Modernizing FastJet Interfaces With pybind11 and interfacing with Awkward Arrays

Aryan Roy

CCE, B.Tech

Manipal Institute of Technology

Mentor: Jim Pivarski

Collaborator: Eduardo Rodrigues

Duration: 4th April 2021 – 30th Aug 2021

Project Description

The experiments conducted in high-energy colliders are known for generating a large amount of data that needs to be processed as fast and efficiently as possible. To enable this, many algorithms and tools are needed, which can sift through the data and do the work on a large scale. One of the sources of data that require such a solution is jets. These are collimated sprays of hadrons that result from the fragmentation of a high-energy quark. These Jets are visible to anyone who looks at the data from the collider. These jets are then later used to reconstruct the parton that produced them. Many algorithms can be used to identify the jets in a given event. All these algorithms have been implemented in FastJET for C++. To bring this functionality to python, a tool named pyjet in SCIKIT-HEP is used. It provides the bindings required in Cython to carry out the required calculations. FastJet has limitations primarily related to its inability to perform columnar analysis. pyjet, on the other hand, also has some limitations, such as its dependence on numpy rather than Awkward Array.

In this work, we propose the overhaul of the pyjet tool to overcome the present shortcomings. The fellowship will be done in two phases; in the first phase, the primary focus will be to convert the Cython bindings into pybind11 because the Cython bindings are unnecessarily complicated. This would involve investigating the whole structure of the tool and designing the new bindings with support from Awkward Array and Uproot in mind, primarily to provide numpy like operations for Jagged Arrays. This will help in future use as further modifications will be much easier. Integration with Awkward arrays would also help with managing jagged arrays in research work. This would mark the end of the first phase. In the second phase, the bindings will also be expanded to include certain functions that have not been exposed yet. Some algorithms like Softdrop and Massdrop are not available now, which can be included as well. The overall aim would be to include the currently unavailable features from FastJet using pybind11.

Timeline

Week 1-2:

- Getting familiar with Awkward Array and uproot and the existing bindings of pyjet.

Week 3:

- Identification of different areas where the support for Awkward Array and uproot can be integrated.

Week 4-6:

- Writing the array handling components of pyjet with awkward array to integrate support for Jagged Arrays

Week 7-9:

- Writing the uproot integration code to felicitate easy access to data stored in ROOT file format.

Week 10-11:

- Buffer week for testing and resolving any unforeseen issues and/or completion of previous weeks work.

Week 12-14:

- Getting familiar with pybind11 and writing the preliminary bindings for existing functions.

Week 15:

- Identifying any causes of inefficiency and streamlining the pybind11 bindings.

Week 16:

- Buffer period for testing and resolving any unforeseen issues.

Week 17:

- Exposing new methods from FastJet using pybind11.

Week 18:

- Completing the code for newly exposed methods and identifying and correcting any sources of inefficiency.

Week 19:

- Buffer week for testing out the newly exposed methods and preparing the documentation.

Week 20:

- Week for preparing the presentation on the work that was done during the fellowship.