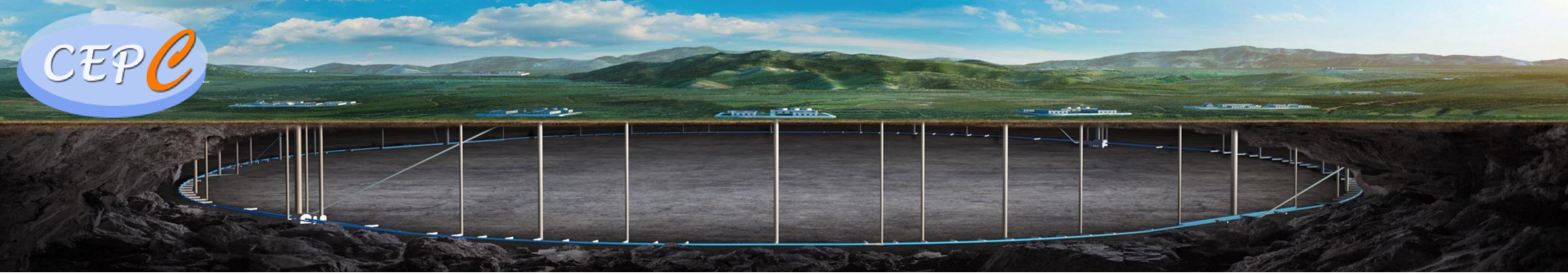


CEPC Software and Computing

Weidong Li

IHEP, CAS

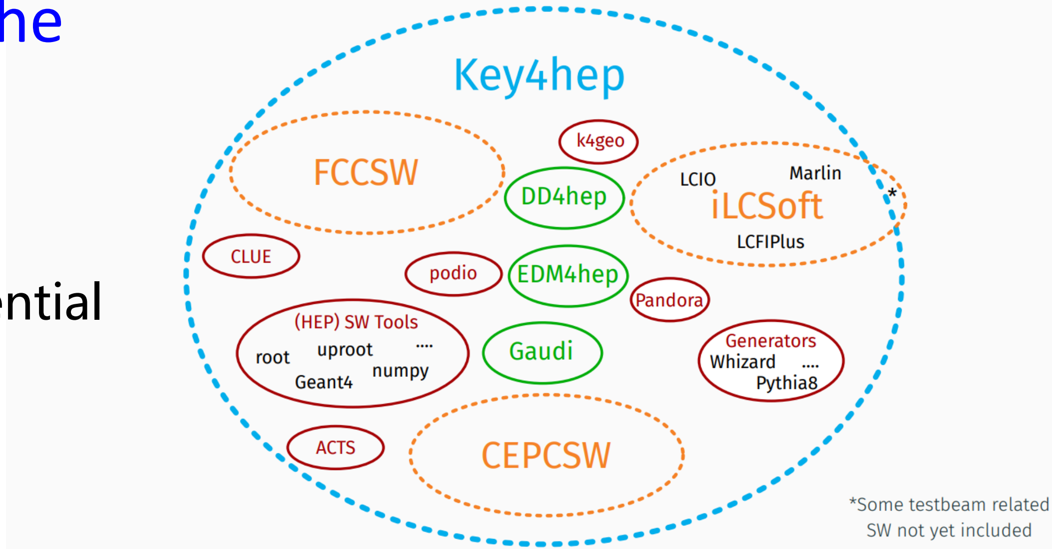
中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

Contents

1. Introduction
2. Requirements
3. Technical Options
   - Technical challenges
   - Technology survey and our choices
4. Software
   - Implementation
   - Performance
5. R&D Activities
6. Computing
7. Summary

Introduction

- The development of CEPC software started with the iLCSoft
  - Developed CEPC components for simulation and reconstruction
  - Produced M.C. data for detector design and physics potential studies
  - CEPC CDR studies done with the iLCSoft



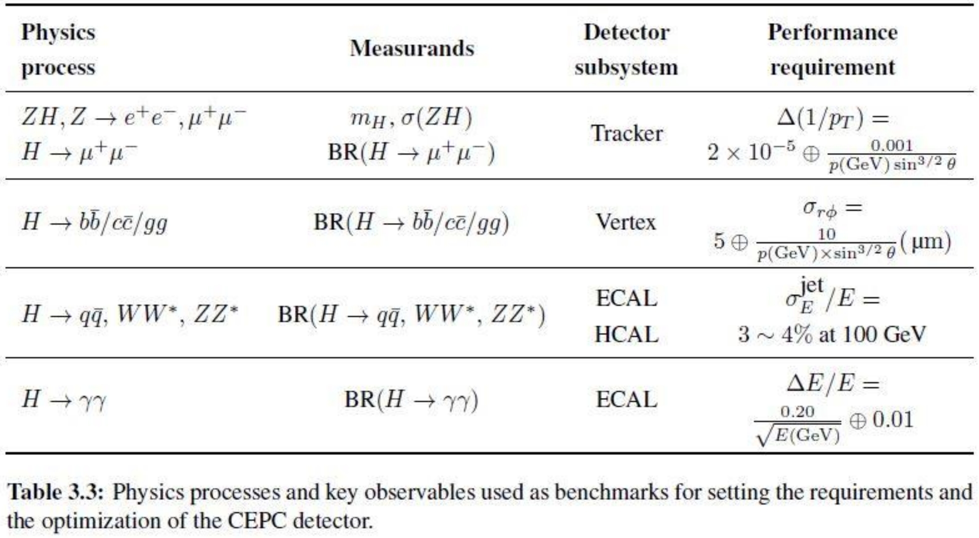
*Some testbeam related SW not yet included

Figure from F. Gaede

- 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 experiment
- Most of our efforts are put on CEPC experiment specific software and workflows

Requirements

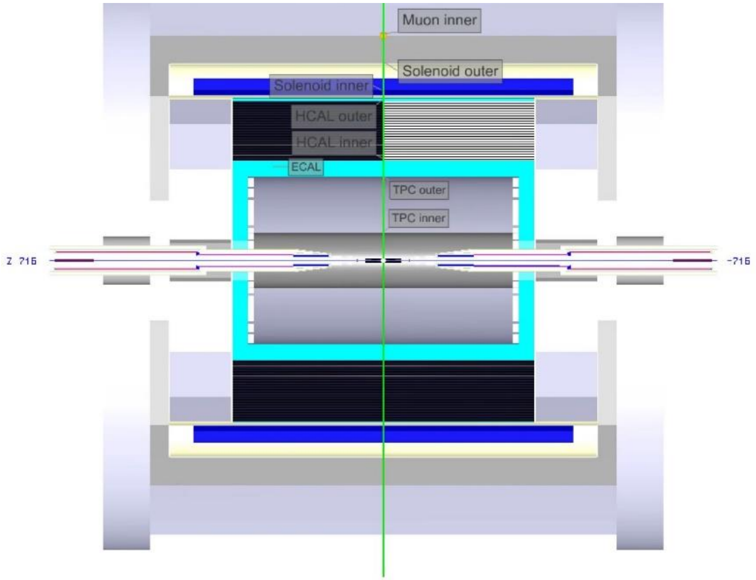
- Physics requirements:
  - high track efficiency (~100%), momentum resolution (<0.1%), PID (2σ p/K separation at P < ~ 20 GeV/c), BMR better than 4% etc.
- The CEPC Reference Detector
  - VTX: Vertex Detector
  - ITK/OTK: Inner Silicon Tracker/Outer Silicon Tracker
  - TPC: Time Projection Chamber
  - ECAL: Crystal-bar Calorimeter
  - HCAL: Scintillator Glass AHCAL
  - Muon: Muon Detector
- Both detector design and physics potential studies need software



Physics process	Measurands	Detector subsystem	Performance requirement
$ZH, Z \to e^+e^-, \mu^+\mu^-$ $H \to \mu^+\mu^-$	$m_H, \sigma(ZH)$ $\mathrm{BR}(H \to \mu^+\mu^-)$	Tracker	$\Delta(1/p_T) =$ $2 \times 10^{-5} \oplus \frac{0.001}{p(\mathrm{GeV}) \sin^{3/2}\theta}$
$H \to b\bar{b}/c\bar{c}/gg$	$\mathrm{BR}(H \to b\bar{b}/c\bar{c}/gg)$	Vertex	$\sigma_{r\phi} =$ $5 \oplus \frac{10}{p(\mathrm{GeV}) \times \sin^{3/2}\theta} (\mu\mathrm{m})$
$H \to q\bar{q}, WW^*, ZZ^*$	$\mathrm{BR}(H \to q\bar{q}, WW^*, ZZ^*)$	ECAL HCAL	$\sigma_E^{\mathrm{jet}}/E =$ $3 \sim 4\%$ at 100 GeV
$H \to \gamma\gamma$	$\mathrm{BR}(H \to \gamma\gamma)$	ECAL	$\Delta E/E =$ $\frac{0.20}{\sqrt{E(\mathrm{GeV})}} \oplus 0.01$

Table 3.3: Physics processes and key observables used as benchmarks for setting the requirements and the optimization of the CEPC detector.

CEPC CDR



Technical Options

Main technical challenges

- How to develop large scale software which can be run on diverse hardware CPU/GPU/FPGA/other accelerators
- How to take advantage of emerging technologies such as AI and Quantum Computing technologies to cope with the tasks that involve complicated data processing

Technology survey and our choices

- Software framework: GAUDI (LHCb/ATLAS/FCC)
- Simulation: Geant 4 (HEP experiments)
- Event Data Model: EDM4hep (ILC/CLIC/FCC)
- Detector Description: DD4hep (ILC/ATLAS/CMS/LHCb)
- Analysis: Root RDataFrame (HEP experiments)
- Event visualization: Phoenix (ATLAS/Belle II/FCC)

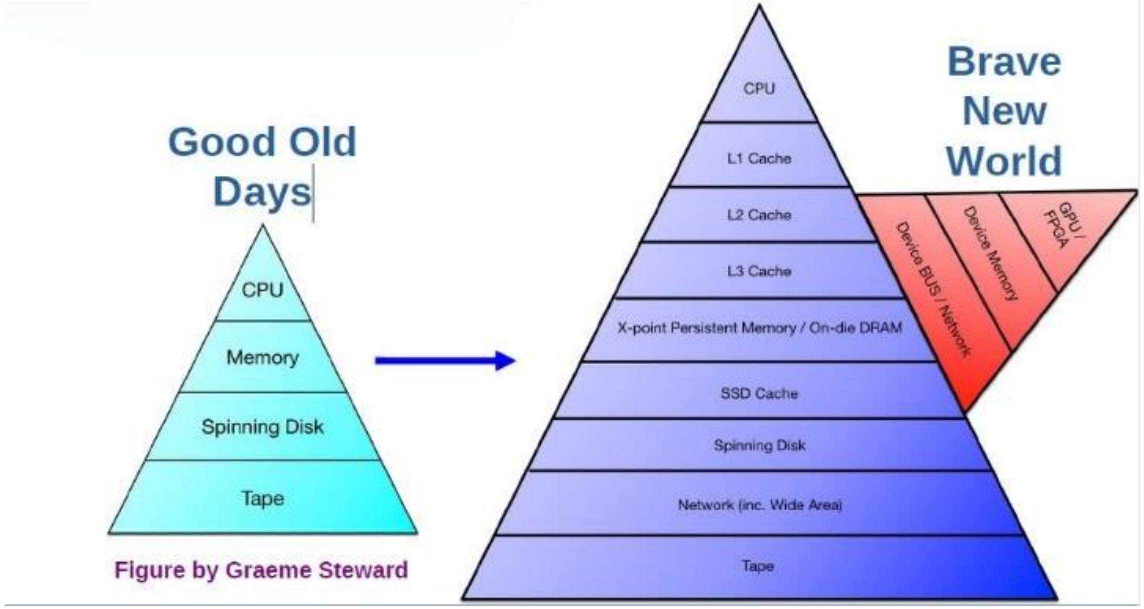
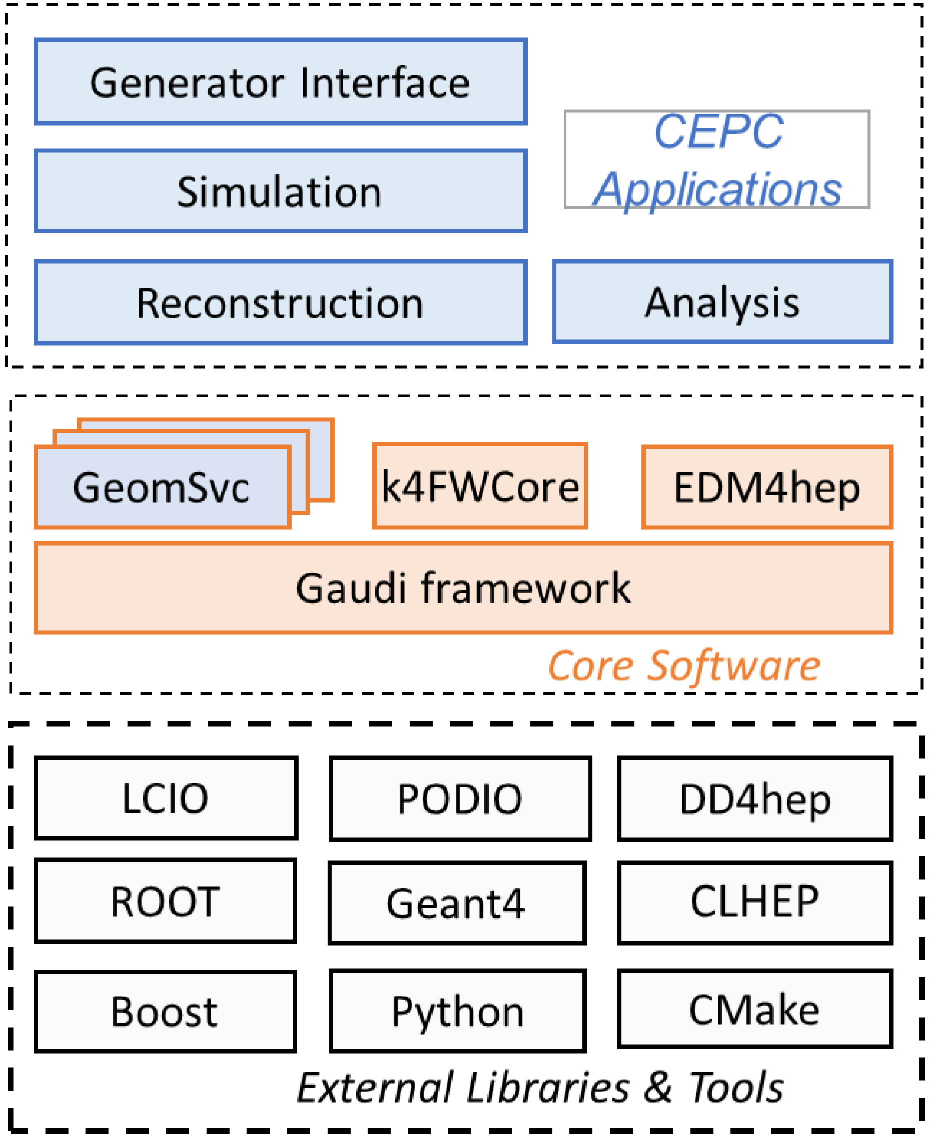


Figure by Graeme Steward

Software: Architecture of CEPCSW

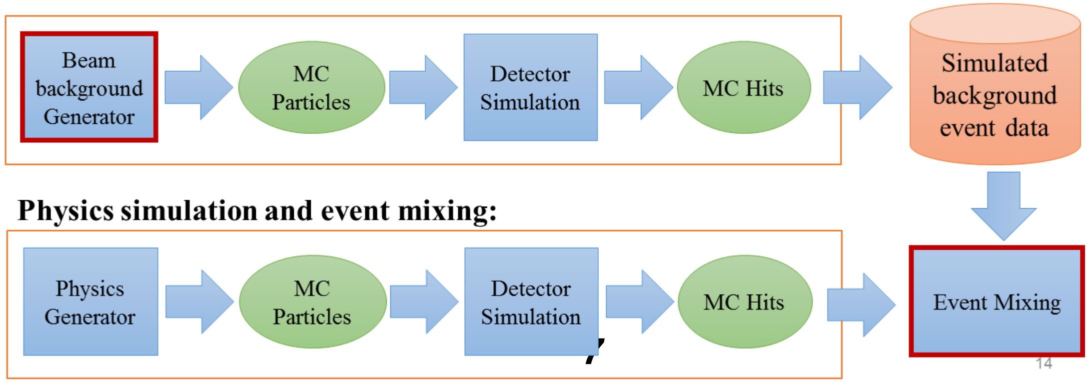
- CEPCSW is organized as a multi-layer structure
  - Applications: simulation, reconstruction and analysis
  - Core software
  - External libraries

- 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.



Software: Simulation Framework

- The simulation framework was developed, based on which the simulation chain for sub-detectors was built
- The region-based fast simulation interface was also developed to integrate different of fast simulation modules into the detector simulation
- An event mixing tool was also provided to mix different types of backgrounds with physics signals at hit level.

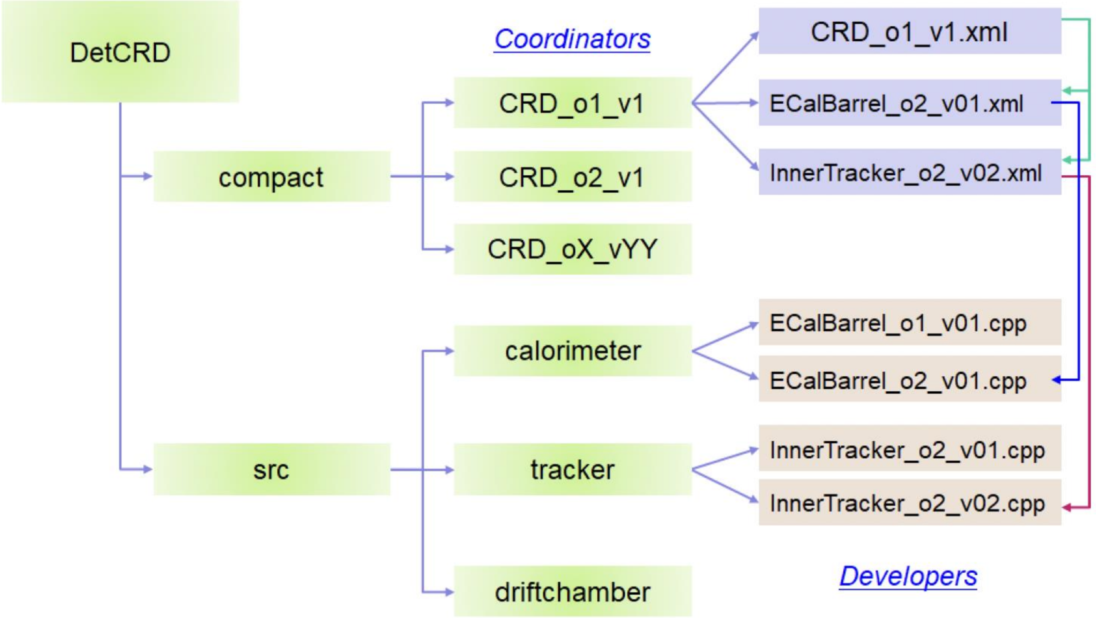


Beam background Generator → MC Particles → Detector Simulation → MC Hits → Simulated background event data

Physics simulation and event mixing:

Physics Generator → MC Particles → Detector Simulation → MC Hits → Event Mixing

- DD4hep was adopted to provide a full detector description, which was generated from a single source
- Different detector design options are managed in the Git repository and a simulation job can be easily configured in runtime



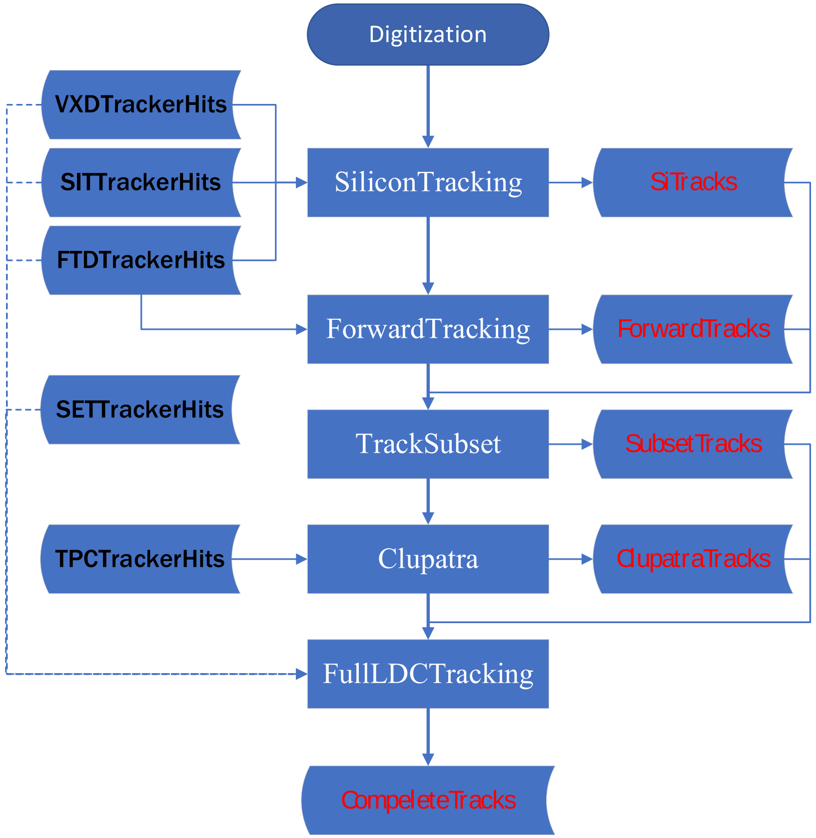
DetCRD
compact
CRD_o1_v1
CRD_o2_v1
CRD_oX_vYY
src
calorimeter
tracker
driftchamber

Coordinators
CRD_o1_v1.xml
ECalBarrel_o2_v01.xml
InnerTracker_o2_v02.xml

ECalBarrel_o1_v01.cpp
ECalBarrel_o2_v01.cpp
InnerTracker_o2_v01.cpp
InnerTracker_o2_v02.cpp
Developers

Software: Tracking (1)

- Track reconstruct tracks in VTX, ITK, TPC, OTK
  - Simple digitization (smearing) is performed on SimTrackerHit→TrackerHit
  - Standalone tracking performs first if possible, and then merging the found track fragments
- Algorithm and tool implementation
  - Based on ILD tracking algorithms imported from Marlin
    - SiliconTracking for VTX, ITK
    - ForwardTracking for ITK
    - TrackSubset to select better from VTX, ITK
    - Clupatra for TPC
    - FullILDTracking to merge silicon and TPC parts of same track
  - Development and optimization
    - Track fitting Tool
- Tracking results
  - edm4hep::Track saves edm4hep::TrackState at location
    - IP, first hit, last hit, calorimeter face
    - track parameters (ω, ϕ_0, $\tan\lambda$, d_0, z_0)
    - reference for PFA
    - linked with PID



Digitization

VXDTrackerHits

SITTrackerHits

FTDTrackerHits

SETTrackerHits

TPCTrackerHits

SiliconTracking

SiTracks

ForwardTracking

ForwardTracks

TrackSubset

SubsetTracks

Clupatra

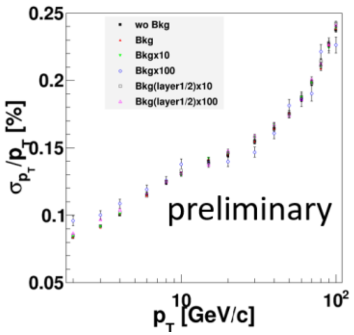
ClupatraTracks

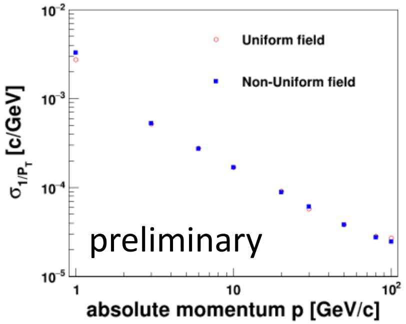
FullLDCTracking

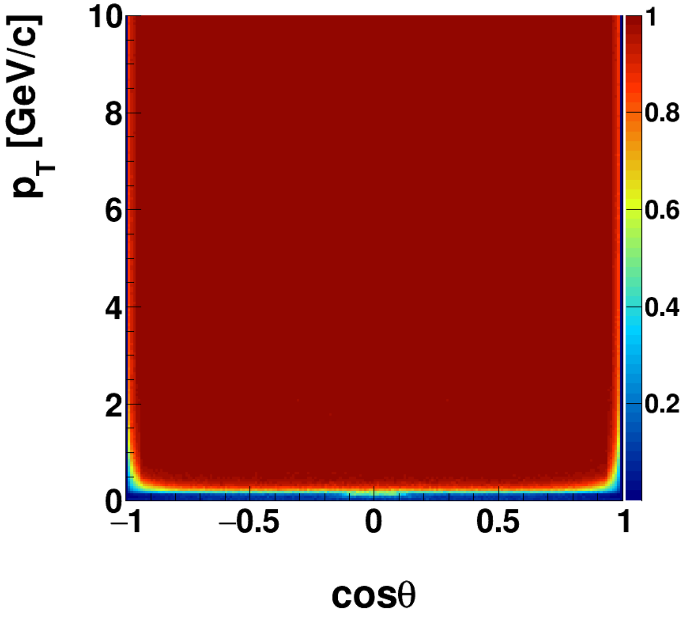
CompeleteTracks

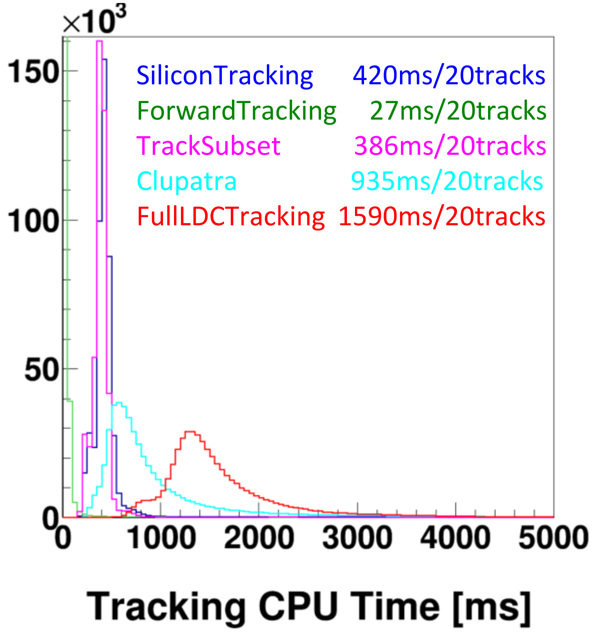
Software: Tracking (2)

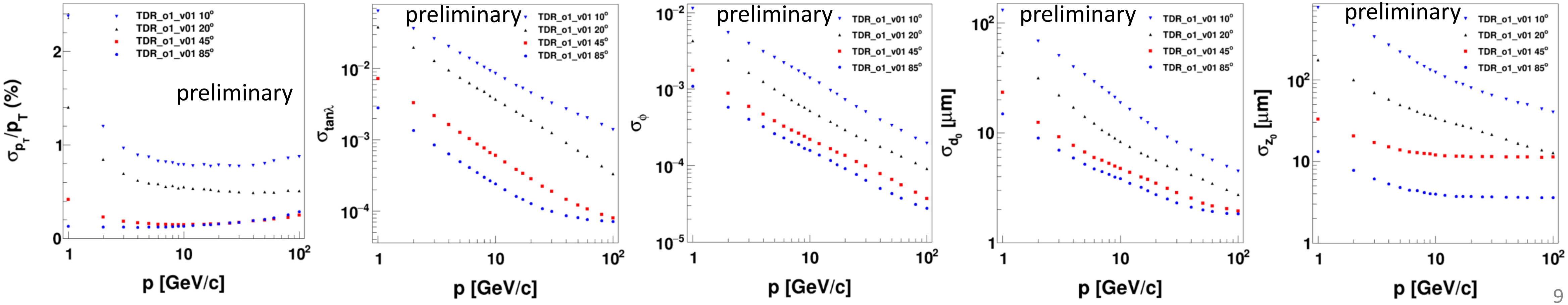
- MC truth used for performance studies
  - MCParticle↔SimTrackerHit↔TrackerHit↔Track
- Track parameter resolution
  - Random backgrounds & non-uniform field
- Computation performance







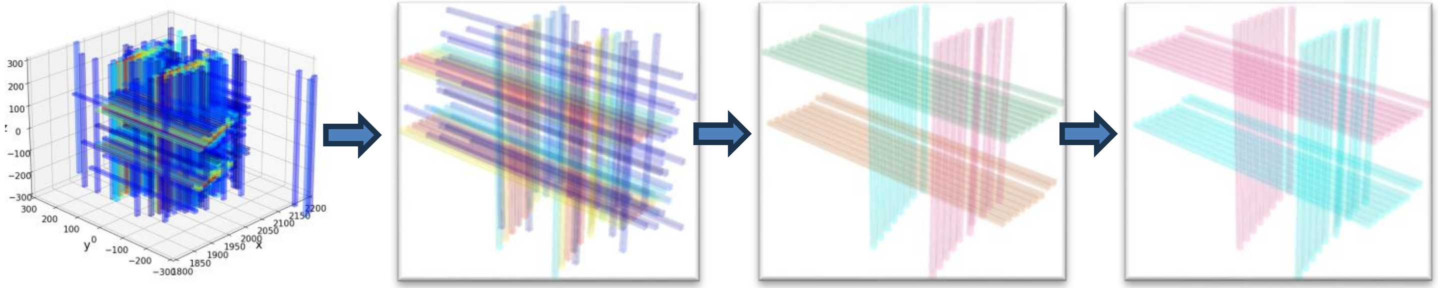


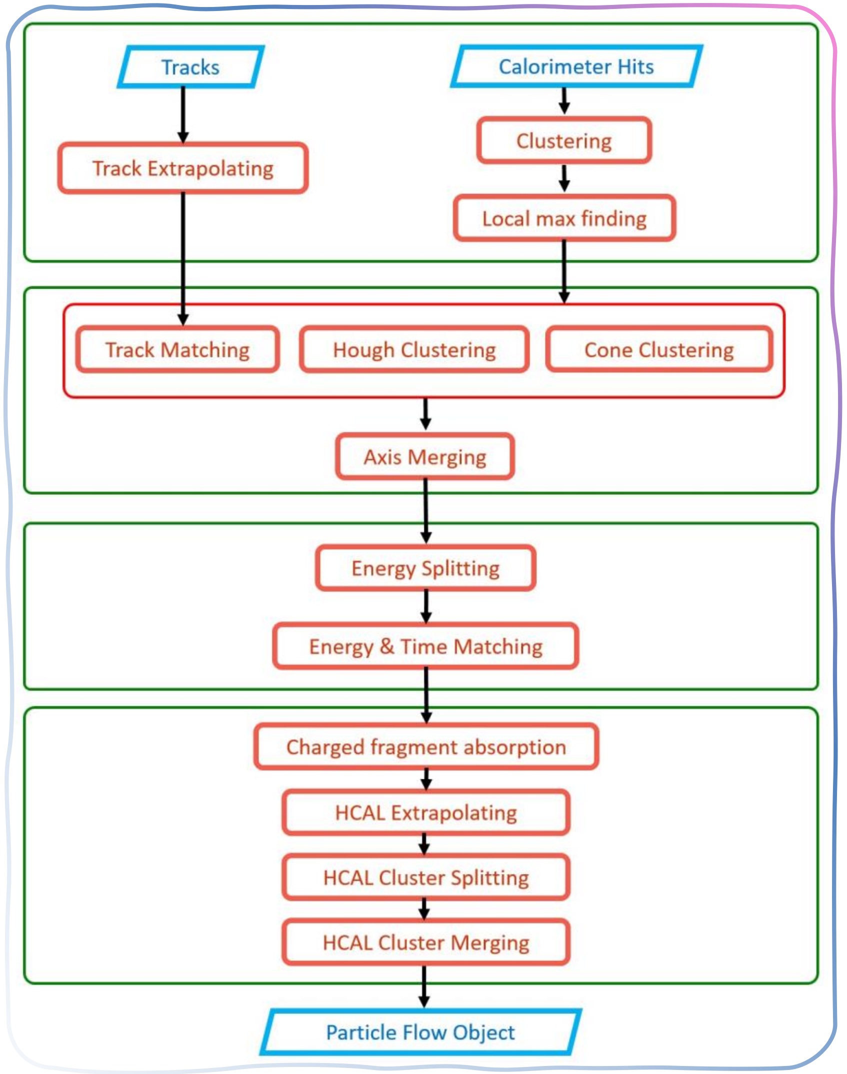


Software: Crystal-bar ECAL (1)

- Development of a new particle flow algorithm for crystal-bar ECAL
  - Clustering
  - Shower recognition
  - Energy splitting and ambiguity removal
  - Re-clustering
- Tuning algorithm parameters for optimized performance

Event display: 2 photons, $E_\gamma = 5$ GeV, distance = 15×15 cm.





Tracks

Track Extrapolating

Calorimeter Hits

Clustering

Local max finding

Track Matching

Hough Clustering

Cone Clustering

Axis Merging

Energy Splitting

Energy & Time Matching

Charged fragment absorption

HCAL Extrapolating

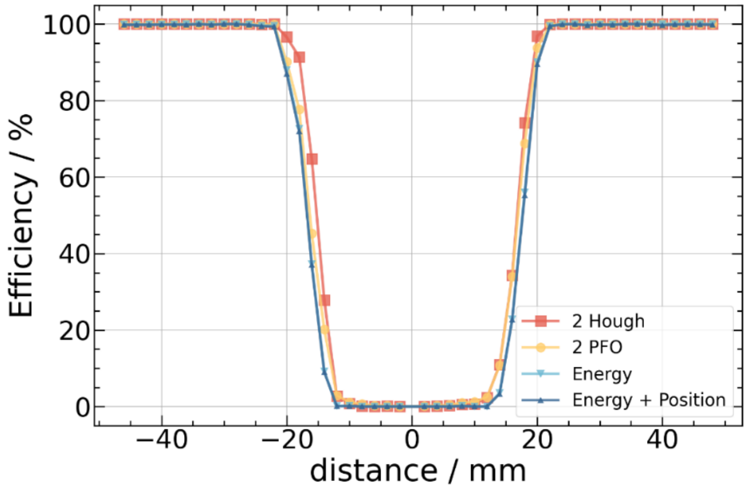
HCAL Cluster Splitting

HCAL Cluster Merging

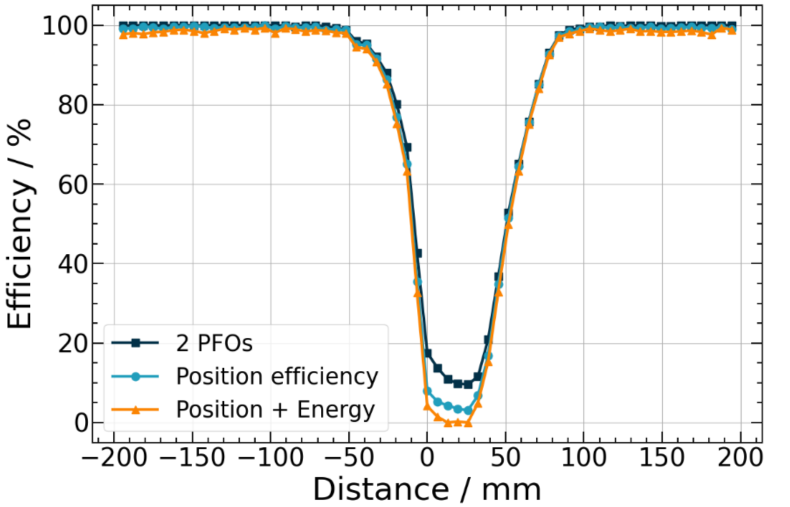
Particle Flow Object

Software: Crystal-bar ECAL (2)

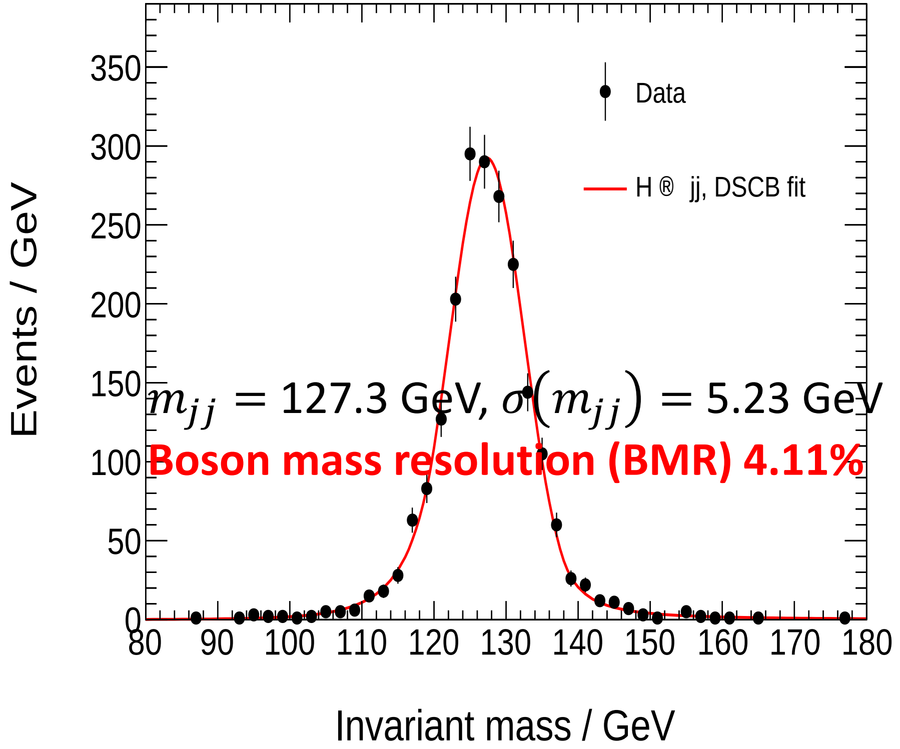
- Separation capability between two particles
  - $\gamma-\gamma$ separation: 2.2 cm @ 100% efficiency
  - $\gamma-\pi$ separation: 10 cm @ 100% efficiency
- Preliminary performance from physics process
  - $\boldsymbol{ee\rightarrow ZH\rightarrow \nu\nu gg}$ in $\sqrt{\boldsymbol{s}}=\mathbf{240}$ GeV



$\gamma-\gamma$ separation for 5 GeV photons



$\gamma-\pi$ separation for 5 GeV γ and π^-



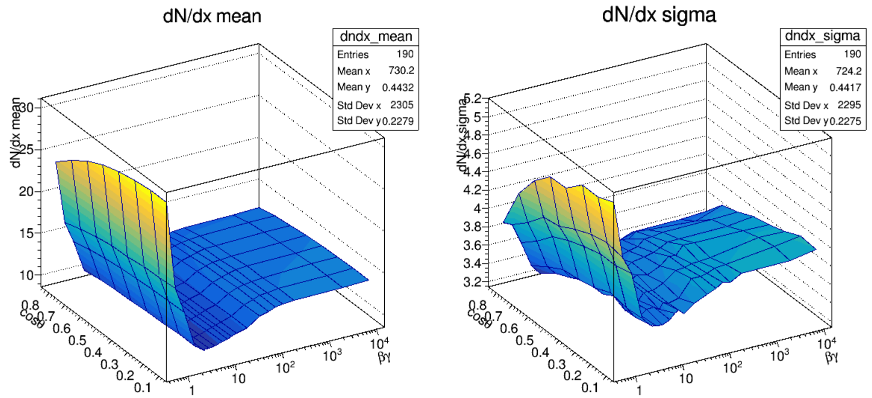
$m_{jj}=127.3$ GeV, $\sigma(m_{jj})=5.23$ GeV

Boson mass resolution (BMR) 4.11%

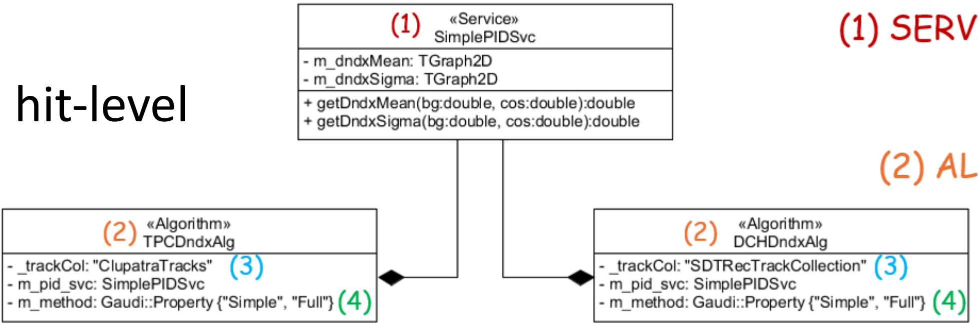
Software: Particle Identification (1)

- dN/dx in gaseous detectors (TPC and DCH)
  - Method:
    - Parameterized model to provide track-level dN/dx based on Garfield++ simulation
  - Event data model:
    - Existed model in EDM4HEP: RecDqdx
  - Reconstruction implementation:
    - Developed 2 Gaudi Algorithms for TPC and DCH dN/dx reconstruction
    - Developed a Gaudi Service for the track-level parameterization
    - An interface reserved for future hit-level reconstruction

TPC



dN/dx mean dN/dx sigma



(1) «Service» SimplePIDSvc
- m_dndxMean: TGraph2D
- m_dndxSigma: TGraph2D
+ getDndxMean(bg:double, cos:double):double
+ getDndxSigma(bg:double, cos:double):double

(2) «Algorithm» TPCDndxAlg
- _trackCol: "ClupatraTracks" (3)
- m_pid_svc: SimplePIDSvc
- m_method: Gaudi::Property {"Simple", "Full"} (4)

(2) «Algorithm» DCHDndxAlg
- _trackCol: "SDTRecTrackCollection" (3)
- m_pid_svc: SimplePIDSvc
- m_method: Gaudi::Property {"Simple", "Full"} (4)

(1) SERVICE for parameterization

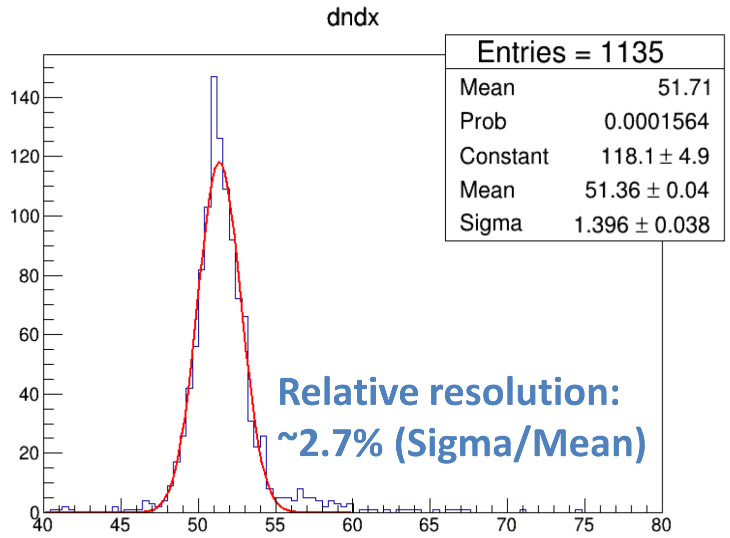
(2) ALGORITHM for dN/dx rec.

(3) Input data collection

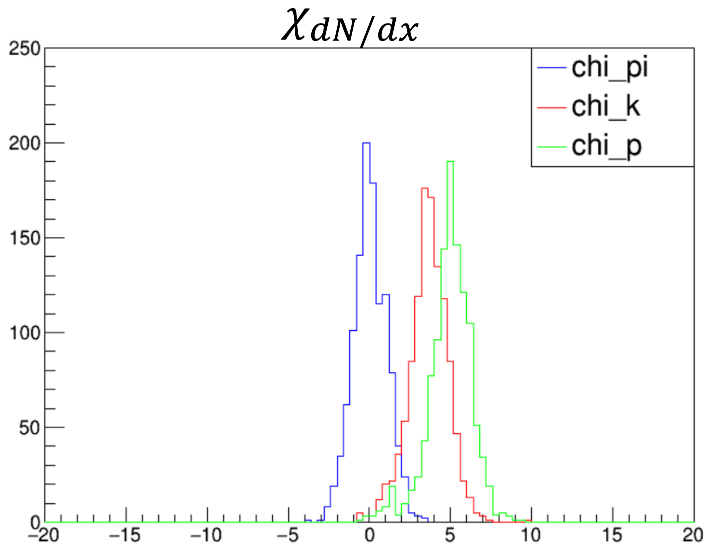
(4) Interface for hit-level rec.

Software: Particle Identification (2)

dN/dx checks w/ pions
- p = 5 GeV/c
- theta = 60 deg

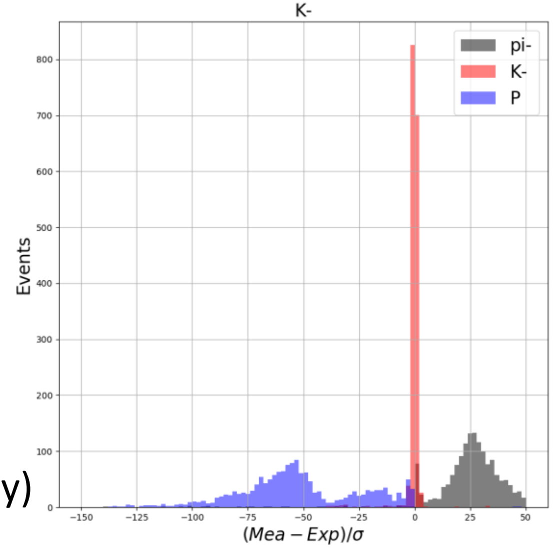


Relative resolution: ~2.7% (Sigma/Mean)



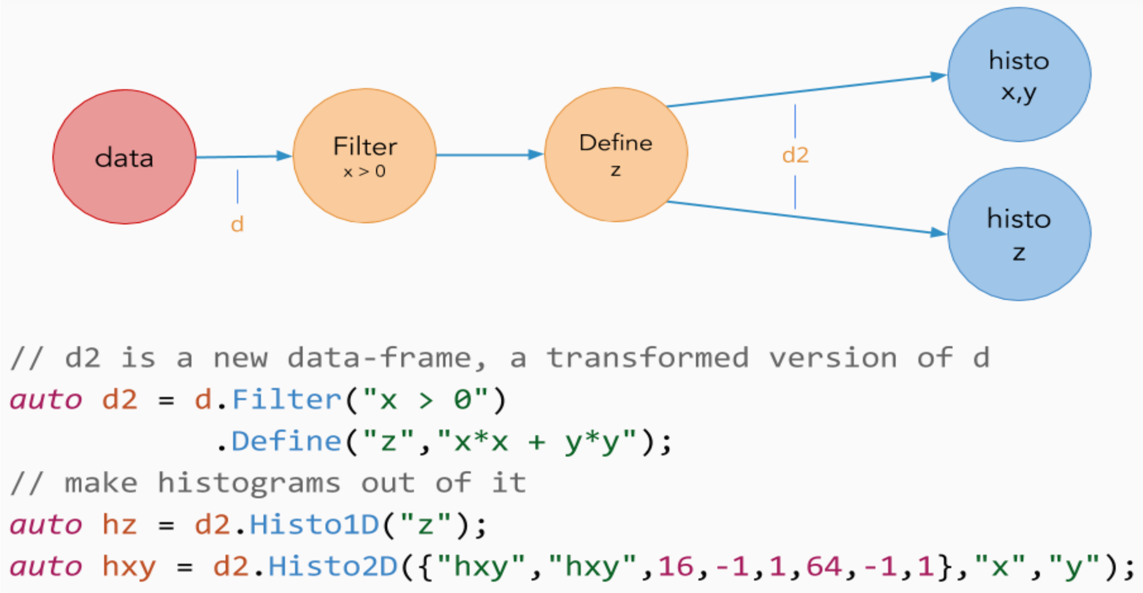
χ with pion (correct) hypothesis is centered around 0

- Time-of-Flight in outer tracker
  - Method:
    - TOF smeared with a fixed timing resolution
  - Event data model: EDM4HEP: RecTof
  - Reconstruction implementation:
    - A Gaudi Algorithm to calculate TOF
      - Geometry: R=1850mm, L/2=2350mm, and B=3T (will be updated based on newest geometry)
      - ToF model: $t_{mea} = tof_{truth} + Gaus(0, t_{bunch} = 20\text{ps}) + Gaus(0, \sigma = 50\text{ps})$

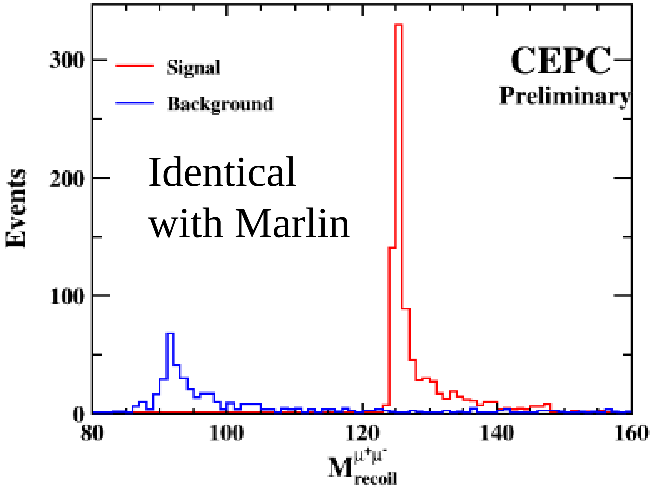


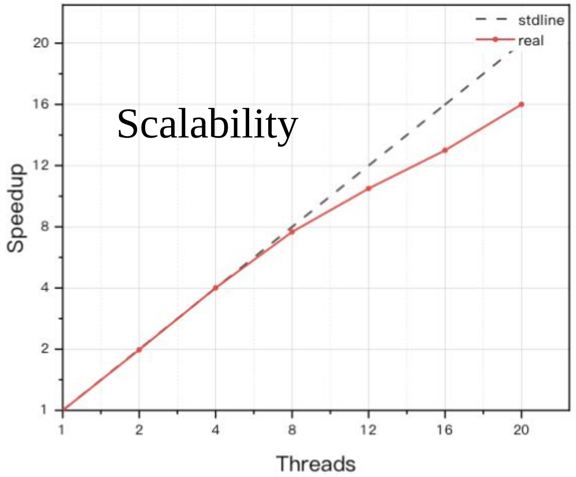
Software: Analysis Framework

- RDataFrame is a powerful tool for parallel data analysis
  - Program language: Python and C++
  - Declarative programming and parallel processing
  - Used by many experiments such as FCC-ee
- Recent developments
  - Developed common data input interfaces to support both LCIO data and EDM4hep data
  - Several algorithms were ported from Marlin
    - JetClustering , KinematicFit
  - More are being implemented
    - VertexFit, JetTagging, PID etc.
  - Performance test with two physics channels
    - e+e- -> Z(mumu)H
    - e+e- ->H(2jet) mumu



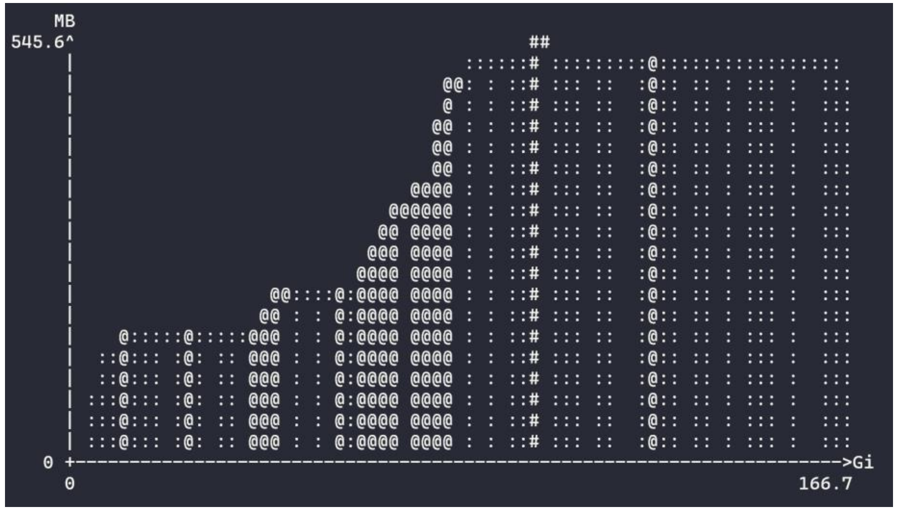
```
// d2 is a new data-frame, a transformed version of d
auto d2 = d.Filter("x > 0")
           .Define("z","x*x + y*y");
// make histograms out of it
auto hz = d2.Histo1D("z");
auto hxy = d2.Histo2D({"hxy","hxy",16,-1,1,64,-1,1},"x","y");
```





R&D: Simulation with Gaussino

- Multi-threading mandatory
  - Reduce memory footprint



```
    MB
545.6^                                          ##
     |                                    ::::::# :::::::::@:::::::::::::::::
     |                                  @@: : ::# ::: ::  :@:: :: : ::: :  :::
     |                                  @ : : ::# ::: ::  :@:: :: : ::: :  :::
     |                                 @@ : : ::# ::: ::  :@:: :: : ::: :  :::
     |                                 @@ : : ::# ::: ::  :@:: :: : ::: :  :::
     |                                 @@ : : ::# ::: ::  :@:: :: : ::: :  :::
     |                               @@@@ : : ::# ::: ::  :@:: :: : ::: :  :::
     |                             @@@@@@ : : ::# ::: ::  :@:: :: : ::: :  :::
     |                            @@ @@@@ : : ::# ::: ::  :@:: :: : ::: :  :::
     |                           @@@ @@@@ : : ::# ::: ::  :@:: :: : ::: :  :::
     |                          @@@@ @@@@ : : ::# ::: ::  :@:: :: : ::: :  :::
     |                  @@::::@:@@@@ @@@@ : : ::# ::: ::  :@:: :: : ::: :  :::
     |                 @@ : : @:@@@@ @@@@ : : ::# ::: ::  :@:: :: : ::: :  :::
     |    @:::::@:::::@@@ : : @:@@@@ @@@@ : : ::# ::: ::  :@:: :: : ::: :  :::
     |  ::@::: :@: :: @@@ : : @:@@@@ @@@@ : : ::# ::: ::  :@:: :: : ::: :  :::
     |  ::@::: :@: :: @@@ : : @:@@@@ @@@@ : : ::# ::: ::  :@:: :: : ::: :  :::
     | :::@::: :@: :: @@@ : : @:@@@@ @@@@ : : ::# ::: ::  :@:: :: : ::: :  :::
     | :::@::: :@: :: @@@ : : @:@@@@ @@@@ : : ::# ::: ::  :@:: :: : ::: :  :::
     | :::@::: :@: :: @@@ : : @:@@@@ @@@@ : : ::# ::: ::  :@:: :: : ::: :  :::
   0 +----------------------------------------------------------------------->Gi
     0                                                                   166.7
```

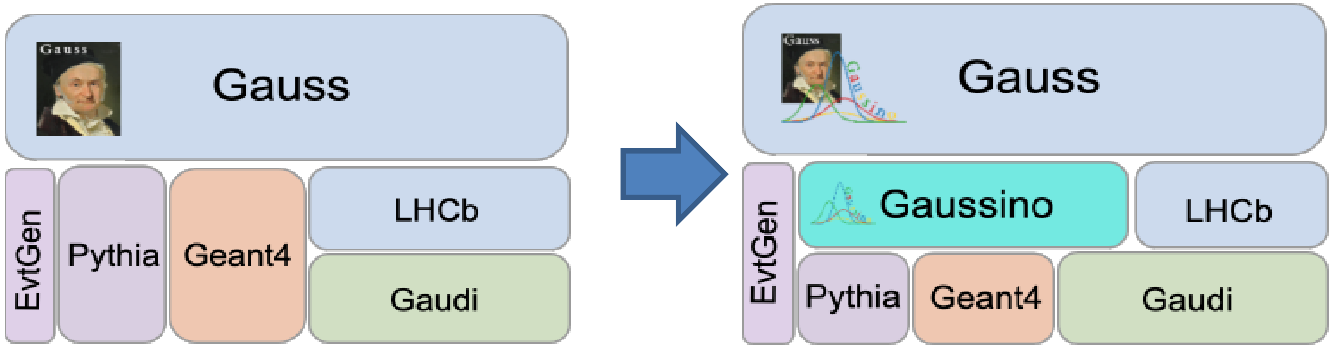
Heap memory usage (serial simulation)

Simulation setup:

- Detector: TDR_o1_v01
- Physics list: QGSP_BERT
- Generation: single muons
- N events: 100

The RSS memory is about 950MB at initialization stage

- Reuse LHCb's Gaussino framework from
- A potential solution in Key4hep. [arXiv: 2312.08152]
  - Efforts are needed to reduce the dependencies

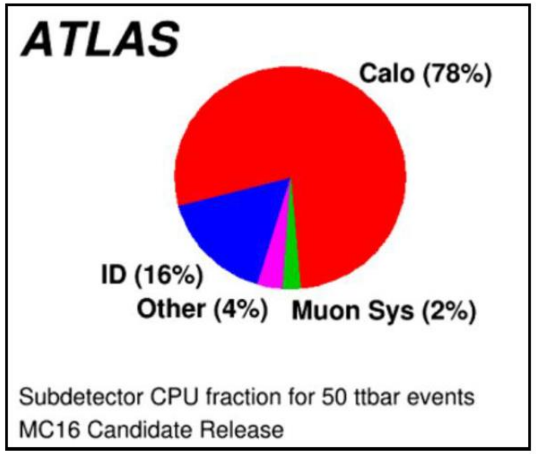


- Implement CEPC-on-Gaussino prototype
  - Cooperate with the software experts from CERN
  - Optimize the packages to reduce dependencies on the LHCb software stack.
  - Develop VXD simulation using the prototype

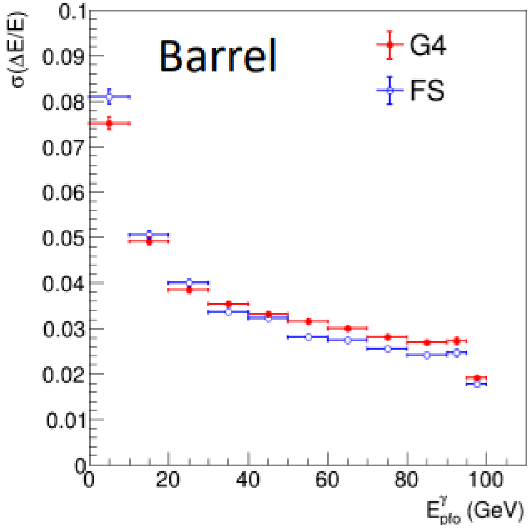
  See ICHEP 2024 Parallel Talk by Tao Lin

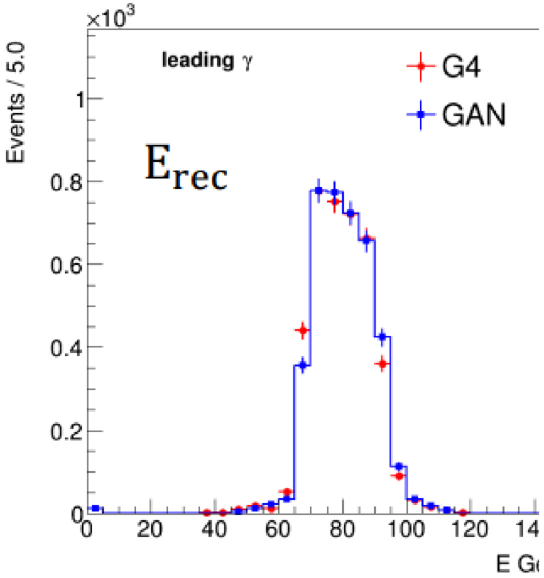
R&D: Fast Simulation

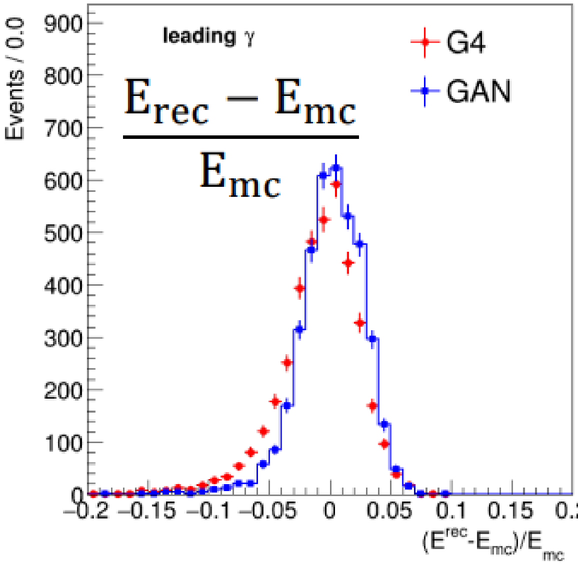
- To save CPU times, fast calorimeter simulation is on the critical path
  - Studies performed with the barrel of silicon tungsten ECAL
- Frozen shower method (GS)
  - Good agreement between FS and Geant4 simulation
  - The simulation speed was doubled
- Machine learning method based on GAN
  - For simplicity, only the simulation of central barrel was studied
  - Good agreement between GAN and Geant4 was reached
  - More than 2 magnitudes speed-up can be achieved



Subdetector CPU fraction for 50 ttbar events
MC16 Candidate Release

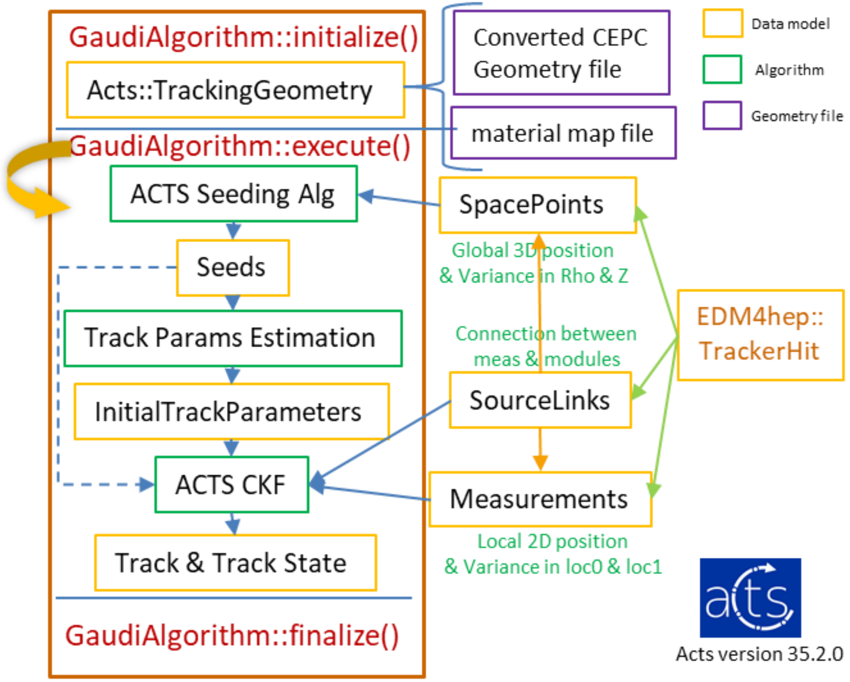






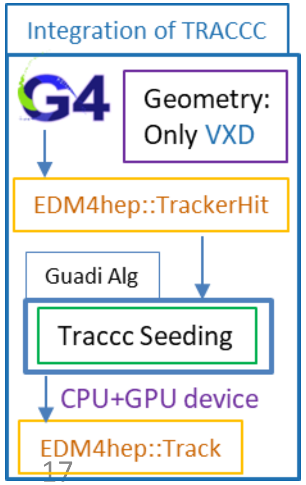
R&D: Tracking with ACTS

- Tracking implementation with ACTS for all silicon trackers (VTX+SIT+FTD)
  - Convert the CEPC’s geometry to ACTS’s geometry
  - Map Geant4 material to ACTS’s material
  - Integration of ACTS (seeding + CKF) and TRACCC (seeding)
  - Now TRACCC seeding algorithm can be run on GPU

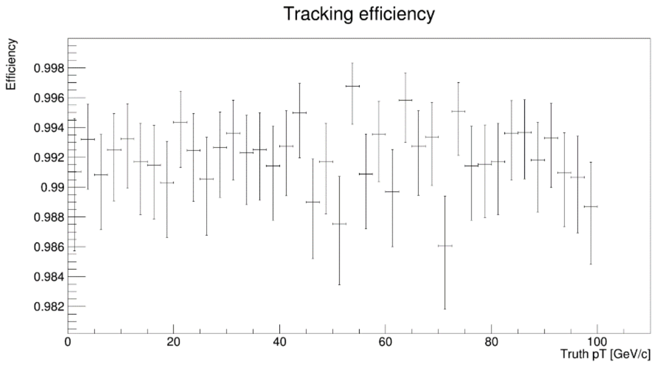

GaudiAlgorithm::initialize()
Acts::TrackingGeometry
Converted CEPC Geometry file
material map file
GaudiAlgorithm::execute()
ACTS Seeding Alg
Seeds
Track Params Estimation
InitialTrackParameters
ACTS CKF
Track & Track State
GaudiAlgorithm::finalize()
SpacePoints
Global 3D position & Variance in Rho & Z
Connection between meas & modules
SourceLinks
Measurements
Local 2D position & Variance in loc0 & loc1
EDM4hep:: TrackerHit
🟨 Data model
🟩 Algorithm
🟪 Geometry file

Acts version 35.2.0

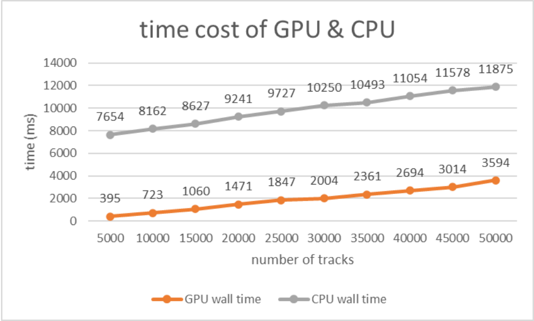
Gaudi Algorithm using ACTS reconstruction


Integration of TRACCC

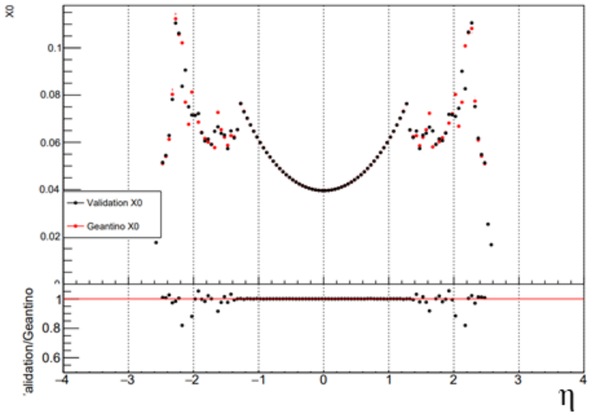
Geometry: Only VXD
EDM4hep::TrackerHit
Guadi Alg
Traccc Seeding
CPU+GPU device
EDM4hep::Track



ACTS Tracking efficiency with pT



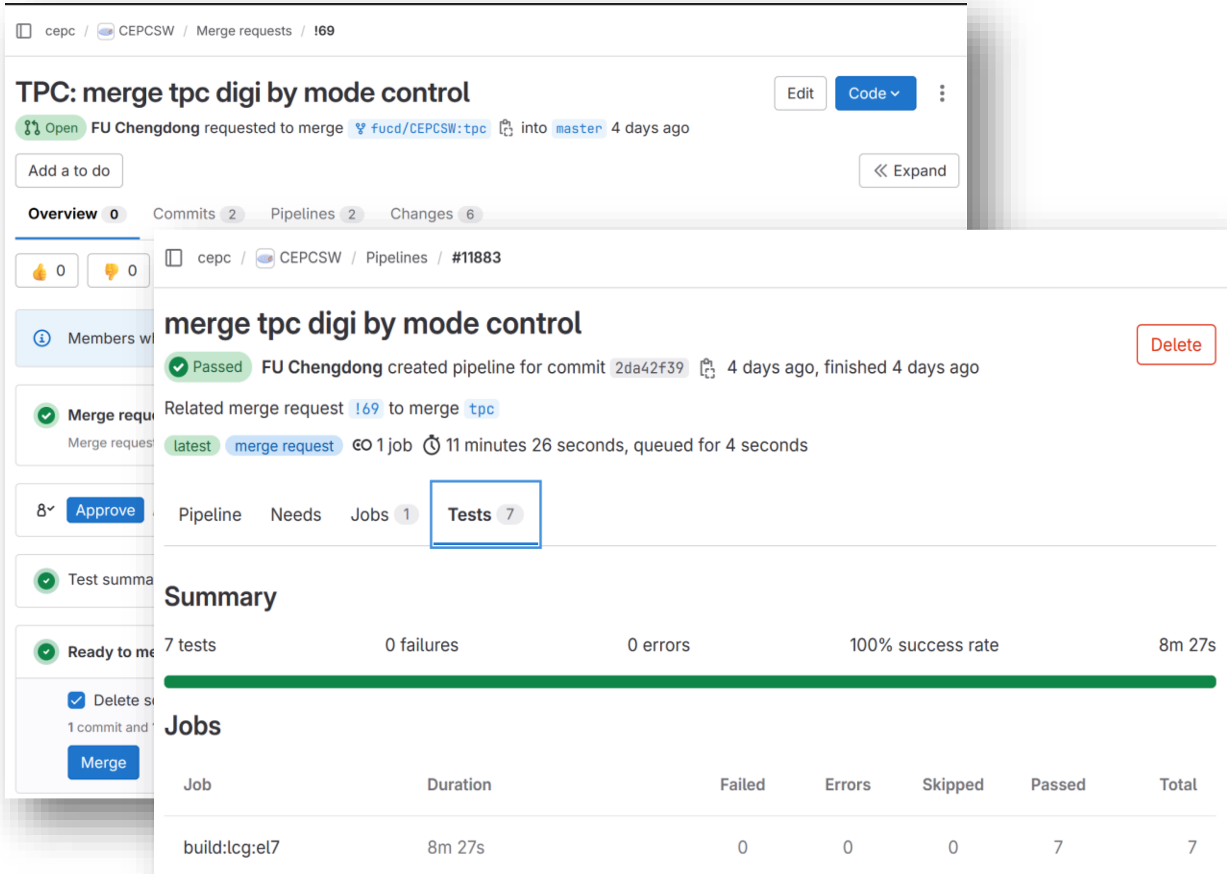
Computing evaluation of TRACCC seeding
CPU: Intel(R) Xeon(R) Silver 4214 CPU @ 2.20GHz
GPU: NVIDIA Corporation TU102GL [Quadro RTX 8000]



Validation of Material Mapping

Software Development Environment

- “Fork - Merge Request” workflow
  - – IHEP GitLab: https://code.ihep.ac.cn/cepc/CEPCSW
  - – Unit test with “ctest” in CI/CD.



cepc / CEPCSW / Merge requests / !69

TPC: merge tpc digi by mode control

Open **FU Chengdong** requested to merge `fucd/CEPCSW:tpc` into `master` 4 days ago

cepc / CEPCSW / Pipelines / #11883

merge tpc digi by mode control

Passed **FU Chengdong** created pipeline for commit `2da42f39` 4 days ago, finished 4 days ago

Related merge request `!69` to merge `tpc`

latest merge request 1 job 11 minutes 26 seconds, queued for 4 seconds

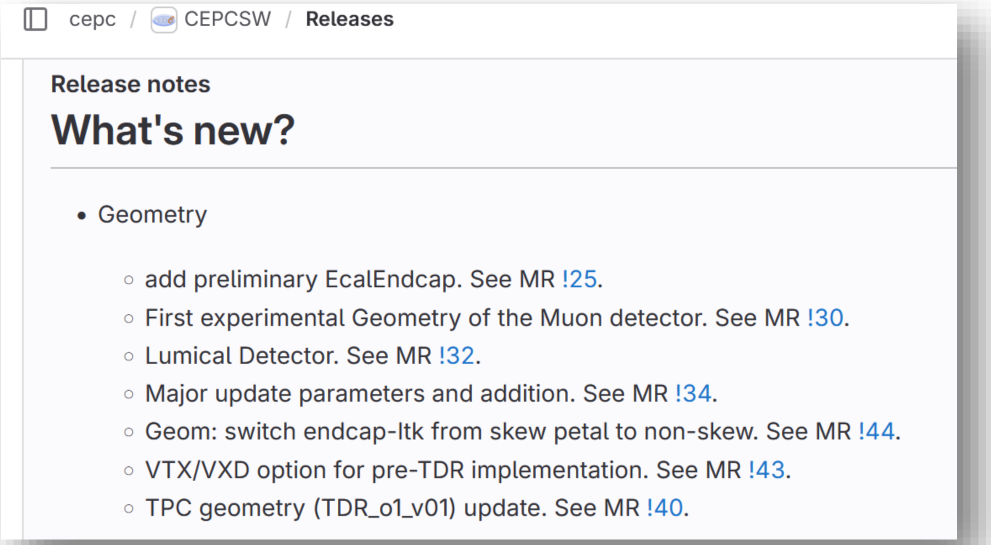
Summary

7 tests 0 failures 0 errors 100% success rate 8m 27s

Jobs

Job	Duration	Failed	Errors	Skipped	Passed	Total
build:lcg:el7	8m 27s	0	0	0	7	7

- “Release early, release often”
  - – To support fast iteration of detector design
- New “time-based” version numbering scheme for TDR: tdrYY.MM.NN
  - – E.g., tdr24.5.0: the first release in May 2024.
  - – Release notes: https://code.ihep.ac.cn/cepc/CEPCSW/-/releases



cepc / CEPCSW / **Releases**

Release notes

What's new?

- Geometry
  - add preliminary EcalEndcap. See MR !25.
  - First experimental Geometry of the Muon detector. See MR !30.
  - Lumical Detector. See MR !32.
  - Major update parameters and addition. See MR !34.
  - Geom: switch endcap-Itk from skew petal to non-skew. See MR !44.
  - VTX/VXD option for pre-TDR implementation. See MR !43.
  - TPC geometry (TDR_o1_v01) update. See MR !40.

CEPC Computing (1)

Technical challenges

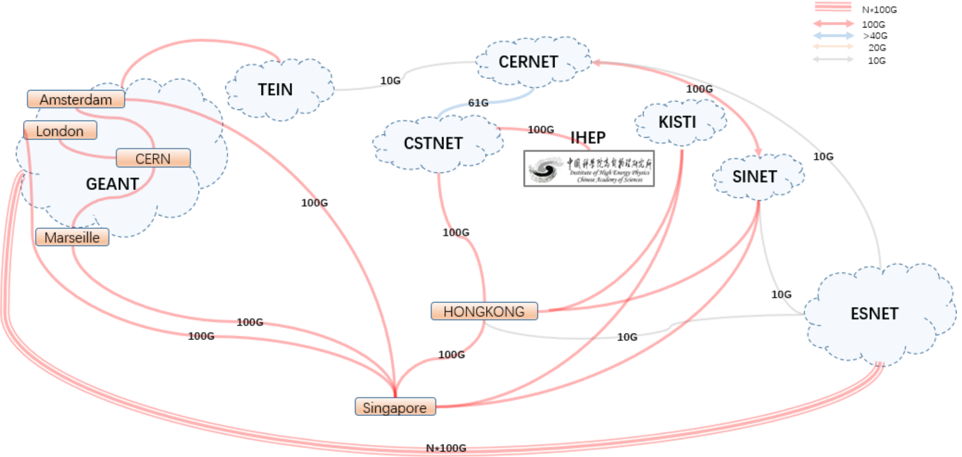
- Diverse resource requirements for both HPC and HTC due to software evolution
- Increased resource requirements caused by growing data volume
- More complicated data provisioning and workflow scheduling requirement
- Growing operation costs and long term sustainability with the system growing big and complicated

Technology survey and our choices

- Choose DIRAC as distributed computing framework
  - Can support large scale data processing activities and has been proved in big HEP experiments including LHCb and BelleII
  - Able to integrate heterogenous resource, eg grid, cluster, cloud, supercomputing center
  - DIRAC consortium allows us to share technology efforts with other experiments
- Choose Rucio as data management system
  - Provide advanced capability for large scale data management and transfer
  - developed into a common standard for scientific data management, and widely evaluated and used by many experiments
  - Rucio plays an important role in the future LHC “Data Lake” data model

CEPC Computing (2)

- Network is important for the CEPC experiment
  - To connect with other data centers in the world
- The international network links (100 Gbps)
  - Built based on the current network condition that IHEP has network connection with the main HEP data centers in the world
- Cyber security is considered on many aspects, including data, resources, network, users, …
  - developed and deployed the SOC that can handle threat discovery -> warning -> response -> defense automatically.
  - CEPC will be one of the experiments that IHEP SOC supports



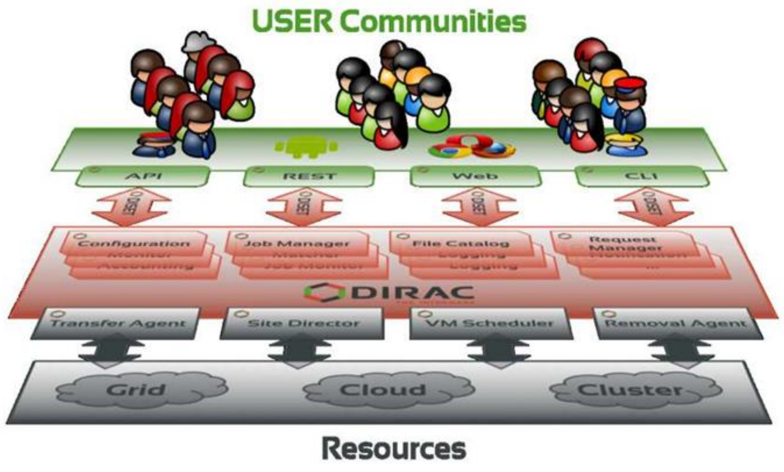
攻击时间:
2023-04-25 17:47:08

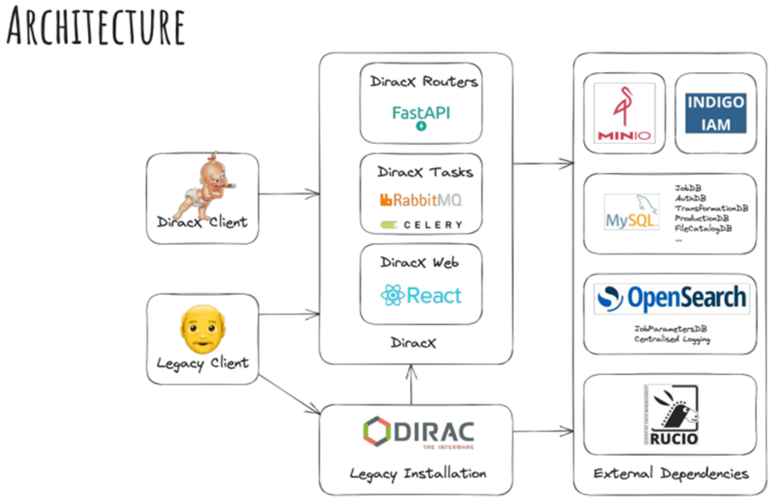
攻击源:
43.154.126.65 (已封锁)

攻击类型:
使用root24小时尝试计算平台登录失败超过5次

CEPC Computing (3)

- Computing Model: a Tier infrastructure to seamlessly integrate heterogeneous and distributed resources
  - Tier 0: IHEP serves as the primary hub for computing and data processing
  - Tier 1: Large sites act as major support centers with strong capabilities of computing and storage system for Tier 0
  - Tier 2: Smaller sites provide supplementary computing power and storage capabilities as cache layers
- A prototype platform established for CEPC R&D
  - Use DIRAC as framework, alongside VOMS, FTS, CVMFS, StoRM, EOS, and other WLCG middleware
  - Users have the capability to access these resources and share data from any locations through this platform
  - The platform encompasses approximately 4,600 CPU cores
- The platform is being improved to align with technology evolvement
  - Migrate into DIRACX - next generation distributed computing framework with advanced capabilities
  - Support token-based authorization and authentication infrastructure
  - Support high performance computing resource and ML activities
  - Use the "Data Lake" model with Rucio and XCache for robust and efficient data management and access





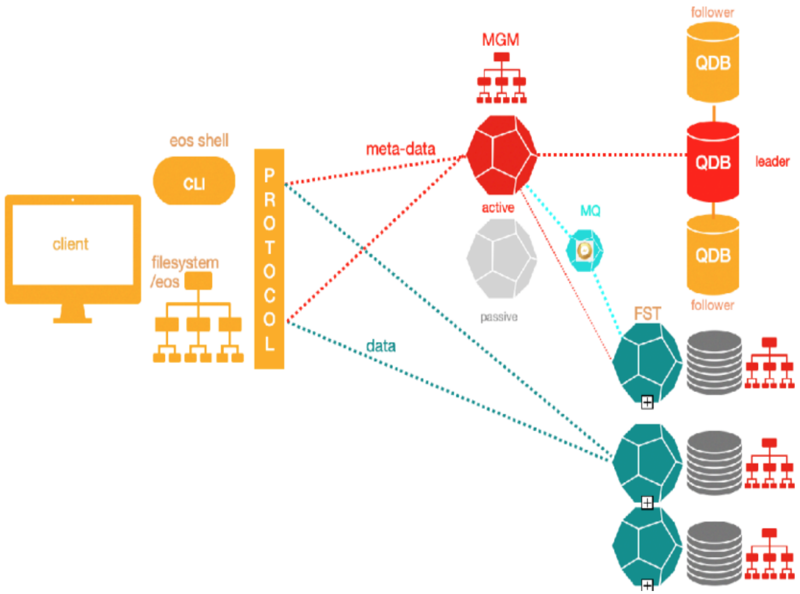
CEPC Computing (4)

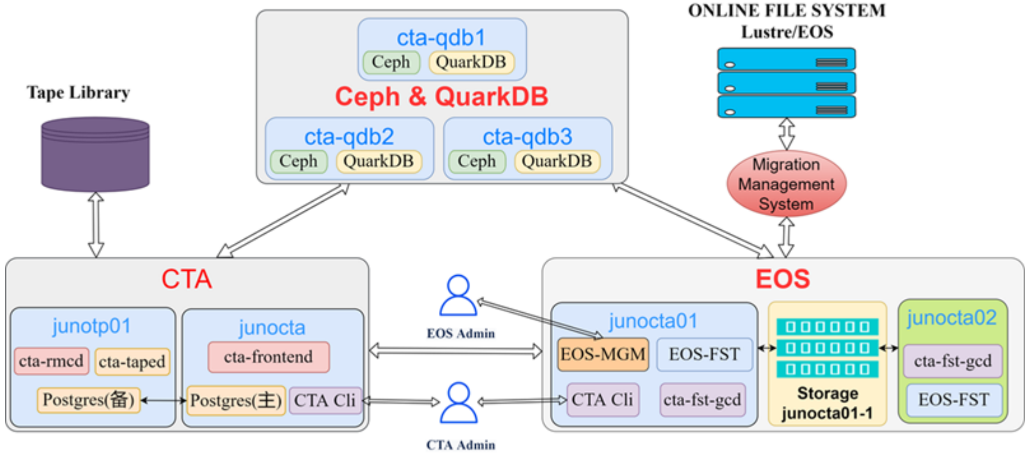
Requirements for storage

- managing the storage space and files with the same instance
- Accessing and processing more data at the same time
- Better management and utilization of systems at the same scale

Four types data storage systems were deployed

- Lustre: an open source parallel file system for industry
  - Adopted by more than 70% of the world's supercomputing centers
  - IHEP and GSI are the largest units deploying LUSTRE in the field of high energy physics
- EOS: open source storage system
  - Based on XRootD protocol, with good scalability and performance
  - IHEP is one of the largest EOS deployment site, approaching 100PB
- CTA: An open source tape library storage system
  - Based on the previous generation system CASTOR
  - It supports EOS/ dCache disk storage system
- CVMFS: software storage system
  - 6 warehouses, 14 terabytes
  - Deploy software repositories for high energy physics experiments





CEPC Computing (5)

- Distributed computing:
  - Xiaomei Zhang, Xuantong Zhang, XiaoHan
  - Work plan
    - Transitioning towards a token-based authorization and authentication infrastructure (AAI)
    - Transform into DIRACX, incorporating advanced technologies for system enhancement
    - Enable support for High-Performance Computing (HPC) and Machine Learning (ML) activities, as well as complex compound workflows
    - Evaluate the “Data Lake” model with Rucio and XCache to enhance data management and access

- Storage:
  - Haibo LI, Yujiang BI ,Yaosong CHENG, Qiuling YAO, Qingbao HU
  - Work plan
    - Improve the performance on server by the way of SSD-optimized, distributed, and object storage
    - Improve the performance with DAOS (Distributed Asynchronous Object Storage): 8GB+ write and 4GB+ read per second on a single node
    - RNtuple supported, offers smaller storage and better parallelization

International Collaboration

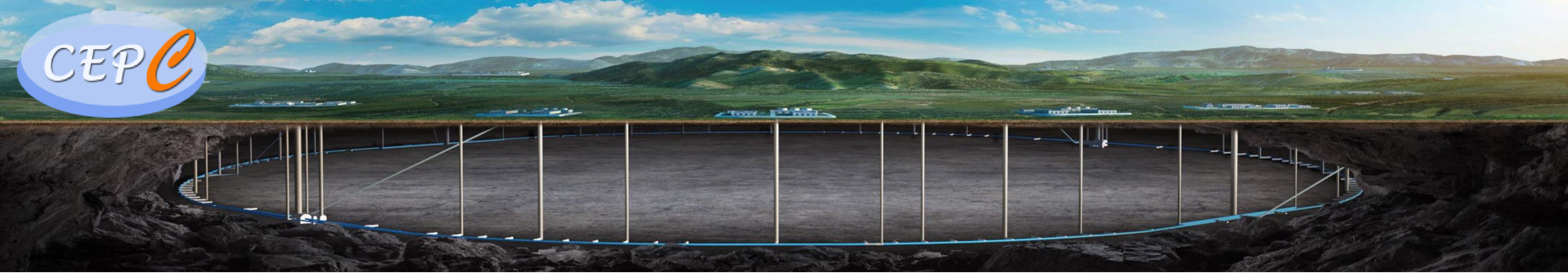
- IHEP and SDU are non-EU members of Key4hep project (AIDAinnova WP12)
  - Bi-weekly Key4hep/Edm4hep meetings and WP12 monthly meeting
- Collaboration with IDEA detector group in R&D of drift chamber
  - Weekly discussion on cluster counting method
  - Further collaboration on tracking is being discussed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

Summary

- The CEPCSW is being developed to
  - build the offline system for data processing and analysis
  - support detector R&D and physics potential studies

- Dirac-based distributed computing platform were deployed and technical R&D is ongoing

- International collaborations help us
  - to join the efforts to resolve some of the common challenges confronted by future HEP experiments



Thank you for your attention!

中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences