Prospective Machine Learning Techniques for Electromagnetic objects reconstruction in the ATLAS detector, Nihal Brahimi

EM shower in ATLAS from charged particles => emits photons that are detected by the detector

Simulation of the detector using Geant4 => differences between simulations and actual detector

Main source of uncertainties: electron identification => DITTO project at LAPP

Likelihood (LH) approach used since 2012 - 13 input variables - assumes no correlation in the input variables - threshold on LH discriminant to define signal

ML-based identification - DNN with the same input variables (but takes advantage of correlations) - CNN input variables + calorimeter images (only tested on MC) - GNN same as CNN but as point cloud

DNN

- 6 outputs: probability to have electron + probability for other backgrounds
- DNN trained on MC, but MC need to be corrected using fudging
- took 10 years to ATLAS to reach something that can be used

CNN

- CNN output concatenated with other input variables then fed to FC layers
- much better performances on MC data

GNN

- advantage: permutation invariant
- 15% improvement compared to CNN

Photons identification

- BDT better than cut-based approach

Data/MC discrepancy

- mismodeling
- detector geometry
- electric fields differences

Corrections: - linear transformation = fudging - cell-based energy correction Other possibilities: - optimal transport (OT) = morph MC distributions to match the data - GANs - propagate modeling uncertainties to the output of the network - train on data

Need: - feedback to test *one* method only (not enough persons to do more)

Questions

- sensitivity to initial conditions?
    - don't know for CNN/GNN