

Contribution ID: 581 Contribution code: **contribution ID 581**Type: **Poster**

An array-oriented Python interface for FastJet.

Analysis on HEP data is an iterative process in which the results of one step often inform the next. In an exploratory analysis, it is common to perform one computation on a collection of events, then view the results (often with histograms) to decide what to try next. Awkward Array is a Scikit-HEP Python package that enables data analysis with array-at-a-time operations to implement cuts as slices, combinatorics as composable functions, etc. However, most C++ HEP libraries, such as FastJet, have an imperative, one-particle-at-a-time interface, which would be inefficient in Python and goes against the grain of the array-at-a-time logic of scientific Python. Therefore, we developed `fastjet`, a pip-installable Python package that provides FastJet C++ binaries, the classic (particle-at-a-time) Python interface, and the new array-oriented interface for use with Awkward Array.

The new interface streamlines interoperability with scientific Python software beyond HEP, such as machine learning. In one case, adopting this library along with other array-oriented tools accelerated HEP analysis code by a factor of 20. It was designed to be easily integrated with libraries in the Scikit-HEP ecosystem, including Uproot (file I/O), hist (histogramming), Vector (Lorentz vectors), and Coffea (high-level glue). Our talk will describe the design of the `fastjet` Python library, integrating the classic interface with the array-oriented interface and with the Vector library for Lorentz vector operations. We will also discuss problems encountered, lessons learned, and user feedback.

Significance

This is the first C++ library to be given Awkward Array bindings for vectorized use in Python. We hope to make the whole ecosystem of commonly used analysis tools usable this way.

References

Previously presented at PyHEP 2021:
<https://indico.cern.ch/event/1019958/contributions/4418484/>
<https://youtu.be/sOM43JfcGgs>

Speaker time zone

Compatible with America

Primary authors: ROY, Aryan; PIVARSKI, Jim (Princeton University)

Presenter: ROY, Aryan

Session Classification: Posters: Broccoli

Track Classification: Track 1: Computing Technology for Physics Research