IRIS-HEP Summer 2025 Fellows Program Project Proposal

Adapting the Coffea Framework for FCC

Applicant: Eglė Lenkaitytė

Mentor: David Lange

Project duration: June 30th 2025 – September 2025 (12 weeks)

Project description

The Future Circular Collider (FCC) is a planned next-generation particle accelerator at CERN, supposed to be a successor to the Large Hadron Collider (LHC). With a potential to reach 100 TeV collision energies, increased data rates and more precision, FCC will open the way to greater understanding of fundamentals and the search for new physics beyond the Standard Model. Even though the FCC is not to start running for the next 15-20 years, it is crucial to develop robust data analysis infrastructure now to ensure good preparation for when the real data begins to flow [1].

The Columnar Object Framework For Effective Analysis (COFFEA) is a Python-based unified analysis toolkit for High Energy Physics (HEP) data. Unlike traditional event loop based HEP analysis, columnar analysis operates on data in a column-oriented, array-programming style. Visualisation of both is shown in Figure 1. Coffea provides tools and components that simplify this method, enabling streamlined and reproducible HEP workflows. This approach offers more efficiency, scalability and is faster and easier to use [2].

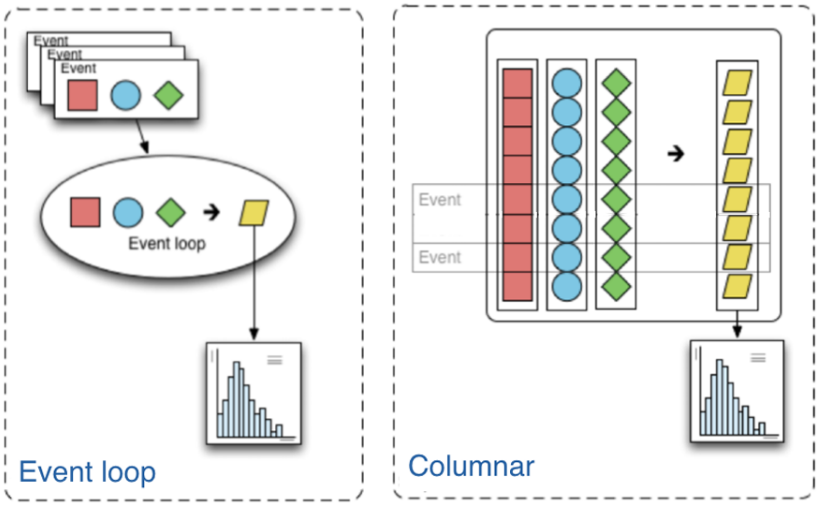

Figure 1. Schematic of the event loop and columnar data processing paradigm [2].

Coffea has already been successfully deployed in CMS (Compact Muon Solenoid). However, its application to FCC data, which is based on Key4hep [3] and FCC-specific software ecosystem (FCCSW) [4], is still being developed. This will allow for Coffea to work smoothly and be compatible with data formats and workflows used in FCC environment.

Planned work

This project aims to adapt the Coffea framework for use with FCC simulation data. Firstly, I will become familiar with the Coffea analysis framework. Secondly, I will create some examples of analysis using Coffea, initially basing them on existing materials, and later try including new examples that I develop. Thirdly, I will develop new types of algorithm components – such as particle identification or jet finding strategies – that are not yet implemented in Coffea.

Timeline

1-2 weeks: get familiar with Coffea.

3-4 weeks: build a series of example analysis for existing materials.

5-7 weeks: include my own examples into the analysis.

8-10 weeks: develop new types of algorithm components.

11-12 weeks: apply the components to my own examples and write documentation.

References

[1] CERN. Future Circular Collider. URL: https://home.cern/science/accelerators/future-circular-collider

[2] N. Smith, et al. Coffea – Columnar Object Framework for Effective Analysis. URL: https://doi.org/10.48550/arXiv.2008.12712

[3] Key4hep documentation. URL: https://key4hep.github.io/key4hep-doc/getting_started/introduction.html#about

[4] FCC Software overview. URL: https://hep-fcc.github.io/FCCSW/