Task 12.2 - Turnkey Software

AIDAinnova, 2nd Annual meeting
Valencia

Thomas Madlener

Apr 24, 2023

HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 101004761.

AIDA innova

CERN

DESY.

Task 12.2 - Turnkey Software

Overview

- Integrated Turnkey Software Stack, for physics and performance studies
- Simplified data model toolkit for modern hardware platforms
- Digitisation extensions for geometry toolkit
- R&D study on frameworks to manage heterogeneous resources

DESY (lead), CERN,
INFN-PI, (INFN-PD, INFN-BA, INFN-BO - unfunded)
IHEP, SDU - associated

D #	Deliverable Name	Lead beneficiary	Type	Due Date (in months)
D12.1	Turnkey software stack (Key4hep)	12 - DESY	Other	46

MS #	Milestone Name	Lead beneficiary	Due Date (in months)	Means of verification
MS47	LC reconstruction prototype in Key4hep	12 - DESY	21	Reproduce similar detector performance as achieved with the current framework (Task 12.2)

Symbols on coming slides:

✅ Done

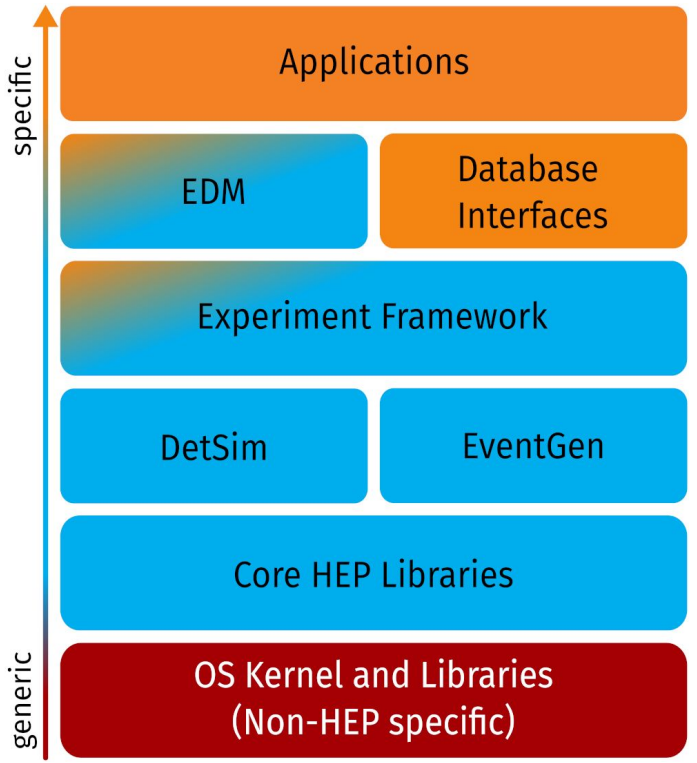
👷 Ongoing or planned

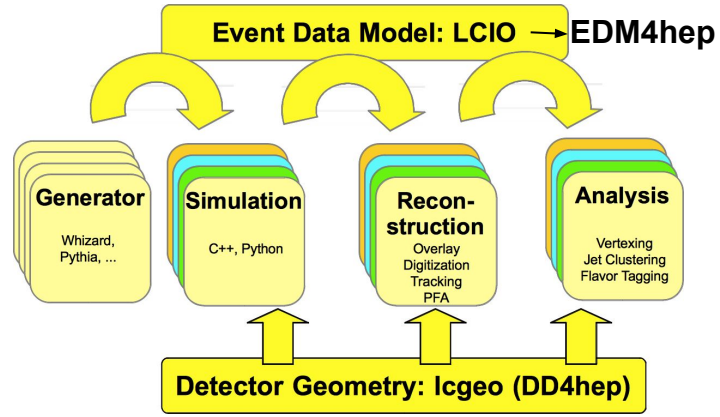
✅

Key4hep

Turnkey software stack for all future collider projects

- Develop a **common turnkey software stack** for future collider studies
- Take existing tools where possible, provide necessary interfaces and contribute to the development
  - A lot of existing software from the shared **iLCSoft** developed by ILC and CLIC for many years
- Develop new tools or libraries where necessary
- All major players involved: CEPC, CLIC, FCC, ILC, EIC, …
- Provide a complete data processing framework
  - Shared components reduce overhead for all users
- Make things as easy to use as possible for everybody (librarians, developers, users)
- Supported by **HSF**, **CERN** and **AIDAinnova**





PODIO

github.com/AIDASoft/podio

The event data model toolkit

- Generate code from simple yaml definition of EDM
- Based on using and storing POD (plain old data) structures
- Make it possible to target different I/O backends

✅ **Frame** class and concept (with accompanying multithreading model)

✅ License change to Apache2

✅ Allow datamodel extensions

✅ Generate code for dumping collections to JSON

✅ Many many changes under the hood

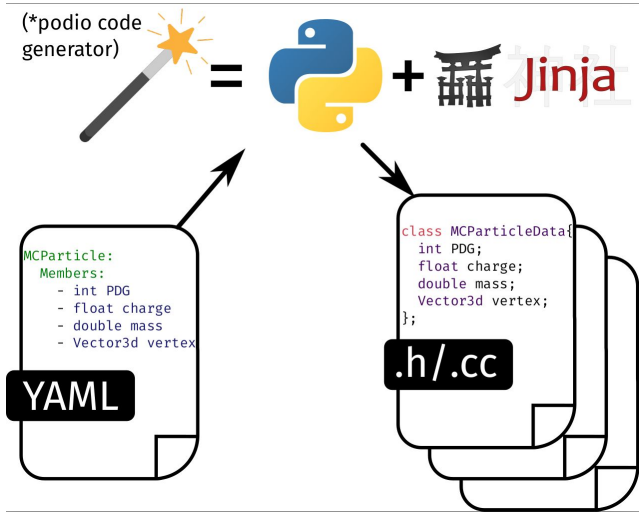
👷 **Schema evolution** of generated EDMs

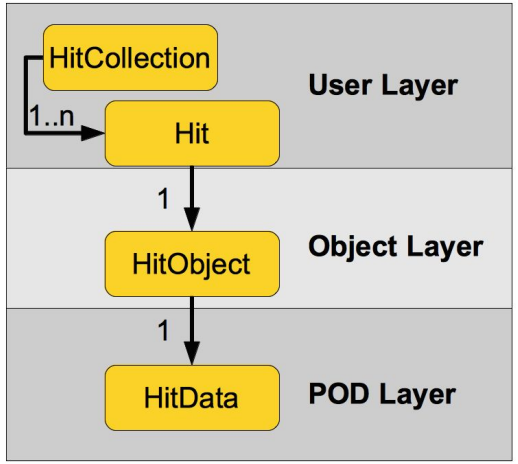
👷 **Version 1.0** (backwards compatibility from then on)

👷 RNTuple based backend (try to merge podio#395 during hackathon)

👷 Some prototyping and testing on heterogeneous resources

👷 Small(-ish) additional features (already a few on the wish list)



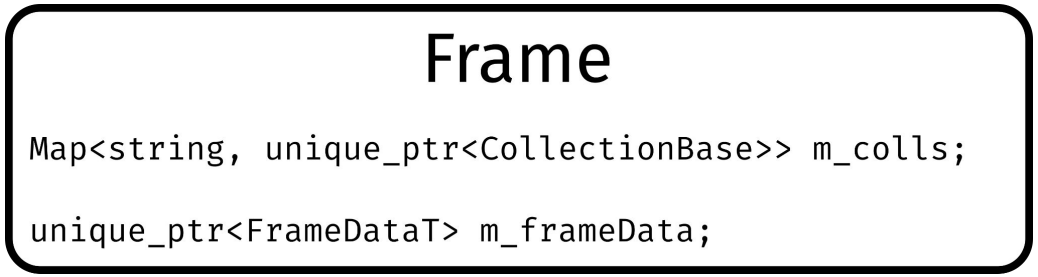


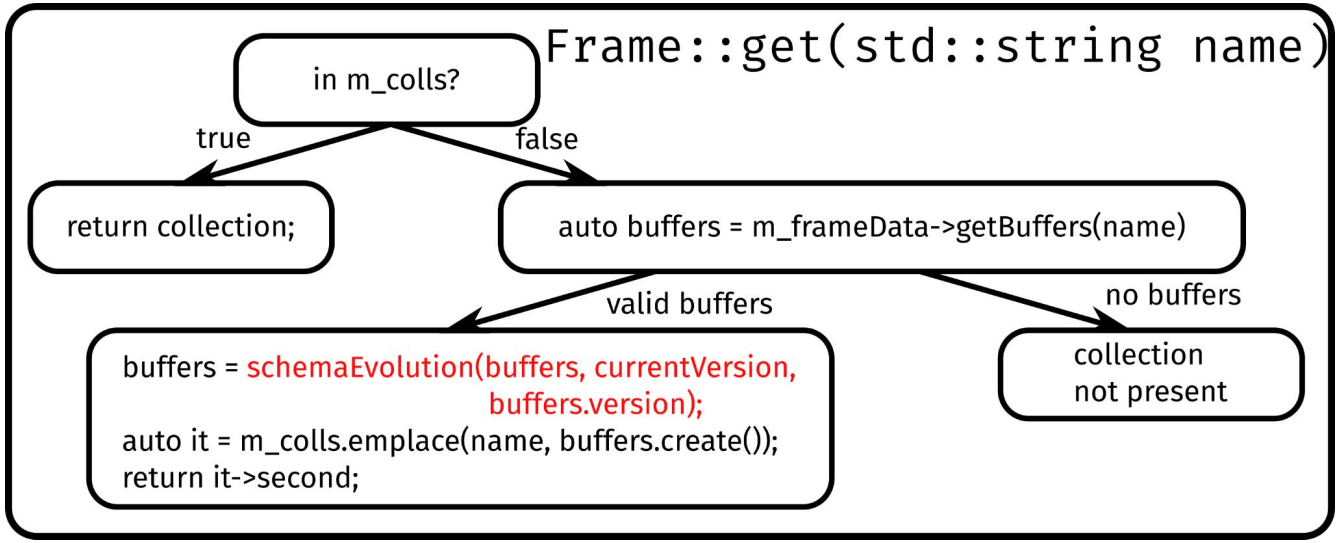
PODIO schema evolution

The feature that keeps on giving

- Biggest missing feature
- Quite large in scope
  - Work with different backends
  - Leverage backend schema evolution if present
  - Support “lazy” unpacking
- Required many under-the-hood changes to podio
- Concept and prototype done
- Currently polishing and refactoring

Conceptual Frame and Frame::get



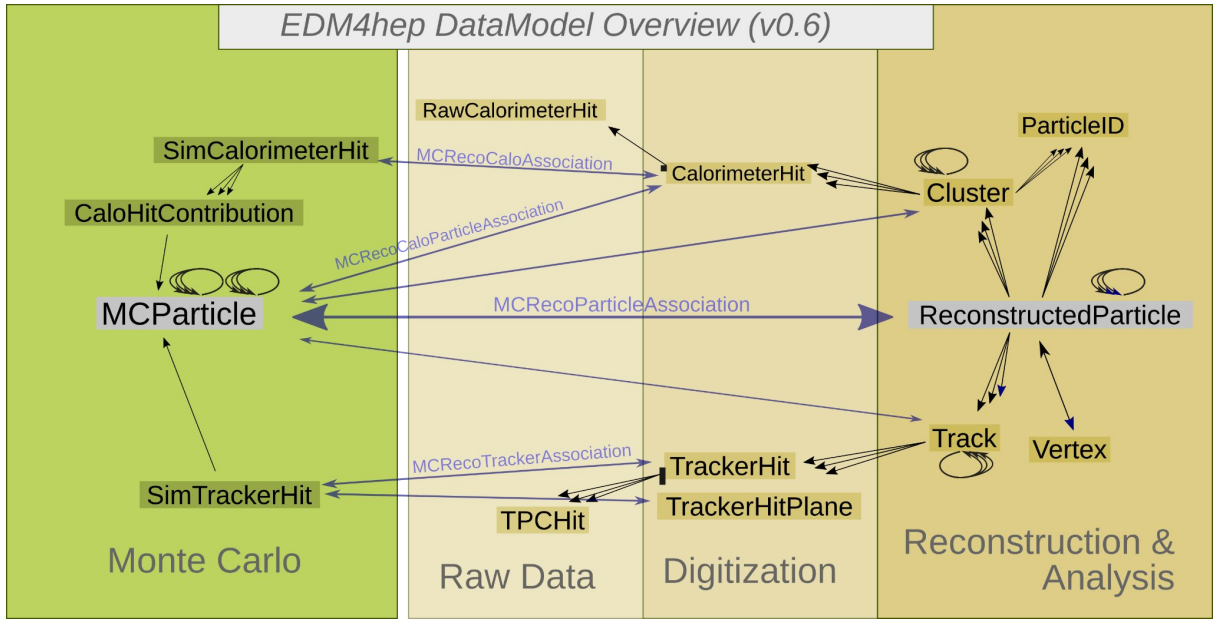


EDM4hep

The common event data model

github.com/key4hep/EDM4hep
edm4hep.web.cern.ch

- EDM4hep defines the common *language* for all Key4hep components to communicate
- Heavily inspired by LCIO that has been successfully shared by ILC and CLIC
  - Additional novel ideas from fcc-edm
- Generated by the PODIO EDM toolkit
  - EDM4hep and EICD main customers of PODIO



✅ Addition of datatypes for drift chamber study

✅ Tool for dumping to JSON for Phoenix event display

✅ Used as “proper” upstream for EICD

👷 **Version 1.0** (backwards compatibility from then on, **needs PODIO schema evolution**)

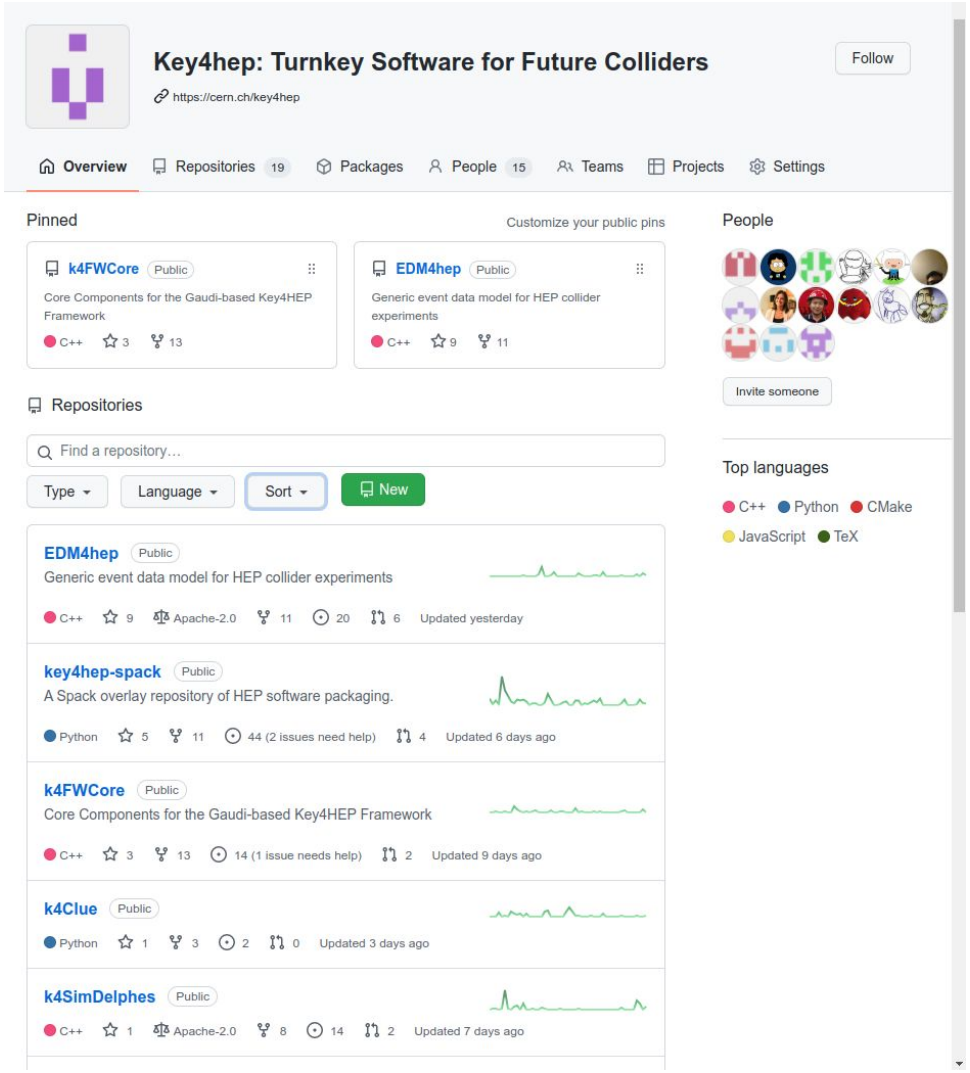
👷 (Standalone) conversion from LCIO

👷 Utility functionality as necessary

Key4hep Framework

github.com/key4hep
key4hep.web.cern.ch

- Gaudi based core framework
  - **k4FWCore** provides I/O for PODIO based EDMs
  - k4SimDelphes for Delphes integration
  - **k4MarlinWrapper** for calling Marlin processors
  - **k4geo** for detector models (rebranded from lcgeo)
  - k4SimGeant4 for Geant4 based simulation
  - k4Gen for generator integration
  - …



✅ **MS47: Comparable performance to iLCSoft via k4MarlinWrapper**

✅ Many small(ish) fixes

👷 Switch FW core to use podio Frames

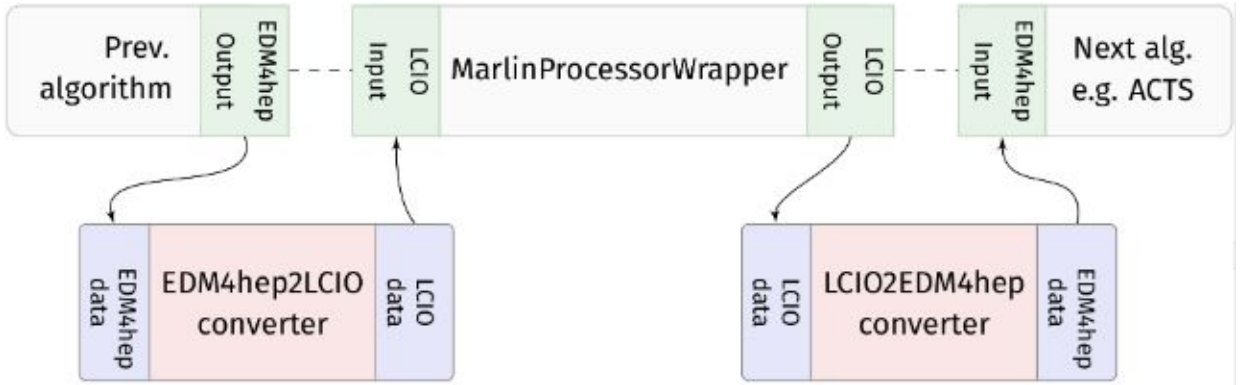
👷 Integration of ACTS, …

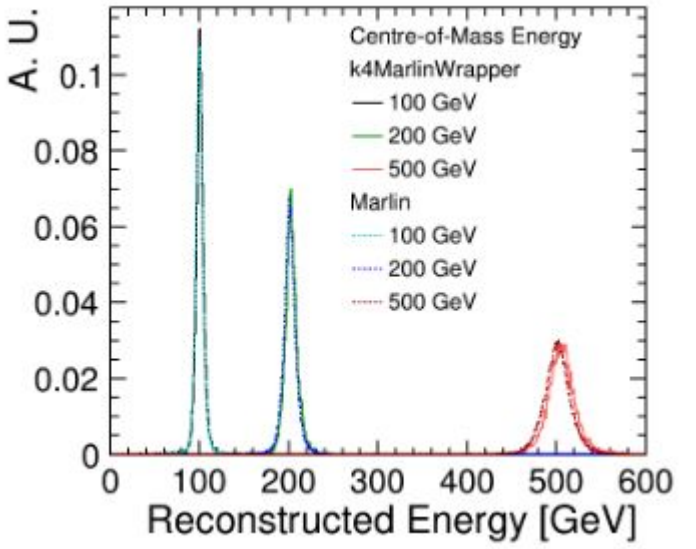
👷 Consolidation of simulation approaches

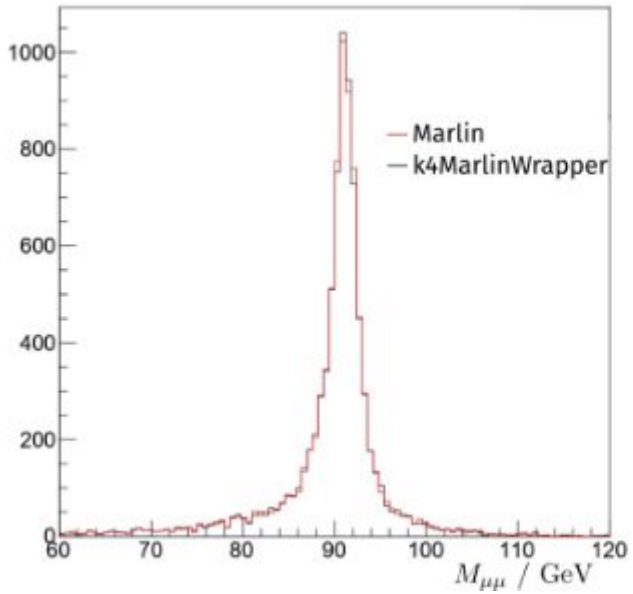
MS47

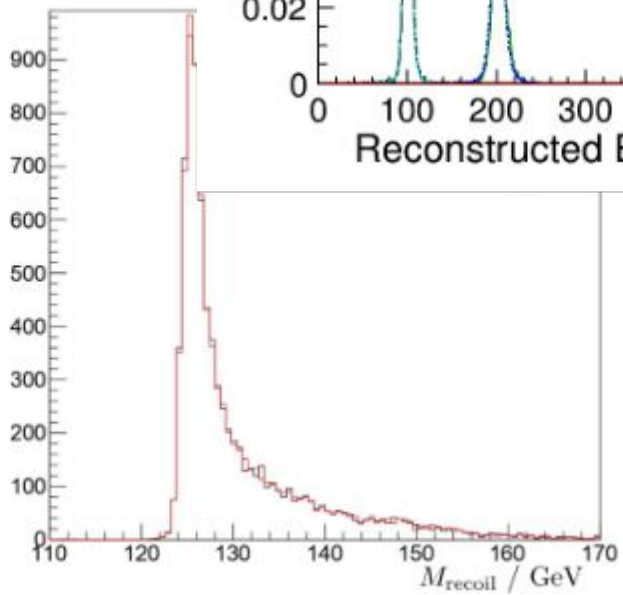
LC reconstruction prototype in Key4hep

- Run unchanged Marlin Processors via Wrapper
- Configurable on-demand conversion between LCIO and EDM4hep (and vice versa)
- Steering file conversion script
- Extended testing uncovered a few smaller issues
- Excellent agreement between Marlin and Gaudi
- CLIC reconstruction run as part of CI for k4MarlinWrapper
- Working horse for k4CLUE studies
- Report online Jan 19, 2023









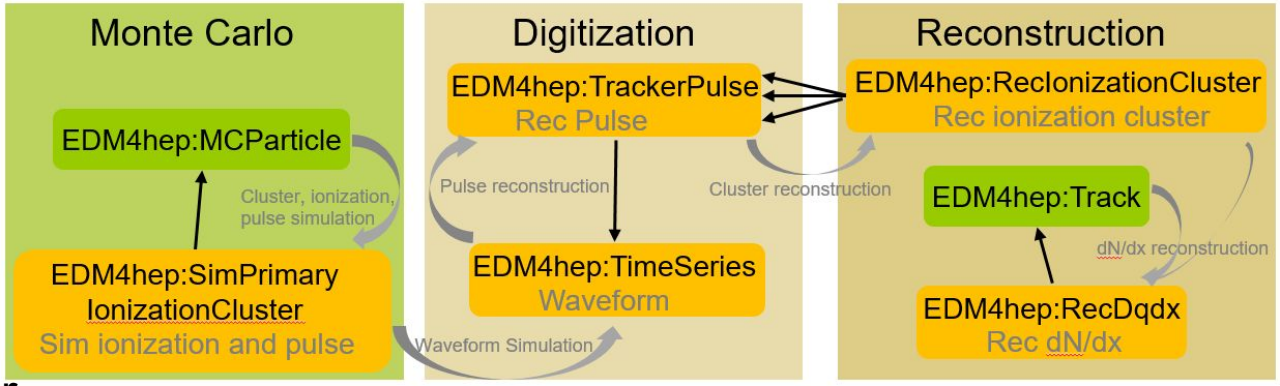
Work at INFN

R&D study on framework to manage heterogeneous resources

- No applicant on second call for dedicated position
  - Despite trying to advertise it as widely as possible
- Competing against numerous other more attractive software/computing positions
  - Both money and career prospects
- Currently figuring out whether a third call is possible
- To be effective would need to be more appealing
  - Funds are rather little

Report from IHEP group

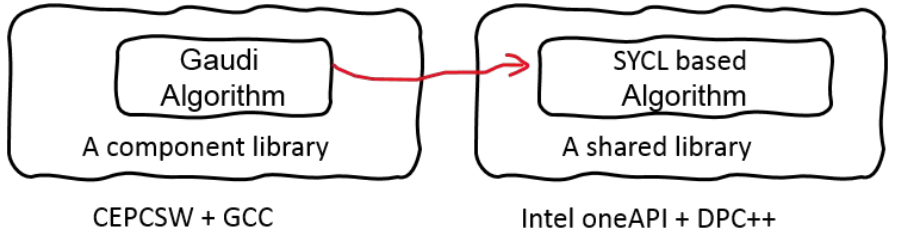
Courtesy Weidong Li



✅ Extending Key4hep to accommodate requirements from simulation of primary ionization in the drift chamber

✅ Integration of ACTS' seeding algorithm (TRACCC) with CEPCS; algorithm using EDM4hep hits as inputs

✅ Prototype for using Gaussino as underlying framework for CEPC detector simulation

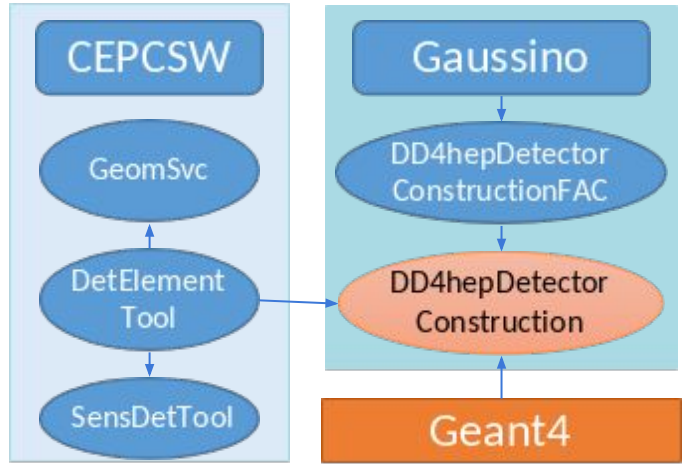


👷 Extend EDM4hep to support ongoing activities in CEPC

- Reconstruction of number of primary ionizations in drift chamber
- Analysis of test beam data for vertex detector prototype

👷 Optimize data flow / communication between CPU and GPU. Investigate possibility of avoiding copies by building a connection between EDM4hep and TRACCCs vecmem

👷 Re-implement CEPC detector simulation with Gaussino

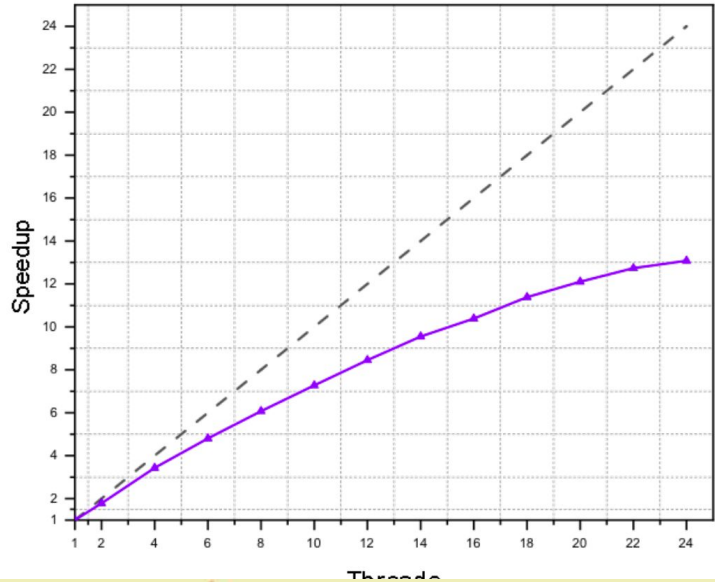


Report from SDU group

Courtesy Xingtao Huang

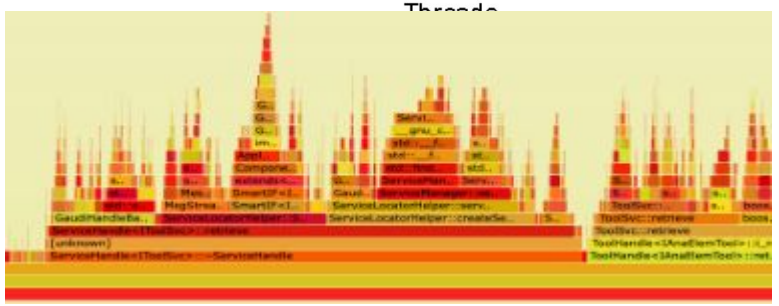
✅ Performance test and development of RDataFrame based Analysis toolkit

- Basic functionality and performance tests using Higgs recoil analysis
- Integration of MarlinKinFit into RDataFrame for CEPC
- Provide example for porting C++ applications to RDataFrame



✅ Developed the Valprod toolkit to support building comprehensive validation jobs

- Integration of Valprod with prmon,
- Enhance Valprod to support disk profiling, drawing on-cpu flame graph,...
- Build basic CI tests for CEPC using Valprod toolkit



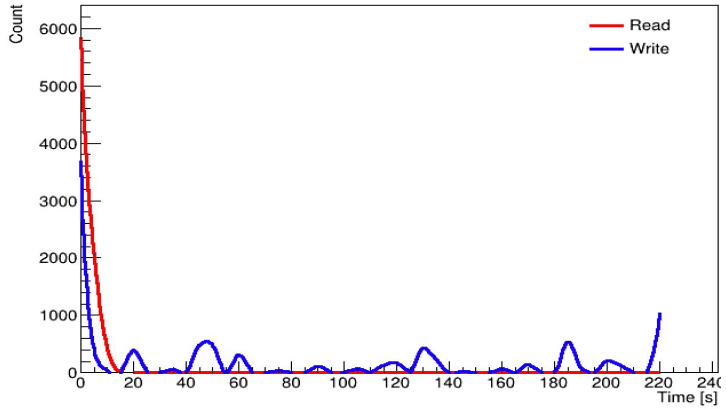
👷 Integrate more missing analysis tools to RDataFrame

👷 Use CEPC beam test data for testing full analysis chain with RDataFrame

👷 Performance tests using multi-threaded RDataFrame and full analysis chain

👷 Continue development of validation system and promote toolkit to other Key4hep components

SimTest IO Operations **CI artifacts**



Summary / Main Goals for the next year

- **Delieverd MS47: LC reconstruction prototype in Key4hep**
- Steady progress in Key4hep
- **First stable releases for PODIO and EDM4hep are in sight**
  - Requires Schema evolution
- First steps on heterogeneous resources with podio generated EDMs
- ACTS integration
- Integration of ML FastSim models into DD4hep
  - Prototype exists
  - Big topic of this years hackathon

Thank you