CEPC Computing and Software

Tommaso Boccali, Gang Li, Weidong Li,

Shengsen Sun

International Workshop on CEPC, Beijing

2019.11.20

Contents

- Software & Computing Session
- CEPC Distributed Computing
- CEPCSW Software Prototype
  - Software framework
  - Generator and detector Simulation
  - Porting of tracking algorithm
- Other software R&D activities
  - ACTS at CEPC
  - Simulation of ECAL with GAN
- Summary

Offline Software & Computing

Two sessions (Monday 14:00-16:00, 16:30-18:30)
10 contributions: 3 invited general talks, 2 talks on distributed computing, 5 talks on framework software, simulation and reconstruction.

https://indico.ihep.ac.cn/event/9960/session/7/?slotId=0#20191118

Speakers	Title
Graeme Stewart	HEP Software R&D
Tommaso Boccali	CMS (and more in general LHC) software stacks – present and future
Martin Barisits	Rucio: Scientific Data Management
ZHANG Xiaomei	Status of CEPC Computing
Suen HOU	Distributed Computing @IPAS Taipei for CEPC
Jiaheng Zou	CEPCSW Prototype and Future Plan
Tao Lin	Simulation framework in the CEPCSW
Chengdong FU	Migration of the silicon tracking algorithm to Gaudi
Yubo Han	Status of ACTS integration for CEPC tracking
Wenxing Fang	Simulation of Calorimeter with GAN

CEPC Distributed Computing

- **CEPC distributed computing system has been built up based on DIRAC and put in production since 2015**
- **7 Active Sites are integrated**
  - from UK, Taiwan, IHEP, China Universities(4)
  - **QMUL from UK and IPAS from Taiwan plays a great role**
  - **LANCASTER from UK was successfully joined two week ago**
- **Resource: ~2500 CPU cores**
  - Resource types include Cluster, Grid ,Cloud
  - ~500 CPU cores and ~500TB storage will be added in IHEP at the end of this year

Site Name	**CPU Cores**
CLOUD.IHEP-OPENNEBULA.cn	24
CLUSTER.IHEP-Condor.cn	48
CLOUD.IHEPCLOUD.cn	200
GRID.QMUL.uk	1600
CLUSTER.IPAS.tw	500
CLUSTER.SJTU.cn	100
GRID.LANCASTER.uk	300
Total (Active)	**2772**

QMUL: Queen Mary University of London
IPAS: Institute of Physics, Academia Sinica

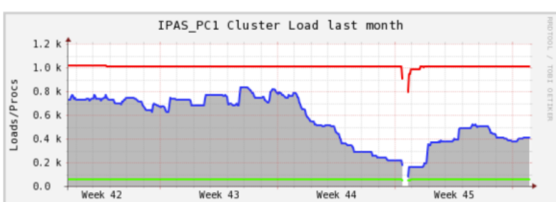
IPAS Taipei cluster for CEPC

- **Two Condor Clusters have been configured for CEPC:**
  - IPAS_PC1(CentOS7), PHYS T3(SL6)
  - IHEPDIRAC provides Singularity mode for IPAS_PC1 to run CEPC jobs in SL6 transparently
- **IPAS is able to increase CPU share for CEPC since demands of other experiments are dropping**
- **10Gbps bandwidth between Taiwan and IHEP, and end-to-end network need to be improved**

Cluster : IPAS_PC1

Nodes	CPU Per node	Memory	Memory/node	Core
16	Xeon Gold 6130 16 Core x 2 1U<2yr	96 GB	3 GB per node	512
4	E5-2686 V4 2.3Ghz 18 Core x 2 2U<3yr	128 GB	3.5 GB per node	144
5	Xeon X5690 3.46GHz 6 Core x 2 blade ~6yr	24 GB	2 GB per node	192
28	Xeon X5345 4corex2 blade ~10yr cold	15 GB	2GB per node	x

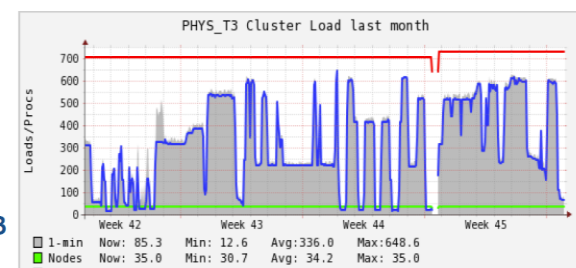
- Total Computing Nodes : 35
- **Total CPU Core : 848 (on Condor)**
- Network : 10 G SFP+ Ethernet
- **OS : CentOS 7 with Singularity**
- **HTCondor Scheduler**
- **Job type: Single Job, Parallel Job**



Cluster : PHYS T3

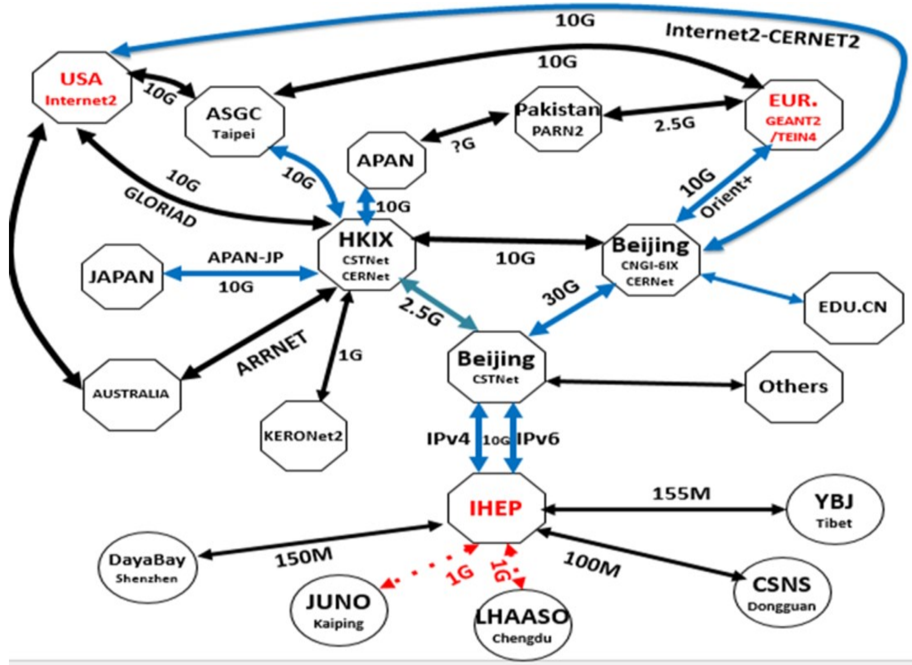
Nodes	CPU Processor	Memory	Memory/node	Core
31	2650L V2,1.7GHz, 10 core x 2 ~5yr	96GB	4.8GB	620

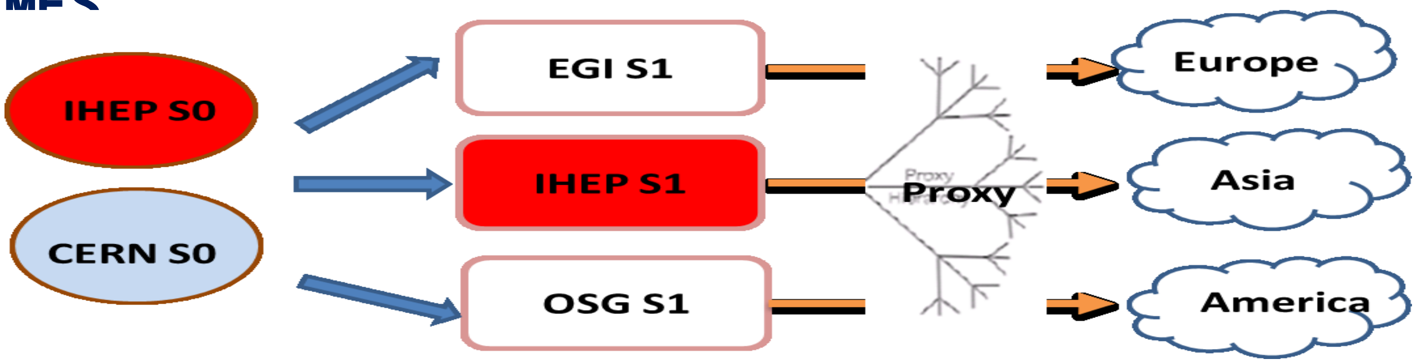
- Total Computing Nodes : 31
- **Total CPU Core : 620**
- Network: Infiniband QDR 40Gb/s
- **OS : Scientific Linux for CERN 6**
- **HTCondor Scheduler**
- **Job type : Single Job, Parallel JOB**



Network and global software deployment

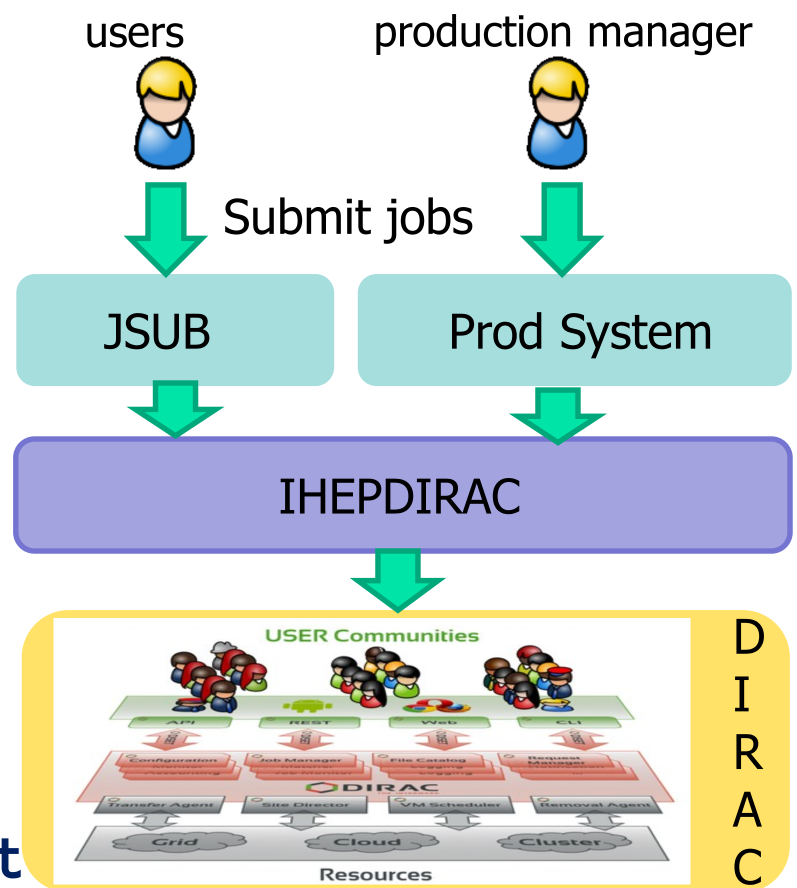
- IHEP international network provides a good basis for distributed computing
  - 20Gbps outbound, 10Gbps to Europe/USA/TaiWan
- IHEP CVMFS service was well established and joined global federation since 2014
- CEPC software was deployed globally via CVMFS





Workload management

- **DIRAC and IHEPDIRAC**
  - Provide a middle layer between CEPC jobs and heterogeneous resources
- **Massive job submission frontend**
  - JSUB – user job
  - Prod system – production job
    - Manage mass workflow and dataflow automatically
    - Provide interface for prod groups to manage production tasks
- **Two systems are adding the support to the CEPC software migration to GAUDI**



Data management

- **Central Storage Element is based on StoRM**
  - Lustre /cefs as its backend
  - Frontend provides SRM, HTTP, gridftp access
- **With EOS becoming main storage system in IHEP, the backend of SE will be changed to EOS**
  - EOS testbed with gridftp protocol is ready
  - Join DOMA TPC (Third Party Copy) to prepare for the future evolution to http and XRooTD
- **Evaluating DIRAC data management and Rucio for future CEPC data management system**
  - The prototype with DIRAC data management is ready
  - RUCIO is also in investigation

Rucio – Scientific data management

- Rucio provides a mature and modular scientific data management federation, developing into a common standard for scientific data management
- Rucio origins from Atlas, and is being adopted and evaluated by many experiments and communities
  - CMS, SKA, BelleII, Dune, ICECube, EGI, GridPP……
- Many experiments including CEPC pay close attention to the integration of DIRAC and Rucio which are two important WMS and DMS in HEP

A growing community

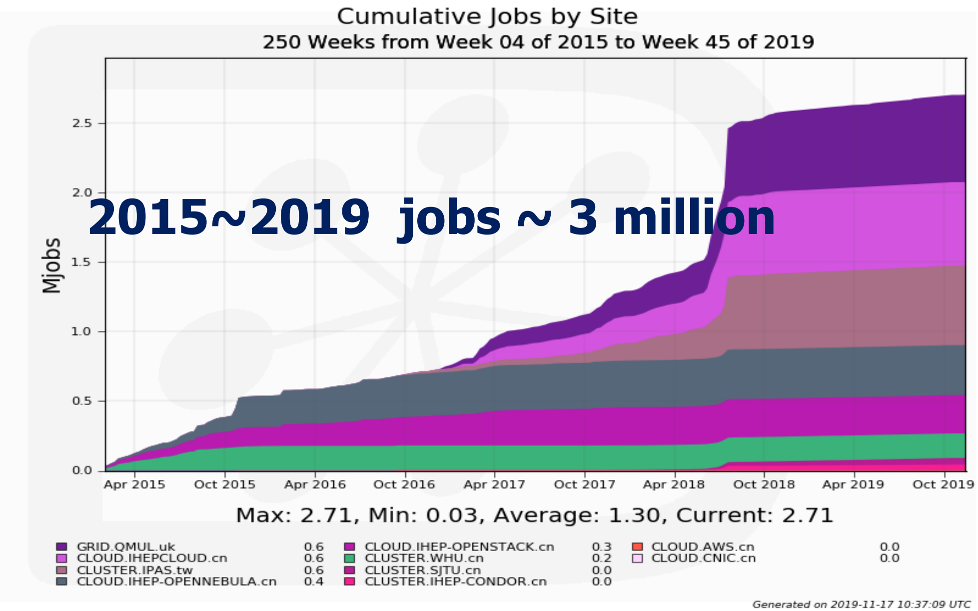
Rucio main functionalities

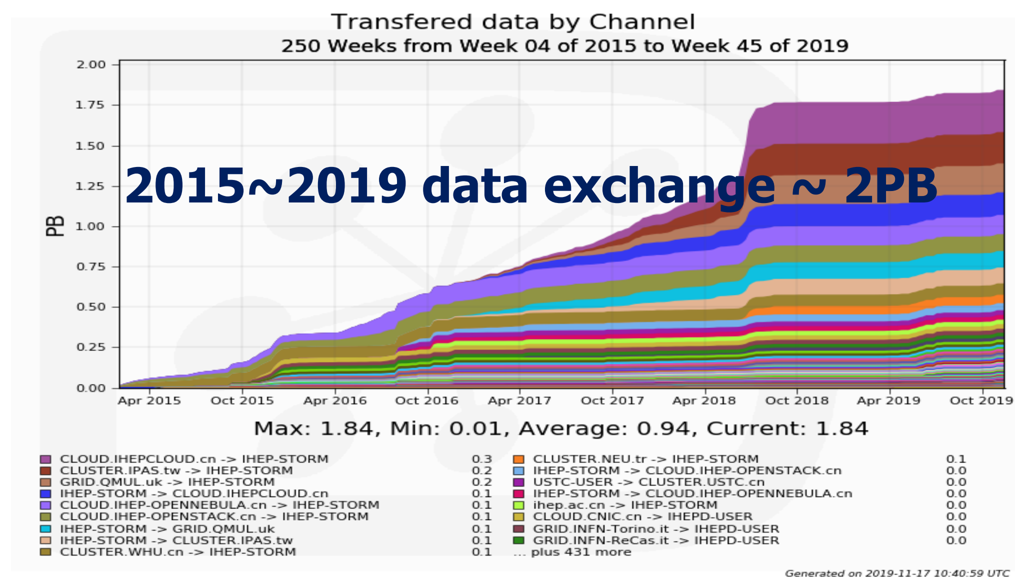
- Provides many features that can be enabled selectively
  - File and dataset catalog
  - Transfers between facilities including disk, tapes, clouds, HPCs
  - Web-UI, CLI, and API to discover/download/upload/transfer/annotate data
  - Extensive monitoring for all dataflows
  - Support for caches and CDN workflows
  - Expressive policy engines with rules and subscriptions
  - Automated corruption identification and recovery
  - Data popularity based replication
  - …

More advanced features

Official production status

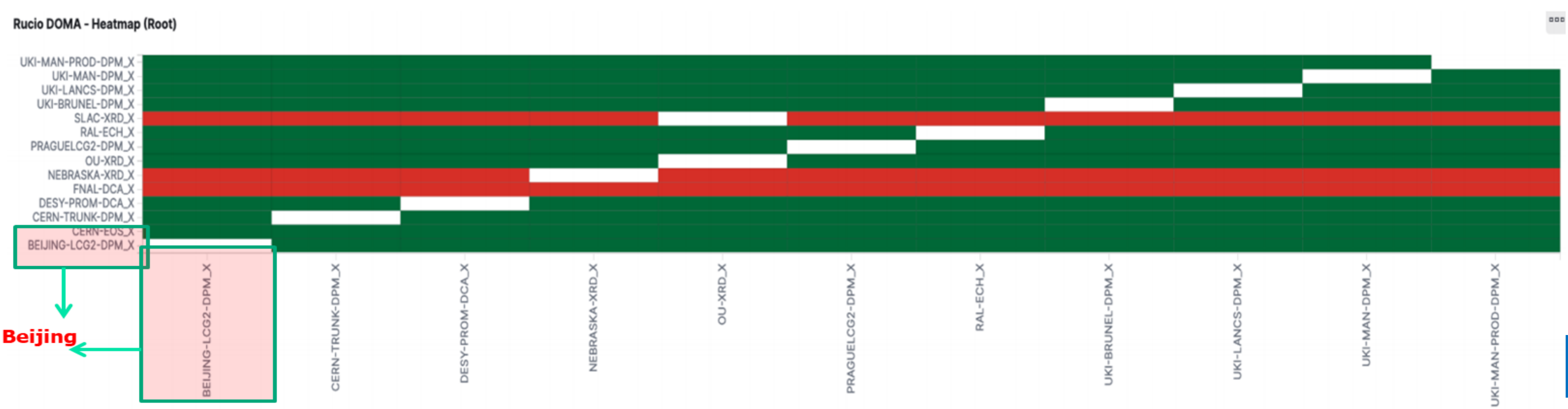
- The distributed computing system is taking full tasks of CEPC massive simulation for the last four years
  - About 3 million jobs, data exchange about 2PB





Cooperation with HSF on future tech evolutions

- HSF provides a platform for expertise in HEP communities to work together for software and computing evolutions to face the challenges in future experiments
- CEPC shares the same interests with other future experiments
- Close cooperation between HSF and IHEP is being established on future technology evolutions
  - WLCG AAI (Authorization and Authentication Infrastructure)
  - DOMA TPC (Third Party Copy)
  - DOMA data federation and access ……

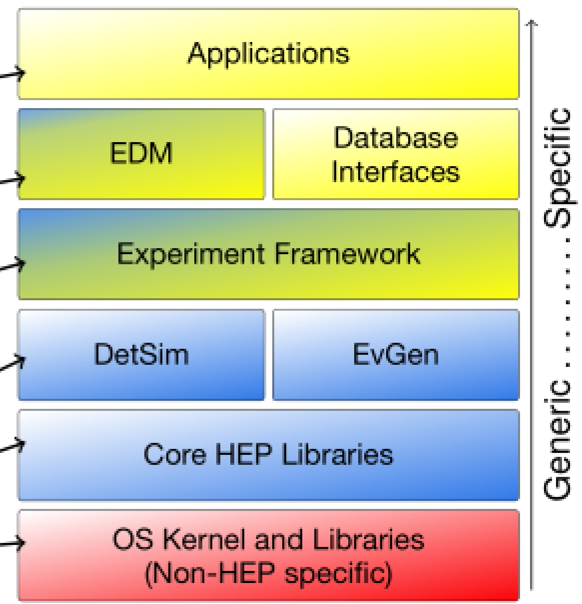


New Software Prototype

- Since Oxford workshop in April, we began to evaluate a new framework for CEPC experiment
- In June, at Bologna workshop, we reached the agreement of one Common Software Stack (Key4HEP) for future collider experiments (CEPC, FCC , ILC, CLIC, STCF, SCT)
  - To maximize the software sharing between experiments

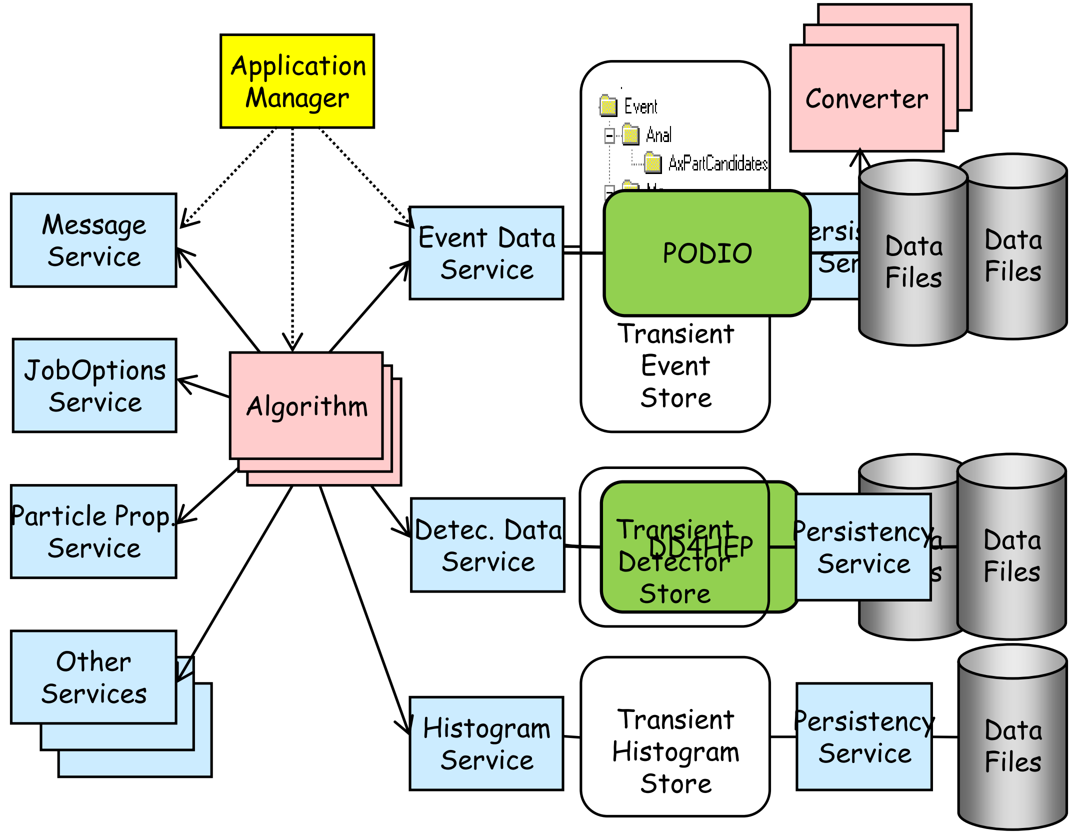
[Ref]: André Sailer, etc. , CHEP2019

- Interfaces to tracking and reconstruction libraries (PandoraPFA, ACTS)
- (More or less) experiment specific event data model libraries
- Experiment core orchestration layer, which controls everything else: Marlin, Gaudi, CMSSW, AliRoot
- Packages used by many experiments: DD4hep, Pythia, . . .
- Usual core libraries (ROOT, Geant4, CLHEP, . . .)
- Non-HEP libraries: boost, python, cmake . . .



Gaudi Framework

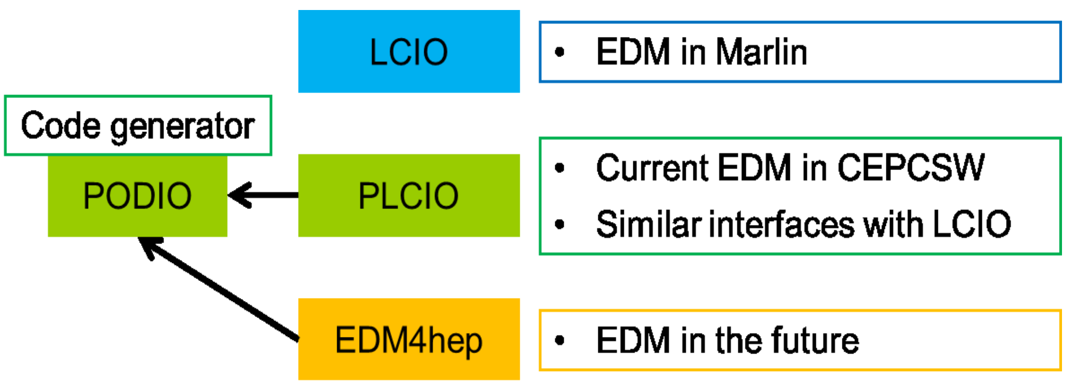
- For the latest releases, a lot has changed
  - External libraries: the minimal set is only 9 (and only ROOT is HEP), the rest are optional.
  - Whiteboard: the requirement of objects in the store to inherit from Data Object has been gone

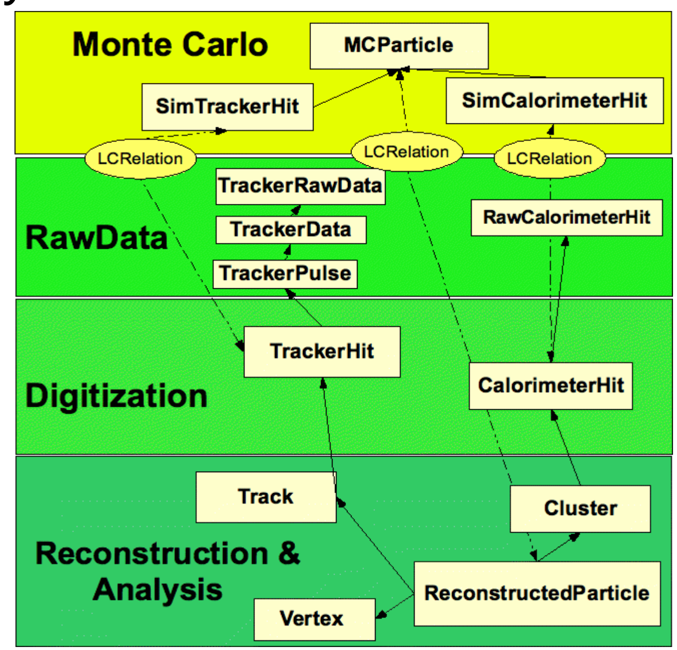


- The core part of the framework is small, key components are:
  - Application Manager, Algorithms, Tools, Services

Event Data Model

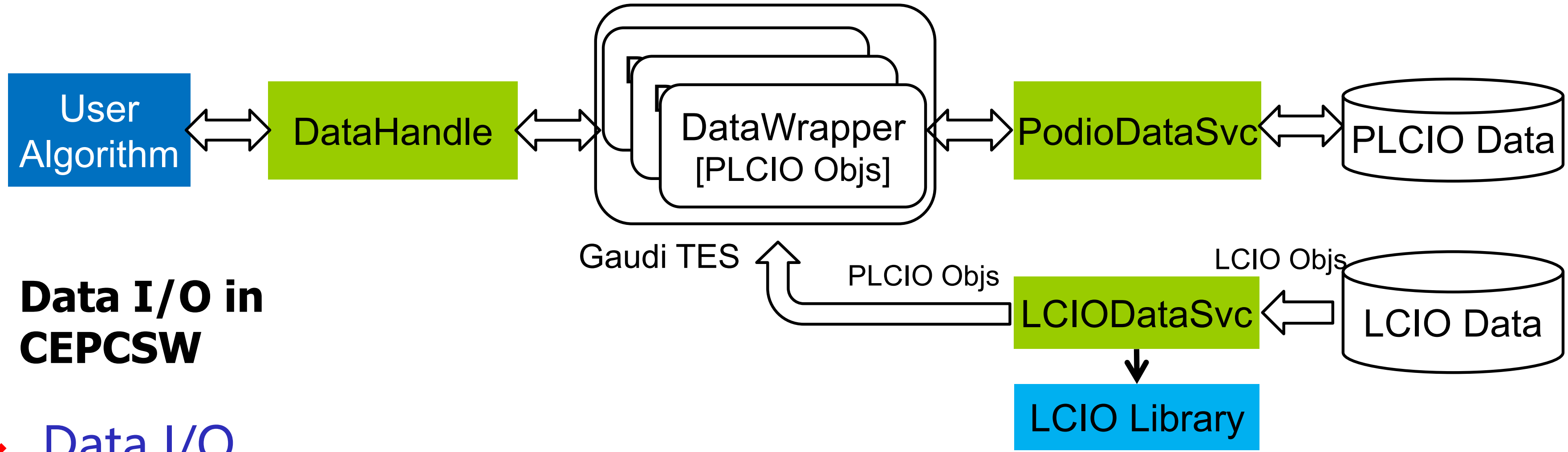
- EDM4hep, a common event data model in the context of CSS
  - Still being discussed
- PLCIO, an implementation of the LCIO data model with PODIO
  - Perfect EDM for our prototype
  - Migration from PLCIO to EDM4hep is easy





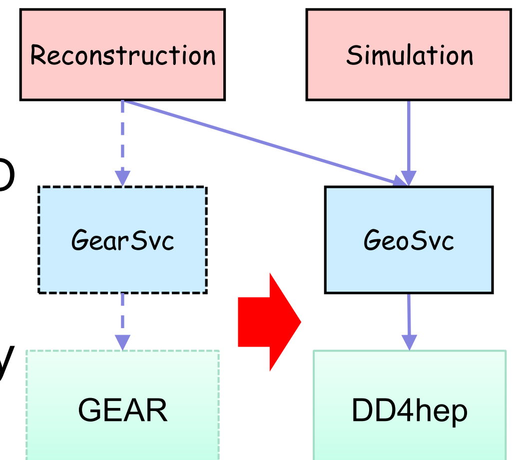
- PLCIO data is in ROOT format
  - Data analysis can be done with ROOT

Event Store and Geometry Service



Data I/O in CEPCSW

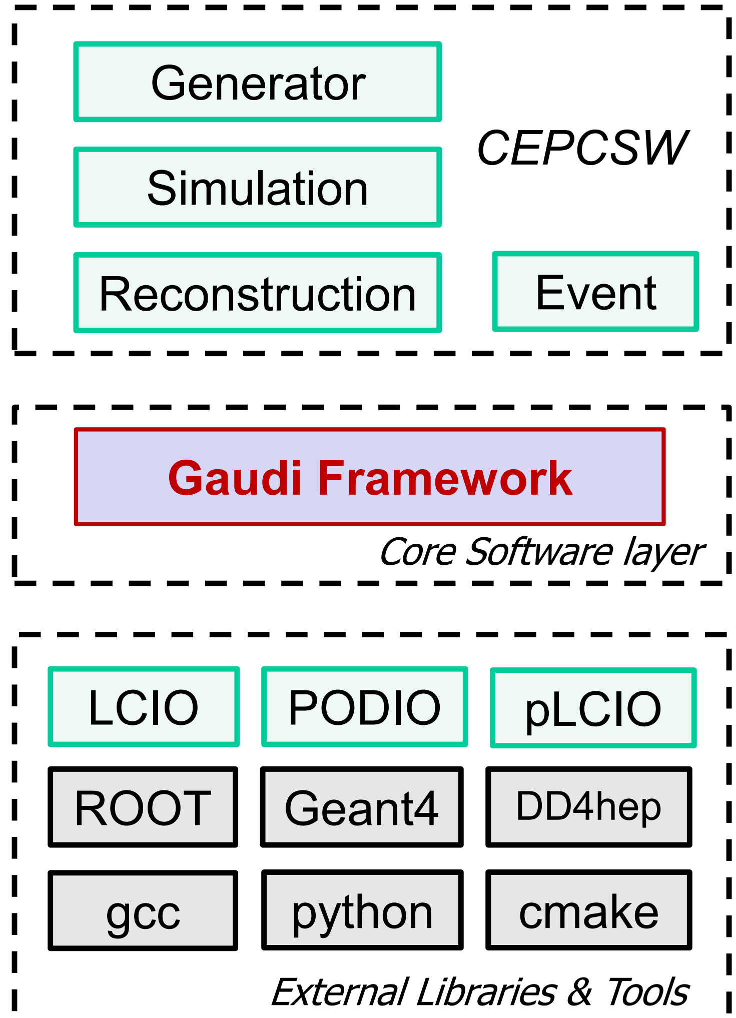
- Data I/O
  - FCCSW FWCore is reused for PLCIO data I/O (many thanks)
  - Extension of LCIODataSvc
    - To read LCIO data generated by Marlin
    - Data object types are converted from LCIO to PLCIO
- Geometry Service
  - GEAR is used for rec during migration temporarily
  - DD4hep will be used for both sim and rec



Software Infrastructure and Building

- Common tools
  - CMake: Build & deployment
    - Gaudi cmake macros
  - Git: version control
    - http://cepcgit.ihep.ac.cn/cepc-prototype
  - CVMFS: software distribution
    - CEPC specific:

      /cvmfs/cepcsw.ihep.ac.cn/prototype
- Software building
  - Reuse the FCCSW & LCG software stack now (many thanks)
  - Move to Key4HEP in the future

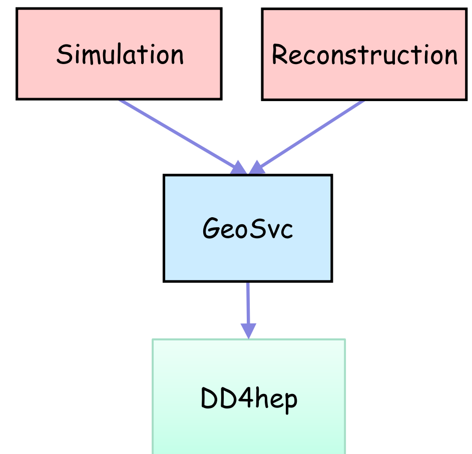


Plans for CEPCSW Framework

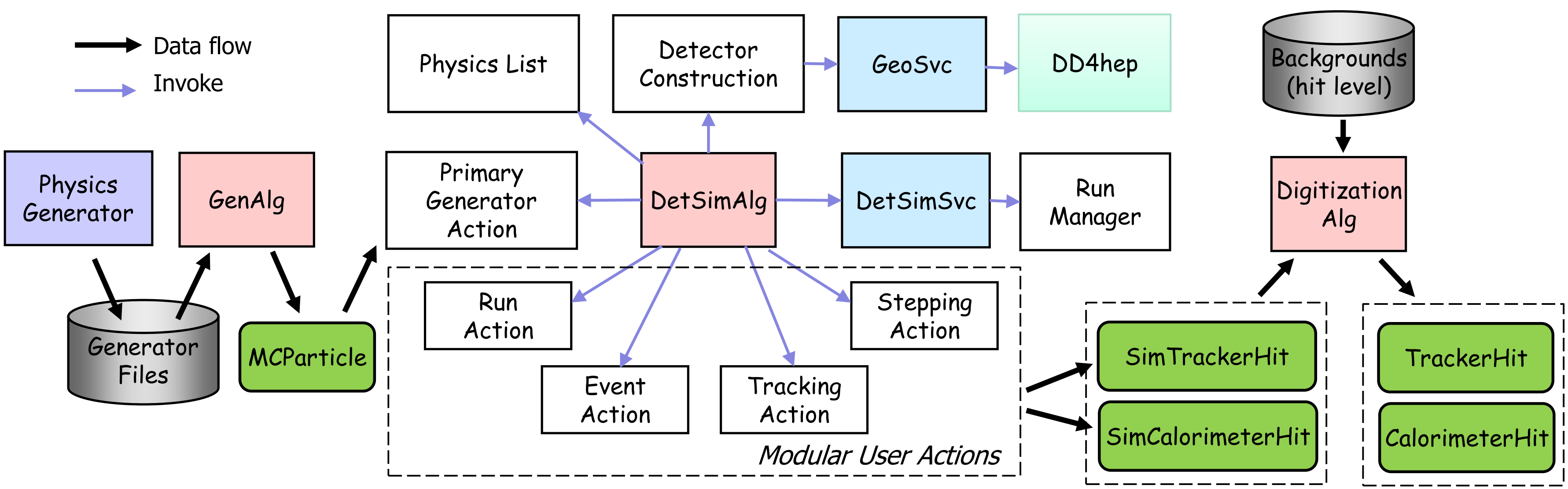
Tasks	Plans
Software migration	More tracking algorithms from Marlin to CEPCSW
Substitution	Geometry service: move from GEAR to DD4hep
	EDM: move from PLCIO to EDM4hep (when it is ready)
Building & Release	Software stack packaging with SPACK & move to Key4HEP finally
	Continuous integration, automatic software building/testing/releasing
Beam tests support	Fulfill the requirements for beam tests data analysis
Parallel computing	Functional and reentrant algorithms
	Execution performance analysis and optimization
New features	Integration with Deep Learning algorithms

Simulation (1)

- **Unified Geometry Service**
  - Interfaced to DD4HEP
  - Used by simulation and reconstruction
  - Configured with XML files



- **Simulation chain**
  - Integrated physics generators & Geant4



Simulation (2)

- Event Data Model
  - MCParticle
  - SimTrackerHit
  - SimCalorimeterHit

  } Hit
  - TrackerHit
  - CalorimeterHit

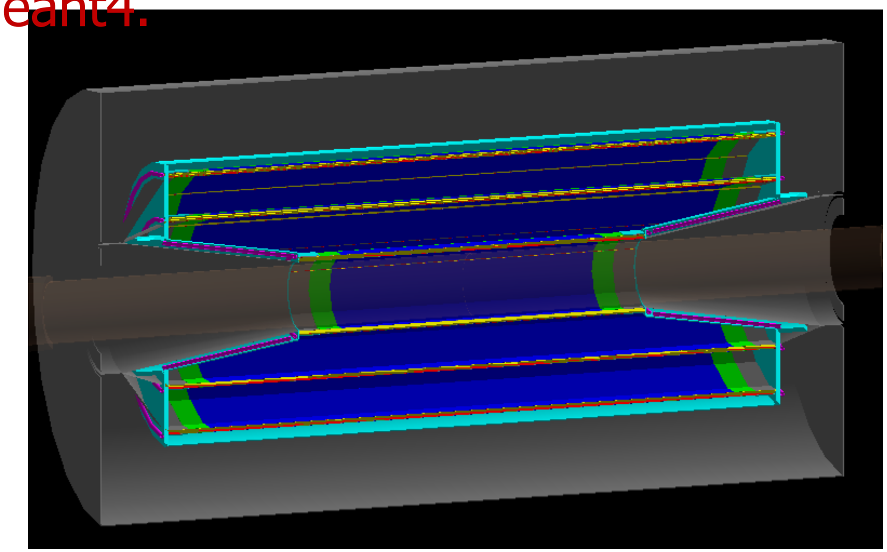
  } Digi
- Following collections are available in output.
  - MCParticle
  - VXDCollection
  - SITCollection
  - TPCCollection
  - SETCollection

  } SimTrackerHit
- Keep compatible with Mokka.

- DD4hep based Geometry
  - Package: Detector/DetCEPCv4

    https://indico.cern.ch/event/783429/contributions/3376689/attachments/1830850/2998281/CEPC2019-Oxford-FullG4DetectorSimulation.pdf

Simulation framework is working and the detector could be visualized using Geant4.



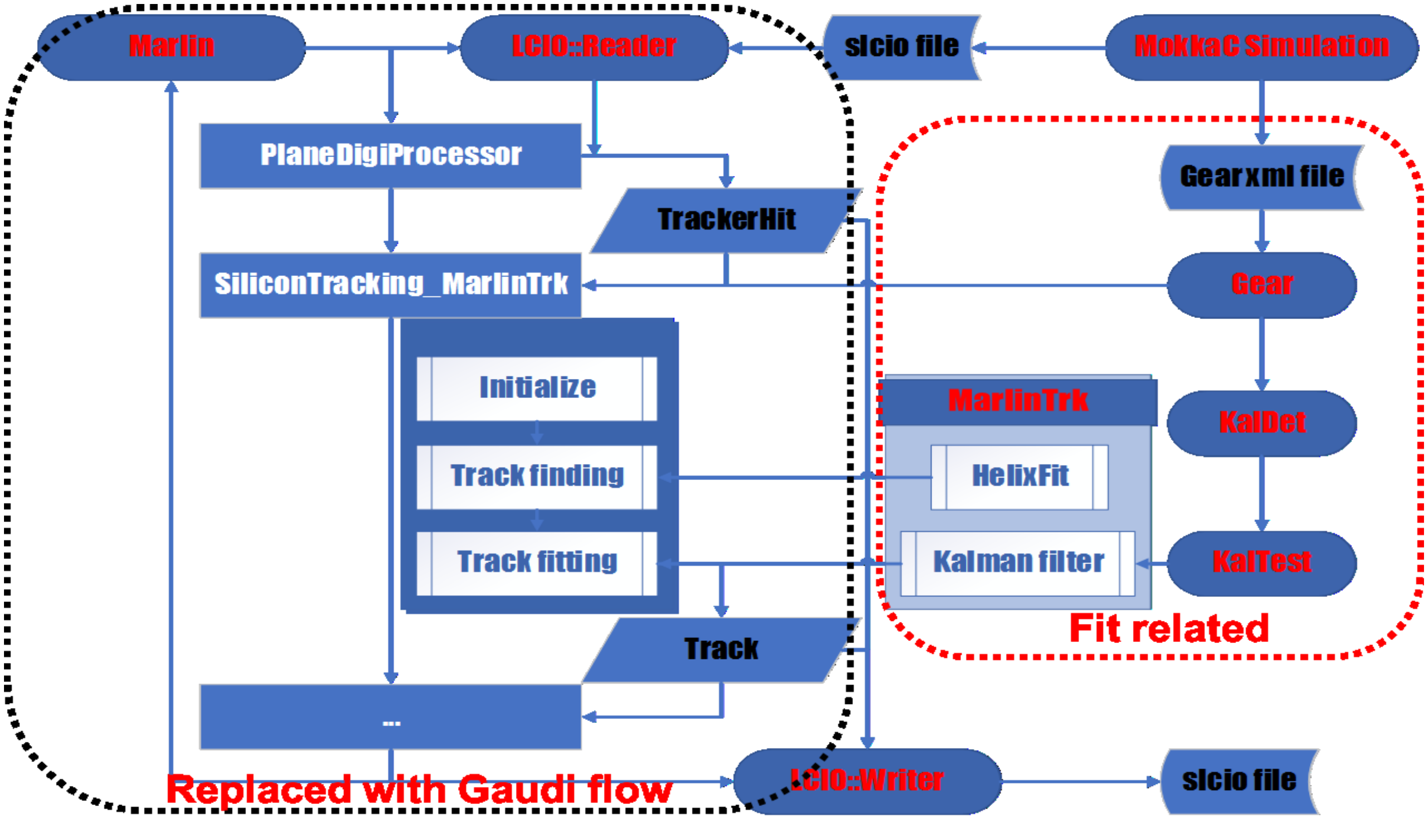
Simulation (3)

- A simulation framework prototype is developed.
  - Configurable Geometry with XML files: support multiple options of detectors and beam test geometry.
  - Physics generator: Integrate with external physics generators easily.
  - Modular user actions to collect data in simulation: Save more information other than the event data model.

Sub-components	Plans
Generators	Generators on the fly
Event Data Model	MCTruth correlation
Geometries & fields	TPC, calo, magnetic field, different options
Digitization	MC hit level event mixing.
Fast simulation	Integration, Parameterization, Machine Learning
Validation & Production	stress testing, performance testing, MC data challenges etc.
Parallelism	Gaudi+Geant4 10

Porting of Tracking Algorithm

SiliconTracking algorithm was chosen as a first porting example from Marlin to CEPCSW

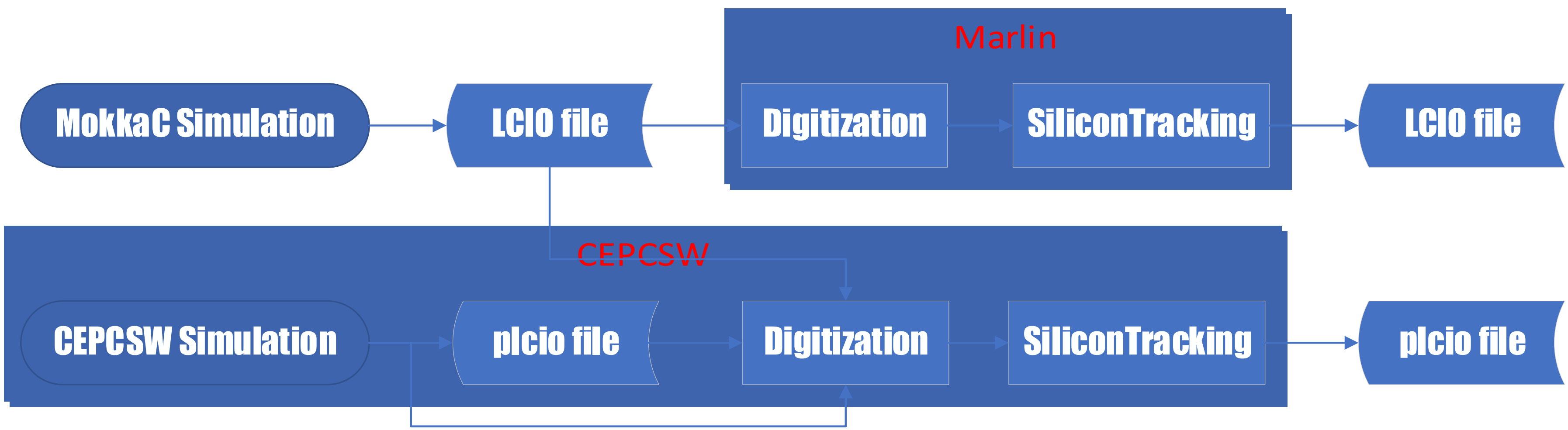


The workflow of Silicon Tracking in Marlin

(replace event looper with Gaudi manager)

Progress of porting

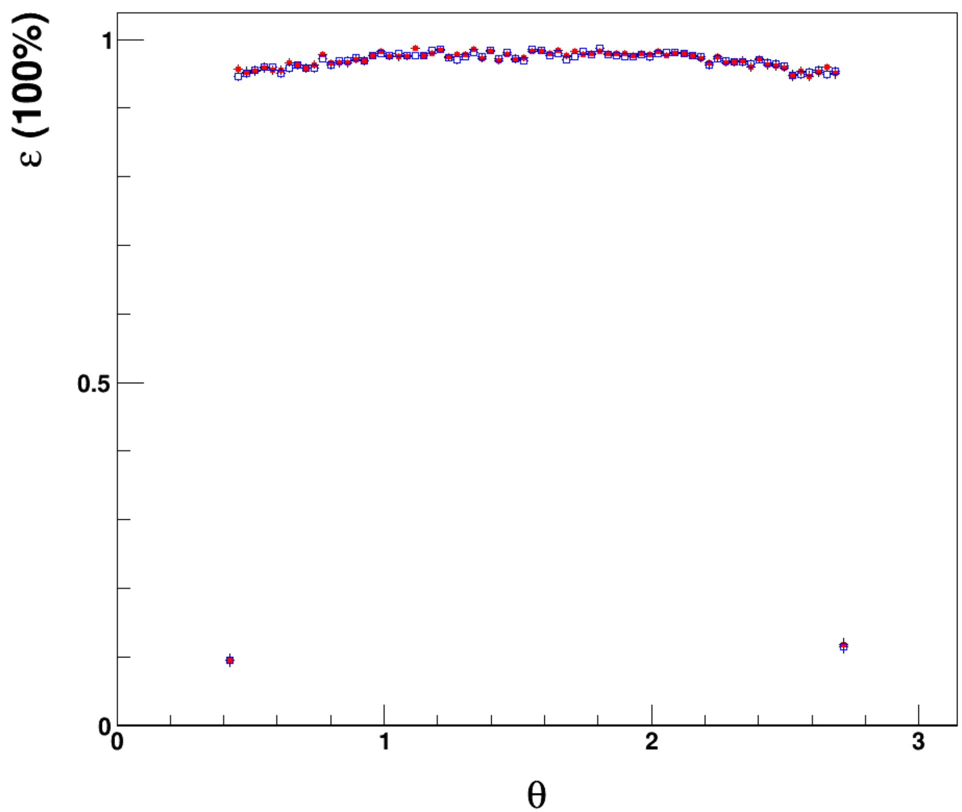
- A chain of tracking for vertex detector has been completed
  - Fitting relative: MarlinTrk, KalTest, KalDet
  - Digitization: PlaneDigiProcessor
  - Reconstuction: SiliconTracking



Tracking Efficiency

- Definition
  - $\varepsilon = N_{macthed\_track} / N_{MC(primary)}$
  - Matching:
    - $|par_{fit} - par_{MC}| < 5\sigma_{par}$ (par=d0, phi0, ω, z0, tanλ)

	R (mm)	$\lvert z \rvert$ (mm)	$\lvert\cos\theta\rvert$	$\sigma(\mu\text{m})$
Layer 1	16	62.5	0.97	2.8
Layer 2	18	62.5	0.96	6
Layer 3	37	125.0	0.96	4
Layer 4	39	125.0	0.95	4
Layer 5	58	125.0	0.91	4
Layer 6	60	125.0	0.90	4



Fake rate:
CEPCSW: (1.23±0.04)%
Marlin: (1.21±0.04)%

⇓

10σ cut:
CEPCSW: (0.56±0.03)%
Marlin: (0.56±0.03)%

The results are consistent, and even identical if using same random number for smearing

ACTS at CEPC (1)

- Increasing software requirement from CEPC detector design
- (Benefit from software upgrade projects for other experiments)

- ACTS: A Common Tracking Software
  - Encapsulate existed code from ATLAS
  - Thread safe/long vector/Modern C++ 17, minimal requirements
    …
  - Experiments independent && Open-source

acts

Tracking

- Geometry
- Surface
- EDM
  - Measurement
  - Parameters
  - TrackState
- Material
- MagneticField
- Utilities
- Propagator
  - ✓ RKN Stepper
  - ✓ StraightLine Stepper
  - ✓ Navigator
- Fitter
  - ✓ Kalman Filter
  - • Gaussian Sum Filter
- Seeding
- Vertexing
- ☐ Track Finding
  - • CKF + HOUGH
- ☐ Calibration – no need
- ☐ Alignment – no need

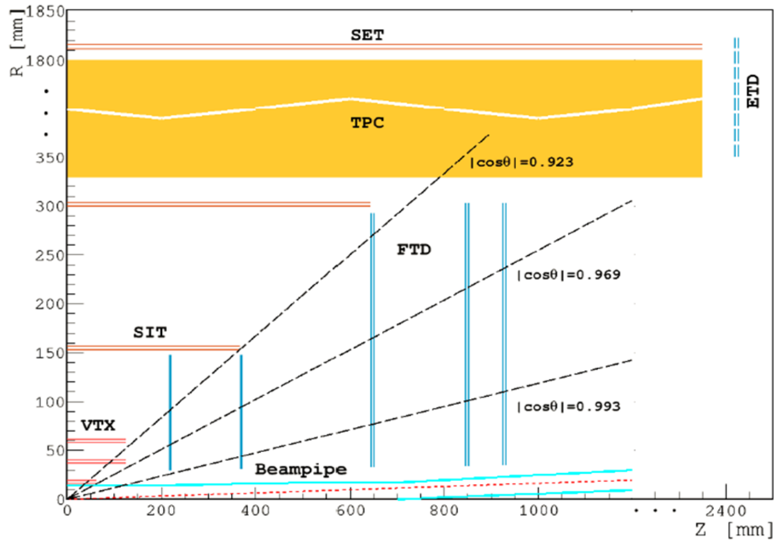
Plugins

- ✓ DD4HEP
- ✓ TGeo
- ✓ Geant
- ✓ Jason
- ✓ Pythia
- …

aits: A light-weight Gaudi framework

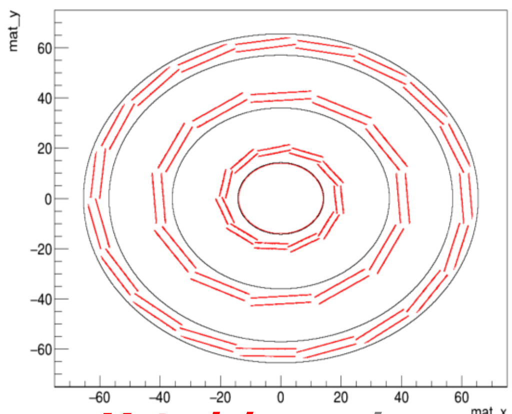
acts: ATLAS fast simulation

ACTS at CEPC (2)

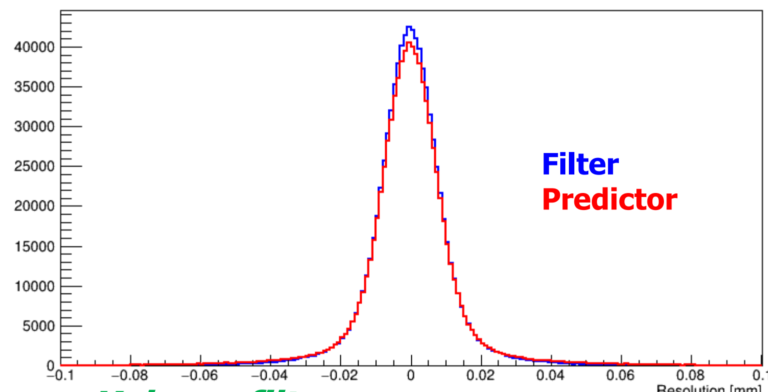


- **CEPC baseline detector tracker (Pre-CDR) preliminary realized in ACTS-FW**
- **Basic functions have been validated**
- [Keep updating: https://gitlab.cern.ch/jinz/acts-framework-cepc](https://gitlab.cern.ch/jinz/acts-framework-cepc)

➢ **To be further studied and migrate into CEPC Framework**

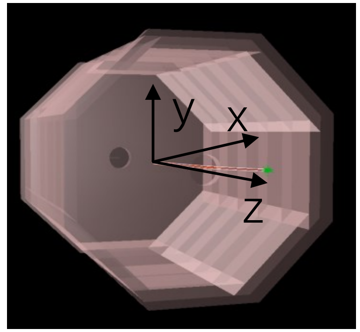


Material mapping

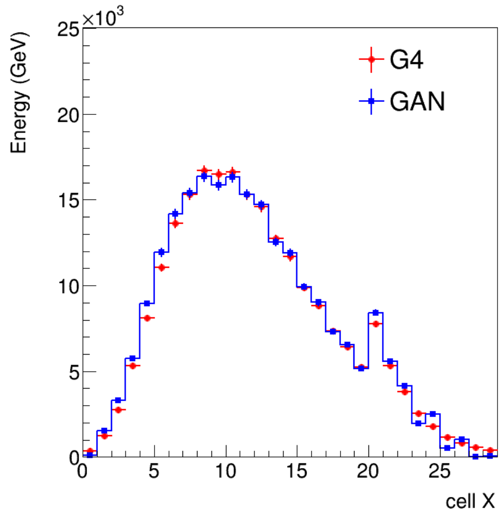


Kalman filter

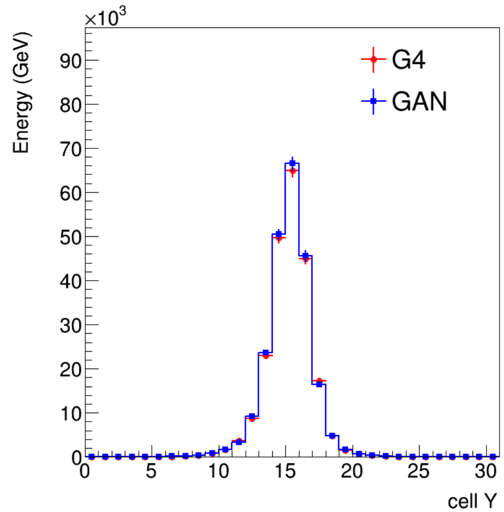
Simulation of Calorimeter with GAN



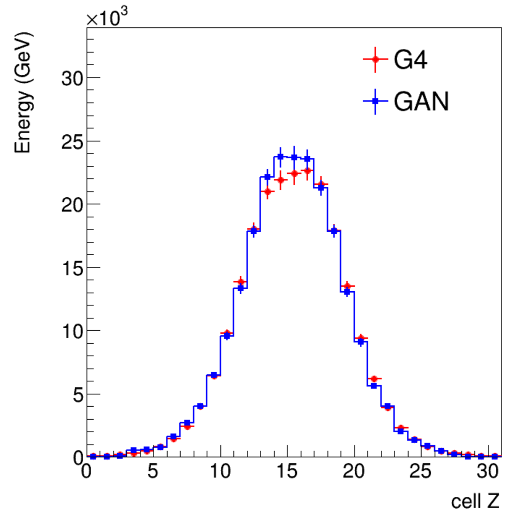
- ❑ Geant4 simulation:
  - ➢ Pro: very precise
  - ➢ Con: requires large computing resources
- ❑ Calorimeter simulation is one of bottlenecks.
- ❑ The Generative Adversarial Networks (GAN) could be used for calorimeter fast simulation.
- ❖ Training data:
  - ○ Single photon gun sample.
  - ○ Energy in [1, 100] GeV uniformly.
  - ○ θ in [50, 140] degree uniformly.
  - ○ ϕ in [-15, +15] degree uniformly.
  - ➢ Hit energy in 31× 31×29 calorimeter cells are considered.



Energy deposited in X(layer) direction



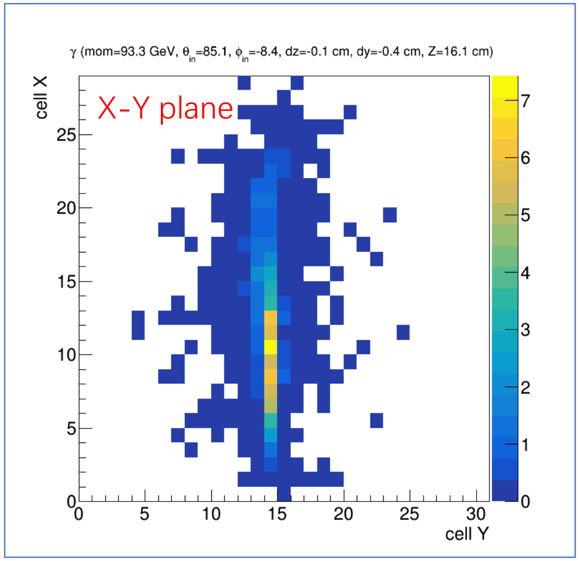
Energy deposited in Y direction

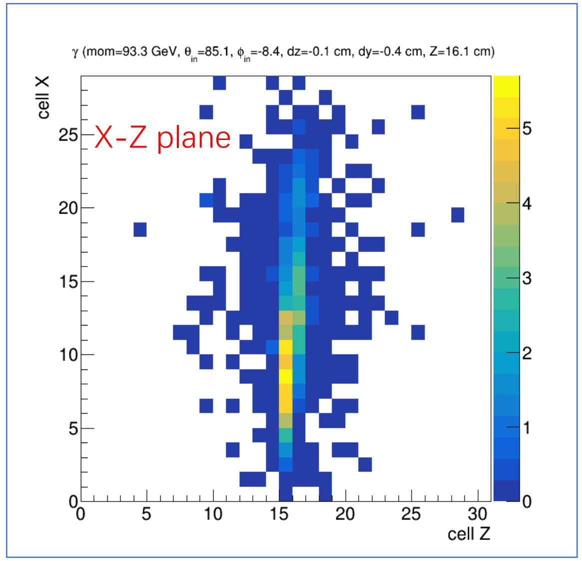


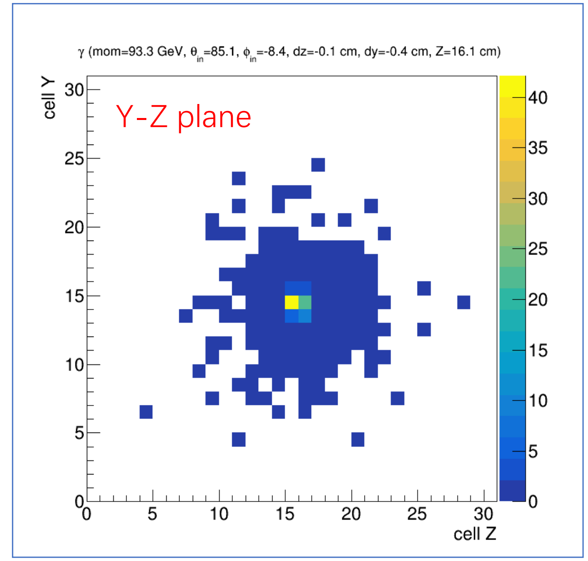
Energy deposited in Z direction

Event display

γ (Mom = 93.3 GeV, θ_{in} = 85.1°, ϕ_{in} = −8.4°, ΔZ^{Pos} = −0.1 cm, ΔY^{Pos} = −0.4 cm, Z = 16.1 cm)

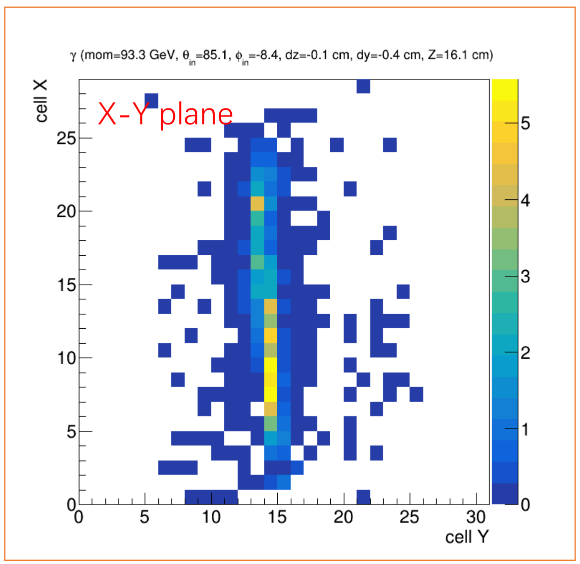


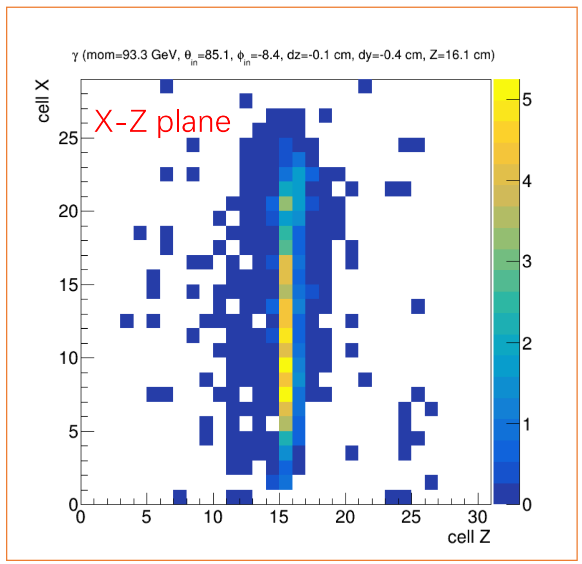


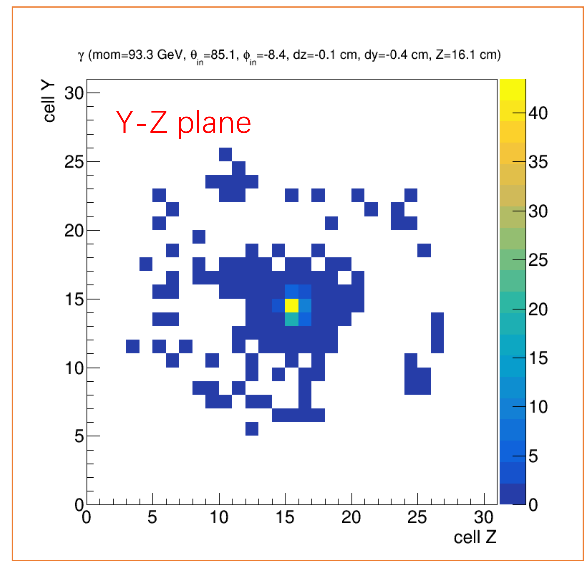


Geant4

GAN

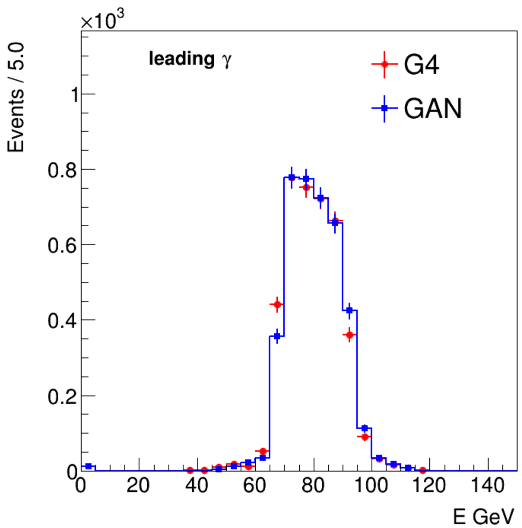




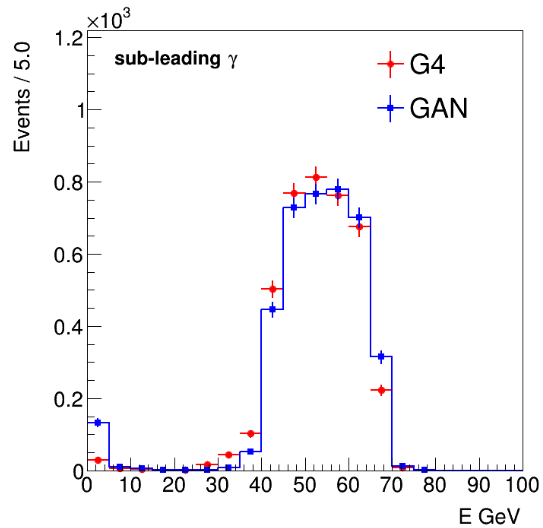


Simulation of Calorimeter with GAN

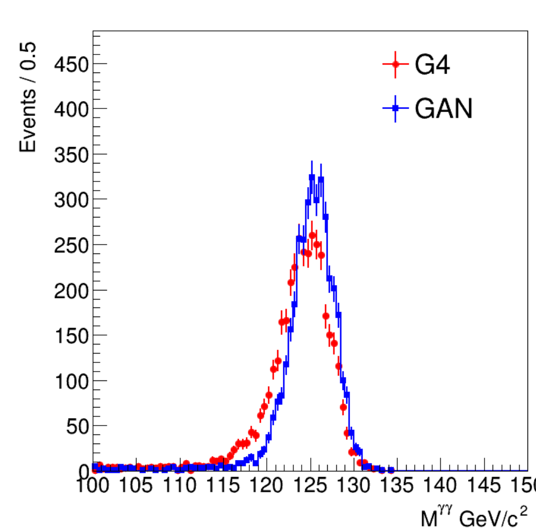
- ❑ Apply GAN and do event reconstruction using mc samples.
  - ➢ $e^{+}e^{-} \rightarrow Z(\nu\nu)H(\gamma\gamma)$ mc samples are used.
  - ➢ Comparing the properties of reconstructed gamma.



Leading γ E_{rec}



Sub−leading γ E_{rec}



$M_{\gamma\gamma}$

- ➢ Looks fine, has room for improvement.

Summary

- CEPC distributed computing system works well and member sites are from:
  - IHEP, IPAS, LANCASTER, QMUL and SJTU
- CEPCSW prototype has been developed using Gaudi, DD4hep, Geant4 and PLCIO, etc.
  - both detector simulation and tracking algorithm can be run successfully
  - ready to add more algorithms to the prototype by following given examples
  - future development will be based on Key4HEP collaborating with CERN

Thank You！

谢谢