

# ADAPTING THE COFFEA FRAMEWORK FOR FCC

Eglė Lenkaitytė<sup>1</sup>, David Lange<sup>2</sup>, Prayag Yadav<sup>3</sup>, Marijus Ambrozas<sup>1</sup>, Andrius Juodagalvis<sup>1</sup>

<sup>1</sup>Vilnius University, Institute of Theoretical Physics and Astronomy, Lithuania

<sup>2</sup>Princeton University, United States of America

<sup>3</sup>University of Hyderabad, India

[egle.lenkaityte@ff.stud.vu.lt](mailto:egle.lenkaityte@ff.stud.vu.lt)

The Future Circular Collider (FCC) is a planned next-generation particle accelerator at CERN and a proposed successor to the Large Hadron Collider (LHC). With the potential to reach collision energies of up to 100 TeV, along with increased data rates and improved precision, it promises significant advances in our understanding of fundamental physics and the search for phenomena beyond the Standard Model. Although the FCC is not expected to begin operation for another 15–20 years, developing robust data analysis infrastructure at an early stage is essential to ensure effective preparation for future data-taking [1].

The Columnar Object Framework For Effective Analysis (Coffea) is a Python-based, unified analysis toolkit designed for high-energy physics data. In contrast to traditional event-loop-based analyses, columnar analysis follows a column-oriented, array-programming style, as illustrated in Figure 1. Coffea provides a collection of tools that simplify this approach, enabling more streamlined and reproducible workflows. As a result, analyses become more efficient, scalable, and easier to implement [2].

Coffea has already been successfully deployed within the CMS (Compact Muon Solenoid) experiment. However, its application to simulation data for the FCC, which relies on the Key4hep framework [3] and a dedicated software ecosystem (FCCSW) [4], is still under active development. Ensuring compatibility with these data formats and workflows is a necessary step toward its broader adoption in this context.

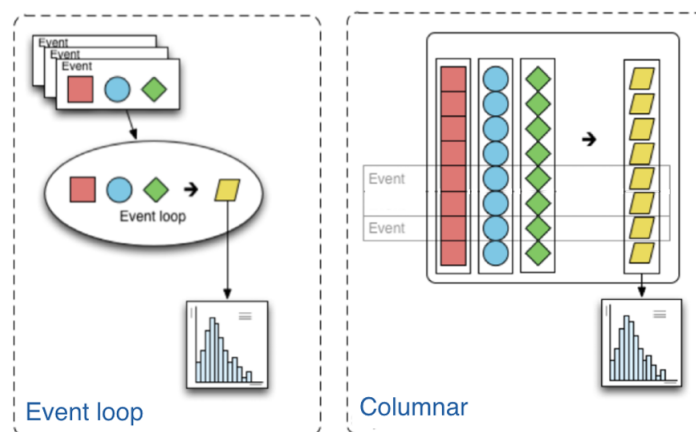


Fig. 1. Schematic of the event loop and columnar data processing paradigm [2].

The primary objective of this study is to adapt the Coffea framework for use with FCC simulation data. In particular, the project evaluates its suitability for the corresponding analysis workflows and explores extensions to its functionality where required.

The work begins with familiarization and practical use of Coffea through existing examples and tutorials, which are then adapted to simulation data produced within the Key4hep-based software environment. Initial results indicate that Coffea's columnar, array-based style can be successfully applied to FCC simulation data when appropriately integrated into the software ecosystem. Analyses adapted from existing workflows are able to process the relevant data formats efficiently, demonstrating clear improvements in scalability and performance compared to traditional event-loop approaches.

Overall, this study shows that Coffea can be effectively adapted and extended to meet the analysis requirements of future FCC studies. This research has been carried out in the framework of the agreement of Vilnius University with the Lithuanian Research Council no. S-IMPRESSU-24-4. The results will be presented at the conference.

[1] "The future circular collider," CERN, Jan. 28, 2026. <https://home.cern/science/accelerators/future-circular-collider>

[2] N. Smith et al., "COFFEA -- Columnar Object Framework for Effective Analysis," Springer Link (Chiba Institute of Technology), Aug. 2020, doi: 10.48550/arxiv.2008.12712.

[3] "Key4hep — Key4hep documentation." <https://key4hep.github.io/key4hep-doc/main/index.html>

[4] "FCC Software." <https://hep-fcc.github.io/FCCSW/>