CEPC Software and Computing

Weidong Li
representing the CEPC software and computing team

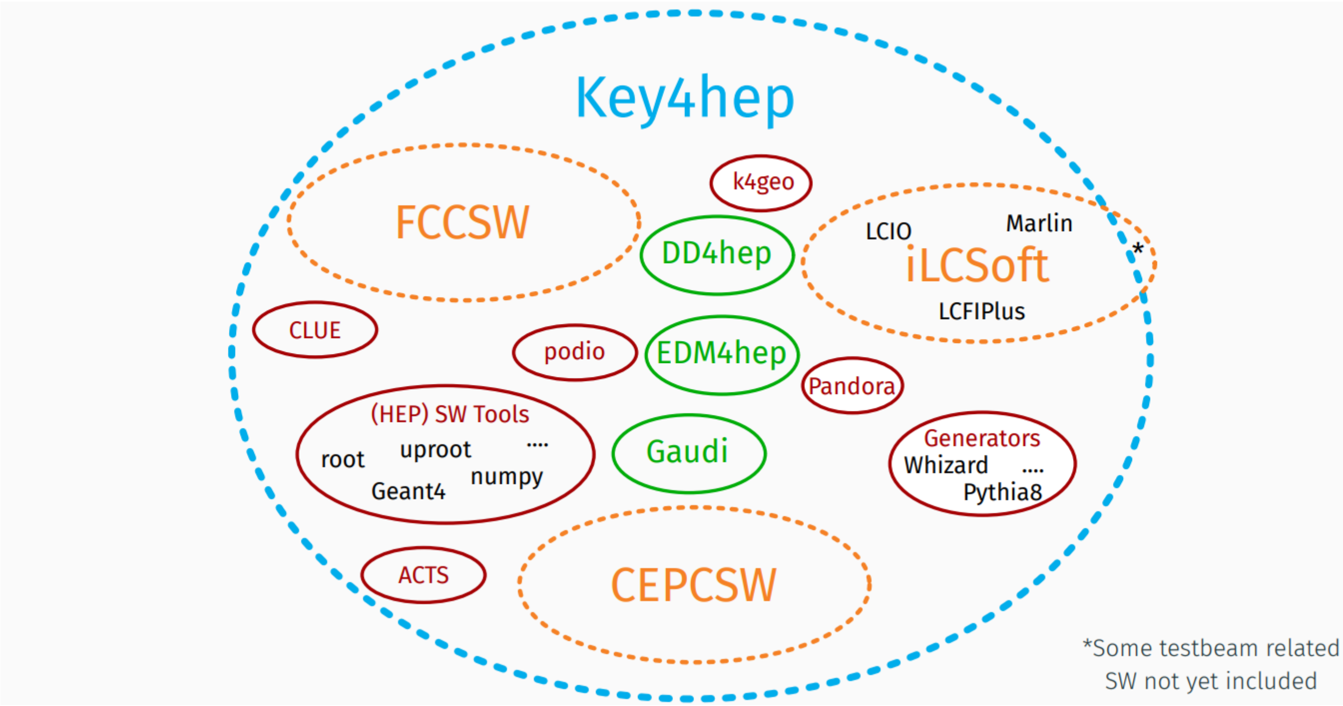
9th CEPC-SppC IAC meeting

IHEP, 31 October 2023

Contents

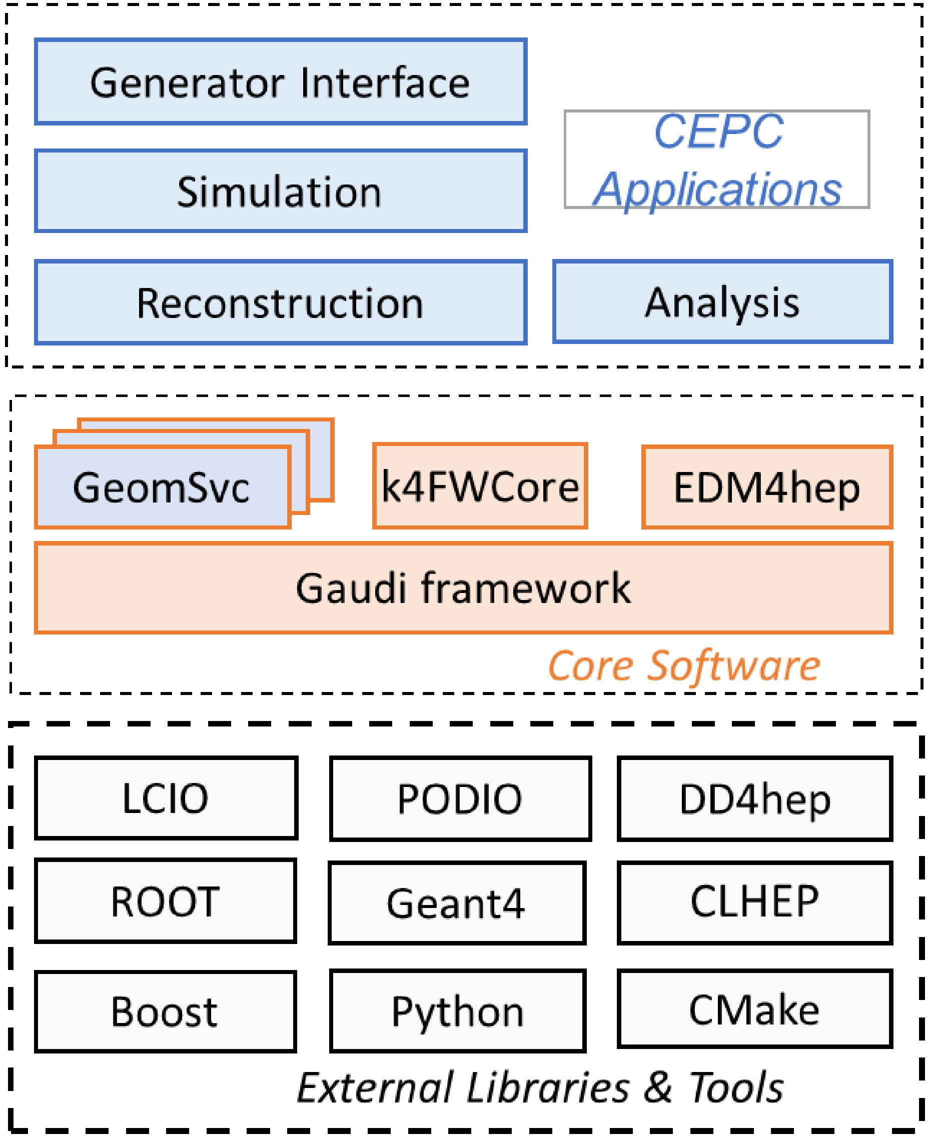
Key4hep

- The consensus among CEPC, CLIC, FCC, ILC and other future experiments was reached at the Bologna workshop in 2019
  - Develop a Common Turnkey Software Stack, Key4hep, for future collider experiments
  - Maximize the sharing of software components among different experiments
- The development of CEPCSW is based on Key4hep and most of our efforts are put on CEPC experiment specific software and workflows



Structure of CEPCSW

- CEPCSW is organized as a multi-layer structure
  - Applications: simulation, reconstruction and analysis
  - Core software
  - External libraries
- The key components of core software include:
  - Gaudi/Gaudi: defines interfaces to all software components
  - Edm4hep: generic event data model
  - k4FWCore: management of event data objects
  - DD4hep: detector geometry description
  - CEPC-specific components : GeomSvc, simulation framework, analysis framework, beam background mixing, fast simulation, machine learning interface, etc.



Packages of CEPCSW

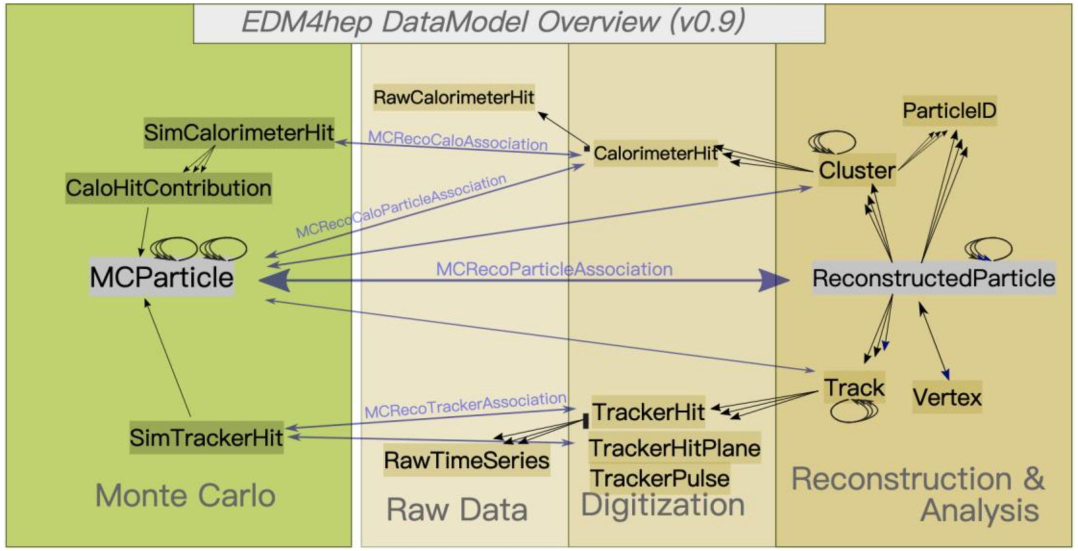
- Detector Options
  - CDR (baseline design)
  - The 4th concept
- MC Generators
  - Multiple formats supported: HepMC, HepEvt, StdHep, LCIO
  - GuineaPig++ for MDI
  - Particle Gun
- Simulation
  - G4 simulation framework
  - Fast simulation algoritrhms e.g.ML-based dE/dx simulation
  - Digitization algorithms for silicon, CALO, drift chamber
- Reconstruction
  - Marlin based tracking algorithms for silicon detector
  - Tracking algorithm for drift chamber
  - Pandora-based PFA
  - Arbor-based PFA
- Analysis tools
  - RDataFrame-based analysis framework
- Examples and docs
  - Menu for EDM4hep, Identifier, etc.

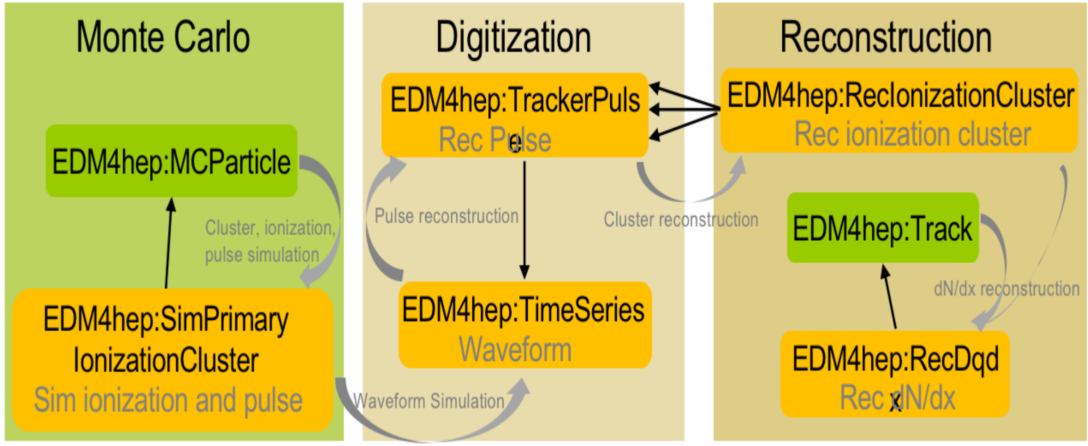
https://github.com/cepc/CEPCSW

~ 50 packages in total

Progress in Event Data Model

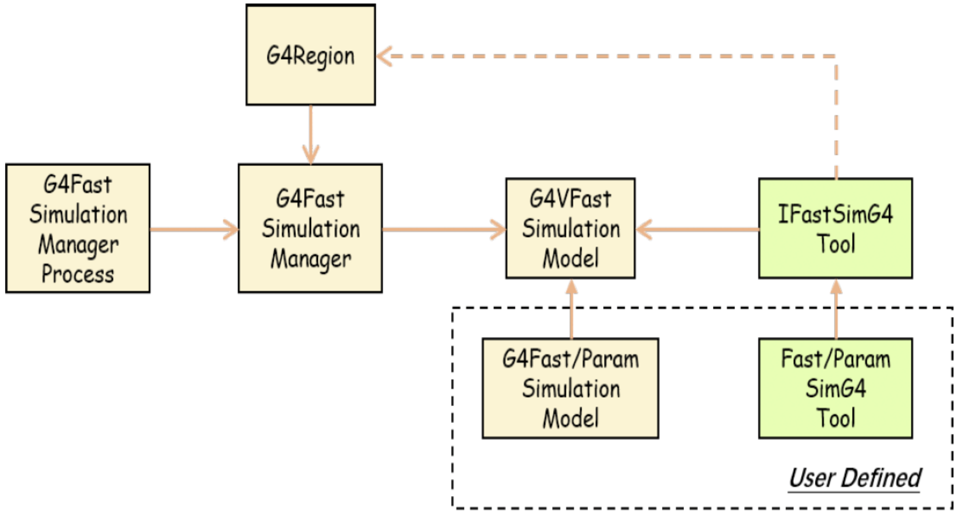
- EDM4hep is the common event data model (EDM) being developed for the future experiments
  - describing event objects created at different data processing stages and also reflecting the relationship between them.
- Extension of the Edm4hep was completed and released
  - By using the upstream mechanism of PODIO, a common EDM was implemented for both TPC and drift chamber

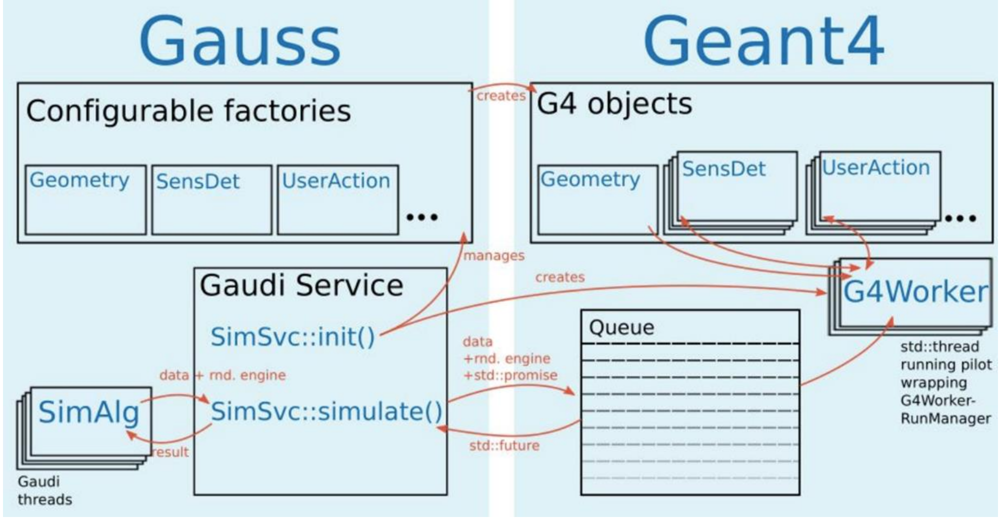




Progress in Simulation Framework

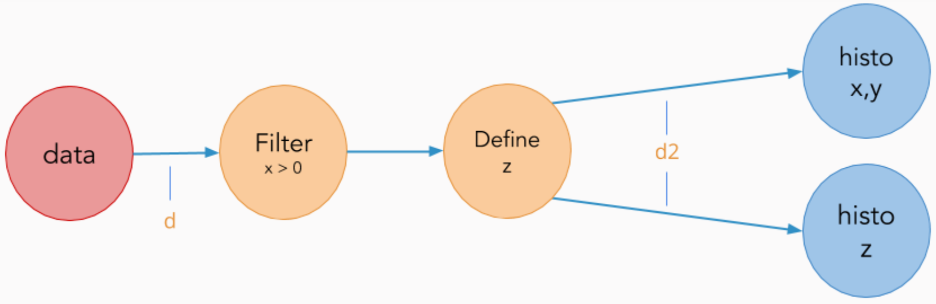
- The simulation framework was developed and currently the simulation chain is complete
- Gaussino, a general simulation framework originally developed by LHCb, is being added into Key4hep as k4Gaussino
- In collaboration with CERN, we are developing simulation prototype for silicon vertex detector using Gaussino as the underlying framework
  - Step 1: Using the original version having the dependency on the LHCb software (Done)
  - Step 2: Creating the modified version in which the LHCb dependency is removed (Done)
  - Step 3: Directly using k4Gaussino (Not started)



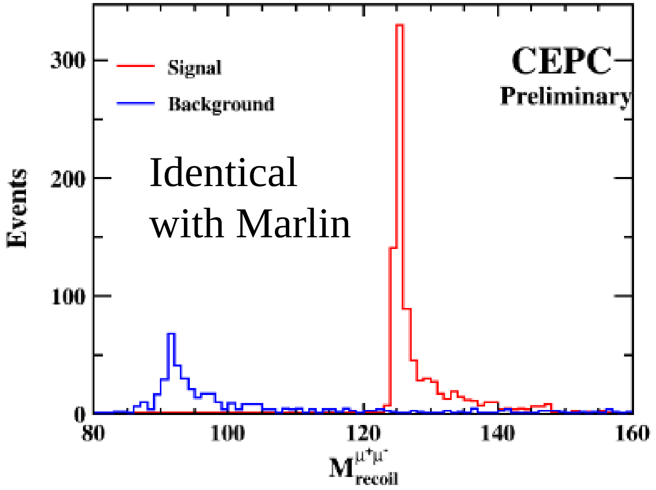


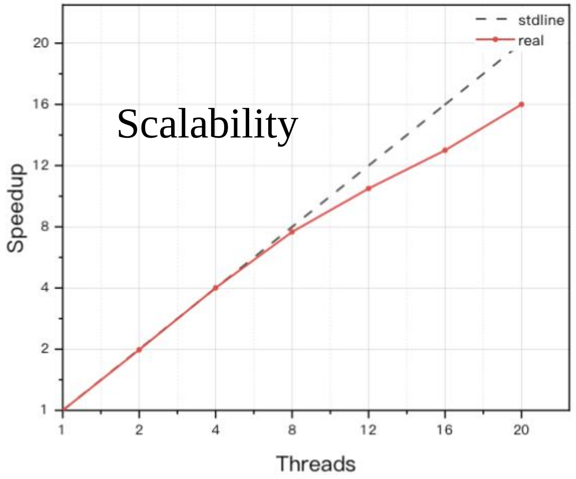
Progress in RDataFrame-based Analysis Framework

- RDataFrame is a powerful tool for data analysis
  - Program language: Python and C++
  - Declarative programming and parallel processing
  - Used by many experiments such as FCC-ee
- New developments since last meeting
  - Several algorithms were ported from Marlin
    - JetClustering, KinematicFit,
  - More are being implemented
    - VertexFit, JetTagging, PID etc.
  - Performance test with two analysis channels
    - e+e- -> Z(mumu)H
    - e+e- ->H(2jet) mumu



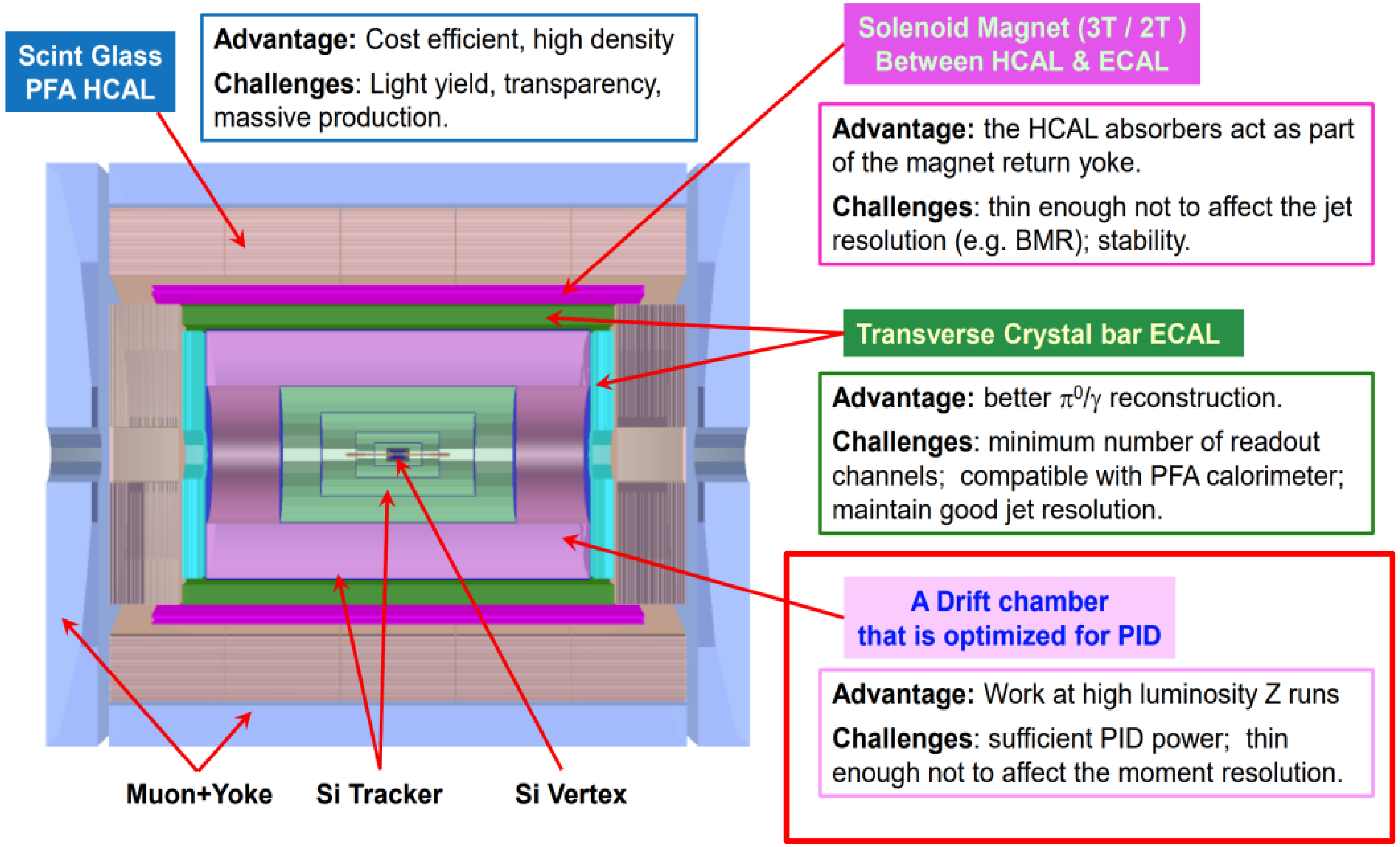

```
// d2 is a new data-frame, a transformed version of d
auto d2 = d.Filter("x > 0")
           .Define("z","x*x + y*y");
// make histograms out of it
auto hz = d2.Histo1D("z");
auto hxy = d2.Histo2D({"hxy","hxy",16,-1,1,64,-1,1},"x","y");
```





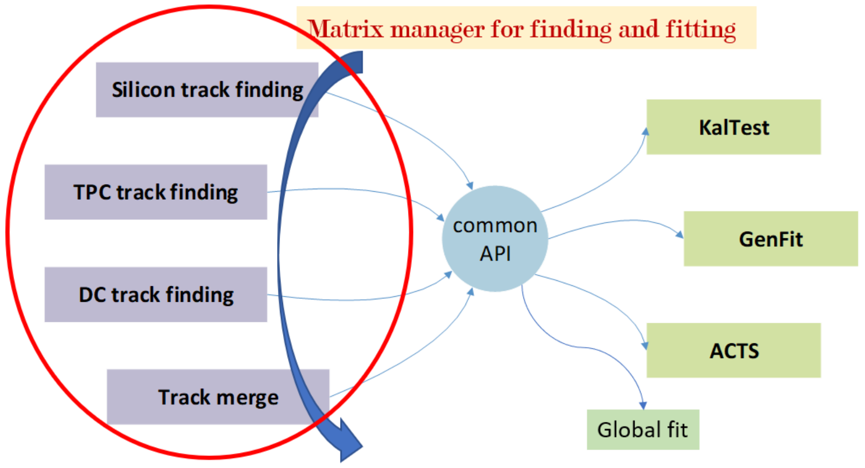
Detector Software Development

- The 4th conceptual detector
  - Silicon Tracker
  - Drift Chamber
  - Transverse Crystal-bar ECAL
  - Scintillator Glass HCAL
- Both detector design and physics potential studies need simulation and reconstruction software
- In the past year, software development was mainly focused on the first three sub-detectors

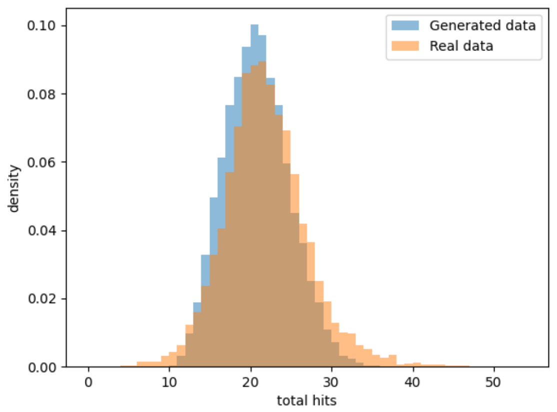


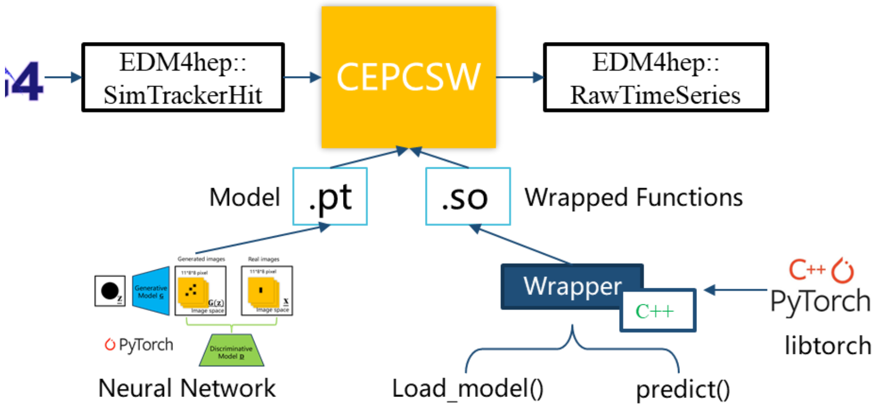
Progress in Silicon Tracker Software (1)

- A common API for different track fitting algorithms was developed
- Performance validation completed with
  - single particles
  - $H \rightarrow \mu\mu$
  - $\tau \rightarrow 3\text{prong}$
  - $b\bar{b}H$
- For the silicon vertex detector, machine leaning based digitization algorithm was developed
  - Hit clusters are generated by a GAN model
  - Training dataset: testbeam data collected at DESY with TaichuPix-3



Comparison of cluster between GAN and data



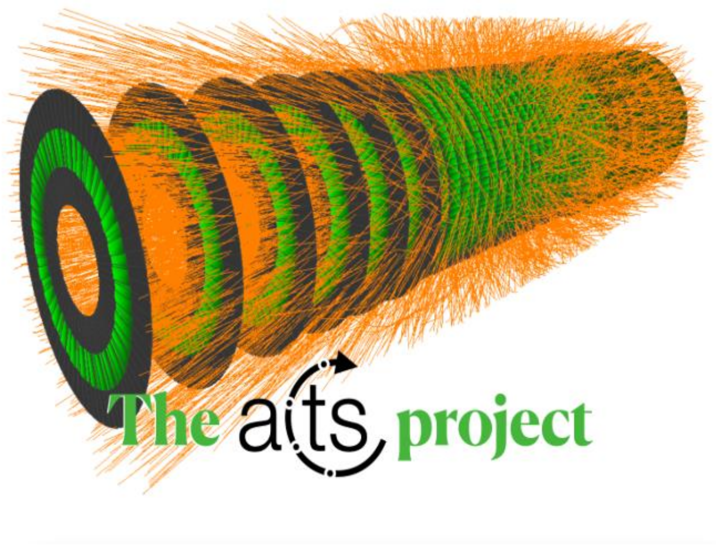


Progress in Silicon Tracker Software (2)

- TRACCC: one of ACTS R&D projects
  - Full chain demonstrator for track reconstruction on CPU/GPU

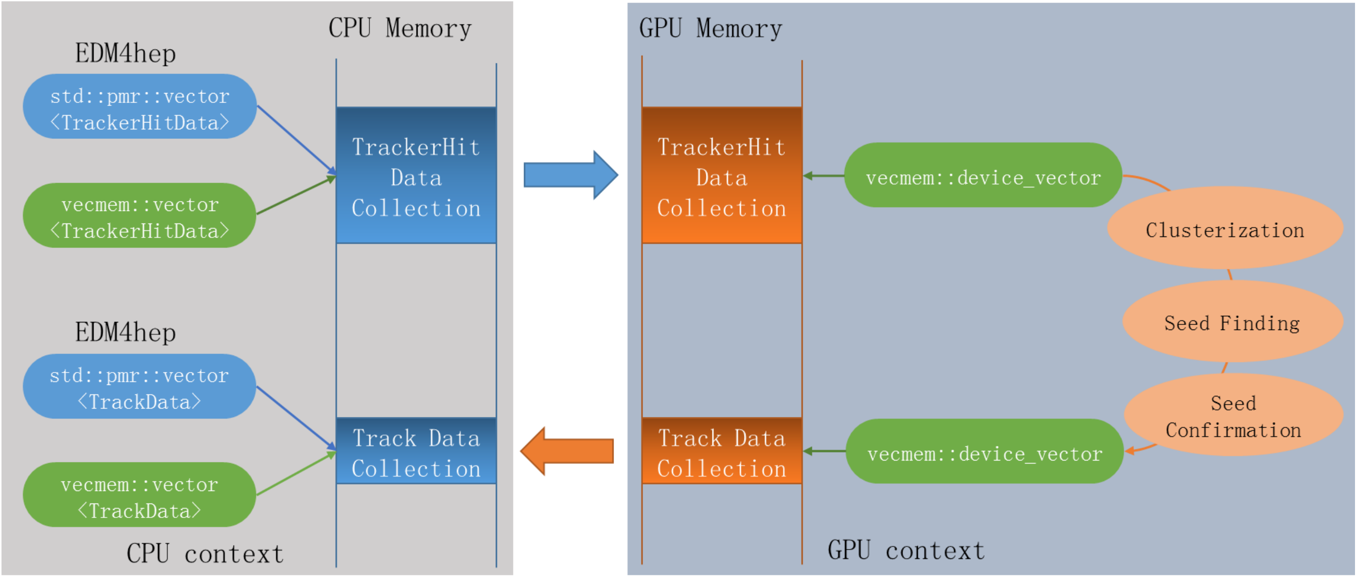
  https://github.com/acts-project/traccc

- Developing CEPC seeding algorithm based on TRACCC
- Building a bridge between EDM4hep and TRACCC
  - Common memory for both EDM4hep and TRACCC
  - No data conversion is needed between them



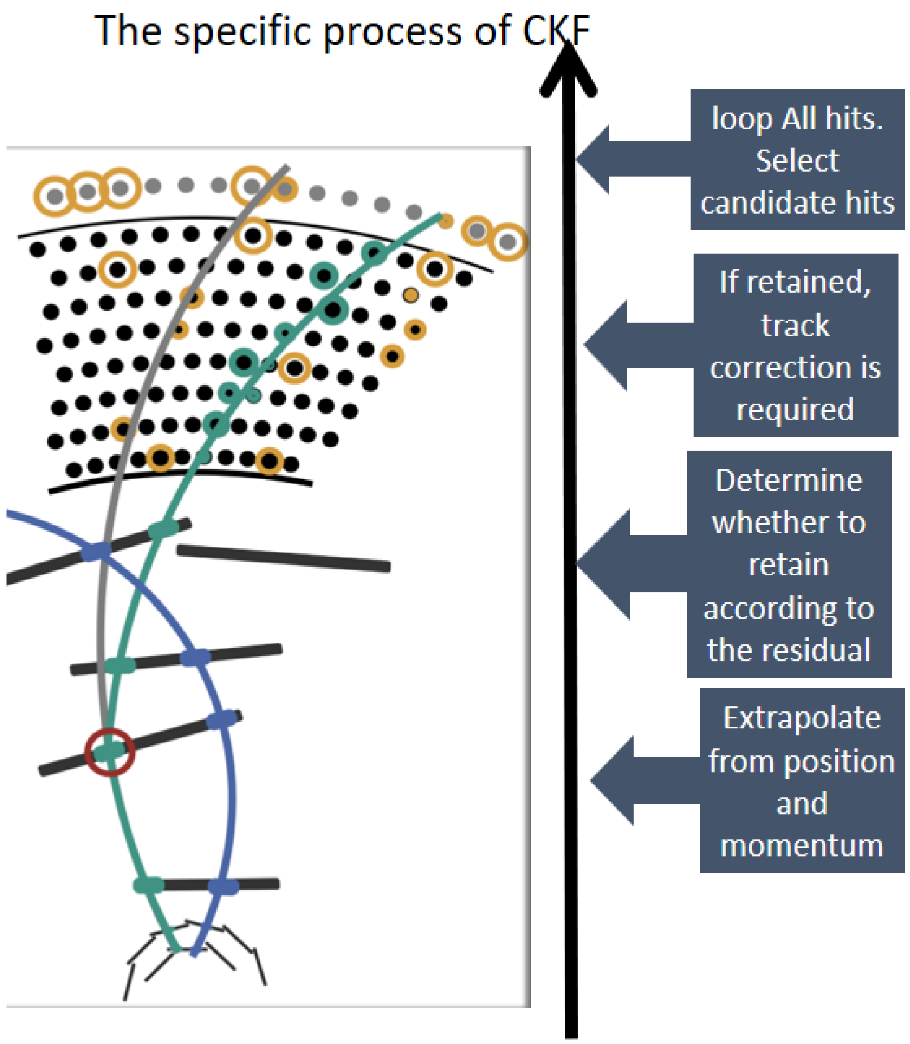
supported by

cooperations



Progress in Drift Chamber Software (1)

- Track finding algorithm
  - Track segments reconstructed in the silicon detector, called seeds, are extrapolated to the DC (Drift Chamber) and then all the DC hits belonging to the track will be collected
- Track fitting algorithm
  - Adopt experiment-independent track fitting toolkit Genfit which is also being used by many other experiments like BelleII, PANDA, COMET etc.
- Completed work includes
  - Porting the BelleII code to the CEPCSW
  - Implementation of geometry and magnetic field
  - Optimization of hit selection criteria
  - Performance validation with simulated data

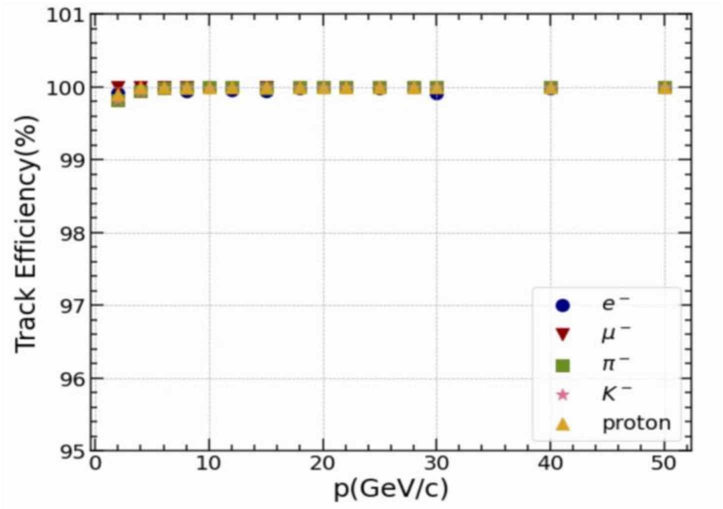


Progress in Drift Chamber Software (2)

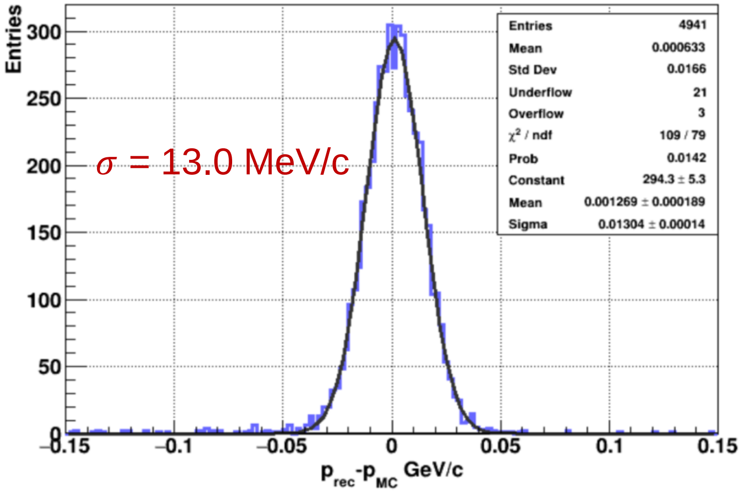
- Tracking Efficiency = N_1/N_2
  - N_1 is the number of tracks satisfying:
    - $chi^2 < 400$
    - $N_{DC\ hits\ on\ track} > 6$
  - N_2 is the numbre of track seeds reconstructed in silicon vertex detector
- The efficiency is close to 99% (inelastic scattering is turn off)

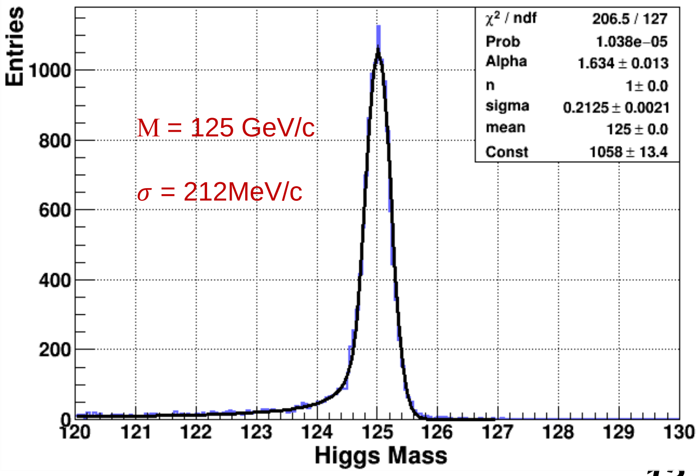
- Momentum resolution of 13.0 MeV/c is consistent with the spatial resolusion, 110 μm, in the simulation
- Impact parameter
  - $\sigma_{d0} = 3.4\ \mu m$ *with* $p_T = 10GeV/c$
  - Consistent with analytics calculation
- Physics event reconstruction of H→ $\mu^+\mu^-$

Tracking efficiency v.s. momentum

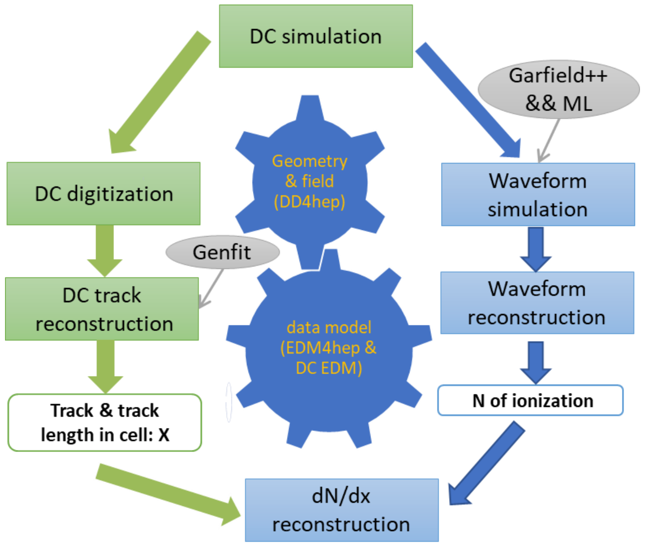


Momentum of 10 GeV/c μ⁻



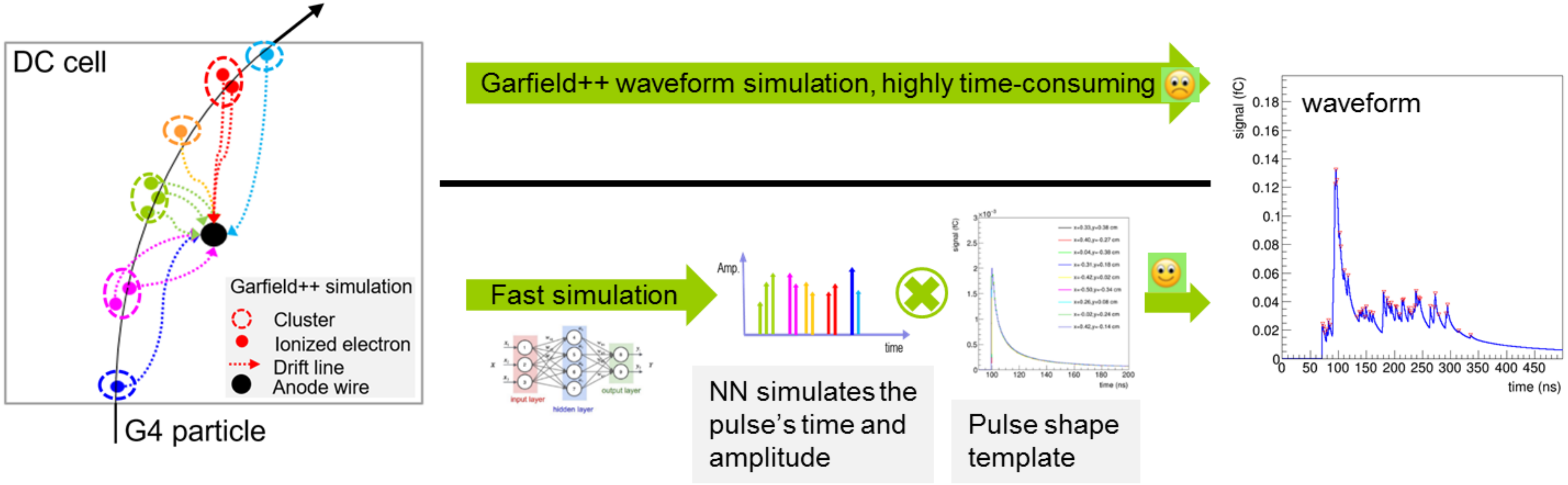


Plan for Drift Chamber Software



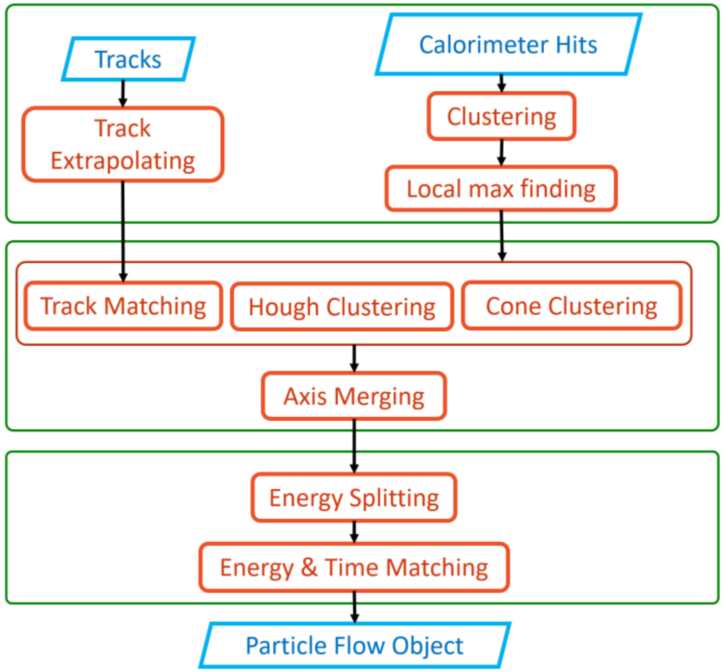
Drift chamber simulation and reconstruction flow

- Use neural network to simulate the time and amplitude of the pulse of an ionized electron
- Add waveform reconstruction algorithm into the CEPCSW
- Develop dN/dx reconstruction algorithm
- Improve tracking performance by using multiple time information obtained from the waveform

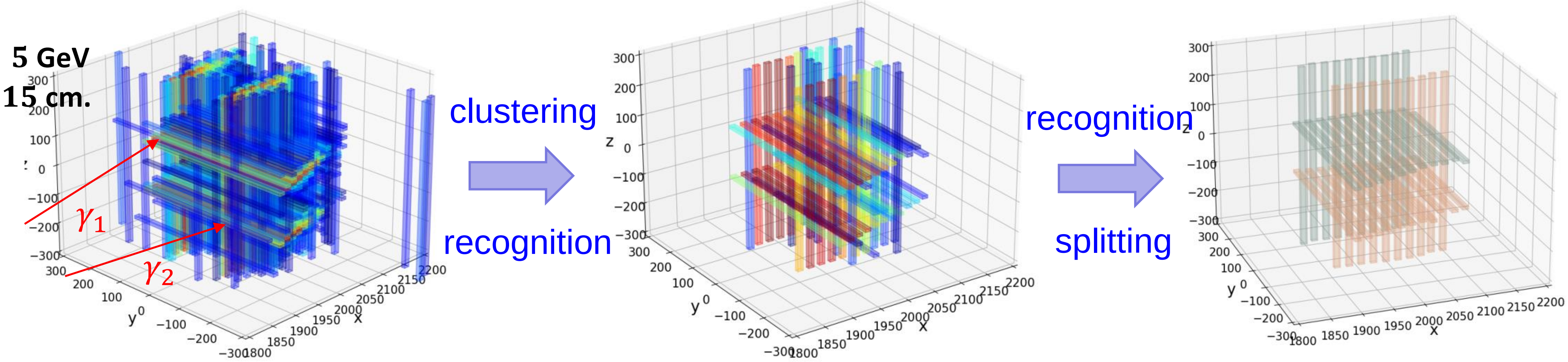


Progress in Crystal-bar ECAL Software (1)

- Software developers worked closely with the detector group
- A prototype of reconstruction algorithm was developed, which mainly consists of 4 functional modules
  - Clustering
  - shower recognition
  - splitting of overlap showers
  - ambiguity removal



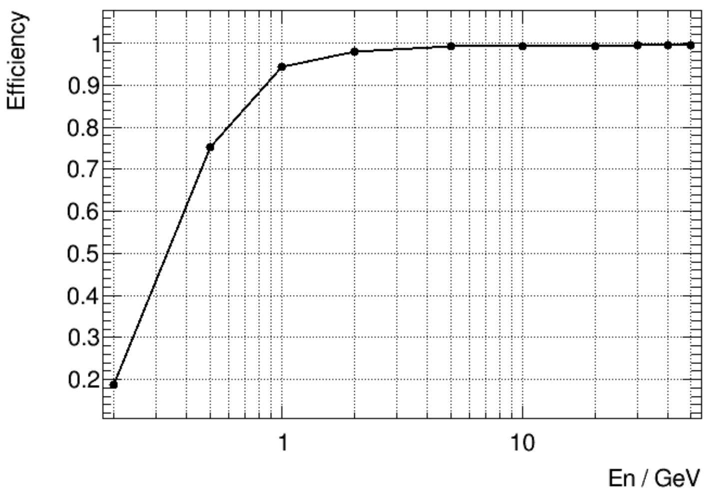
2 photons, $E_\gamma = 5$ GeV
Distance = 15×15 cm.



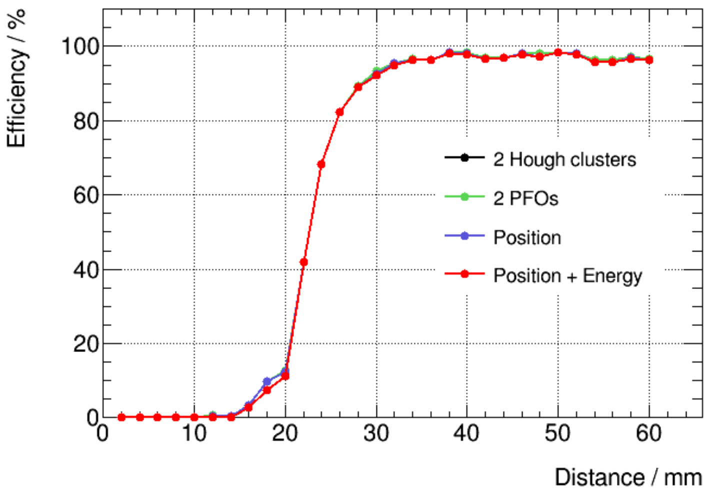
Progress in Crystal-bar ECAL Software (2)

- Reconstruction performance
  - Photon recognition efficiency is ~100% for photons with $E > 1GeV$,
  - Separation efficiency is ~95% for $\gamma - \gamma$ and $\gamma - \pi$ when distance $> 30\ mm$
- The algorithm is still being optimized for better performance

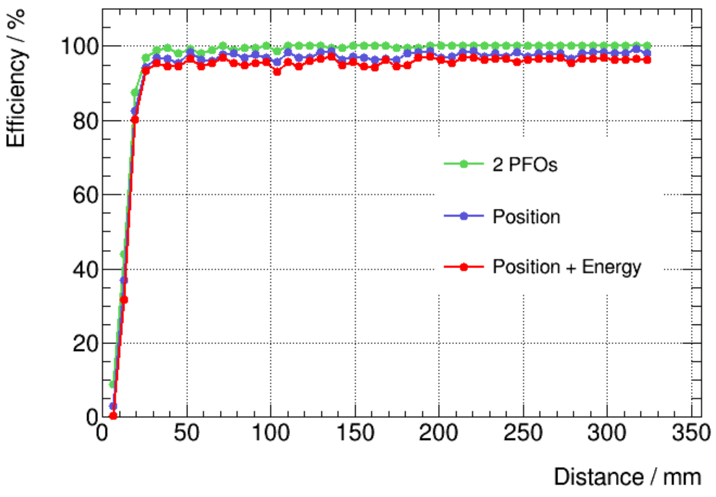
γ recognition efficiency



$\gamma - \gamma$ separation efficiency

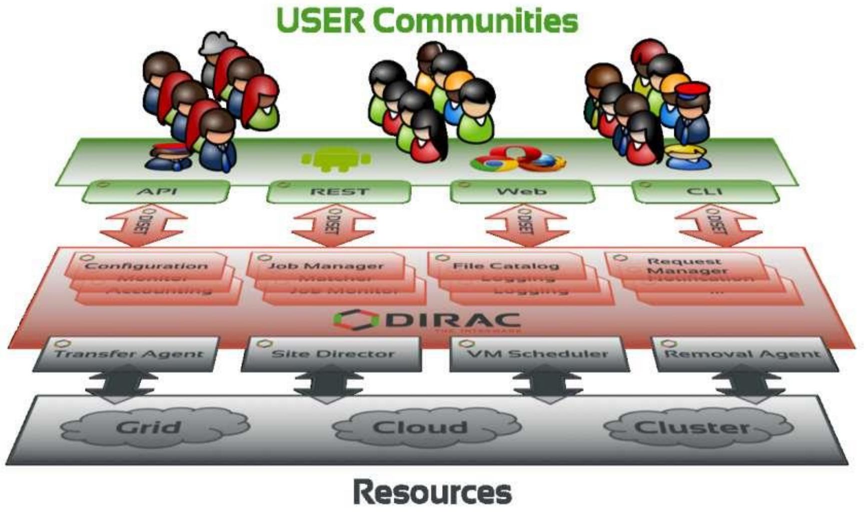


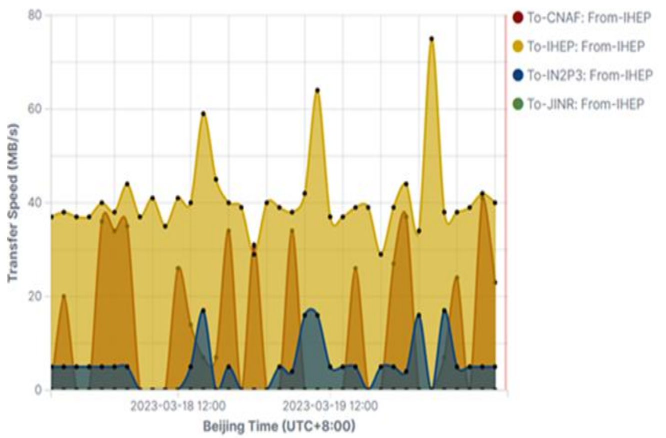
$\gamma - \pi$ separation efficiency

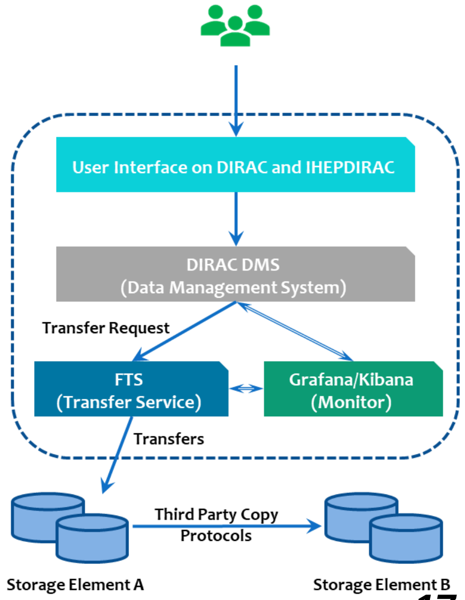


Progress in CEPC Computing Platform

- CEPC has established a distributed computing platform for detector R&D
  - DIRAC as middleware , also using VOMS, FTS, CVMFS, StoRM, EOS, WLCG middleware
  - IHEPDIRAC contains tools developed for CEPC applications
- There are about 4,600 CPU cores in this platform
  - IHEP holds 2,000 dedicated cores including 640 cores shared with ILC
  - Other sites contribute the rest 2600 cores
- The prototypes for workload management and data management were developed
- Regular TPC test system was established and monitoring was implemented with ElasticResearch+Kibana/Grafana



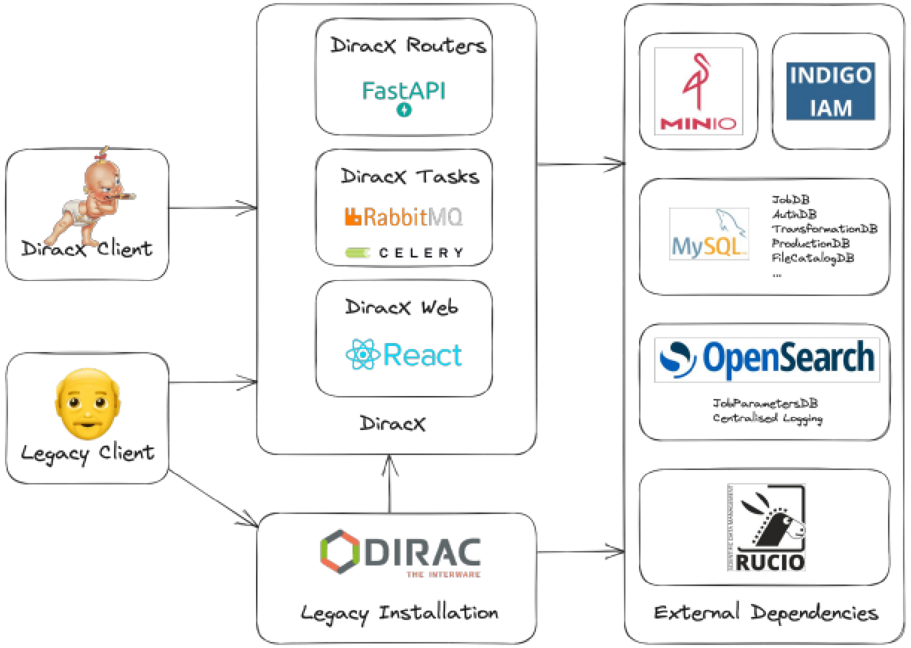




Plan for CEPC Computing Platform

- Complete the implementation of token-based authorization and authentication infrastructure
- Migration to DiracX, which is the next incarnation of Dirac re-implemented with emerging IT technologies
- Evaluate the "Data Lake" model with Rucio and XCache for smart data access and management

ARCHITECTURE



- Integration with HPC resource and adding support for machine learning activities

International Collaboration

- IHEP and SDU are non-EU members of Key4hep project (AIDAinnova WP12)
  - Bi-weekly Key4hep/Edm4hep meetings and WP12 monthly meeting
- Collaboration with IDEA detector group in R&D of drift chamber
  - Weekly discussion on cluster counting method
  - Further collaboration on tracking is being discussed
- Collaboration with the ACTS group established this year
  - Adding precise simulation of drift chamber into the Open Data Detector
  - Implementation of tracking in drift chamber with ACTS
- IHEP is an official member of the DIRAC consortium
  - Weekly DIRAC development meeting and DIRAC workshop

Summary

- The CEPCSW is being developed based on Key4hep
- Significant progress has been made since the last IAC meeting
  - Core software: event data model, simulation framework and RDataFrame-based analysis framework
  - Detector software: silicon tracker software, drift chamber software and crystal-bar ECAL software
  - CEPC distributed computing platform
- International collaboration and cooperation are being promoted with HEP R&D teams

Thank You！

谢谢