Configurable data analysis workflow for HEP experiment

Scientific Achievement

Novel workflow designed to efficiently use HPC resources enables physicists to generate, simulate and analyze billions of Monte Carlo events in a time scale of hours.

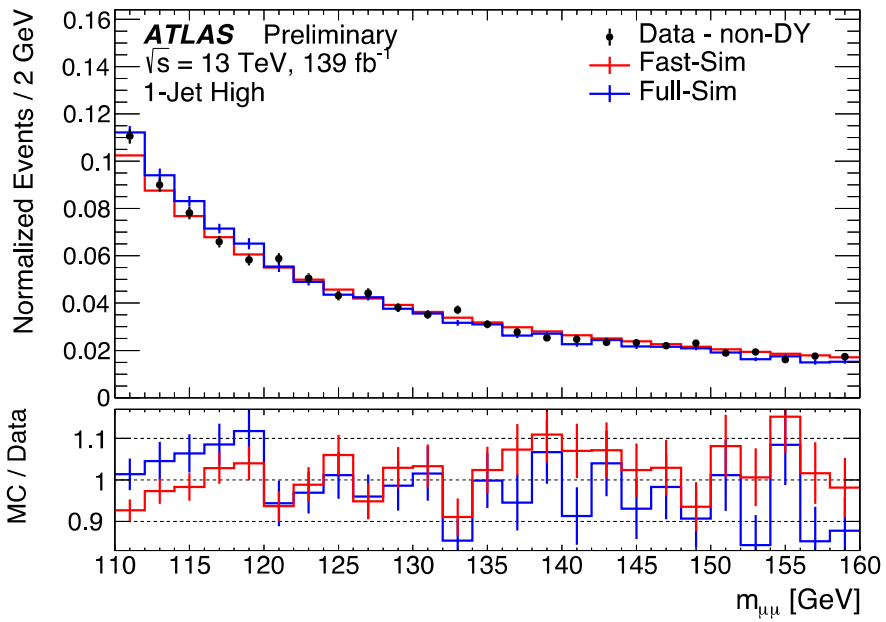
Significance and Impact

Boosted the discovery sensitivity for the search of the Higgs boson decaying to a pair of muons using the ATLAS detector, applicable to other data analysis at LHC too

Research Details

- The main challenge is the precise estimation of the background modeling uncertainty. Controlling the uncertainty requires massive Monte Carlo simulated events
- Traditional text-based storage of event files cannot be efficiently parallelized. We first introduced the HDF5-based storage
- Simulating the ATLAS detector takes very long time, making it impossible to generate billions of events; we introduced a fast detector simulation package enabled the task
- Data analysis is performed by using MPI-enabled parallelization

ATLAS-CONF-2019-028



Normalized dimuon invariant mass spectra, comparing data to ATLAS full simulation and fast simulation. Previously the fast simulation was generated by PowHeg.

Our fast simulation, using Sherpa generator, renders better agreements and will be used in the peer-reviewed publication. The better agreement improves the discovery sensitivity by 3-5%.