

“Analysis Software in Key4HEP and beyond”

David Lange

Princeton University



**PRINCETON
UNIVERSITY**

Existing RDataFrame based framework

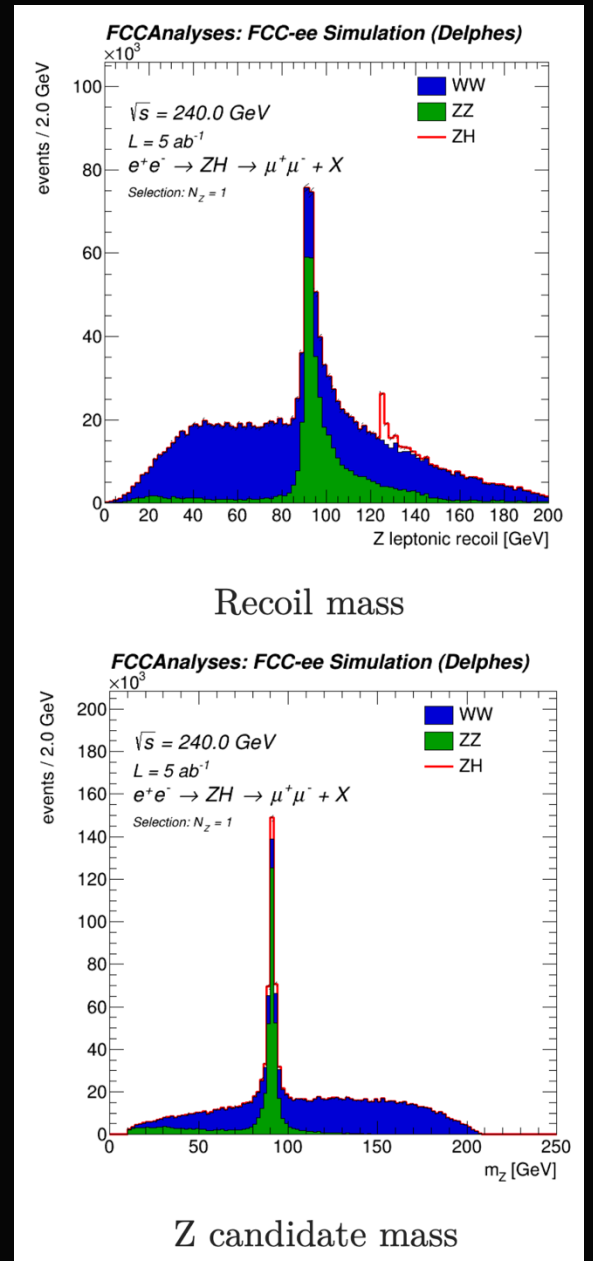
Python-based RDataFrame code w/ C++ helper

Supports various selection and plotting stages and has lots of examples

<https://github.com/HEP-FCC/FCCAnalyses/tree/master>

```
fccanalysis run examples/FCCee/higgs/mH-recoil/mumu/analysis_stage1.py
fccanalysis run examples/FCCee/higgs/mH-recoil/mumu/analysis_stage2.py
fccanalysis final examples/FCCee/higgs/mH-recoil/mumu/analysis_final.py
fccanalysis plots examples/FCCee/higgs/mH-recoil/mumu/analysis_plots.py
```

We wanted to add support for doing similar analyses using Coffea and other Python tools that have grown out of the US LHC community



COFFEA and the Python Ecosystem

- **Uproot** to read and write root files.
- **Awkward Array** to manipulate HEP data in a numpythonic way.
- **Dask** to for scaling-out
- **Hist** for histogramming
- **MPLHEP** for plotting
- **Numba** for just-in-time compilation
- and other useful packages ...



Project Goals

1. Add support for EDM4HEP schemas to Coffea
 - Coffea schemas play the role of building logical groups of branches -- collections and relationships between branches to represent physical objects and their connections-- and helper functions that reduce boilerplate code
 - There is a recent major break in compatibility, but in any case, one aspect of this support is supporting different EDM4HEP schemas
2. Build and document several demonstrators based on Coffea and EDM4HEP simulation samples

* Prayag Yadav is doing much (essentially all) of this work as part of a NSF HSF-India fellowship project (with Princeton/Univ. of Hyderabad)

Project Status

- "Old"-style EDM4HEP files (including Winter2023 FCC samples) are now supported by recent Coffea versions (from 2024.10.0)

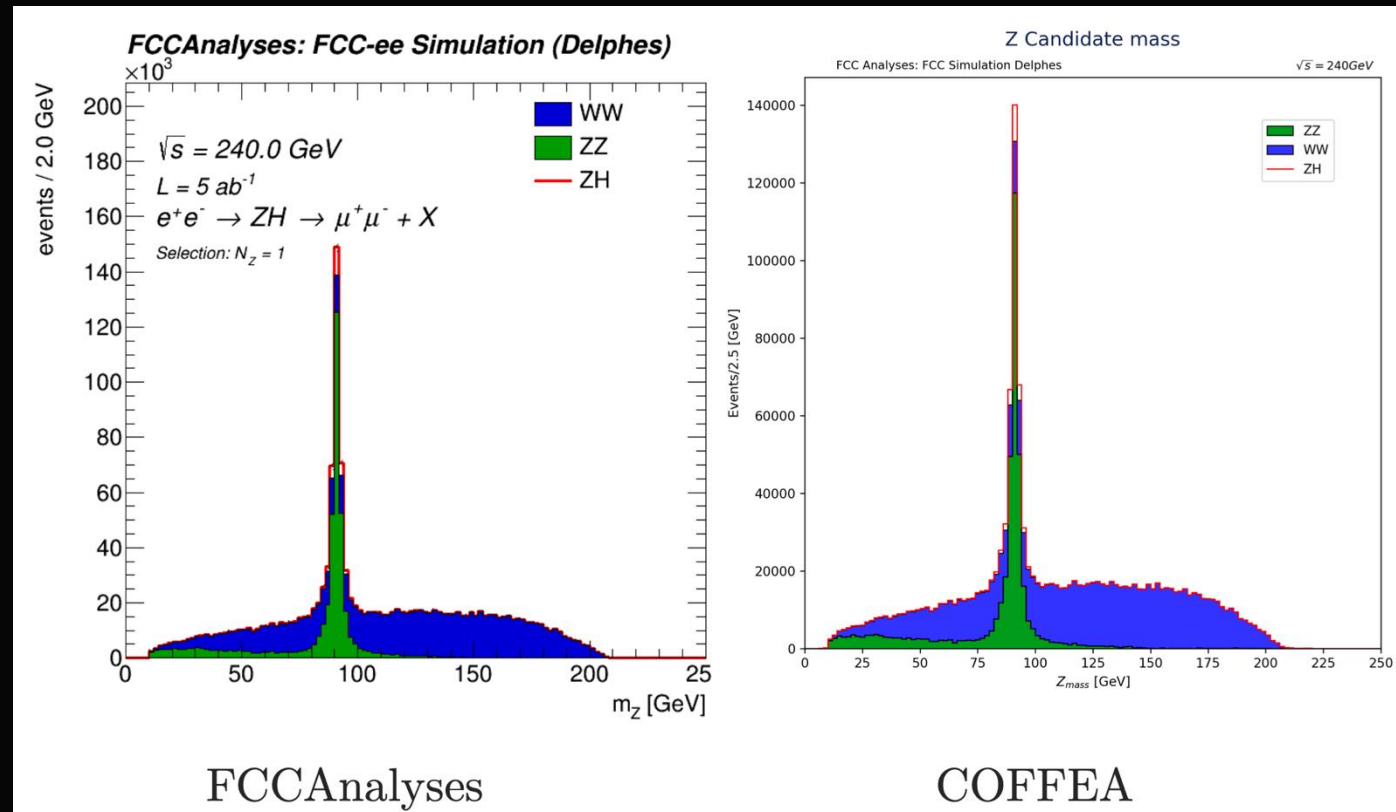
```
from coffea.nanoevents import NanoEventsFactory, BaseSchema, FCC
```

```
fcc = FCC.get_schema("pre-edm4hep1")  
fcc
```

```
coffea.nanoevents.schemas.fcc.FCCSchema
```

```
events = NanoEventsFactory.from_root(  
    '../coffea-fcc-analyses/data/wzp6_ee_mumuH_Hbb_ecm240/events_159112833.root:events',  
    schemaclass=fcc,  
    entry_stop=100,  
    metadata={'dataset': 'ZH'},  
    delayed=False,  
    uproot_options={"filter_name": lambda x : "PARAMETERS" not in x}  
).events()
```

Technical performance and Physics results are as expected



- Only simple benchmarking done, but no significant performance differences (Coffea was faster on our test case)

Work in progress

- Working through support for “New”-style files as well as a general solution building the Coffea schema with help of the Podio YAML specification (future proofing)
- Finishing up documented examples to include in Key4Hep (
 - First two examples are clones of the Z(mumu)H recoil analysis and a jet clustering analysis from fccanalysis.
 - Key4Hep stack is not required for use but is a convenient way to make this discoverable
 - Some more details at <https://indico.cern.ch/event/1487970/#1-coffea-and-fcc>

”Beyond”

- Developer/User environment
 - Constantly improving but considerable space to improve (and in particular towards getting new students started)
- Make existing full sim algorithms better
 - Add threading support
 - Transition from wrappers to native Gaudi algorithms
 - Technical performance
- Bring algorithms into key4hep and make them work for proto detectors
- Longer term – Apply new approaches to future collider detector design
 - Eg, differentiable pipelines or simulation based inference techniques?

Proposed workshop on “Software & Computing: From LHC to Future Colliders”

- Workshop and training event proposed for late winter (possibly week of February 24-28) hosted at Princeton University.
- Possible goals for questions to answer:
 - Needs: What does the US community need in terms of software and computing for Future Collider studies?
 - Resources: Are there opportunities for the future collider effort to align with and/or leverage existing S&C projects (HEP-CCE, IRIS-HEP, etc.)? What similar collaborations are possible with CERN and other international efforts?
 - Near term goals: What concrete things could be accomplished over the next 12 months?
 - An associated training event could cover modern software tools, Key4Hep, etc.
- Partial support can be provided by the NSF-funded IRIS-HEP software institute as an “ecosystems/blueprint” workshop and training event.