



LHCb Jet Separation Challenge

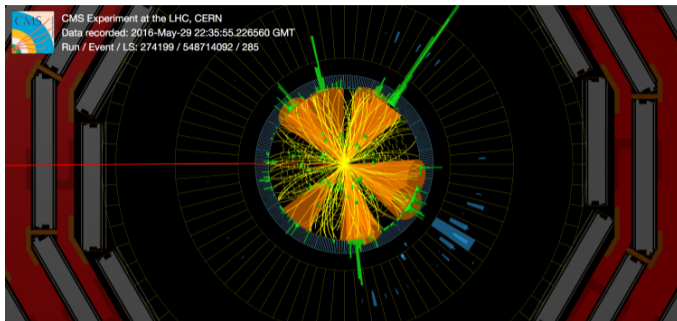
Stephen Farry

STFC School on Data Intensive Science

October 16, 2020

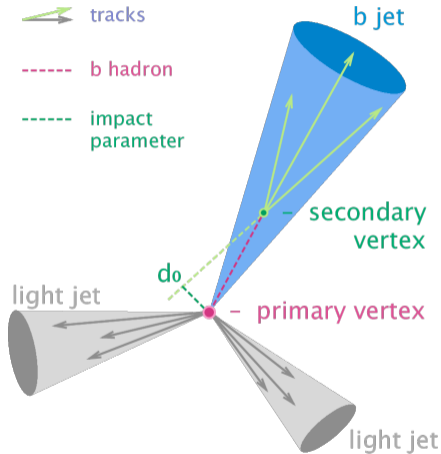


Anatomy of a proton proton collision



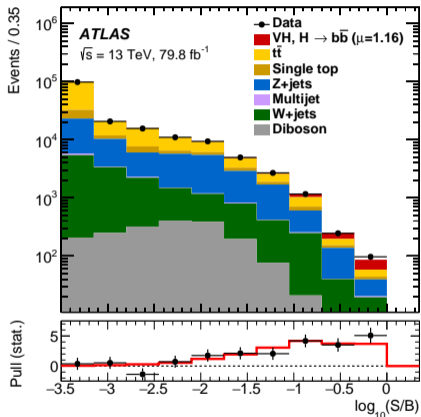
- The collision of protons at high energies at the Large Hadron Collider (LHC) is a messy business...
- For any given measurement, the signal events must be extracted from a huge volume of background events
- A very common signature is a narrow stream of high energy particles, known as a jet
- The study and classification of jets is a core part of the physics program of all experiments

Quarks, Gluon and Jets



- Quarks and Gluons are fundamental particles that cannot exist in isolation due to the principle of colour confinement
- They subsequently produce “jets” of particles which can be identified and classified
 - The jets arising from the heavy beauty and charm quarks, can be classified as *b* and *c*-jets
 - The other jets, which can arise from up, down and strange quarks or gluons, are classified as light jets
- The jets produced by these jets have subtle differences in their substructure that can be used to distinguish them at a particle detector
 - heavy quarks are more massive - larger multiplicity, larger momentum
 - heavy quarks travel a short distance before decaying - can find a “secondary vertex” within the jet

Why is jet tagging important?



Example: The Higgs Boson

- The most common way a Higgs boson decays is to pairs of b -jets
- However, it was only discovered through this signature in 2018, 6 years after the initial discovery.
- Why? Many processes at the Large Hadron Collider produce jets, both light and heavy
 - The signal is swamped by background events
- Need machine learning techniques to “see” our signal

Your dataset

- You have simulated datasets of a common process at the LHC, the production of high energy pairs of jets as reconstructed by the LHCb detector, one of the four main detectors on the LHC ring
- The jets have already been pre-selected to contain a secondary vertex, but the abundance of light jets mean that a significant number are still present in any data sample
- Available is a number of observables related to the jet, its constituents, and the properties of the secondary vertex
- Can you separate the b jets from both c jets and light jets, in order to make a measurement at the LHC?
- <https://www.kaggle.com/c/lhcb-jet-data-separation/>

backup