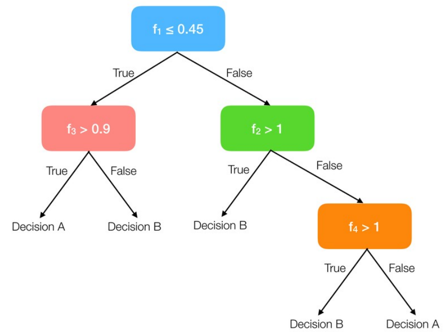
PHYSICS PHENOMENON







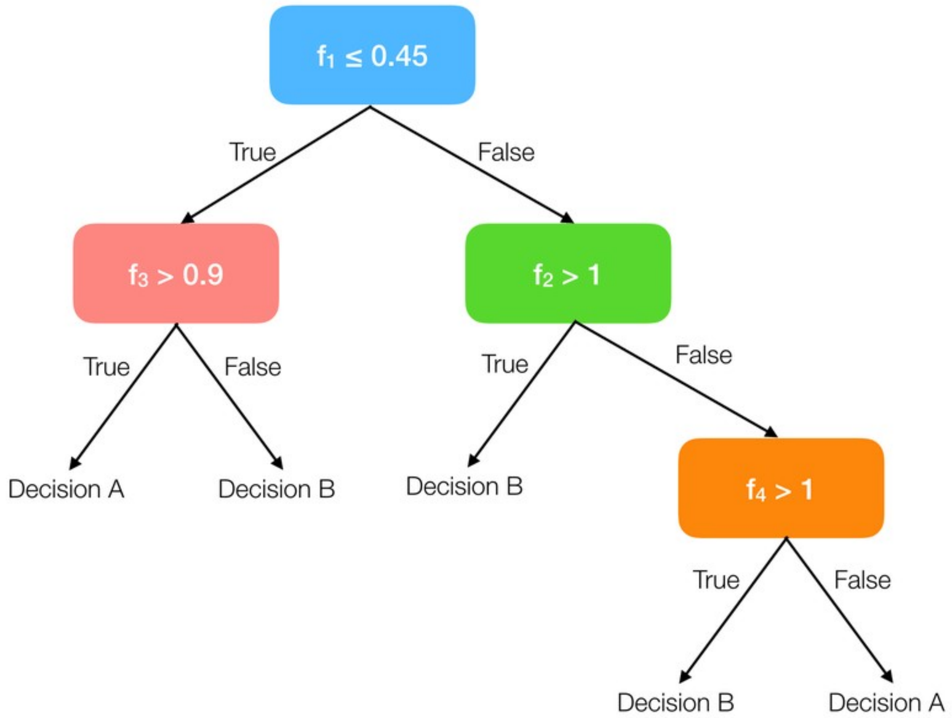
ATLAS EXPERIMENT



BDT ALGORITHM

Boosted decision tree (BDT) model

- Split samples in 50% for training and 50% for testing the model
- Set hyperparameters for the model (num_leaves, early_stopping, etc…)
- Train a BDT model using classifier class from LightGBM package, developed by Peking University and Microsoft corporation



Training data: simulated signal candidates

- Use simulated data because every aspect of generated events is known! (*truth level* data)

Simulate Higgs production and $H \rightarrow \gamma\gamma^*$ decays in pp collisions → Simulate the interaction with ATLAS detectors → Run reconstruction algorithm and identification

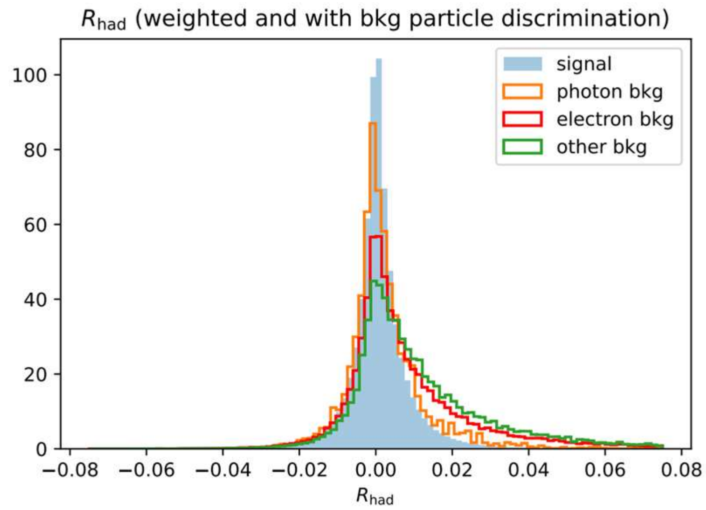
Same procedure applies for simulated background samples: use objects that could be falsely reconstructed as merged electrons

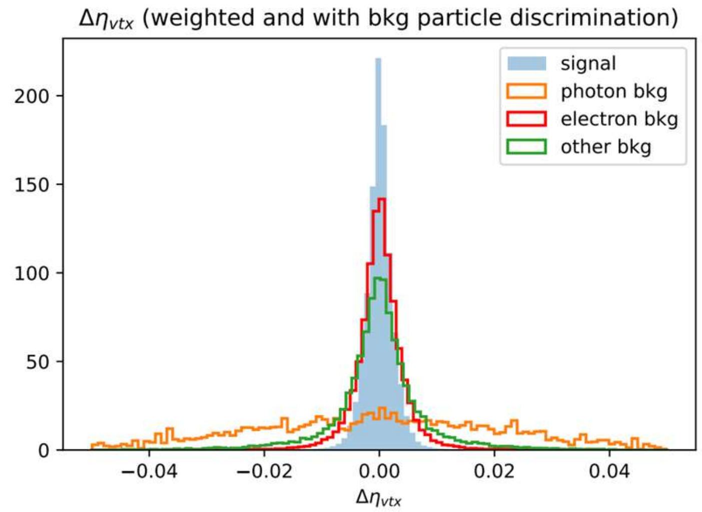
- Jet Filter (JF) sample of hadronic di-jets
- $H \rightarrow \gamma\gamma$ photon decay samples

Feature distributions

Selected samples on kinematics and Loose ID

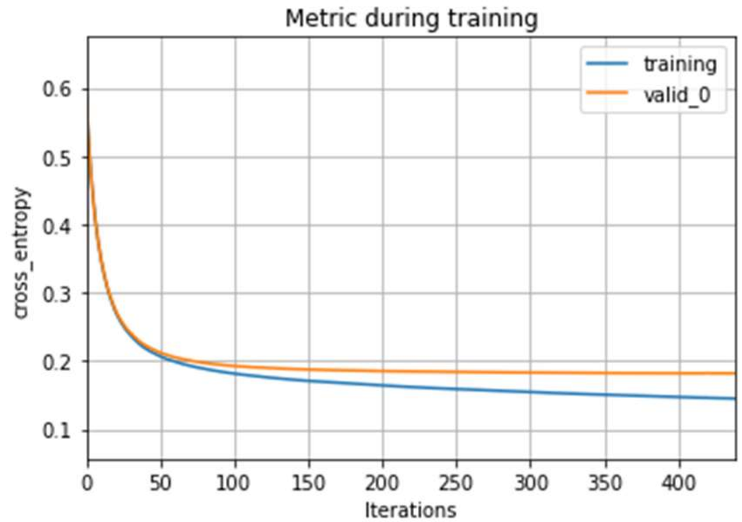
- 17 discriminating features used: shower shape variables from the EMC, track related variables from the inner detector
- The more feature distributions are different, the more the BDT model will be capable of discriminating signal and bkg candidates

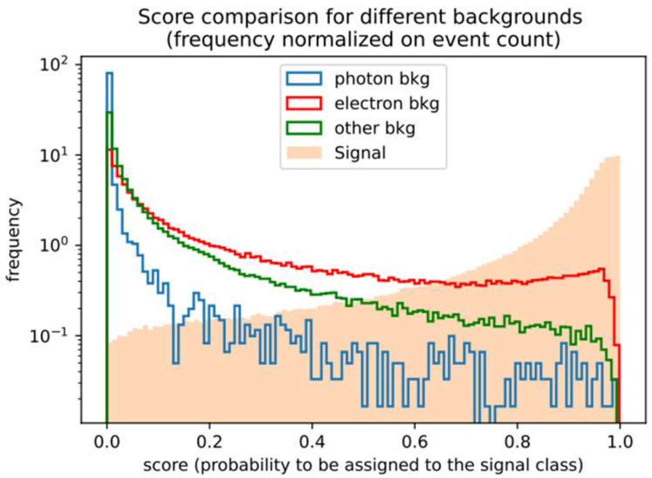


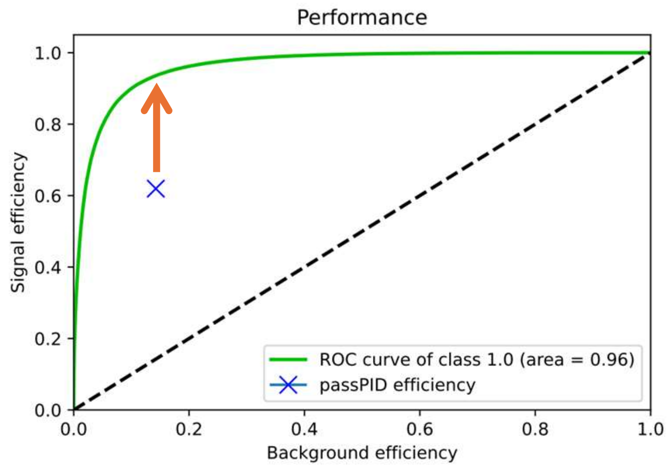


Training results on test sample

- Each candidate has been assigned a prediction score from 0 (bkg) to 1 (signal)
- Set a threshold value for the score to define a BDT identification
- Compare the signal efficiency of the BDT ID with the identification developed in the previous Run 2 analysis (passPID)

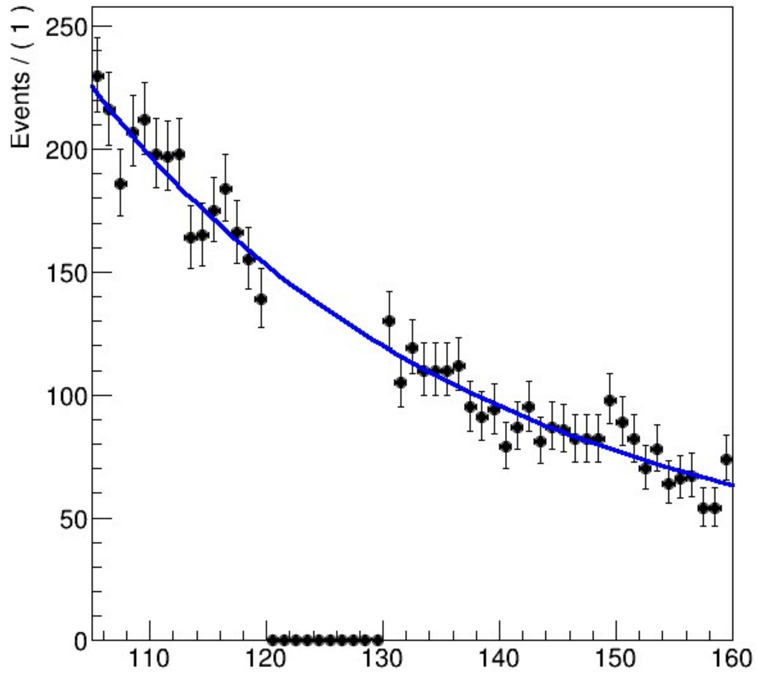






Significance on sidebands

- Bkg samples used for training have an arbitrary composition of particles, so they do not reflect real data
- Approximated expected significance Z_{90} is estimated from events in data sidebands and signal Monte Carlo sample
- The data sample is blinded: events in a range of invariant mass near 125 GeV are removed, calculate the number of events B_{90} through a fit
- The number of signal events S_{90} in the same range is computed from MC



$m_{\ell\ell\gamma}$ $[GeV]$

$$Z_{90} = \frac{S_{90}}{\sqrt{B_{90}}}$$

Significance Z_{90} results

- Data is classified in a total of 9 categories, distinguished by flavor, by kinematics and by the production mode of the Higgs
- Fits and calculation of significance are performed to each category separately
- The expected significance Z_{90} is calculated repeating the full analysis with the Run2 PID and the BDT ID
- The threshold value is set to have the same B_{90} in category 3 for the two IDs

Category name	Category code
$\mu\mu$ low-p_{T_t}	1
ee resolved low-p_{T_t}	2
ee merged low-p_{T_t}	3
$\mu\mu$ VBF	4
ee resolved VBF	5
ee merged VBF	6
$\mu\mu$ high-p_{T_t}	7
ee resolved high-p_{T_t}	8
ee merged high-p_{T_t}	9

passPID identification

Cat	B_{90}	S_{90}	Z_{90}
1	2013	59.37	1.32
2	1061	22.54	0.69
3	1390	33.52	0.90
4	6.93	1.64	0.62
5	3.04	0.55	0.32
6	3.61	1.07	0.57
7	45.99	5.68	0.84
8	35.48	4.10	0.69
9	35.48	4.10	0.69

BDT model identification

Cat	B_{90}	S_{90}	Z_{90}
1	2013	59.37	1.32
2	1061	22.54	0.69
3	1374	43.36	1.17
4	6.93	1.64	0.62
5	3.04	0.55	0.32
6	4.86	1.37	0.62
7	45.99	5.68	0.84
8	35.48	4.10	0.69
9	33.23	5.19	0.90

Conclusions

- A new ID algorithm for merged electrons has been implemented using a BDT
- A working point ensuring the same background efficiency of the one from the passPID merged electron ID has been tuned on the test sample showing an improvement on signal efficiency
- The performance of the BDT algorithm has been tested on real data in $m_{\ell\ell\gamma}$ sidebands reporting a potential improvement of the 26% in significance (from 1.27 σ to 1.60 σ)
- When accounting also for muon and resolved electron candidates, that keep their identification method unchanged, the overall increase in expected significance is 8%, 2.28 σ to 2.48 σ
- The new BDT ID has shown sizeable improvements in identification of merged electrons, and will be used in new analyses

Thank you for listening!

Grazie per l’ascolto!

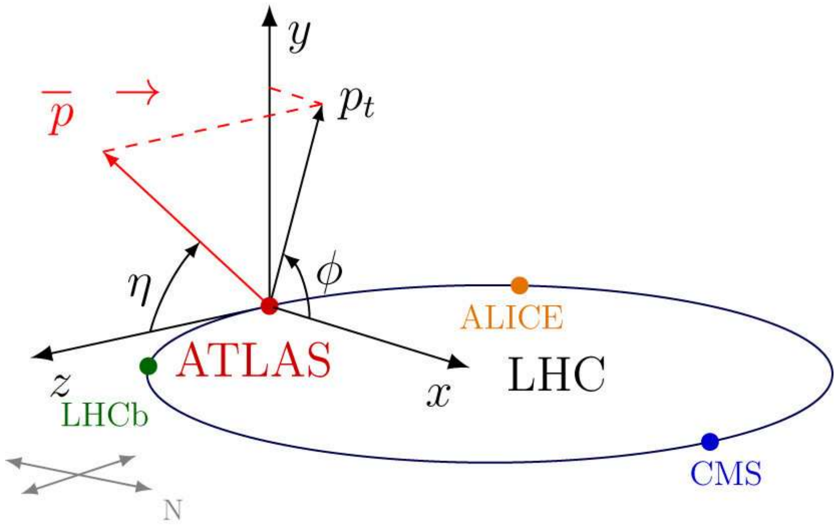
BACK-UP

Extensions of the SM

- Supersymmetry (SUSY)
- Extensions of the Higgs sector (for example higgs triplets)

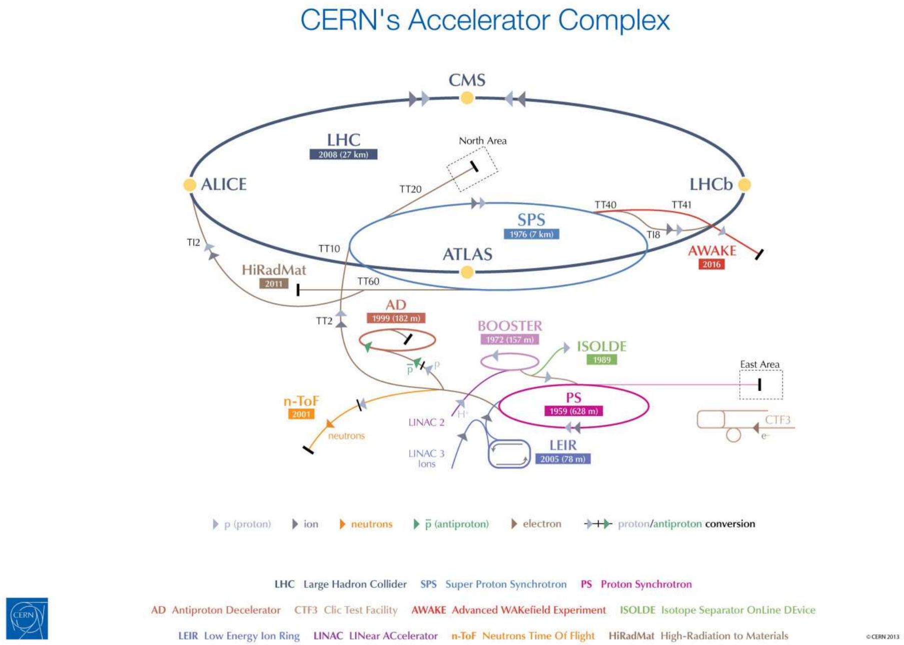
Coordinate system

- p_T: transverse momentum, defined as the component of momentum transverse to the interaction plane
- $\eta = -\ln(\tan(\theta/2))$: pseudorapidity, calculated from the polar angle



The Large Hadron Collider (LHC)

- LHC is the largest particle accelerator in the CERN experimental complex at Geneve, Switzerland (27 km long ring)
- It is designed to have proton-proton (pp) collisions at a center-of-mass energy of 13.6 TeV at Run 3 (2022-2026)



Simulated background samples

- Fake candidates: objects reconstructed similarly to merged electrons by the detectors
- Jet Filter (JF) sample: it consists of di-jets of hadrons from Higgs decay
- $H \rightarrow \gamma\gamma$ photon decay samples: one sample with Higgs produced via gluon-gluon fusion (ggH), one via vector boson fusion (VBFH)

Selections on training data

Signal sample:

- Kinematic selections, $p_T > 10\ GeV$ and $|\eta| < 2.37$
- Select candidates matching to γ^* from Higgs decay
- Loose identification

JF bkg sample:

- Kinematic selections, $p_T > 10\ GeV$ and $|\eta| < 2.37$
- Loose identification
- Select electrons coming from γ: this is done to exclude non-resonant $\gamma\gamma^*$ irreducibile bkg

Photon bkg samples:

- Kinematic selections, $p_T > 10\ GeV$ and $|\eta| < 2.37$
- Loose identification
- Select γ coming from Higgs: this is done to exclude dalitz decays from sample

Weighting samples

Kinematic reweight to address p_T and η distributions