Offline Software for High Energy Physics Experiments

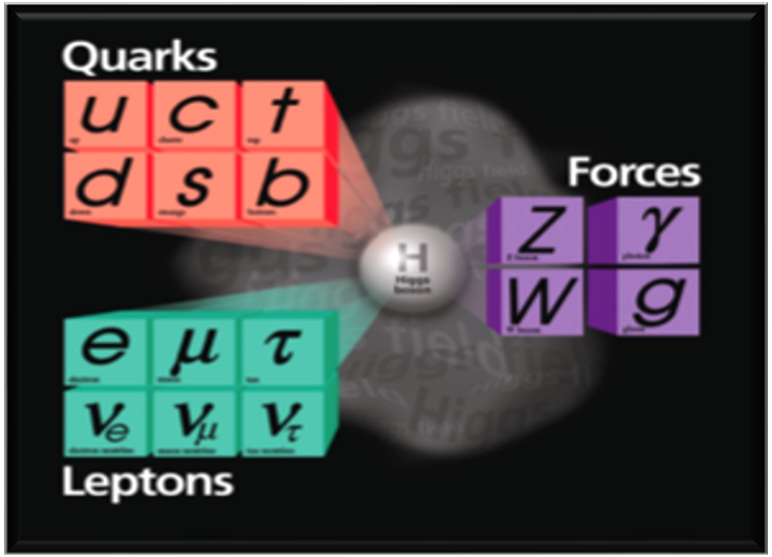
黄性涛
山东大学 前沿交叉科学青岛研究院

第十一届华大QCD讲习班
2023年10月23-27日

Outline

- Introduction to HEP Experiments and Software System
- BESIII Offline Software System (BOSS)
- JUNO Offline Software System (JUNOSW)
- STCF Offline Software System (OSCAR)
- CEPC Offline Software System (CEPCSW)
- Introduction to Particle Physics of SDU
- Summary

The Standard Model (SM)



- Fundamental Particles
  - Three generations of quarks
  - Three generations of leptons

- Fundamental Interactions
  - Electromagnetic Interaction
  - Weak Interaction
  - Strong Interaction

Electromagnetic Interaction + Weak Interaction ⇒ Electroweak Theory (EM)

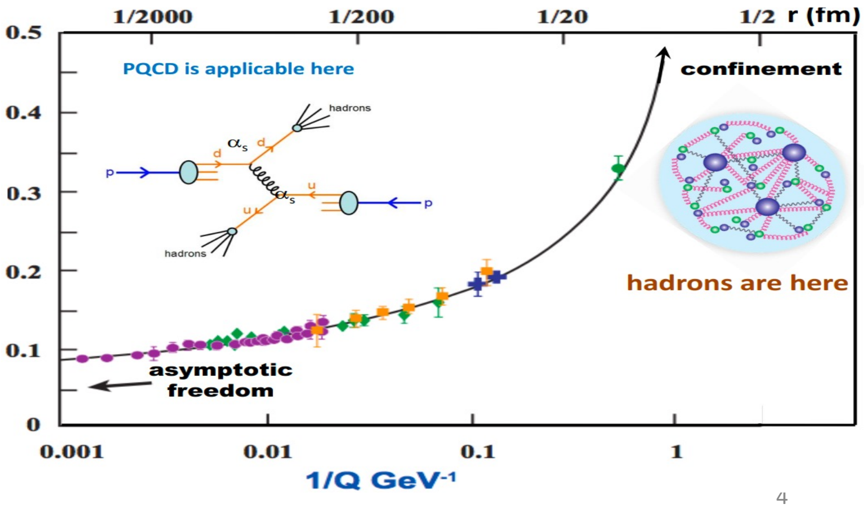
Strong Interaction ⇒ Quantum Chromodynamics (QCD)

SM: Verified by Experimental measurements

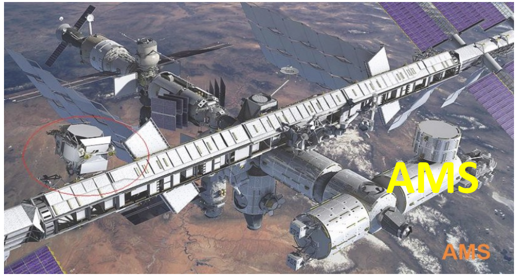
- Electroweak Theory (EM)
  - Predicted the W and Z
  - Predicted the Higgs

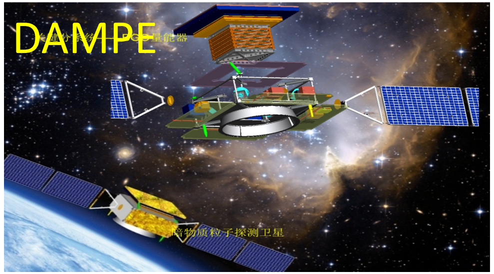
QUANTITY	MEASURED (GeV)	SM PREDICTION (GeV)
Mass of W boson	80.387 ± 0.019	80.390 ± 0.018
Mass of Z boson	91.1876 ± 0.0021	91.1874 ± 0.0021

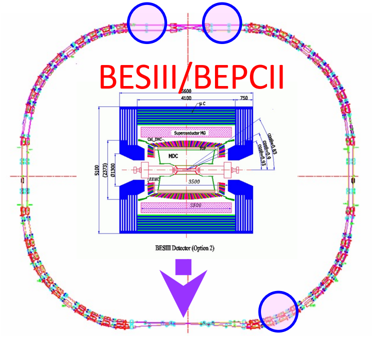
- Quantum Chromodynamics (QCD)
  - Asymptotic freedom
  - Confinement

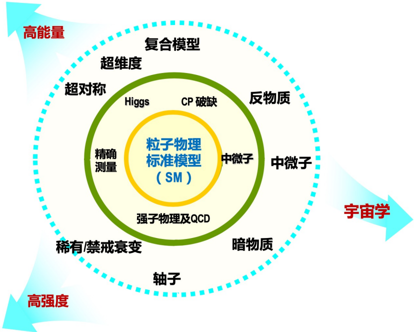


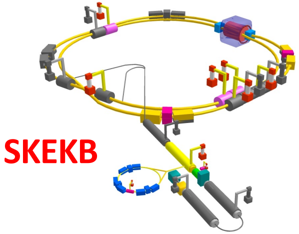
Big-Science Facilities

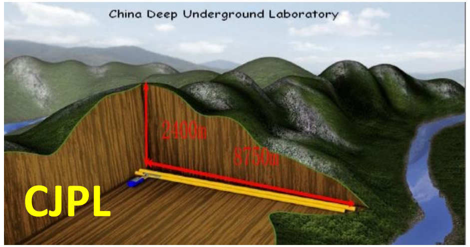




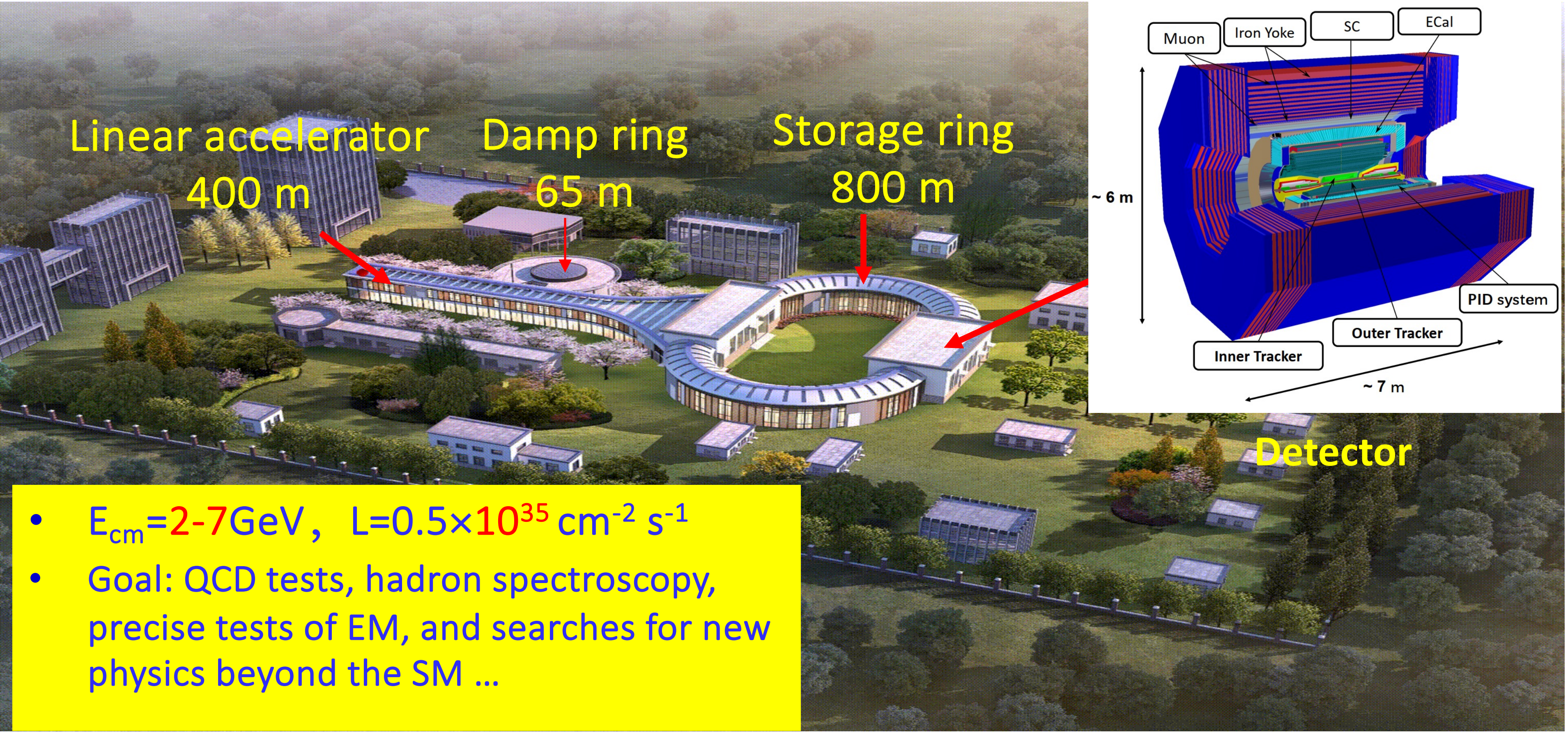






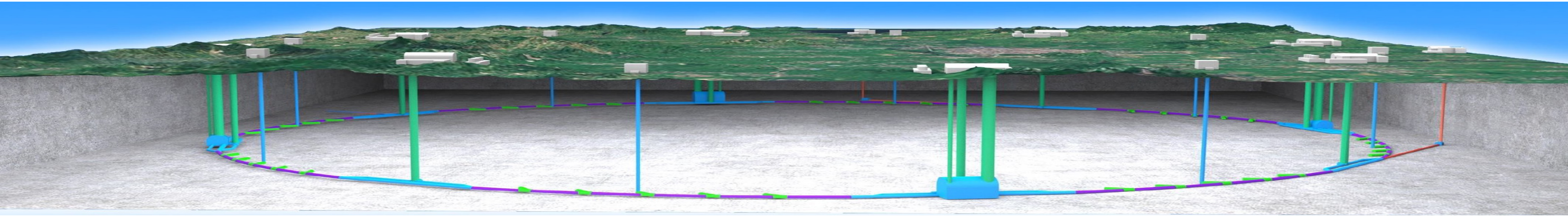


Super Tau-Charm Facility (STCF)



- E_{cm}=2-7GeV， L=0.5×10^{35} cm^{-2} s^{-1}
- Goal: QCD tests, hadron spectroscopy, precise tests of EM, and searches for new physics beyond the SM …

Circular Electron Positron Collider (CEPC)



New Physics
~10 TeV

- ✓ Dark Matter
- ✓ Extended Higgs
- ✓ Composite Higgs
- ✓ Supersymmetry
- ……

High Precision
~1%

- ✓ Higgs: 1%-0.1%
- ✓ EW : O(10^2-10^3) vs current
- ✓ Flavor
- ……

Detector
Particle Flow

- ✓ High Granularity
- ✓ Good Resolution
- ✓ Reliable PID
- ……

Design & Implementation
Detector & Accelerator

- ✓ Multi-purposed Acc.
- ✓ New concept Detector
- ✓ Cost, added values
- ✓ Spill-over
- ……

- ✓ Management
- ✓ Collaboration
- ……

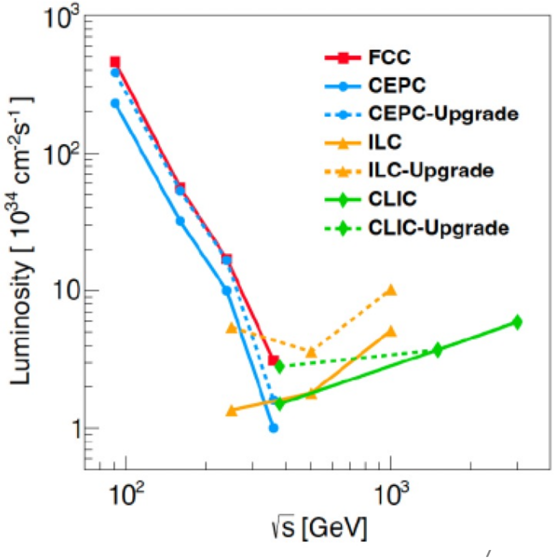
High Lumi.
~10$^{34-36}$cm^{-2}s^{-1}

- ✓ Higgs: 20 ab^{-1}
- ✓ Z: 100 ab^{-1}
- ✓ W: 6 ab^{-1}
- ✓ Top: 1 ab^{-1}

Accelerator
2 IPs/100 km

- ✓ Switchable Operations
- ✓ Upgradability
- ✓ Large Piwinski angle with Crab-waist
- ……

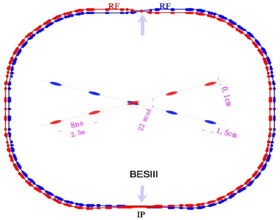
Operation mode	Z	W	Higgs
Center-of-mass energy (GeV)	91	160	240
Operation time (year)	2	1	10
Instantaneous luminosity/IP (10^{34}cm$^{-2}s^{-1}$)	115	16.0	5.0
Integrated luminosity (ab^{-1}, 2 IPs)	60	3.6	12
Event yield (30 MW)	2.5×10^{12}	1.0×10^8	2.5×10^6
Event yield (50 MW)	4.0×10^{12}	1.6×10^8	4.0×10^6

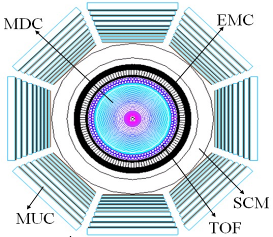


Research works at HEP Experiments

加速器 ⇨ 探测器 ⇨ 数据 ⇨ 事例产生子 | 探测器模拟 | 离线刻度 | 事例重建 | 事例显示 | 物理分析 ⇨ 物理

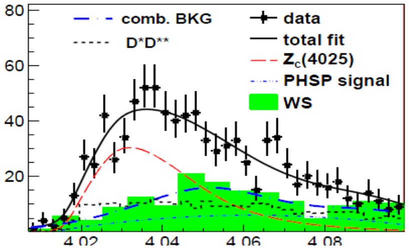
离线核心软件平台





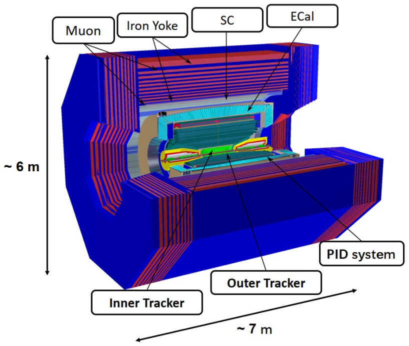
Hardware

Computing and Software

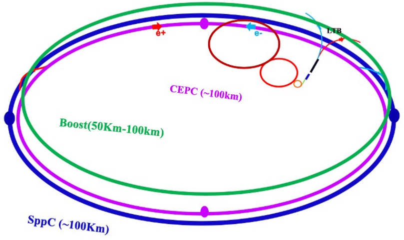


HEP Experiments: PB -> EB eta

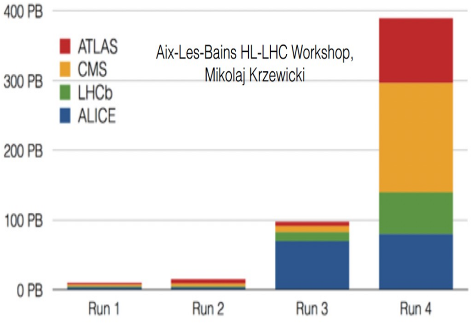
Experiments	Running Time	Data/Year (PB)	Total Data（PB）
DayaBay	2011-2020	0.2	**2**
BESIII	2008-2028	0.5	**10**
LHAASO	2019-2030	1.2	**12**
JUNO	2022-2030	2.0	**20**



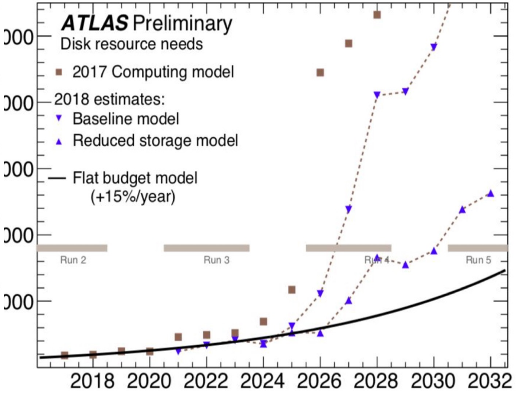
STCF: 每年0.5 - 1 EB



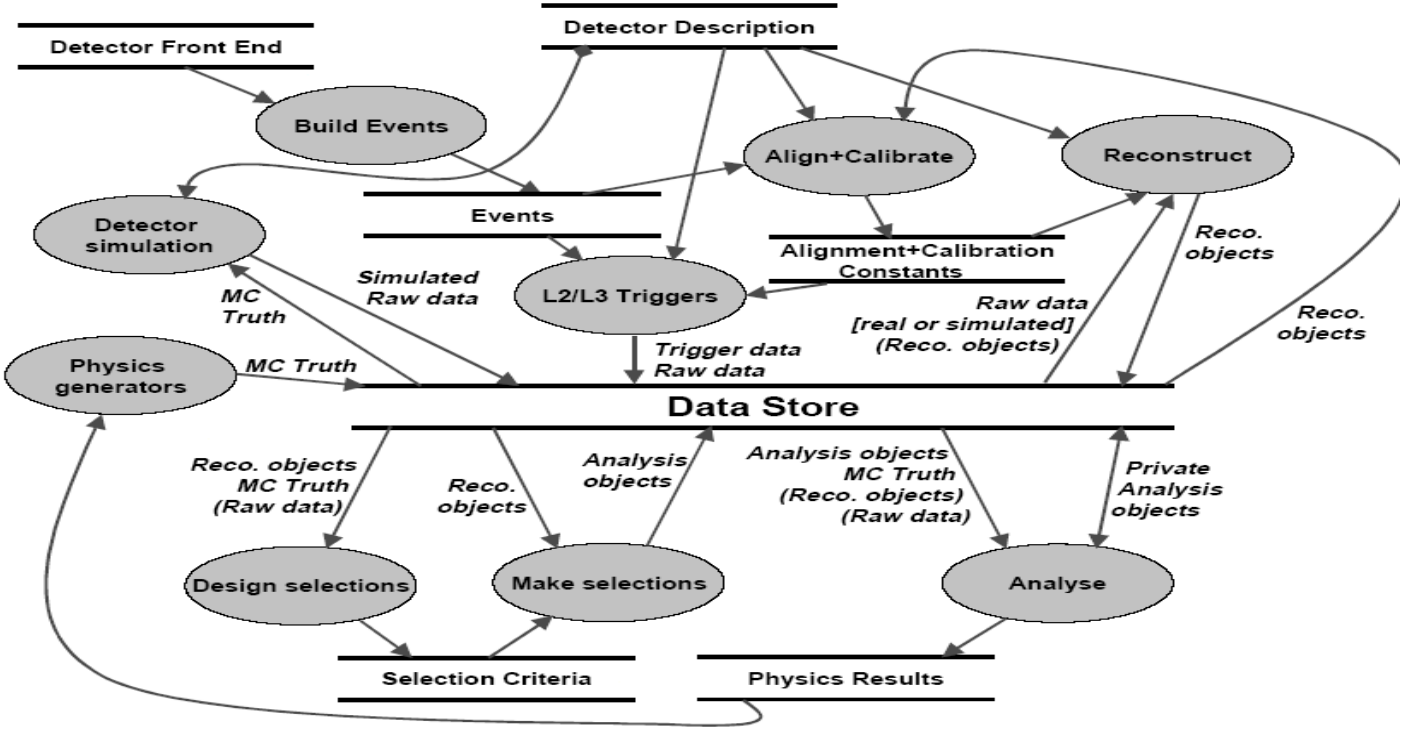
CEPC/SppC：
- ■Higgs/W Factory 每年 1.5~3 PB；
- ■Z factory 每年 0.5~5 EB.



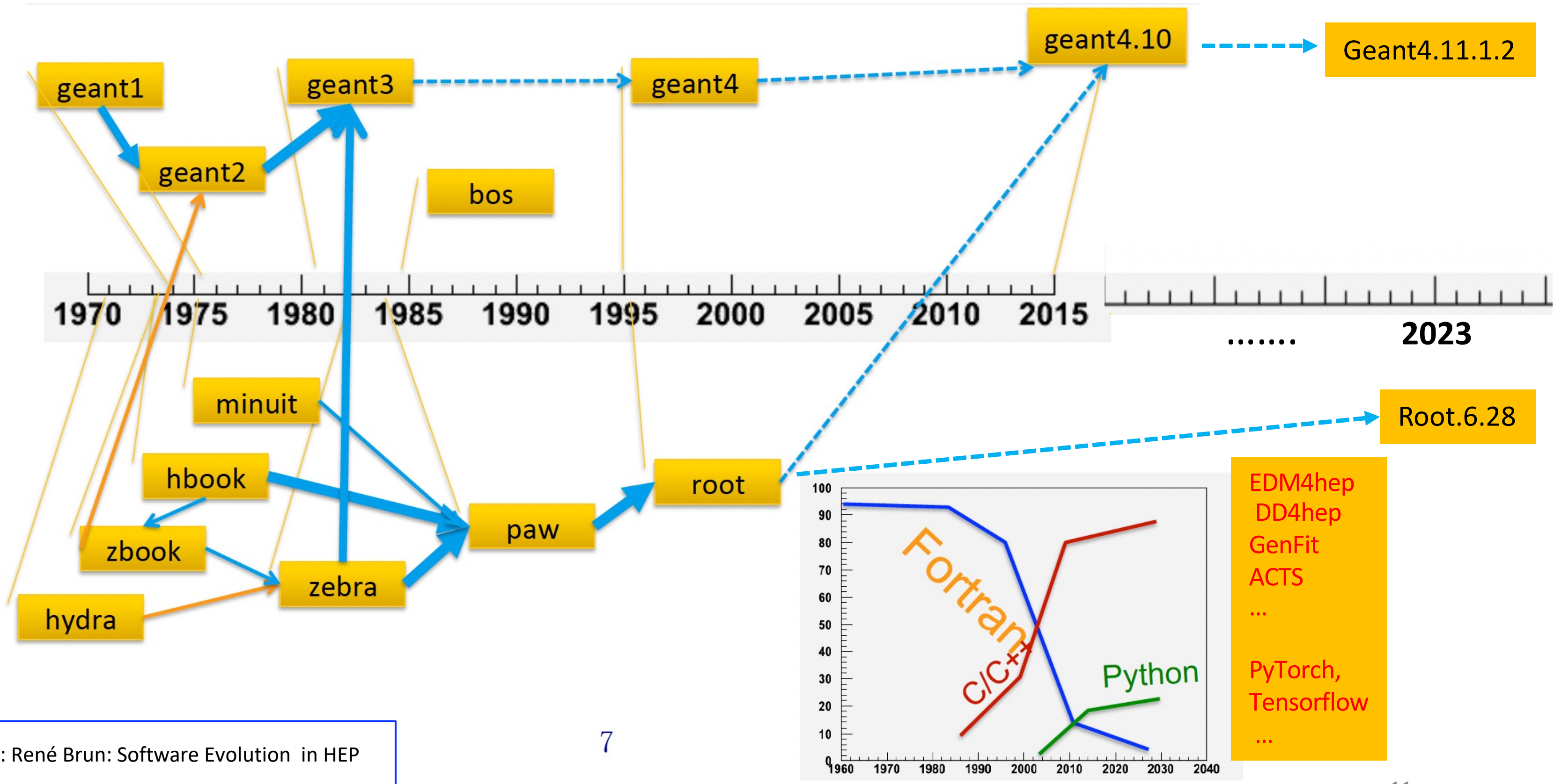
RAW data to offline for LHC experiments



Offline Data Processing and Software

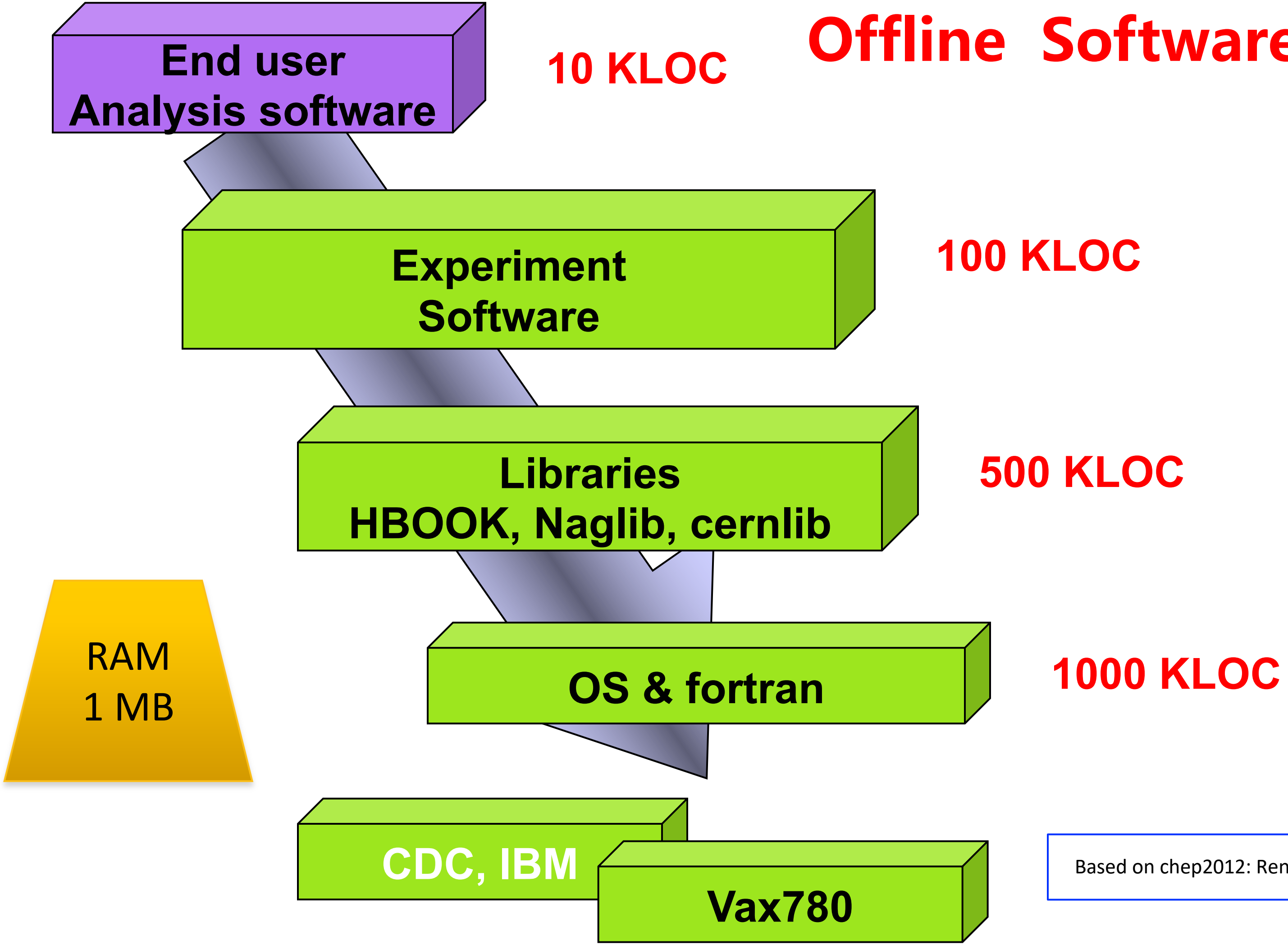


Evolution of Libraries for HEP



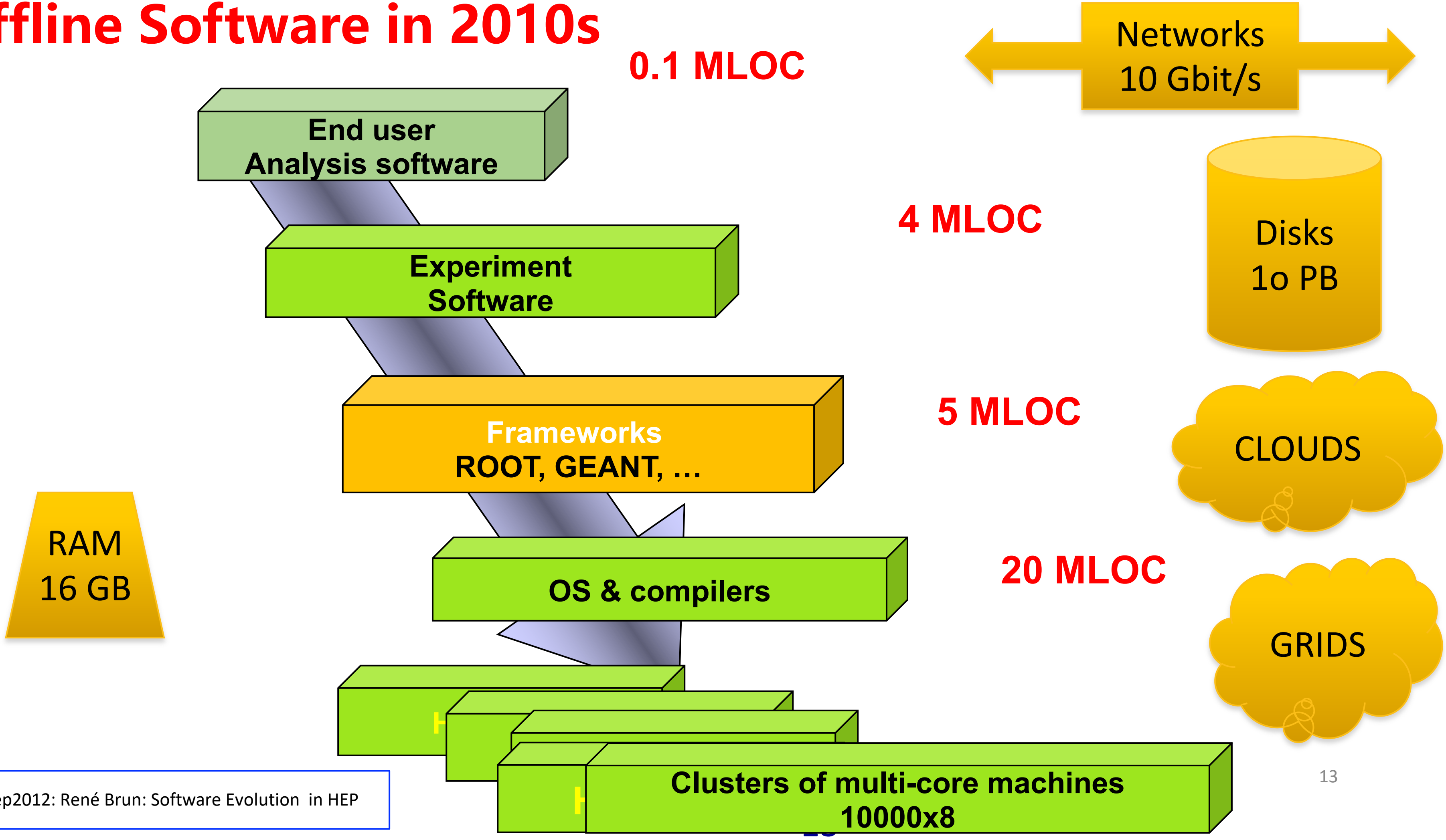
chep2012: René Brun: Software Evolution in HEP

Offline Software in 1980s

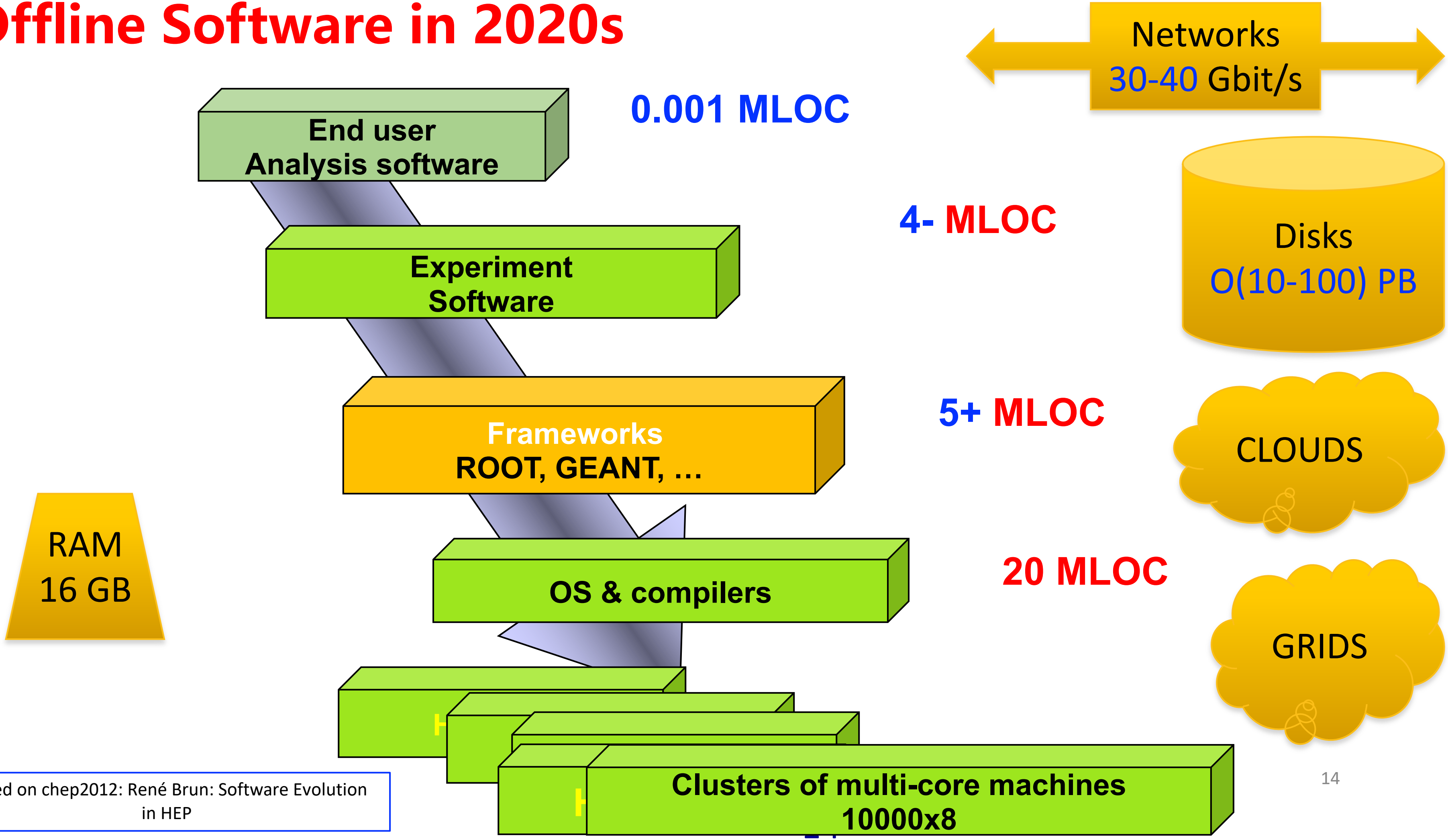


Based on chep2012: René Brun: Software Evolution in HEP

Offline Software in 2010s



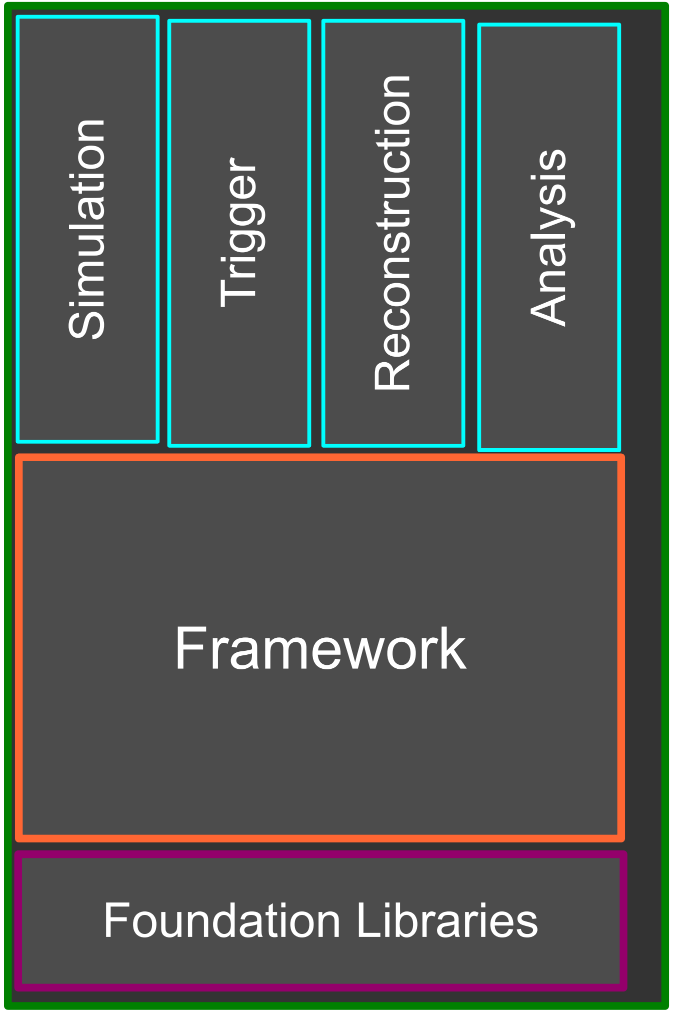
Offline Software in 2020s



Based on chep2012: René Brun: Software Evolution in HEP

Offline Software Organization

- Applications using framework components (Algorithms, Services, etc)
- Framework: Provides basic services, common interfaces, data exchange and persistency mechanisms, interactivity
- Basic libraries (STL, ROOT, GSL, CLHEP, AIDA, Python etc)



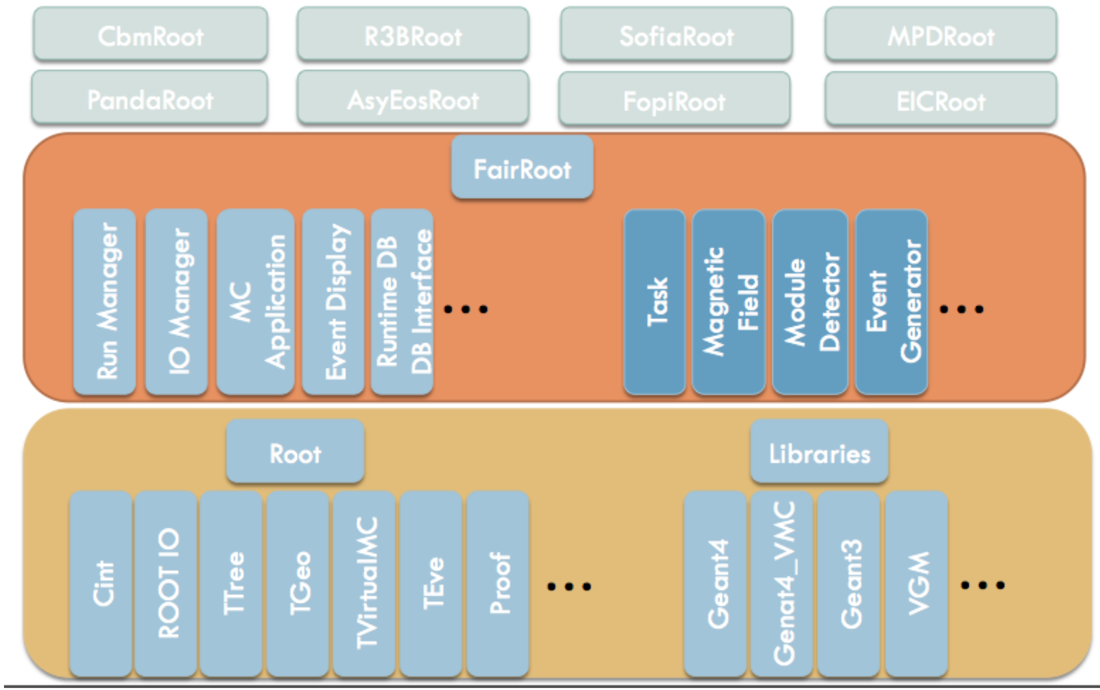
Main Frameworks

- FairRoot (Germany)
- BASF2 (Janpan)
- Art (U.S.)
- Gaudi /GaudiHive *(Cern)
  - BOSS (China)
  - NUWA (China)
  - CEPCSW (China) *
- SNiPER (China)
  - LodeStar (China)
  - OSCAR (China)
  - JUNOSW (China)
  - HERDOS (China)

FairRoot

M Al-Turany et al 2012 J. Phys.: Conf. Ser. 396 022001

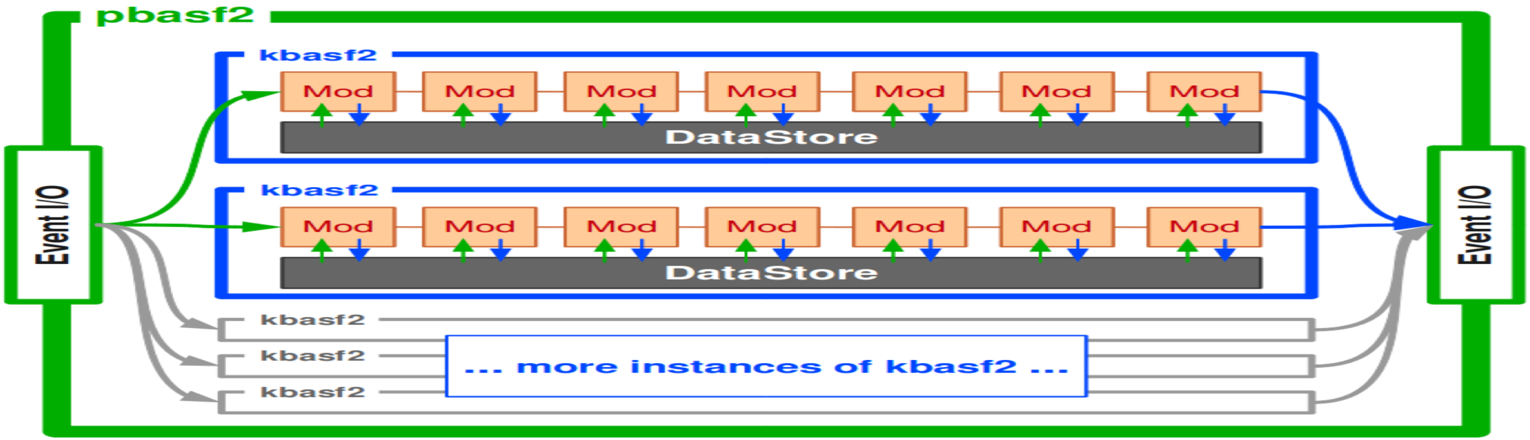
- Developed in 2003 by GSI-IT fully based on ROOT for CBM , now used by CBM, PANDA, R3B, ASYEOS
- Use plug-in mechanism from Root to load libraries
- Use a dynamic event structure based on Root TFolder and Ttree
- Task , TGeoManager, TEve, TSQLServer are used for management and services
- Root macros for the configuration



https://software.belle2.org/development/sphinx/index.html

BASF2

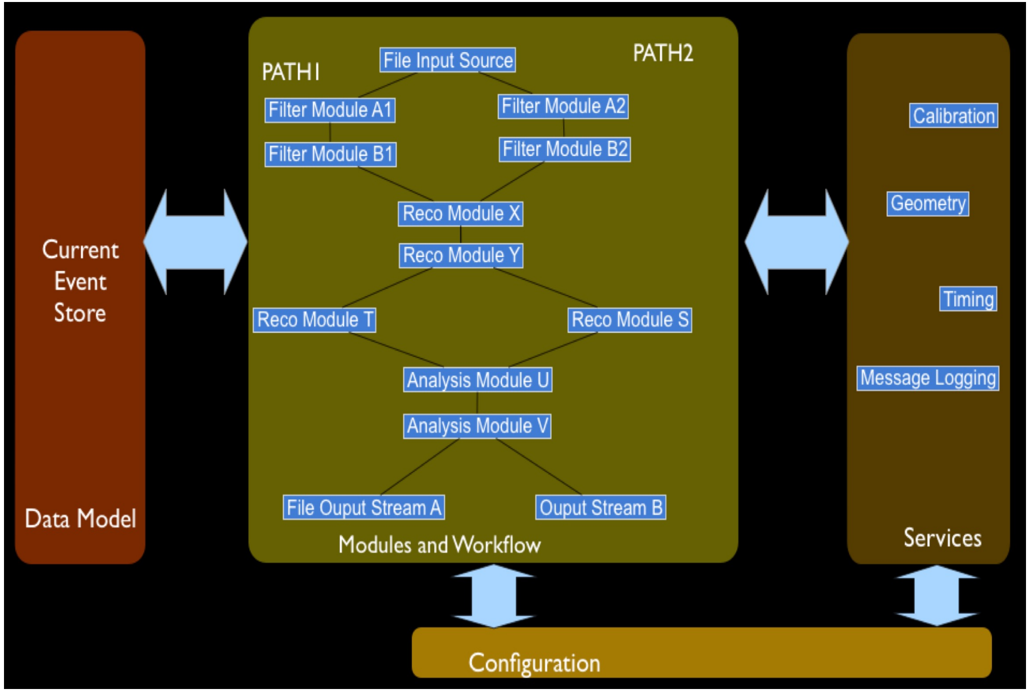
- Developed for Belle II for both online and offline
- Data are managed by ROOT and shared with DataStore
- ROOT IO for (de)serialisation of Objects
- Path is a linear arrangement of modules
- Multipaths (processes) are controlled with conditions
- Python for configuration



Art

https://art.fnal.gov

- Grew from CMSSW and used by g-2, Mu2e, NovA and LArSoft
- Modules include inputs, producers, filters, analyzers and outputs
- I/O and work schedule are handle by a state machine
- Products are managed with DataStore and shared between modules



GAUDI

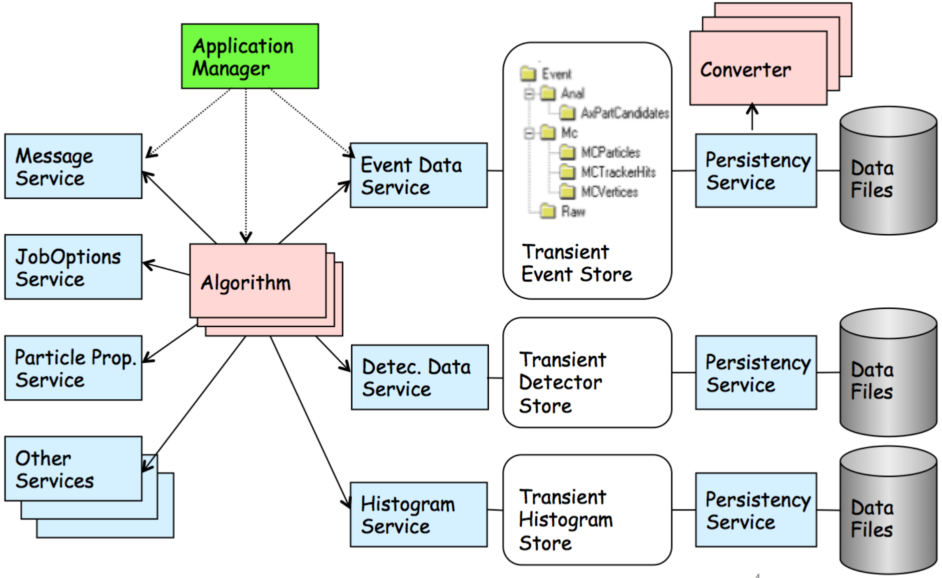
https://gaudi-framework.readthedocs.io/en/latest/
Computer Physics Communications 140 (2001) 45–55

- Developed by LHCb in 1999
  - A general data processing framework
  - Provide common and basic services and tools
- Used by ATLAS, Fermi, BESIII, Daya Bay, MINERVA,LBNE
- Advantages
  - Data Store-centered architectural style
  - Clear separation between data and algorithms
  - Clear separation between persistent data and transient data
  - Encapsulated user code
  - Dynamic loading libraries
  - Well defined generic interfaces

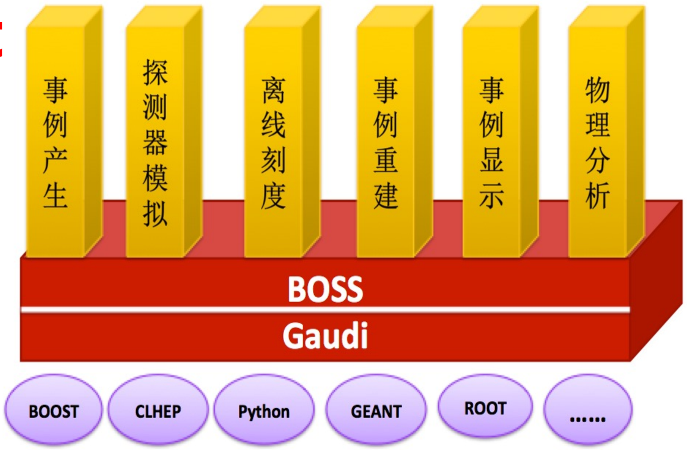
LHCb

GAUDI

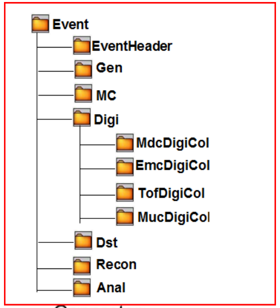
LHCb Data Processing Applications Framework

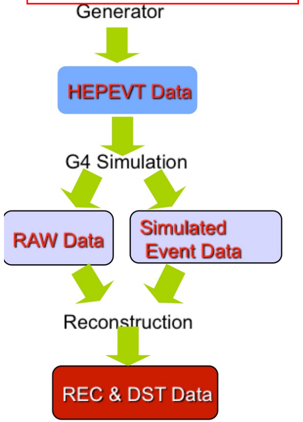


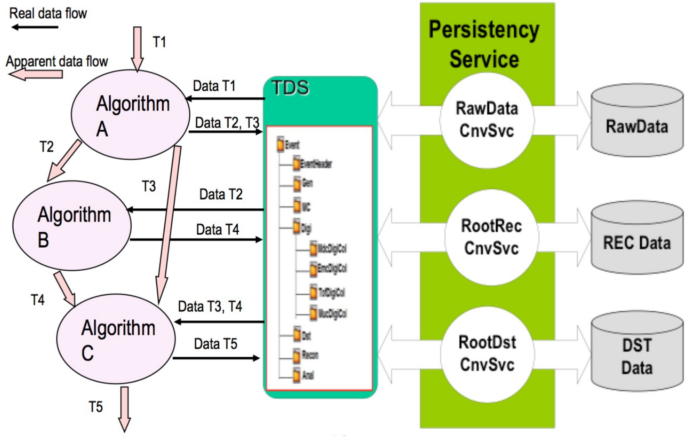
BOSS： BESIII Offline Software Syst



- BESIII Framework
- BESF based on Belle/BASF, LHCb/Gaudi ,BaBar/ProxyDict
- BOSS based on Gaudi in 2003
  - Event Data Model
  - File I/O
  - Services
  - Algorithms for Sim., Calib., Rec. and Analysis





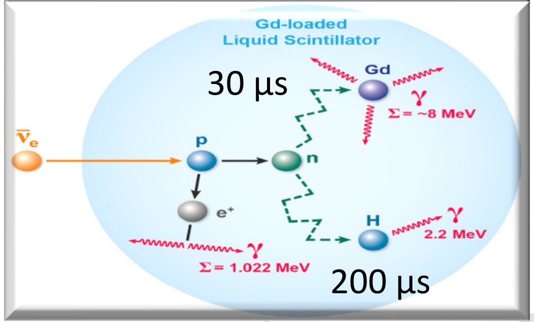


- China 1st Framework designed with OO and C++ for the whole offline data processing and analysis

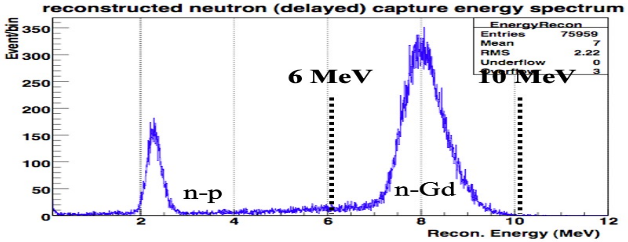
NuWa: Daya Bay Software System

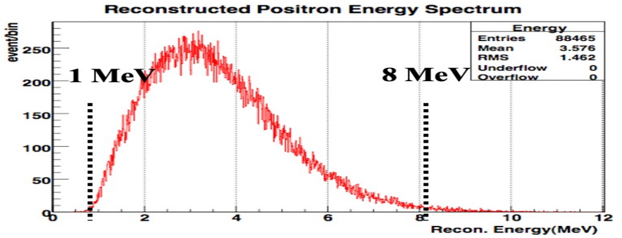
- Developed based on Gaudi in 2006
  - New challenge :Correlation Analysis
  - Event Data Model
  - Archive Event Store
  - Event objects relationship in diff. stages
  - Lightweight Analysis Framework

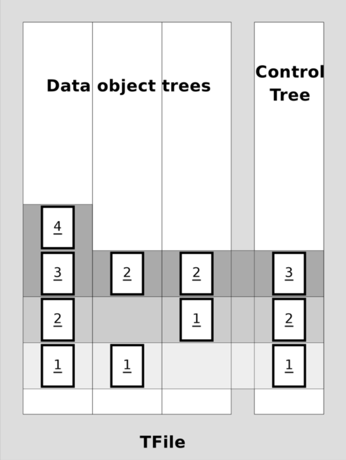
$$\bar{\nu}_e + p \rightarrow e^+ + n$$

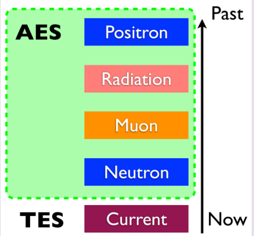


NuWa: Neutrino at Daya Wan



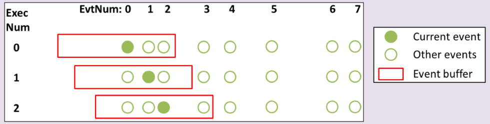






Extending Gaudi Transient Event Store (TES) to Archive Event Store (AES) for prompt-delayed analysis

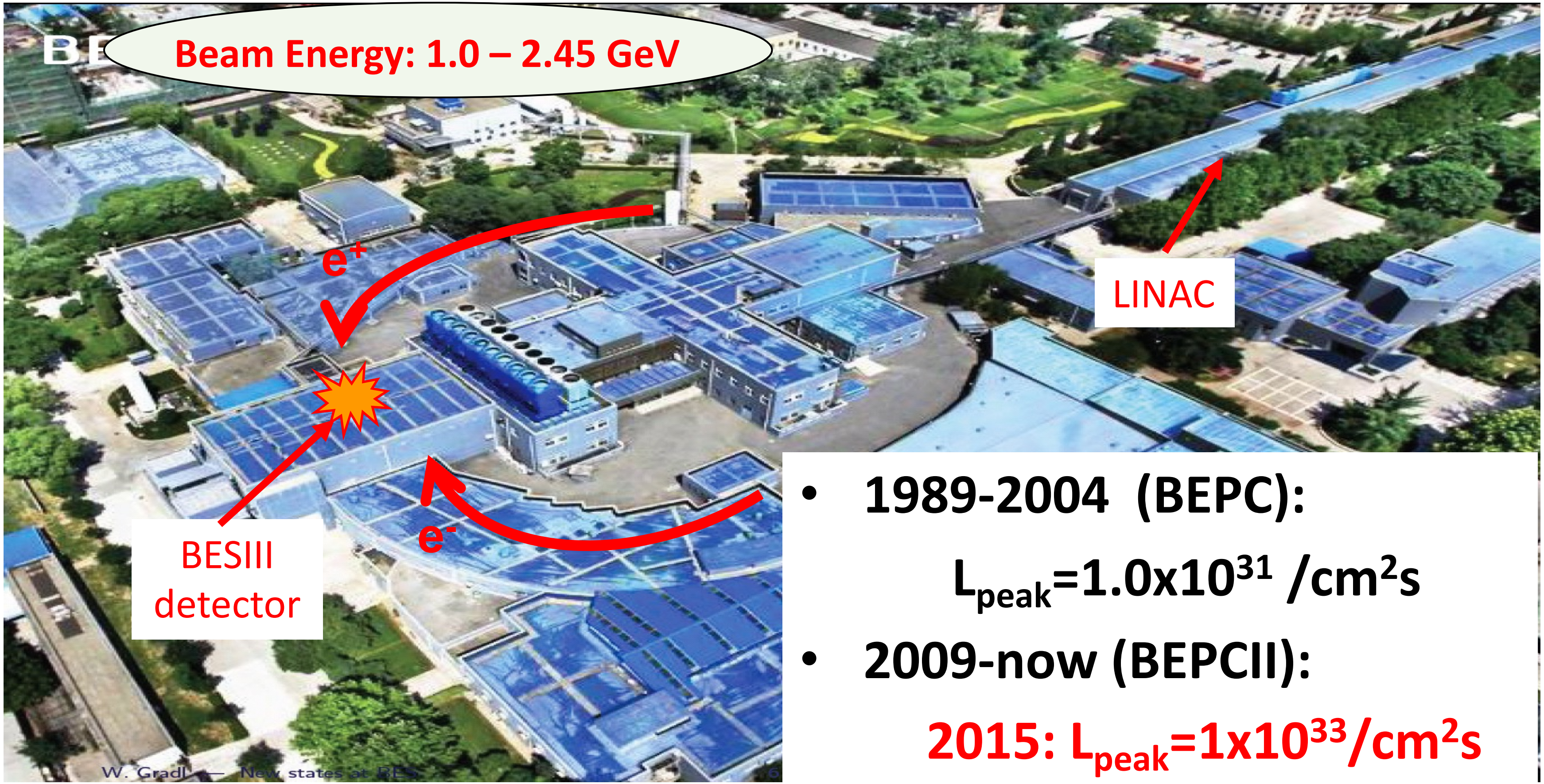
- Keeps data objects in memory across execution cycles.
- Allows users to look for correlated events in past.
- Configurable based on TES location.



A Lightweight Analysis Framework (LAF) with high I/O performance and flexible event buffer

BESIII Offline Software System

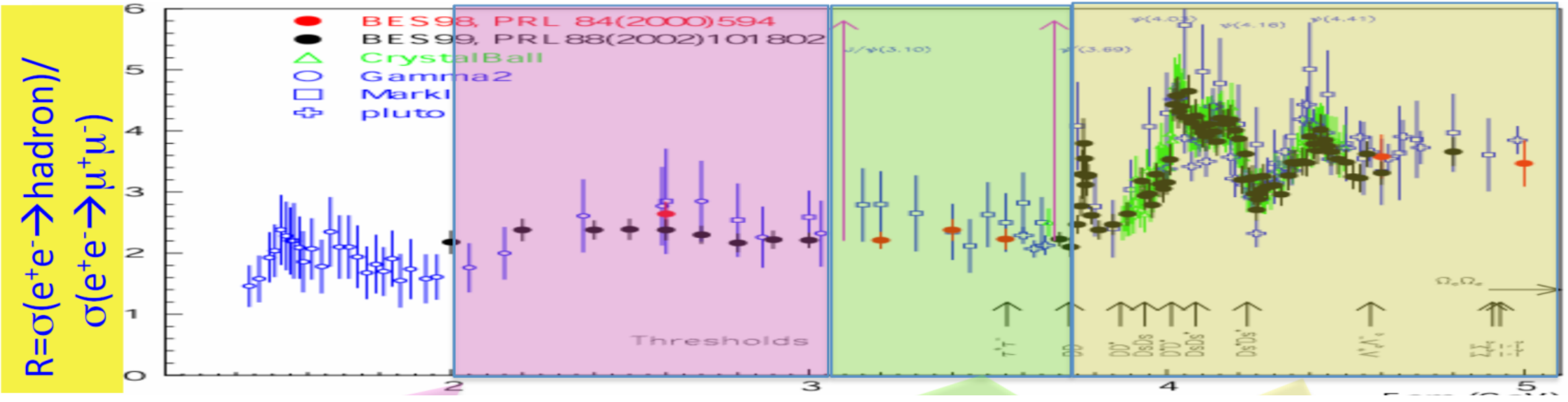
Beijing Electron Positron Collider II (BEPCII)



- 1989-2004 (BEPC):
  L_{peak}=1.0x10^{31} /cm^2s
- 2009-now (BEPCII):
  2015: L_{peak}=1x10^{33}/cm^2s

Physics at τ - charm Energy Region

- Rich of resonances: charmonia, charmed mesons, charmed baryons
- Threshold characteristics (pairs of τ, D, D_s, …) -- low BG at threshold, high X-section -- indirect probe of NP
- Transition between pQCD and non-pQCD
- Energy location of the new forms of hadrons



BESIII Data Samples

Data sets collected so far include

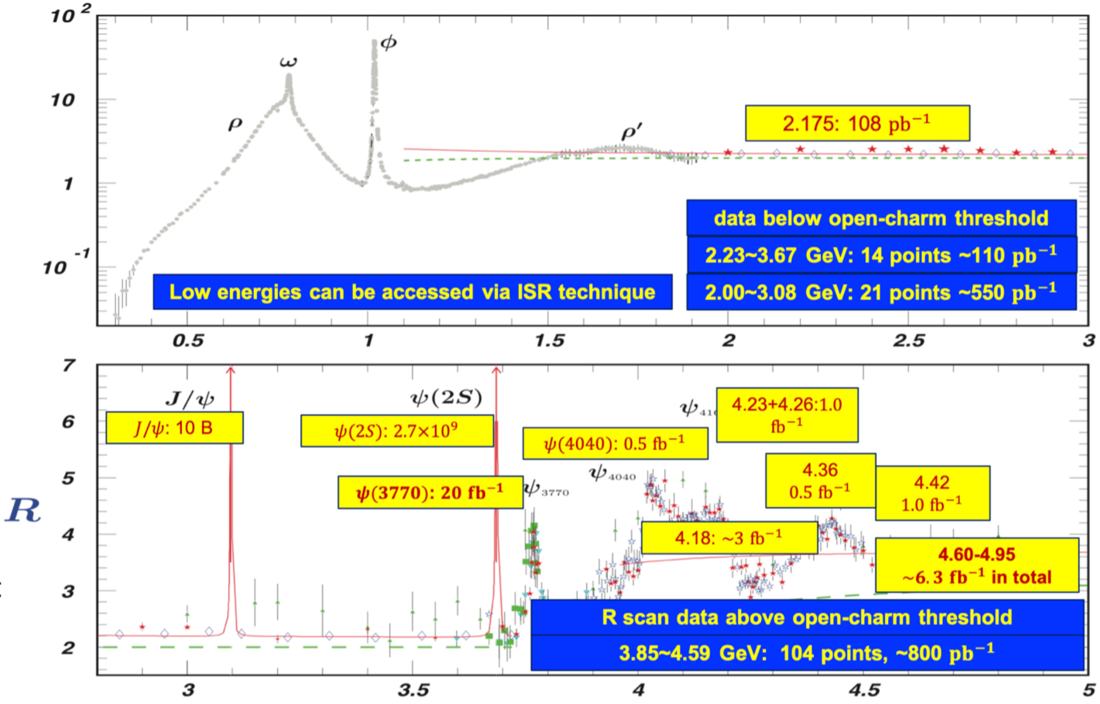
- 10×10^9 J/ψ events
- 2.7×10^9 $\psi(2S)$ events
- 17 fb^{-1} $\psi(3770)$
- Scan data between

  2.0 and 3.08 GeV,

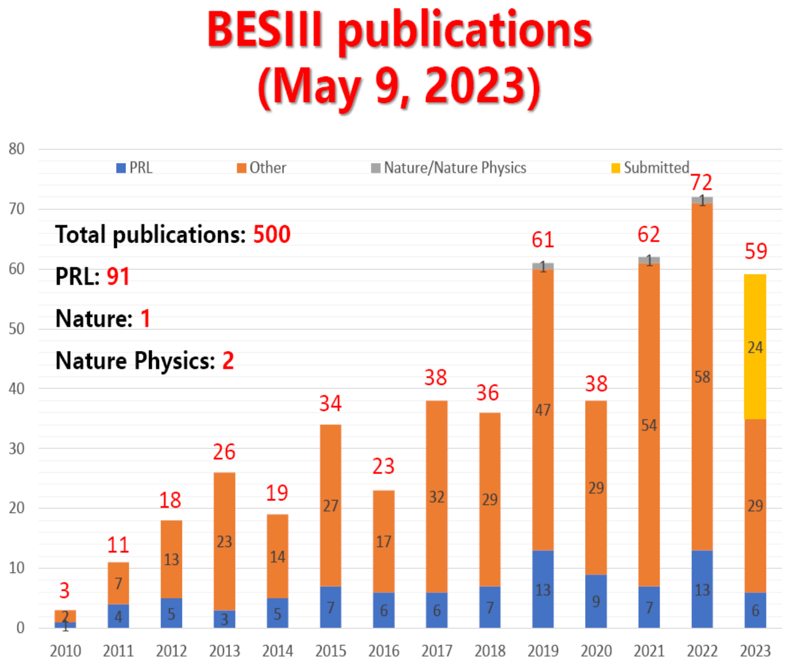
  and above 3.74 GeV
- Large datasets for XYZ studies:

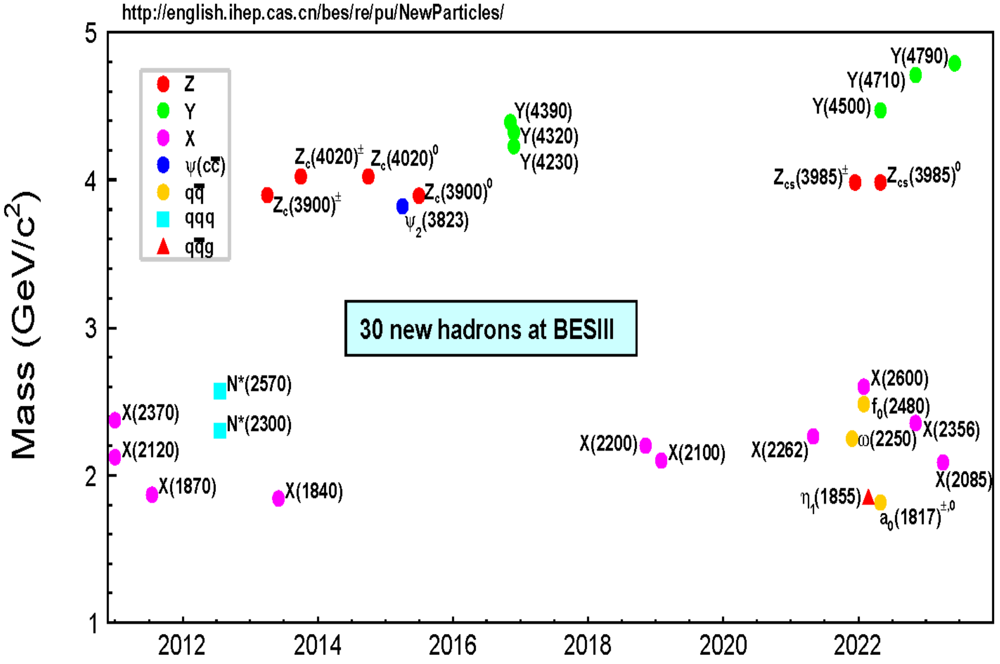
  scan with >500 pb-1 per energy point

  space 10 – 20 MeV apart



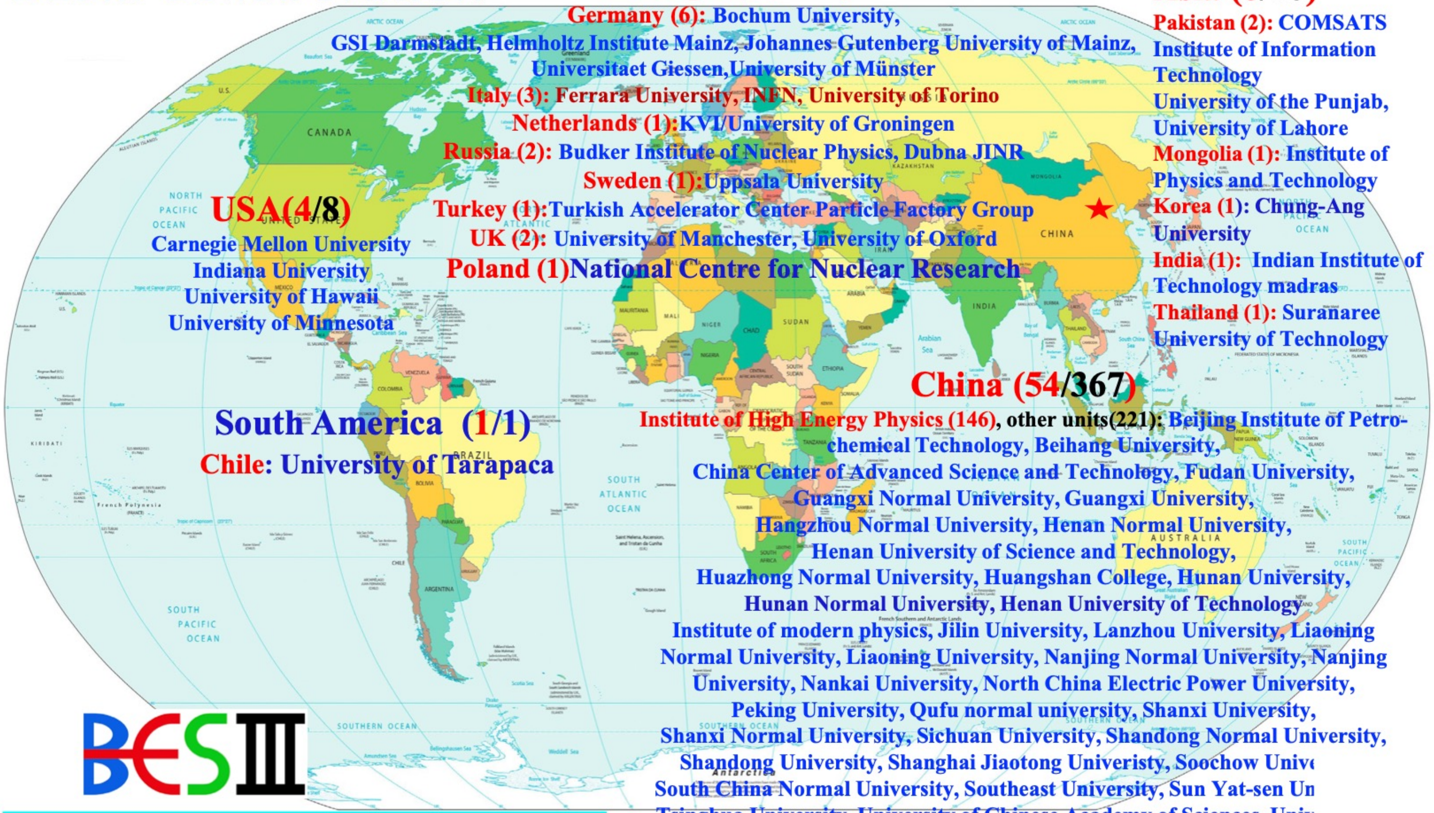
BESIII Physics





May 31， 2023: Celebration Ceremony of the 500 Publications of BESIII Collaboration

BESIII Collaboration



Europe (17/115)

Germany (6): Bochum University, GSI Darmstadt, Helmholtz Institute Mainz, Johannes Gutenberg University of Mainz, Universitaet Giessen, University of Münster

Italy (3): Ferrara University, INFN, University of Torino

Netherlands (1): KVI/University of Groningen

Russia (2): Budker Institute of Nuclear Physics, Dubna JINR

Sweden (1): Uppsala University

Turkey (1): Turkish Accelerator Center Particle Factory Group

UK (2): University of Manchester, University of Oxford

Poland (1) National Centre for Nuclear Research

Asia (6/10)

Pakistan (2): COMSATS Institute of Information Technology, University of the Punjab, University of Lahore

Mongolia (1): Institute of Physics and Technology

Korea (1): Chung-Ang University

India (1): Indian Institute of Technology madras

Thailand (1): Suranaree University of Technology

USA(4/8)

Carnegie Mellon University
Indiana University
University of Hawaii
University of Minnesota

South America (1/1)

Chile: University of Tarapaca

China (54/367)

Institute of High Energy Physics (146), **other units(221):** Beijing Institute of Petro-chemical Technology, Beihang University, China Center of Advanced Science and Technology, Fudan University, Guangxi Normal University, Guangxi University, Hangzhou Normal University, Henan Normal University, Henan University of Science and Technology, Huazhong Normal University, Huangshan College, Hunan University, Hunan Normal University, Henan University of Technology Institute of modern physics, Jilin University, Lanzhou University, Liaoning Normal University, Liaoning University, Nanjing Normal University, Nanjing University, Nankai University, North China Electric Power University, Peking University, Qufu normal university, Shanxi University, Shanxi Normal University, Sichuan University, Shandong Normal University, Shandong University, Shanghai Jiaotong Univeristy, Soochow Univ Soochow Univ South China Normal University, Southeast University, Sun Yat-sen Un Tsinghua University, University of Chinese Academy of Sciences, Univ Jinan, University of Science and Technology of China, University of Science and Technology Liaoning, University of South China, Wuhan University, Xinyang Normal University, Zhejiang University, Zhengzhou University, YunNan University , China University of Geosciences

BESIII

~500 members
From 82 institutions in 16 countries

BESIII Detector

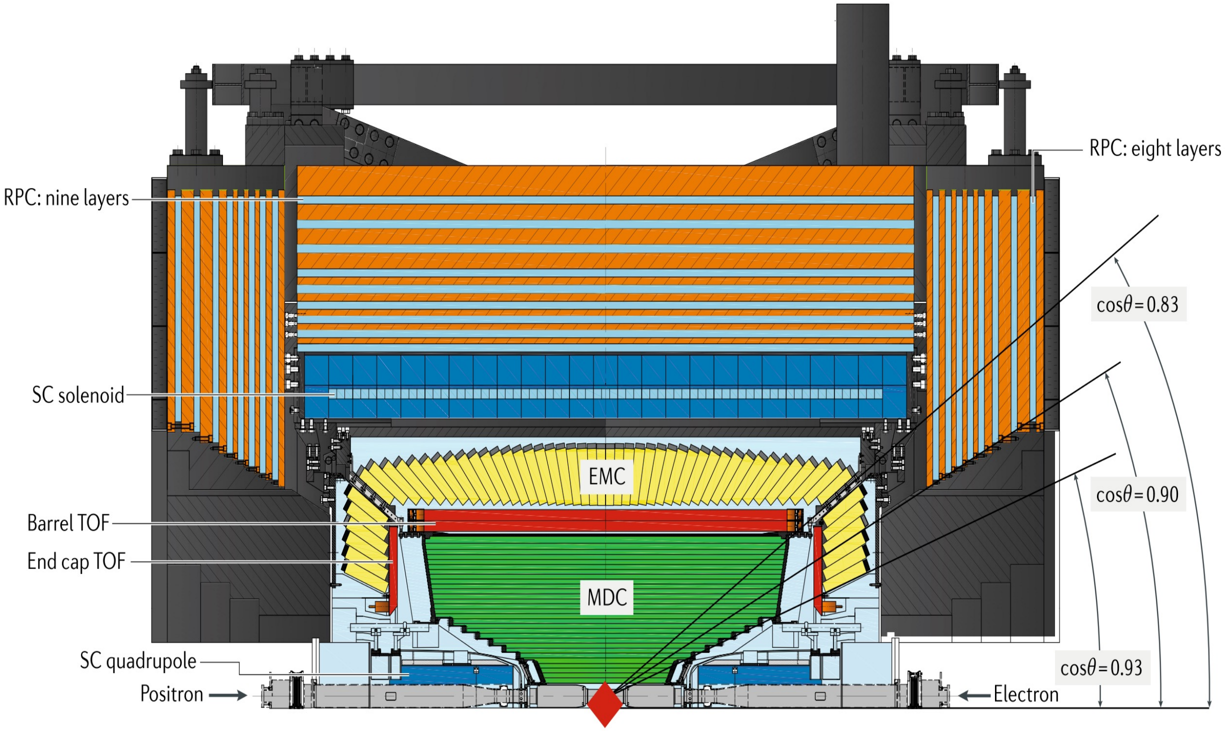
Super-Conducting Magnet
1.0 T (2009)
0.9 T(2012)

Muon Counter (MUC)
8 - 9 layers RPC
$\delta_{R\Phi} = 1.4\ \text{cm} \sim 1.7\ \text{cm}$

Main Drift Chamber (MDC)
$\sigma_P/P = 0.5\%$ (1 GeV)
$\sigma_{dE/dx} = 6\%$

Time of Flight (TOF)
σ_T: 90 ps (barrel)
60 ps (endcap)

Electromagnetic Calorimeter (EMC)
CsI (Tl)
$\sigma_E/\sqrt{E} = 2.5\%$ (1 GeV)
$\sigma_{z,\varphi} = 0.5 - 0.7\ \text{cm}/\sqrt{E}$



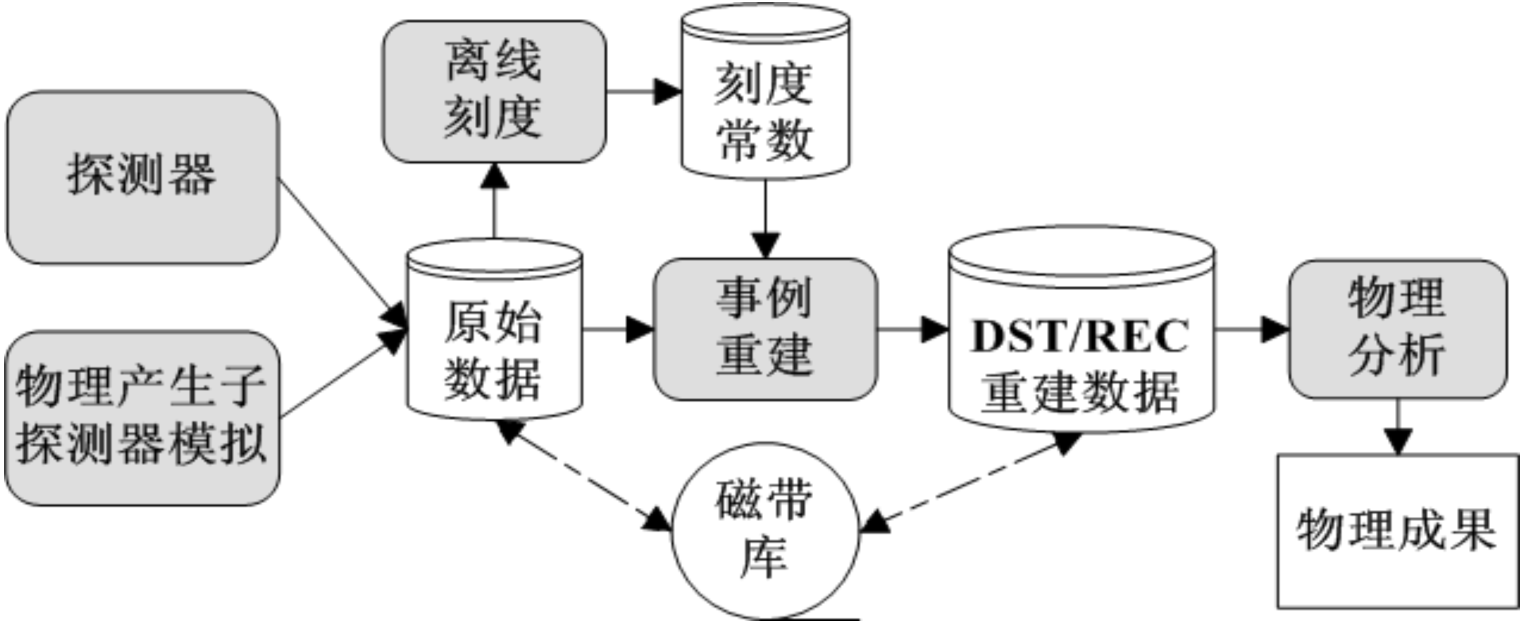
BESIII探测器性能指标

Exps.	MDC Spatial resolution	MDC dE/dx resolution	EMC Energy resolution
CLEOc	110 μm	5%	2.2-2.4 %
Babar	125 μm	7%	2.67 %
Belle	130 μm	5.6%	2.2 %
BESIII	115 μm	<5% (Bhabha)	2.4%

Exps.	TOF Time resolution
CDFII	100 ps
Belle	90 ps
BESIII	68 ps (BTOF) 60 ps (ETOF)

MUC:
Efficiency ~ 96%
BG level:
< 0.04 Hz/cm^2(B-MUC),
< 0.1 Hz/cm^2(E-MUC)

BESIII Offline Data Processing and Analysis



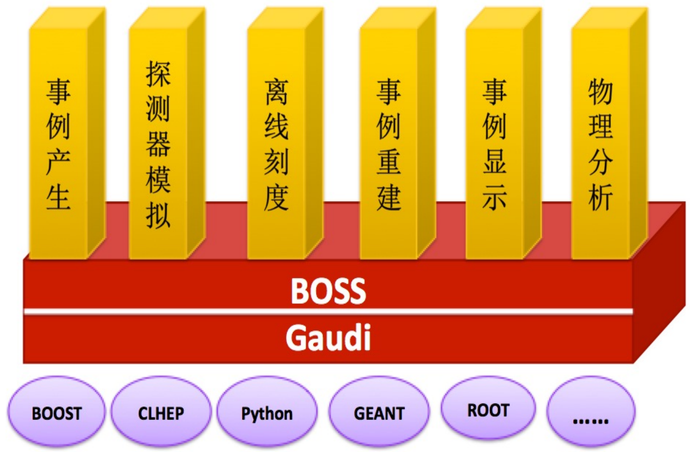
Offline Software for BESI,BESII,BESIII

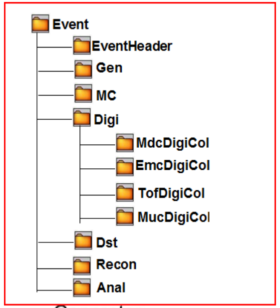
From Huaimin Liu

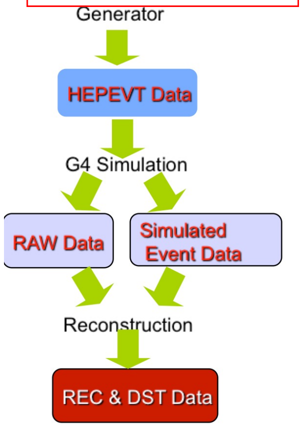
MARKIII	…	BESI	BESII	BESIII
MOAN	**…**	**DRUNK**	**DRUNK**	**BOSS**
GASP	**…**	**SOBER**	**SIMBES**	**BOOST**
EGS	…	*EGS*	*GEANT3*	*GEANT4*
F(M)ORTRAN	*…*	*FORTRAN*	*FORTRAN*	*C++*
1980s		*1980s*	*1990s*	*2000s*

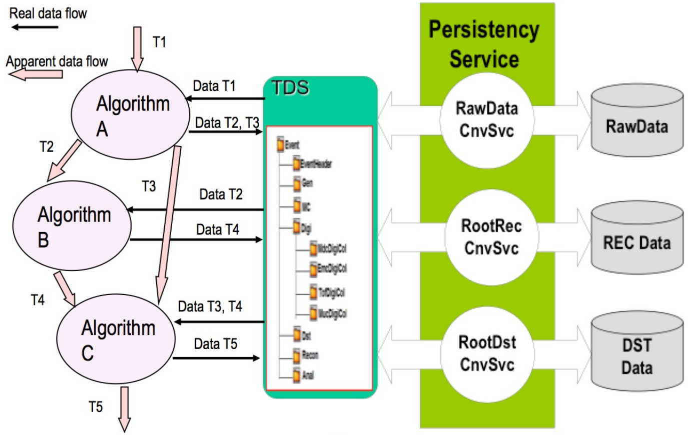
BOSS

- BESIII Framework
- BESF based on Belle/BASF, LHCb/Gaudi ,BaBar/ProxyDict
- BOSS based on Gaudi in 2003
  - Event Data Model
  - File I/O
  - Services
  - Algorithms for Sim., Calib., Rec. and Analysis
- China 1st Framework designed with OO and C++ for the whole offline data processing and analysis









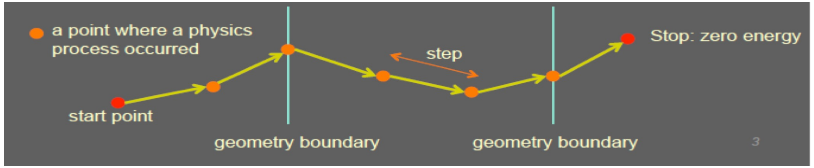
Detector Simulation

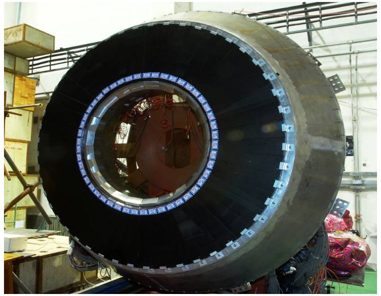
❖ very useful and essential procedure for HEP Experiment

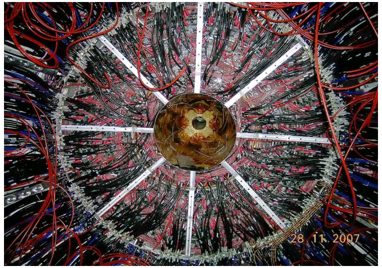
- designing an experiment
- analyzing the data

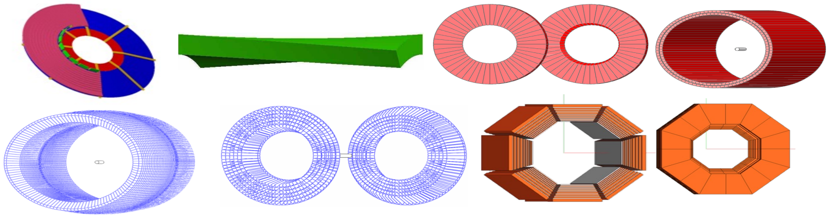
❖ key components

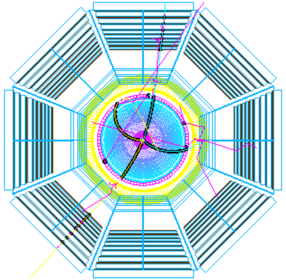
- Physics Generators
- Detector Description
- Physics processes
- Detector response → Hits -> Digis

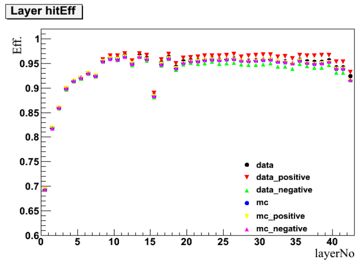


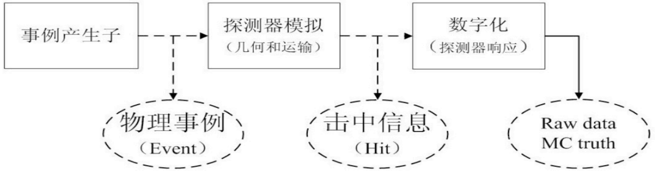




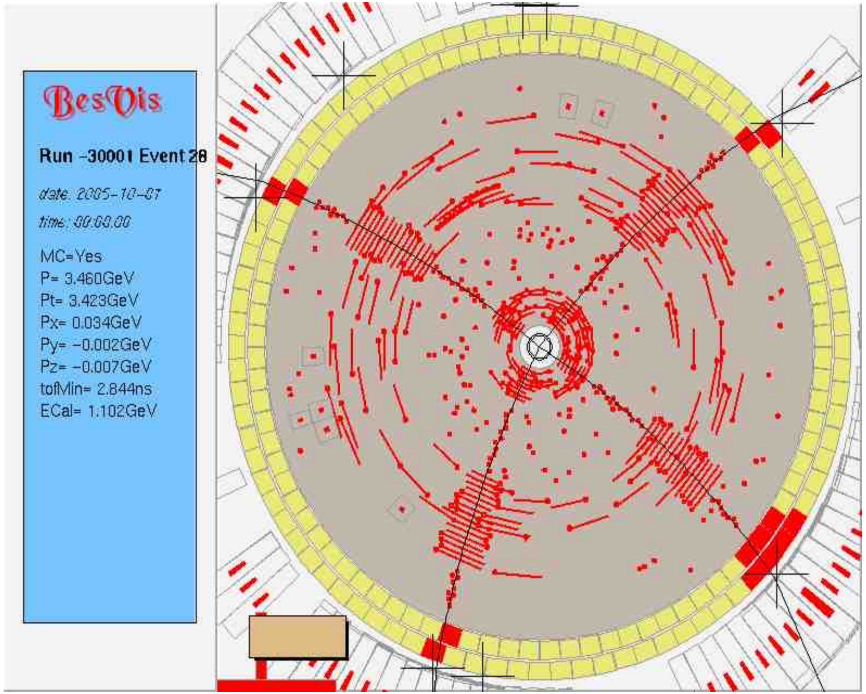


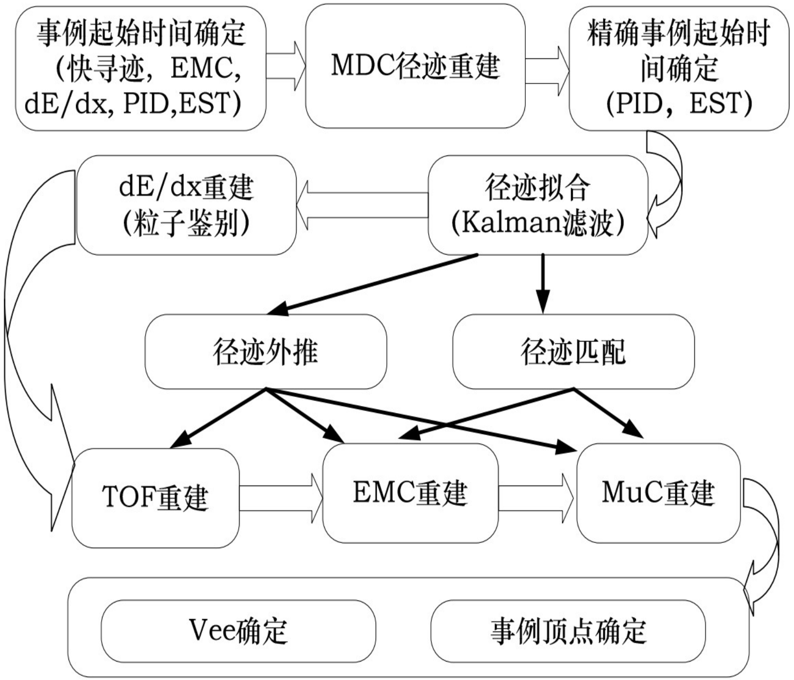






Event Reconstruction

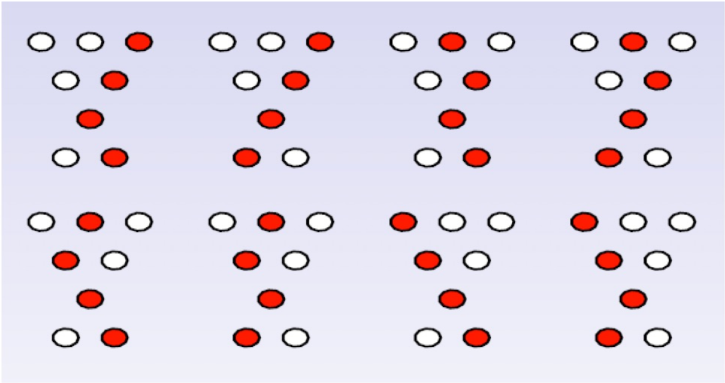


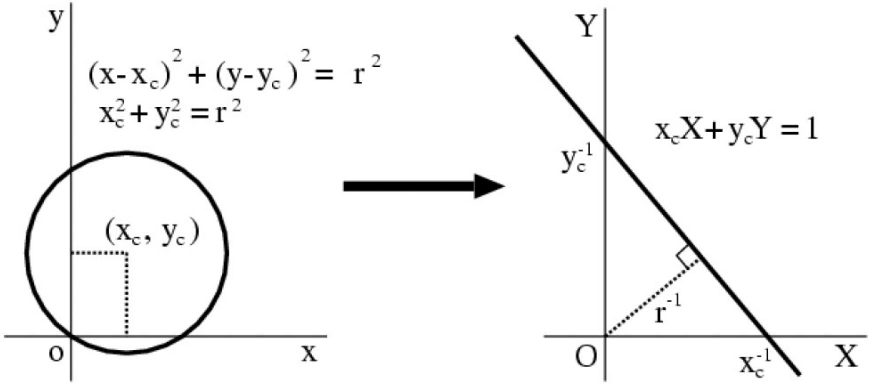


漂移室(MDC)重建

- 根据MDC原始击中进行径迹寻找和径迹拟合，计算带电粒子的动量和方向等物理信息。
  - 寻迹方法：模板匹配法、共形变换法、Hough变换法
  - 径迹拟合：最小二乘法、Kalman滤波法

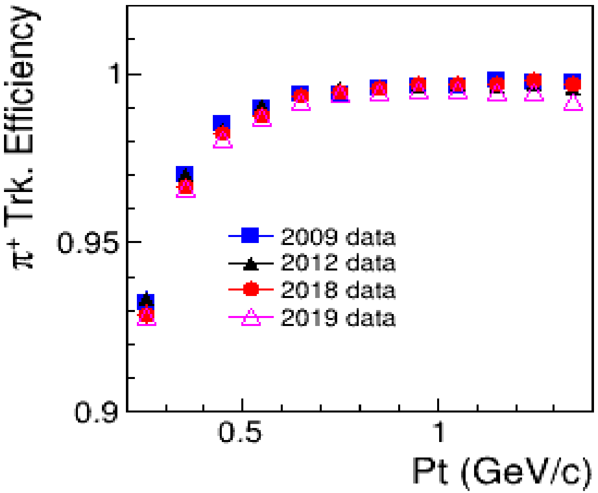
高能物理实验	空间分辨(μm)	dE/dx分辨
CLEOIII （美国）	110	5.0%
Babar （美国）	125	7%
Belle（日本）	130	5.6%
BESIII	135	5.0%



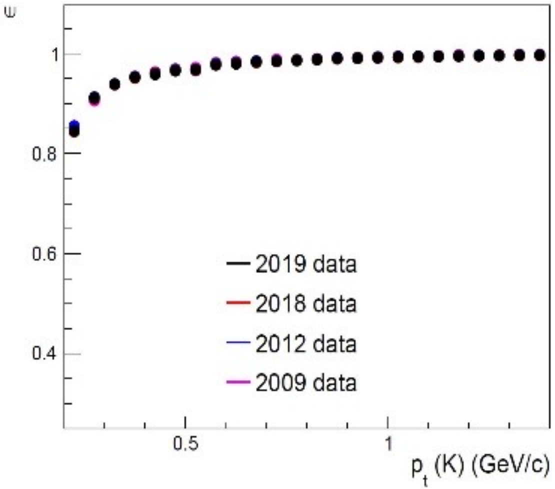


漂移室(MDC)重建

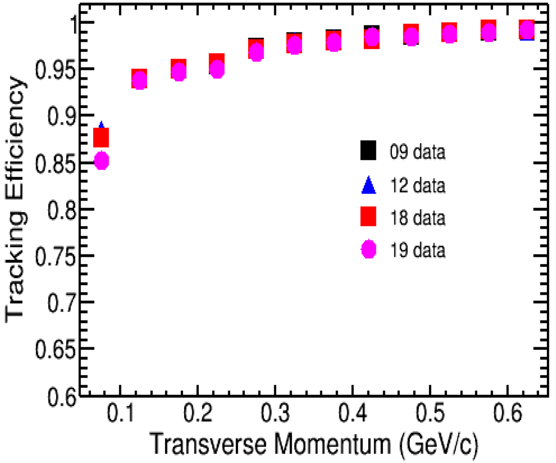
Pion



Kaon

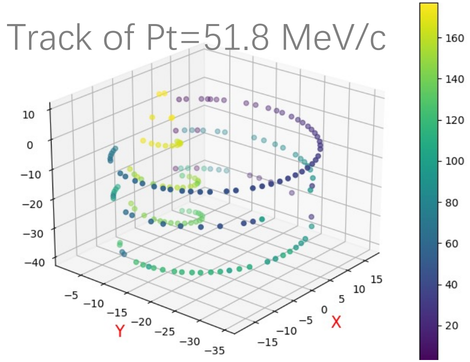


Proton

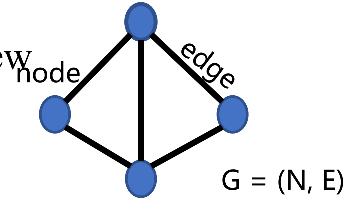


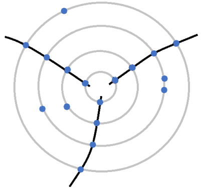
- 高横动量带电径迹重建效率接近100%
- 但低横动量（<200MeV）径迹的重建效率还有待提高
- 发展基于机器学习的径迹重建方法是进一步提高径迹重建效率一种可能方案

Tracking algorithm based on ML (1)

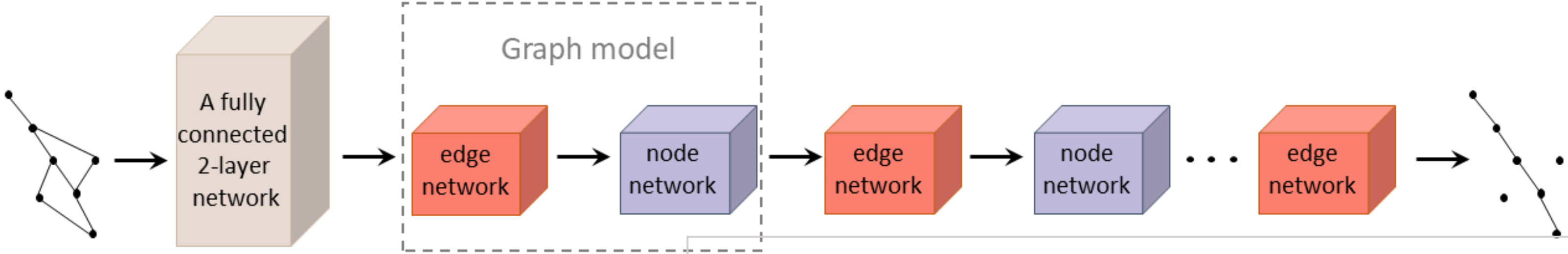


- With the upgrade of BEPCII, higher Background and noise will bring big challenges
- Also an optimizations is needed to further increase the tracking efficiency and performance for special events
  - Low transverse momentum, large dip angle, 2nd vertex …
- Lots of efforts on novel technologies e.g. GNN, DBSCAN…, to explore new tracking algorithms





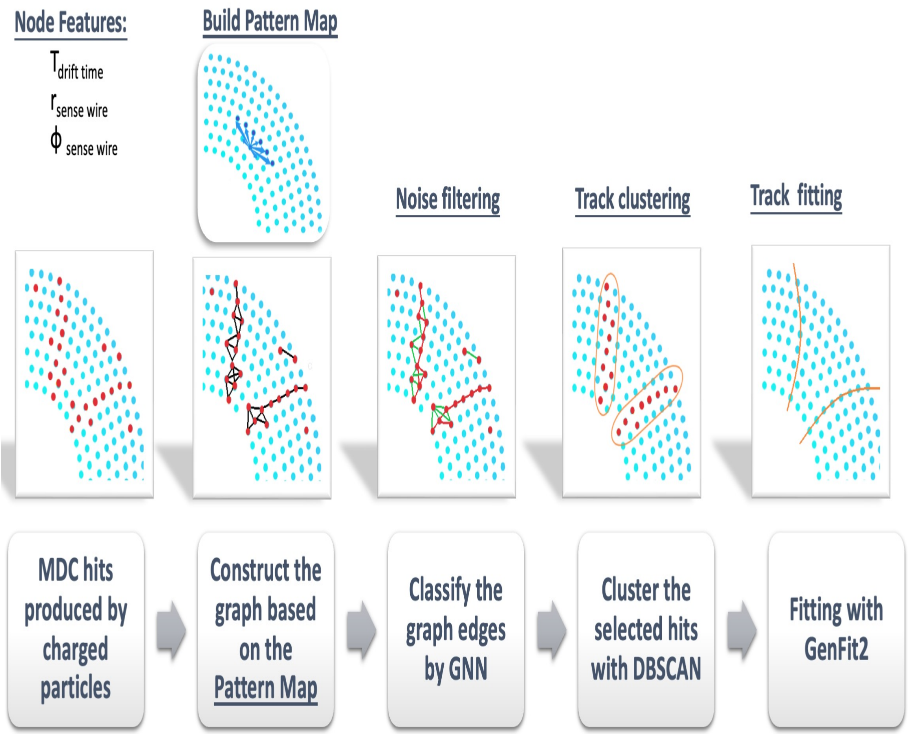
- A Graph model is build with Multi-Layer Perceptions (MLPs)
  - Edge network computes weights for edges using the features of the start and end nodes
  - Node network computes new node features using the edge weight aggregated features s of the connected nodes and the nodes' current features



CHEP2023: Xiaoqian Jia et al : BESIII track reconstruction algorithm based on machine learning

Tracking algorithm based on ML (2)

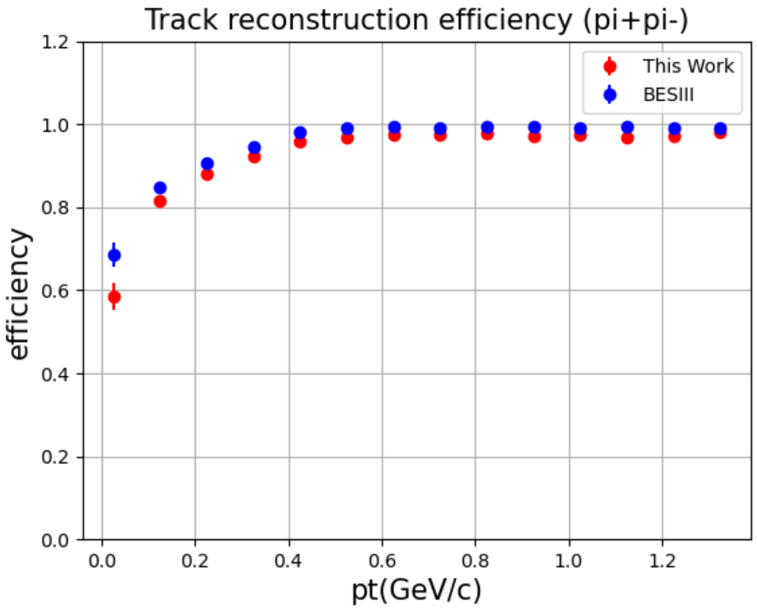
- A hit pattern map representing the connectivity between drift cells with large MC samples produced with BESIII offline software (BOSS)
  - Two million single for 5 types of charged particles ($e^{\pm}$, $K^{\pm}$, $\mu^{\pm}$, $p^{\pm}$, $\pi^{\pm}$)
  - 0.05 GeV/c < P < 3 GeV/c
- An edge-classifying Graph Neural trained to distinguish the hit-on-track from noise hits.
- A clustering method based on DBSCAN is developed to cluster hits from multiple tracks.
- A track fitting algorithm based on GENFIT2 is developed to obtain the track parameters, where DAF are implemented to deal with ambiguities and potential noises.

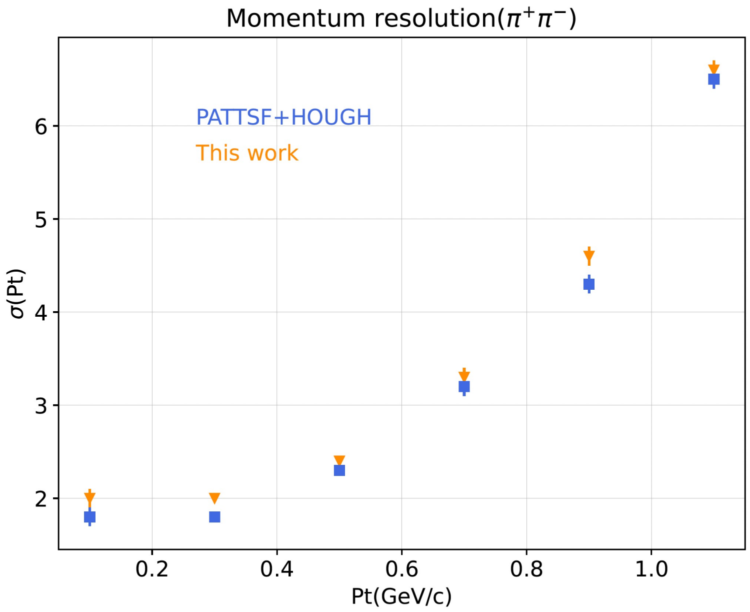


chep2023: Xiaoqian Jia et al : BESIII track reconstruction algorithm based on machine learning

Tracking algorithm based on ML (3)

- The preliminary results presents promising performance
  - $J/\Psi \rightarrow \rho^0 \pi^0 \rightarrow \gamma \gamma \pi^+ \pi^-$ from MC simulation

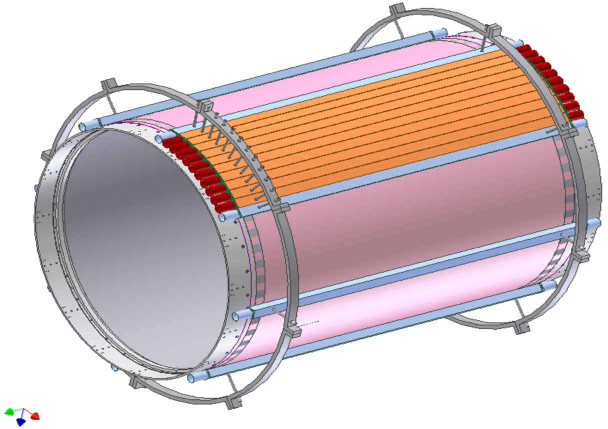


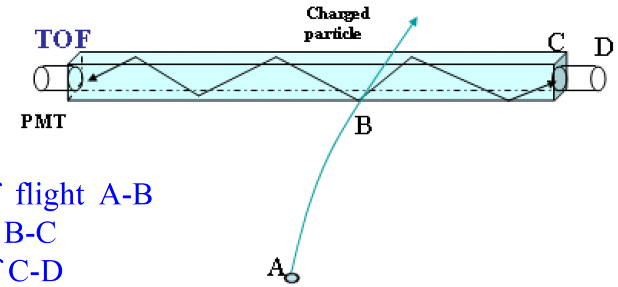


chep2023: Xiaoqian Jia et al : BESIII track reconstruction algorithm based on machine learning

飞行时间计数器(TOF)重建

❖ 计算粒子的飞行时间等物理量，进行粒子鉴别。
  - TOF测量的带电粒子的原始时间和脉冲幅度
  - 漂移室重建得到的带电径迹的动量以及通过径迹外推得到的击中位置，飞行距离等信息以及事例起始时间等





T_{tof}: Time of flight A-B
T_{pro}: Time of B-C
T_{pmt}: Time of C-D
T_{elc}: delay time at electronic

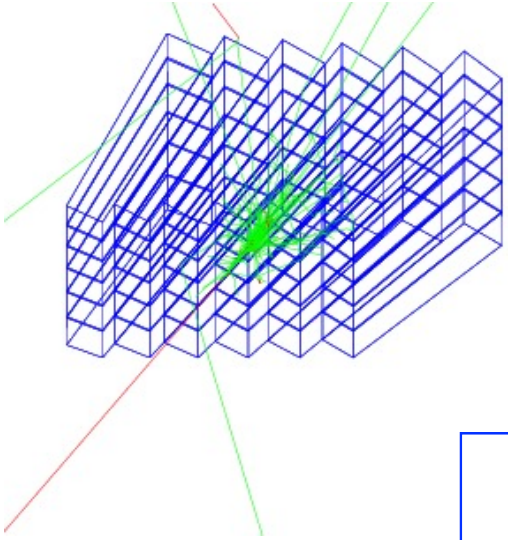
$$T_{est} = TDC - T_{ev}$$

$$T_{ev} = T_{tof} + T_{pro} + T_{pmt} + T_{elc}$$

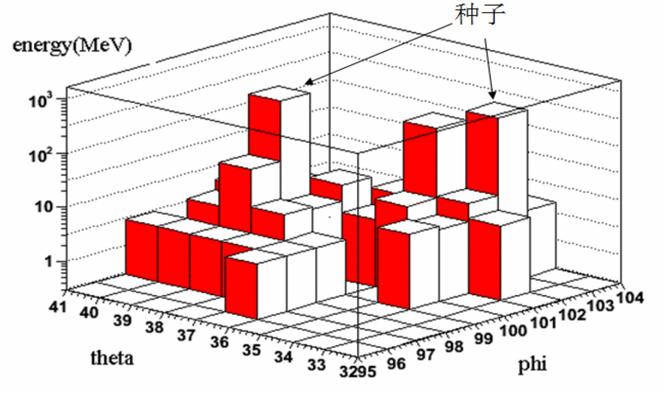
Experiments	TOF Time resolution
CDFII	100 ps
Belle	90 ps
BESIII	68 ps (BTOF) 60 ps (ETOF)

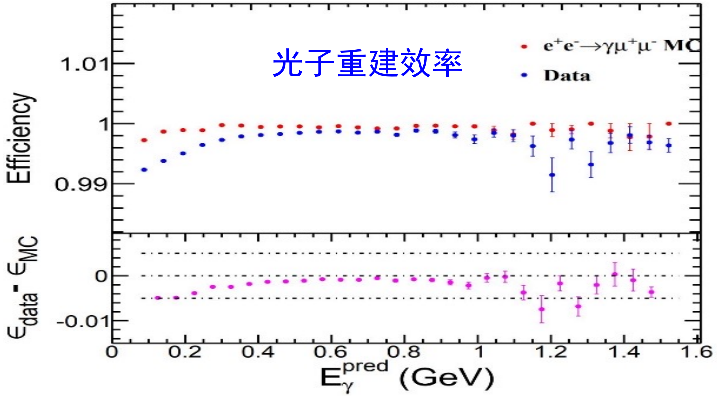
量能器(EMC)重建

❖ 通过寻找电磁簇射在晶体中沉积能量形成的簇团，计算出入射粒子的总能量和击中位置。



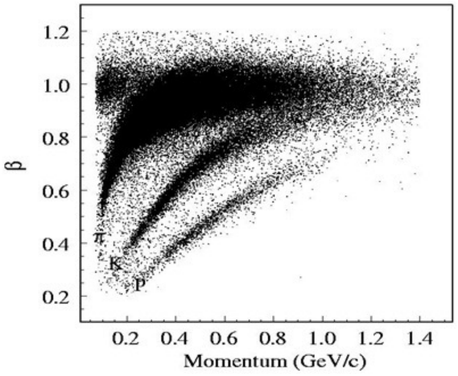
Cluster search → **Seed search** → **Clusters split into Showers** → **Calculation and correction of shower energy and position**

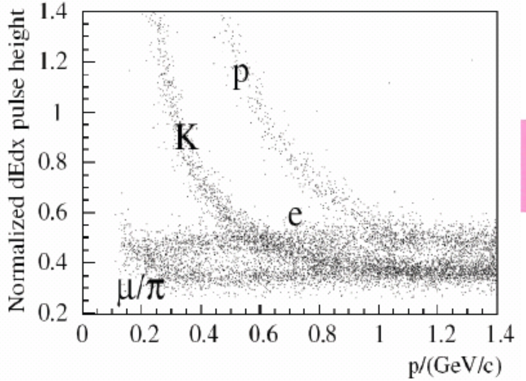


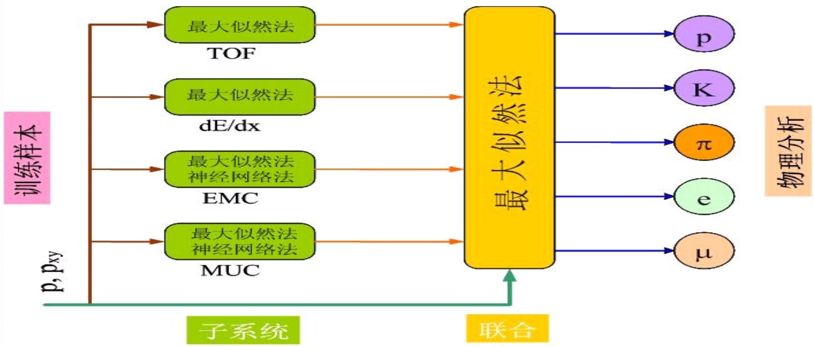


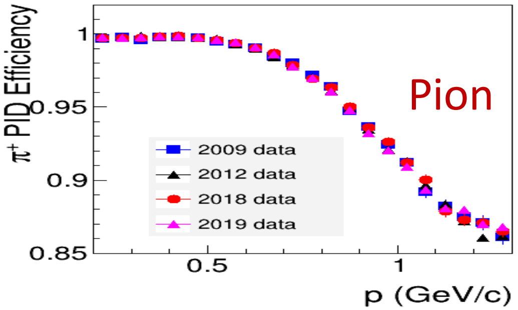
粒子鉴别（PID）

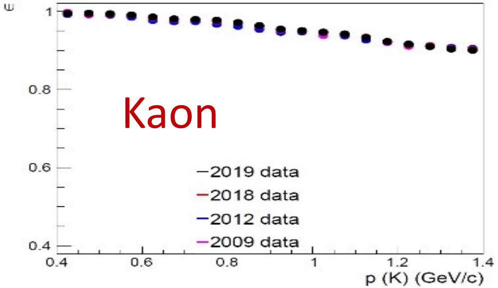
❖ 通过综合利用各子探测器给出的特征物理量，对粒子进行分类

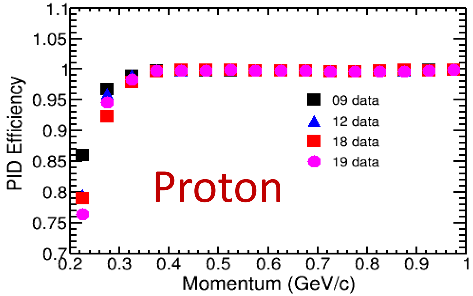










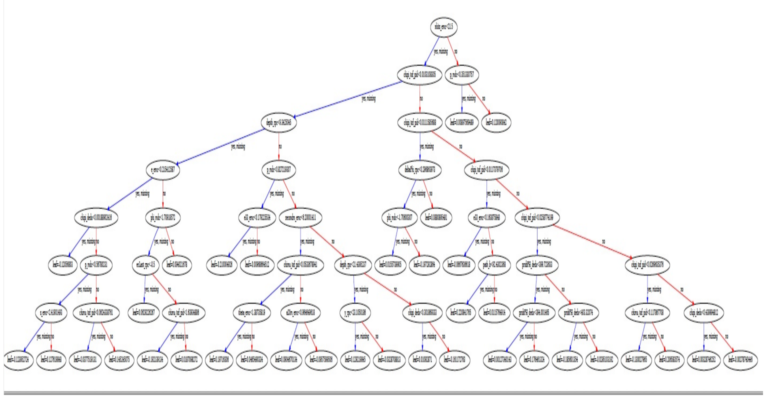


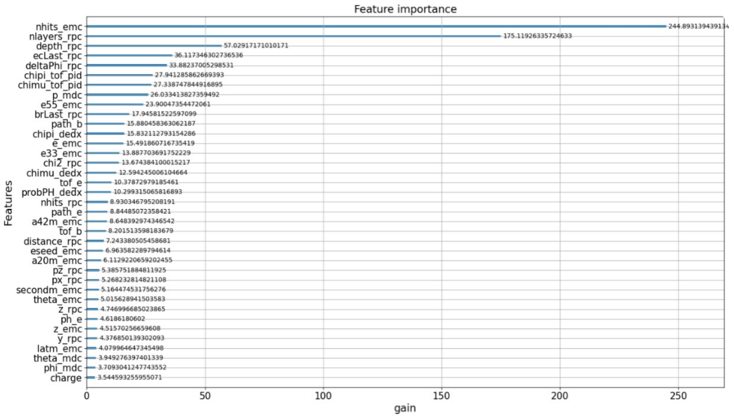
❖ 利用先进的机器学习/深度学习技术提高粒子鉴别效率

New Muon/Pion PID algorithm based on the BDT

❖ BDT model

- Based on XGBoost with the max_depth （8), n_estimators (300)
- All 108 features from MDC, dE/dX, TOF, EMC, MUC information
- Finally 37 features selected according to their importance.





❖ Data sample with high purity and good distribution

- MC/data: $J/\psi \to \pi^+\pi^- \pi^0 \to \pi^+\pi^-\gamma\gamma$ (P = 99.37%)

  $J/\psi \to \gamma\, \mu^+\mu^-$ (P = 97.97%)

- Different processes:

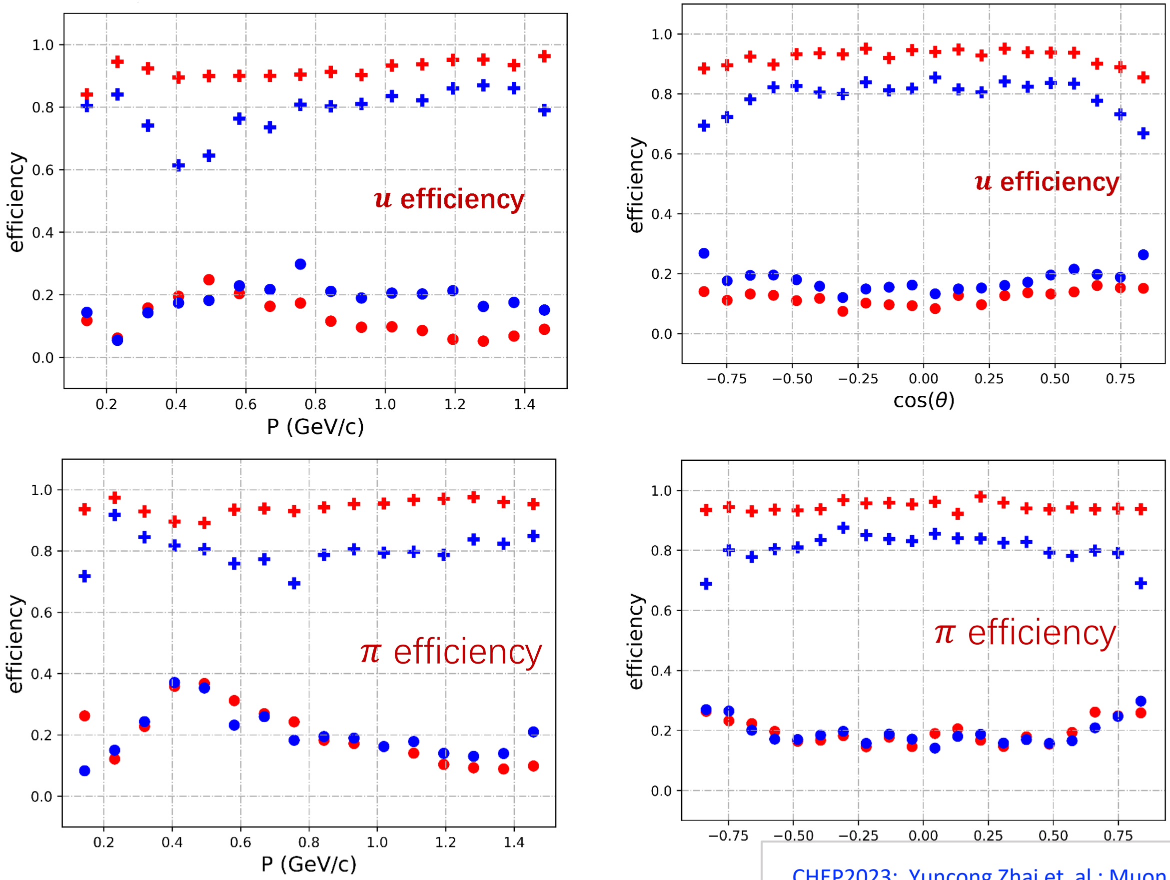
  $\psi(2s) \to \pi^+\pi^- J/\psi \to \pi^+\pi^- \mu^+\mu^-$ (P = 99.13%)

- 0.1 GeV/c < p < 1.5 GeV/c (bin numbers :16)

  abs(**cosθ**)< 0.88 (bin numbers :20)

CHEP2023: Yuncong Zhai et al : Muon/Pion Identification Based on Machine Learning Algorithm at BESIII

New Muon/Pion PID algorithm based on the BDT



Signal efficiency

XGboost model | Traditional PID algorithm

Background efficiency

XGboost model | Traditional PID algorithm

- *Signal efficiency* :

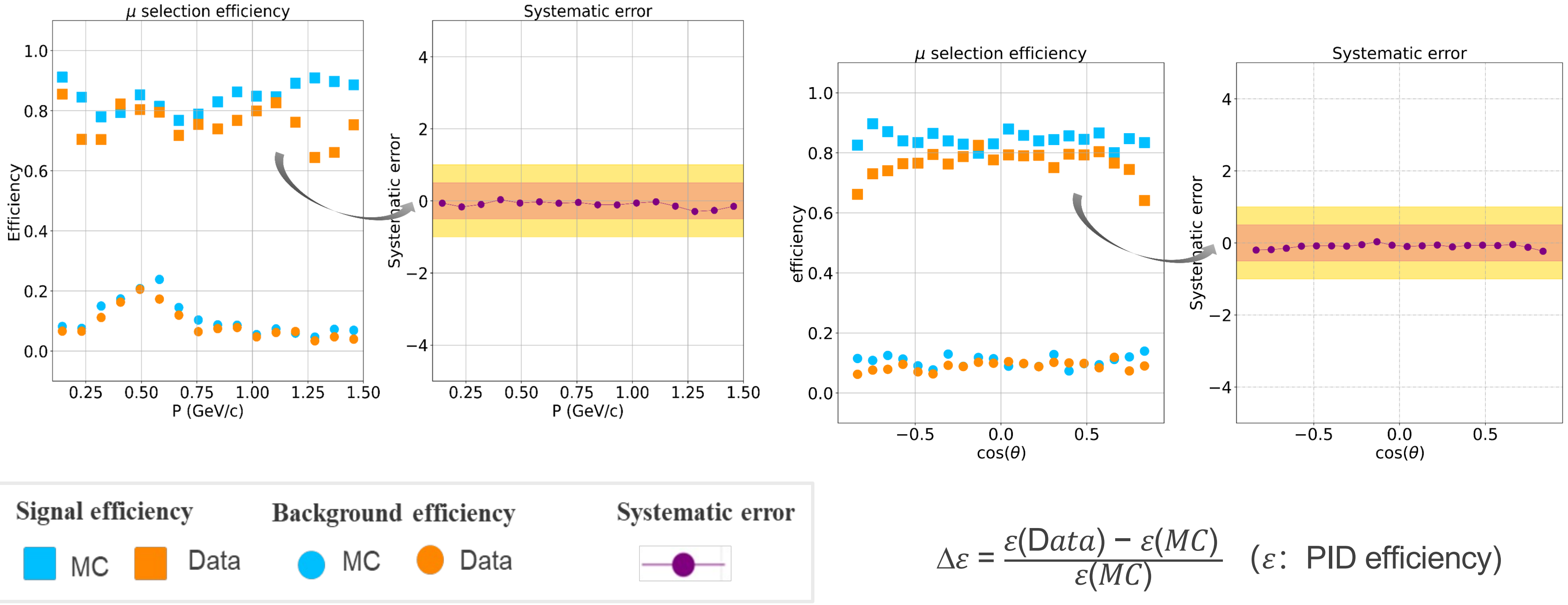
$$\frac{\text{Number } of \text{ signal } selected\ correctly}{Total\ number\ of\ \text{signal}}$$

- *Background efficiency* :

$$\frac{number\ of\ background\ \text{selected}\ l}{Total\ number\ of\ background}$$

CHEP2023: Yuncong Zhai et al : Muon/Pion Identification Based on Machine Learning Algorithm at BESIII

New Muon/Pion PID algorithm based on the BDT

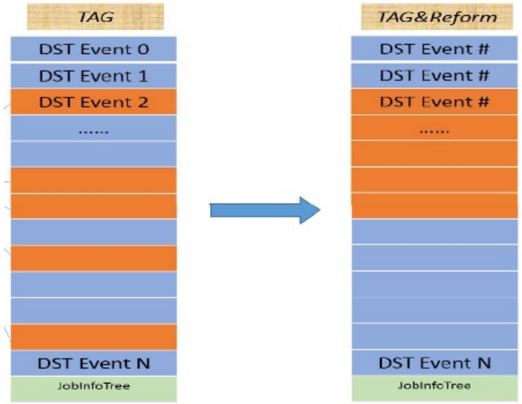


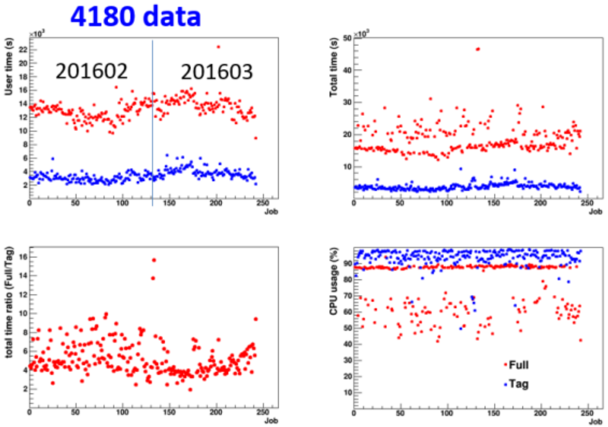
$$\Delta\varepsilon = \frac{\varepsilon(Data) - \varepsilon(MC)}{\varepsilon(MC)} \quad (\varepsilon: \text{ PID efficiency})$$

CHEP2023: Yuncong Zhai et al : Muon/Pion Identification Based on Machine Learning Algorithm at BESIII

分析技术：事例特征标记

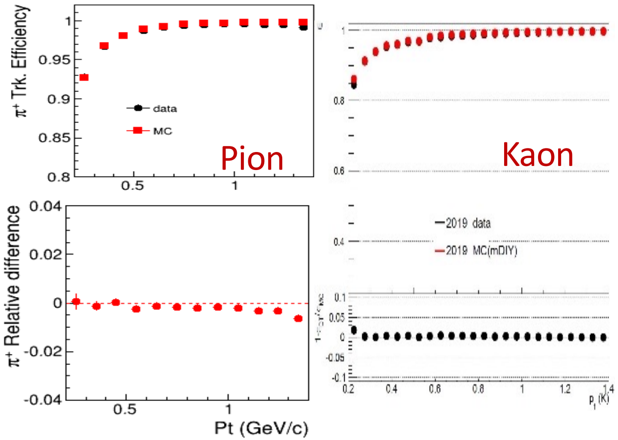
- 高统计量实验数据对计算资源的使用效率提出了更高的要求
- 频繁的磁盘访问是目前限制计算效率的瓶颈问题，随着数据的不断获取，将日益严重
- 对特定衰变过程的事例选择，重建数据文件中的绝大部分事例通过简单选择条件就可以排除，如带电径迹数目，粲介子或粲重子的标记等
- 事例特征标记的重建数据提供了一种不通过访问磁盘获得事例特征的方法
- 结果显示，基于事例特征标记的方法能够大幅加速物理分析速度



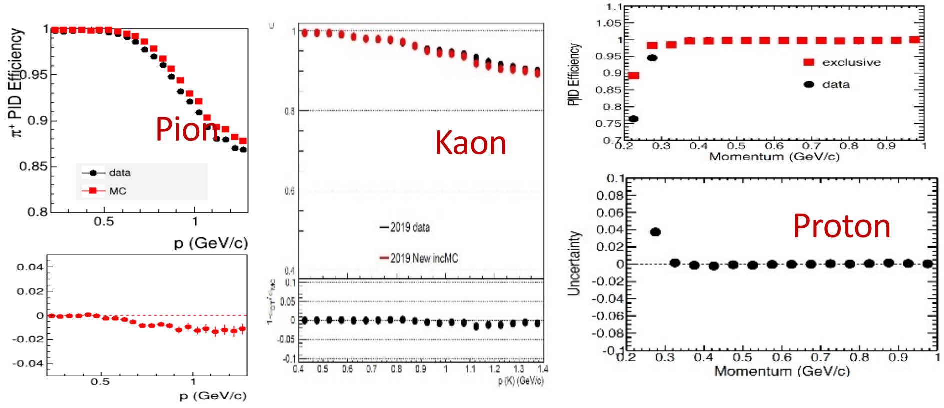


系统误差

- BESIII已经获取和即将获取的前所未有的高统计量实验数据大幅降低了物理测量结果的统计误差。进一步降低系统误差达到与统计误差相匹配的需求十分迫切
- 发挥BESIII实验高统计量的优势，获得随粒子种类、电荷、横动量和出射角度依赖的效率修正因子，实现系统误差 <0.5%的目标。



带电径迹重建效率系统误差

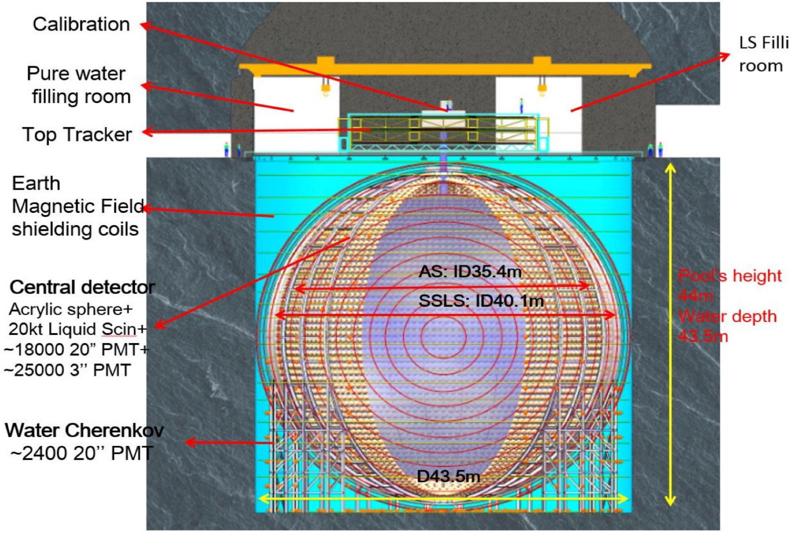


粒子鉴别效率系统误差

JUNO Offline Software System

Jiangmen Underground Neutrino Observatory (JUNO)

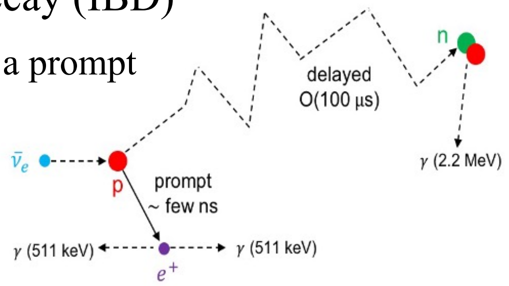
- JUNO is a multiple-purpose neutrino experiment currently under construction in southern China
  - Large fiducial volume: 20 kton liquid scintillator detector
  - Excellent energy resolution :3% energy resolution at 1MeV
  - Rich physics programs:
    - Reactor neutrinos: Mass Hierarchy and precision measurement of oscillation parameters
    - Supernova neutrinos
    - Geoneutrinos
    - Solar neutrinos
    - Atmospheric neutrinos
  - Status:
    - Finished civil construction in Dec. 2021
    - Detector assembling goes smoothly now
    - Expected to start data taking in 2024



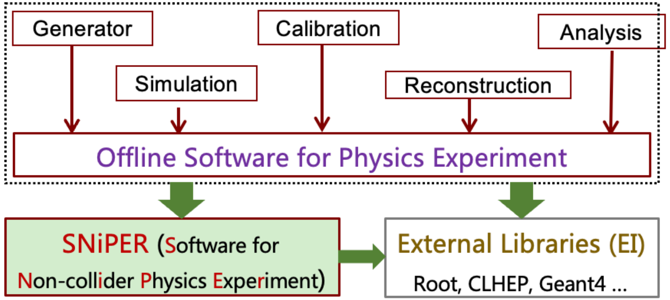
AS: Acrylic sphere; SSLS: stainless steel latticed shell

JUNO Software (JUNOSW)

- Data processing and analysis of neutrino experiments is different from the one of the collider experiments, i.e. reactor neutrino is detected with the Inverse Beta Decay (IBD)
  - M.C. production: Event splitting and mixing -> one neutrino event splits a prompt and a delayed event -> complicate relations
  - Physics analysis : time and vertex coincidence between the prompt and delayed events -> event buffering
  - Large data volume and rare signal: 2 PB/year * 20 years; only ~ 60 reactor neutrinos per day



- JUNOSW (originally called as JUNO offline) has been developed based on NuWa (the offline software of the Daya Bay experiment) from scratch since 2012.
  - Core Software: Framework, Event Data Model, Geometry Management, Database System, Analysis Framework
  - Applications: simulation, calibration, reconstruction and analysis
  - Modern management technologies: C++17, python3, cmake, gitlab

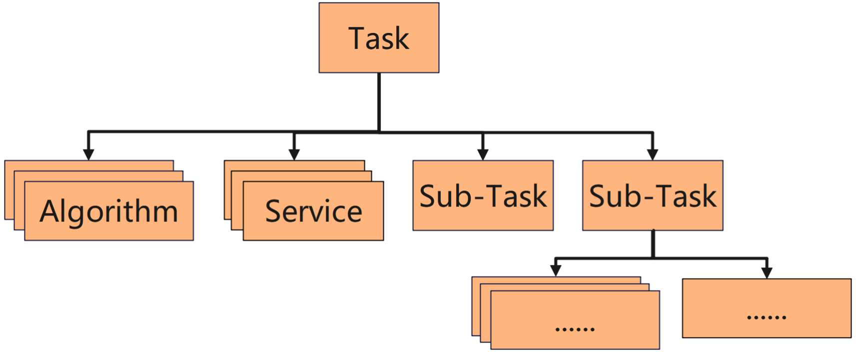


CHEP2023: X.T. Huang et al : JUNO Offline Software for Data Processing and Analysis

SNiPER Framework

- SNiPER (Software for Non-Collider Physics Experiment) : a general-purpose software framework
  - Light weighted: only dependent on Boost.Python
  - Highly modularized and extensible
  - Originally developed for JUNO, but also adopted by several experiments: LHAASO, STCF, HERD, nEXO, etc.

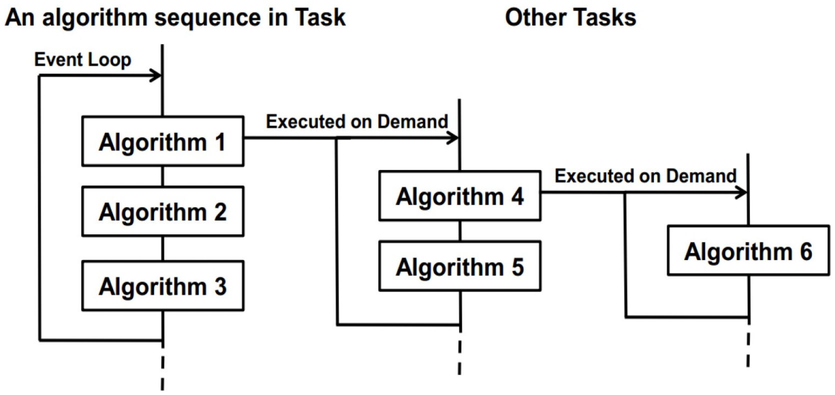
- Key components
  - Algorithm: an unit of event data proceeding
  - Service: an unit for common functions that can be called by users, anywhere when necessary.
  - Task: a lightweight application manager to assemble specific algorithms, service as well as sub-tasks.
  - Data Buffer: a central place in memory for holding and sharing multiple events



SNiPER Framework

- SNiPER (Software for Non-Collider Physics Experiment) : a general-purpose software framework
  - Light weighted: only dependent on Boost.Python
  - Highly modularized and extensible
  - Originally developed for JUNO, but also adopted by several experiments: LHAASO, STCF, HERD, nEXO, etc.

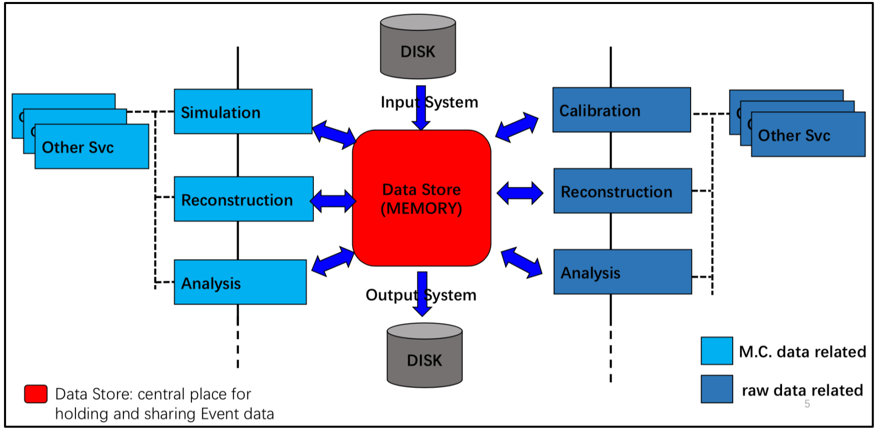
- Key components
  - Algorithm: an unit of event data proceeding
  - Service: an unit for common functions that can be called by users, anywhere when necessary.
  - Task: a lightweight application manager to assemble specific algorithms, service as well as sub-tasks.
  - Data Buffer: a central place in memory for holding and sharing multiple events



SNiPER Framework

- SNiPER (Software for Non-Collider Physics Experiment) : a general-purpose software framework
  - Light weighted: only dependent on Boost.Python
  - Highly modularized and extensible
  - Originally developed for JUNO, but also adopted by several experiments: LHAASO, STCF, HERD, nEXO, etc.

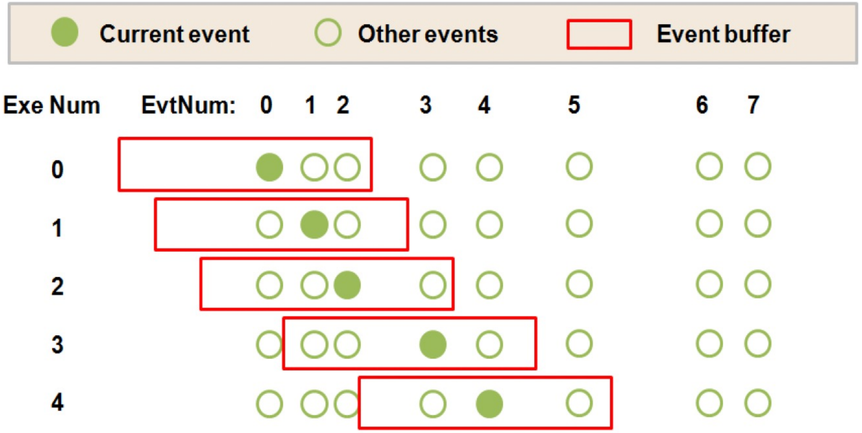
- Key components
  - Algorithm: an unit of event data proceeding
  - Service: an unit for common functions that can be called by users, anywhere when necessary.
  - Task: a lightweight application manager to assemble specific algorithms, service as well as sub-tasks.
  - Data Buffer: a central place in memory for holding and sharing multiple events



SNiPER Framework

- SNiPER (Software for Non-Collider Physics Experiment) : a general-purpose software framework
  - Light weighted: only dependent on Boost.Python
  - Highly modularized and extensible
  - Originally developed for JUNO, but also adopted by several experiments: LHAASO, STCF, HERD, nEXO, etc.

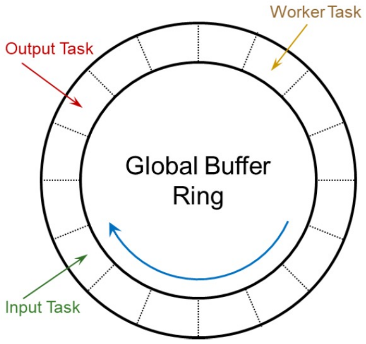
- Key components
  - Algorithm: an unit of event data proceeding
  - Service: an unit for common functions that can be called by users, anywhere when necessary.
  - Task: a lightweight application manager to assemble specific algorithms, service as well as sub-tasks.
  - Data Buffer: a central place in memory for holding and sharing multiple events

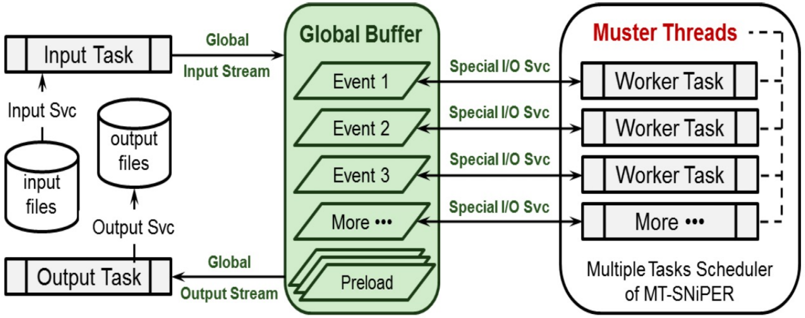


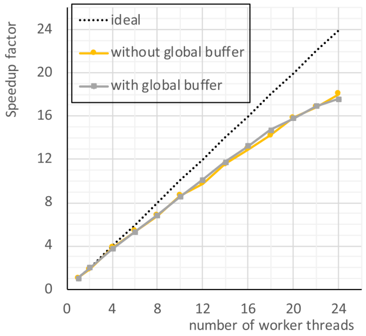
Multi-threaded SNiPER (MT.SNiPER)

- MT.SNiPER supports concurrent task of event processing based on the TBB
  - A thread-safe Global Buffer is designed and implemented with a ring and configurable capacity and cordon
  - All events in the Global Buffer are sorted by their time stamps
  - Allow the dedicated Input Task, Worker Task, Output Task to put, take, and popup events at the same time in different threads
  - Events are concurrently processed in different Worker Tasks
- The multi-threaded simulation and reconstruction have been working very well

J.H. Zou et al., J. Phys. Conf. Ser. 1085, 032009 (2018)
T. Lin et al., EPJ Web of Conferences 214, 02008 (2019)



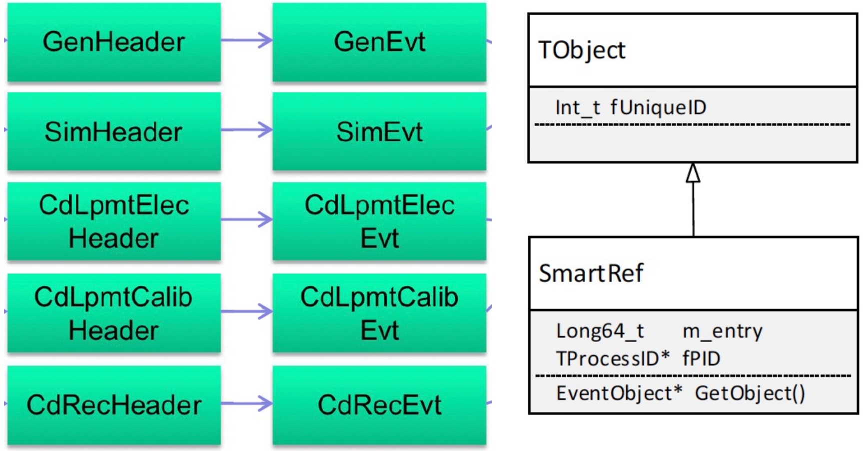




Event Data Model

- Event Data Model (EDM) defining event objects for different processing stages and correlations between different event objects takes very important roles on the whole data processing and analysis
- JUNO EDM is based on ROOT and takes advantage of its intrinsic powerful functions, persistency, IO streamer, scheme evolutions, run time type Information etc.
  - All EDM classes are derived from TObject
  - Two layer design (Header and Event) of EDM classes are adopted to speed up event selection
    - The Header holds light-weighted features (tag) of events while the Event holds heavy data
    - A smart references (SmartRef) based on GUUID is developed to build correlations between Header and Event
    - SmartRef also provides a lazy-loading mechanism to dramatically reduce I/O burden

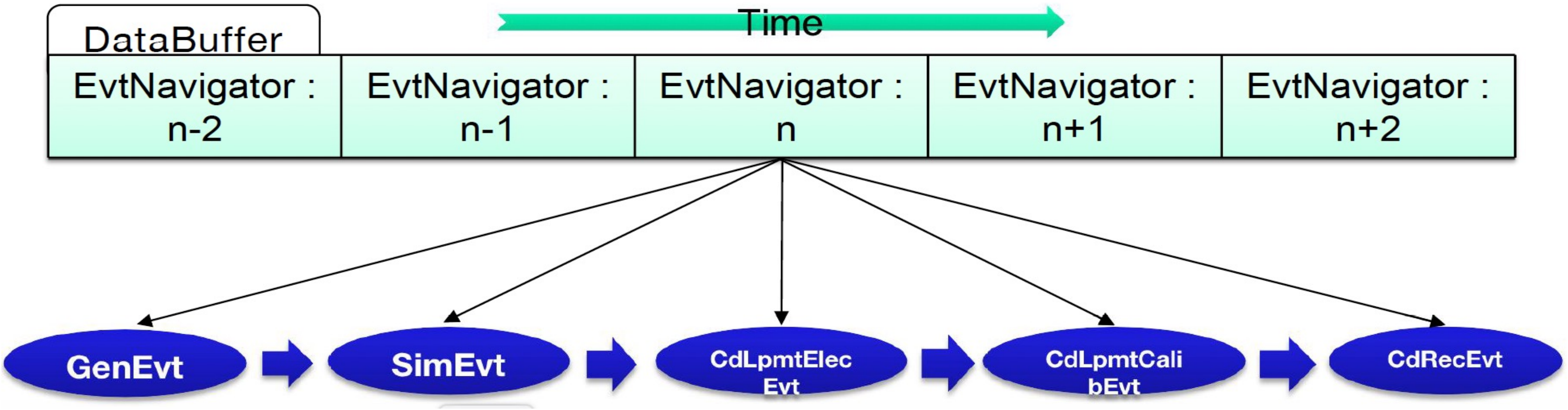
T. Li, X. T. Huang, Journal of Physics: Conference Series 762 (2016) 012001



Event Data Model

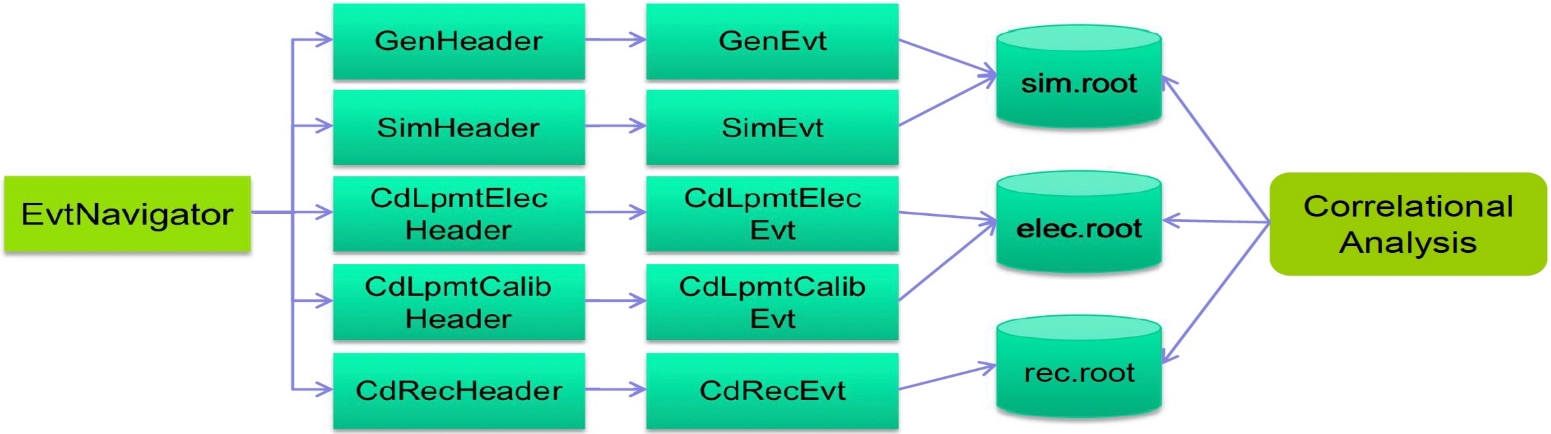
❖ EvtNavigator is developed to serve as the catalogue for EDM objects

- EvtNavigator also uses SmartRef to correlate EDM objects
- A event-index mechanism based on Global UUID is implemented to facilitate cross-file correlation analysis
- Provide the way to trace the MC truth of readouts at the event/particle/hit level with the help of external associations



Event Data Model

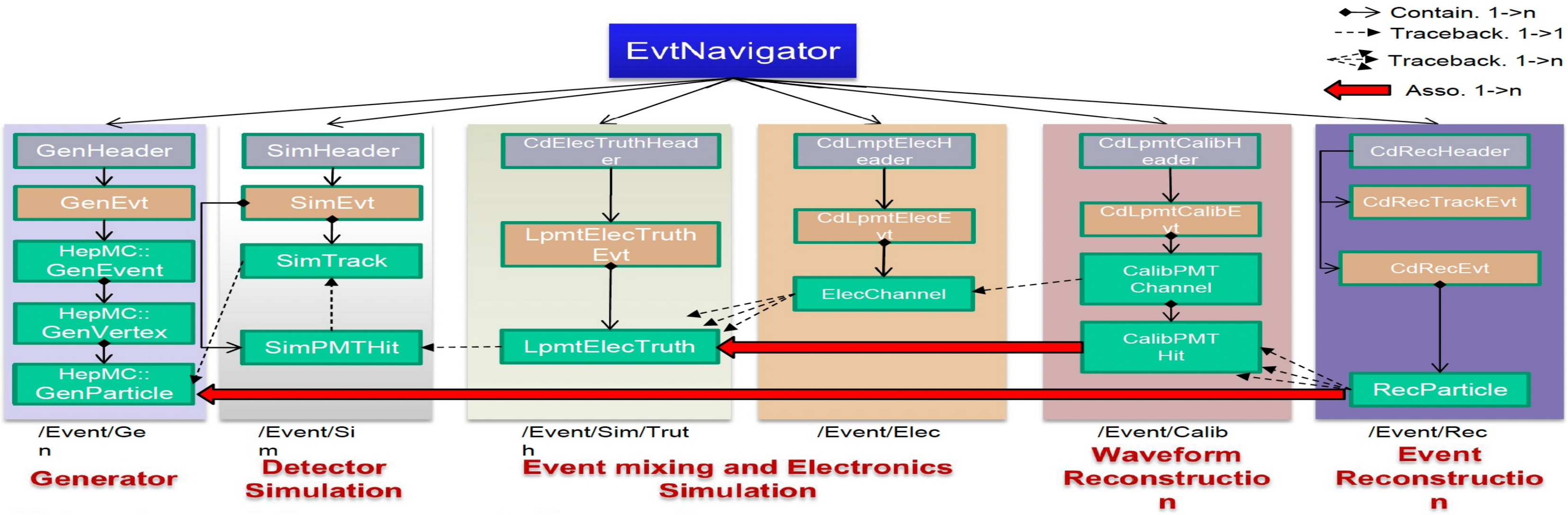
- EvtNavigator is developed to serve as the catalogue for EDM objects
  - EvtNavigator also uses SmartRef to correlate EDM objects
  - A event-index mechanism based on Global UUID is implemented to facilitate cross-file correlation analysis
  - Provide the way to trace the MC truth of readouts at the event/particle/hit level with the help of external associations



Event Data Model

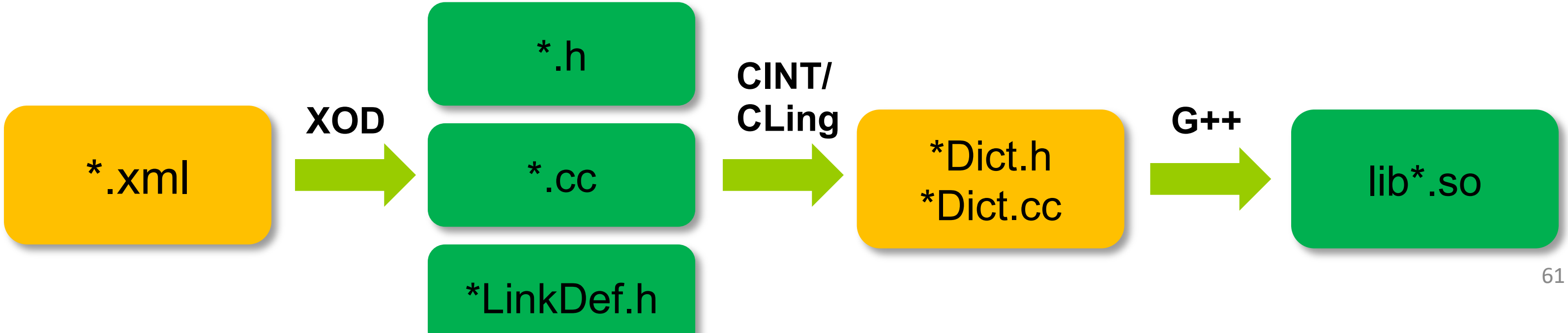
❖ EvtNavigator is developed to serve as the catalogue for EDM objects

- EvtNavigator also uses SmartRef to correlate EDM objects
- A event-index mechanism based on Global UUID is implemented to facilitate cross-file correlation analysis
- Provide the way to trace the MC truth of readouts at the event/particle/hit level with the help of external associations



XML Object Description (XOD) Tool

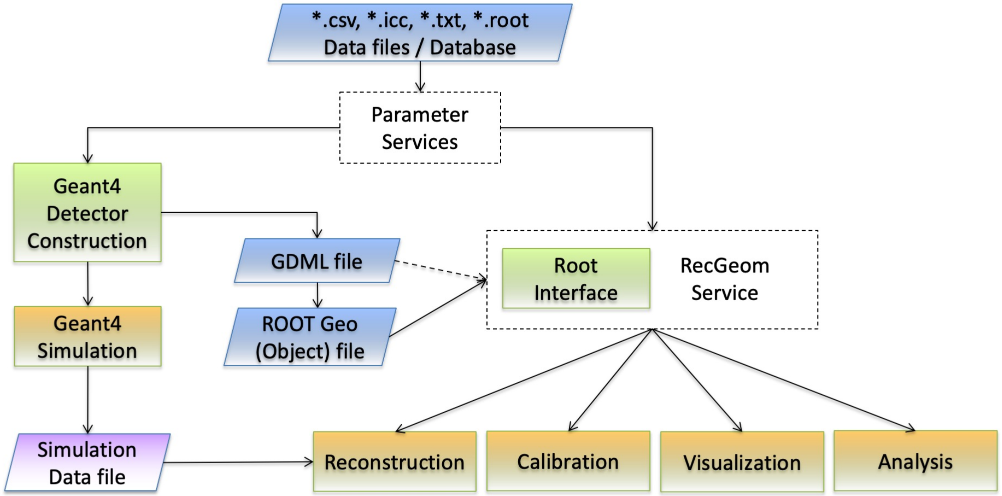
- Traditionally writing C++ Code manually
  - Many repeatable work such as Getters and Setters
  - Difficult to be maintained
- Use XML file to define EDM
  - Strong syntax (DTD, XML Schema)
  - More readable, easier to maintain
  - Automatically generate the Get-, Set-functions, ROOT I/O Streamers
- A tool, XmlObjDesc (XOD), is developed to automatically generate class codes



Geometry Management System (GMS)

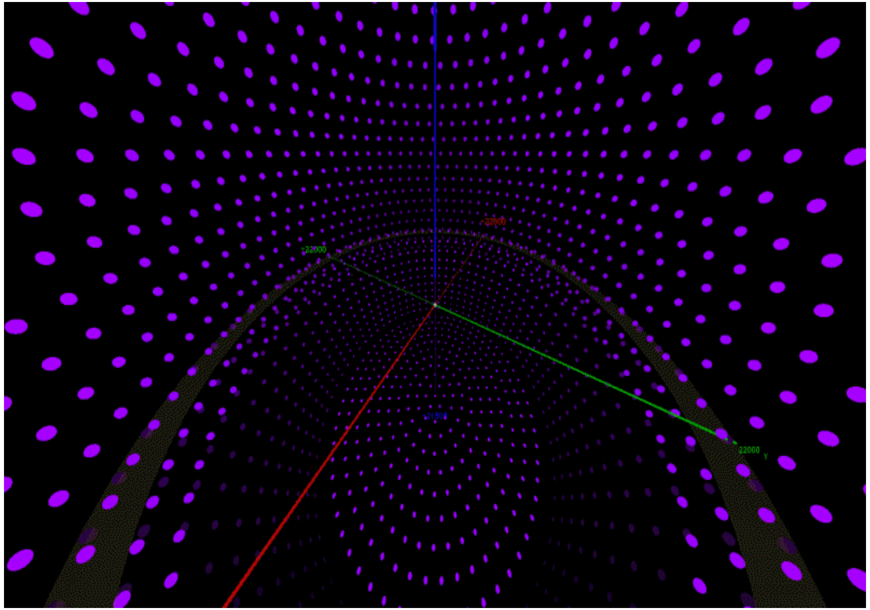
- GMS is developed to provide a consistent and detailed detector description for different applications
  - Input from single source of geometry data : currently from ROOT file, plan to read from DB
  - Export from Geant4 detector construction into GDML/ROOT TGeo file format
  - Perform automatic conversion with Files – Geant4 – GDML – ROOT – Applications with different services
    - Parameter Service
    - RecGeom Service
  - Geometry information depending on different level of request
  - Some information not supported by GDML can be retrieved from Parameter Services
    - Optical Surface
    - Matrixes

K. J. Li et al., NIM A 908 (2018) 43–48

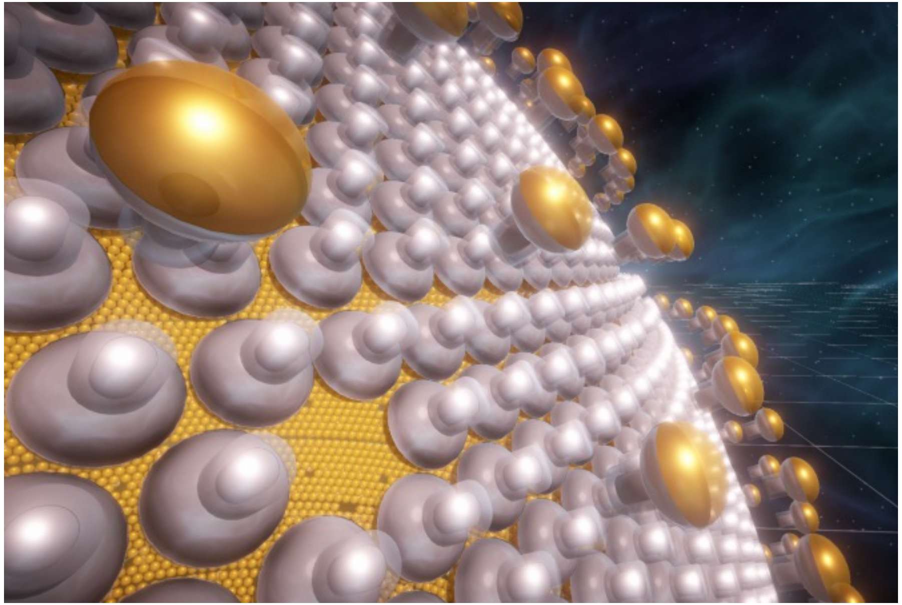


Event Display

- ❖ Root based Event Display: SERENA (Software for Event display with Root Eve in Neutrino Analysis)



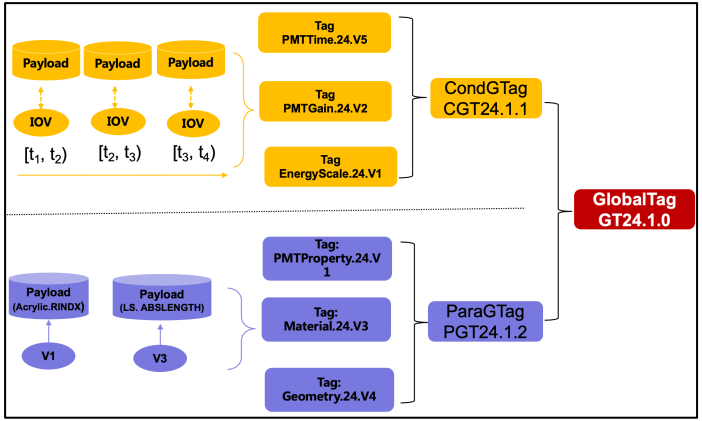
- ❖ Unity based Event Display: ELAINA (Event Live Animation with unIty for Neutrino Analysis)

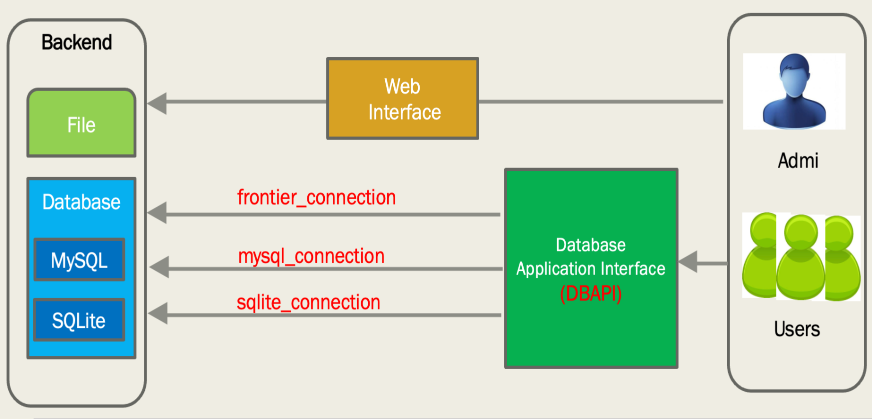


J. Zhu et al., JINST 14 (2019) 01, T01007

Database System

- OfflineDB is developed to unify management of Conditions Data and Parameter data
  - Two types of data are stored in Payloads in their suitable formats and deployed in CVMFS
  - Meta data are managed with DB tables: Payload, IOV, Version, Tag and GlobalTag
  - A GlobalTag combines the tag of Conditions data and Parameter data
  - Three backups are implemented to meet different applications
    - MySQL for quickly testing of database
    - Frontier+Squid for massive data production
    - SQLite for online event classification (OEC)



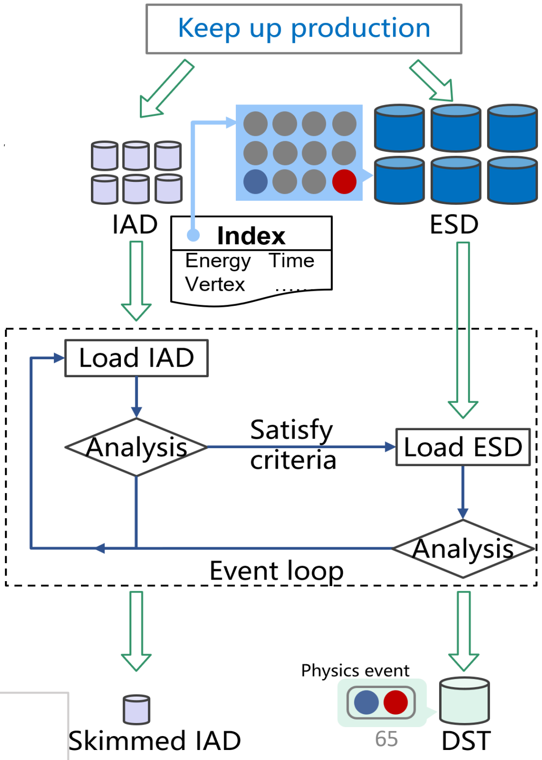


Analysis Software Framework

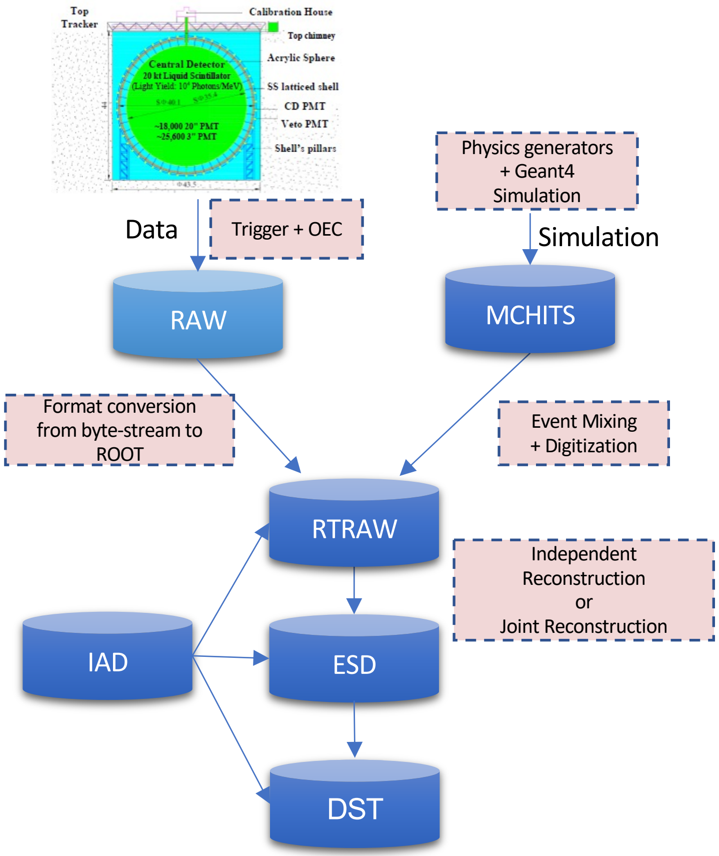
- Challenges of data analysis
  - Time correlation analysis (a prompt signal and a delayed signal within varying time window)
  - Rare signals (1kHz event rate, but only ~60 reactor neutrinos per day)

- Developed analysis software framework with an Indexing Technique
  - The Indexed Analysis Data (IAD) consists of two parts:
    - A collection of Key Reconstructed Variables (KRV) needed by the analysis
    - A pointer to the associated Event Summary Data (ESD)

- Analysis work flow for Time correlation analysis
  - Step1: perform event pre-selection within the IAD stream based on its key variables
  - Step2: navigate back to the associated ESD and perform further selection in the ESD stream



Event Data Flow



- MCHITS (Monte-Carlo Simulated Hits Data)

  In ROOT format, contains GenEvt, SimEvt

- RAW (Raw Data from DAQ and OEC)

  In byte-stream format. Also contains the OEC result.

- RTRAW (ROOT RAW)

  In ROOT format (with EvtNavigator).

  Sorted; contains ElecEvt, TrigEvt, OecEvt; may contain GenEvt and SimEvt for simulated data.

- ESD (Event Summary Data)

  In ROOT format (with EvtNavigator), contains CalibEvt, RecEvt

- DST (Data Summary Type)

  Contains information of physical events, Flexible definition and file format, Does not contain relationships

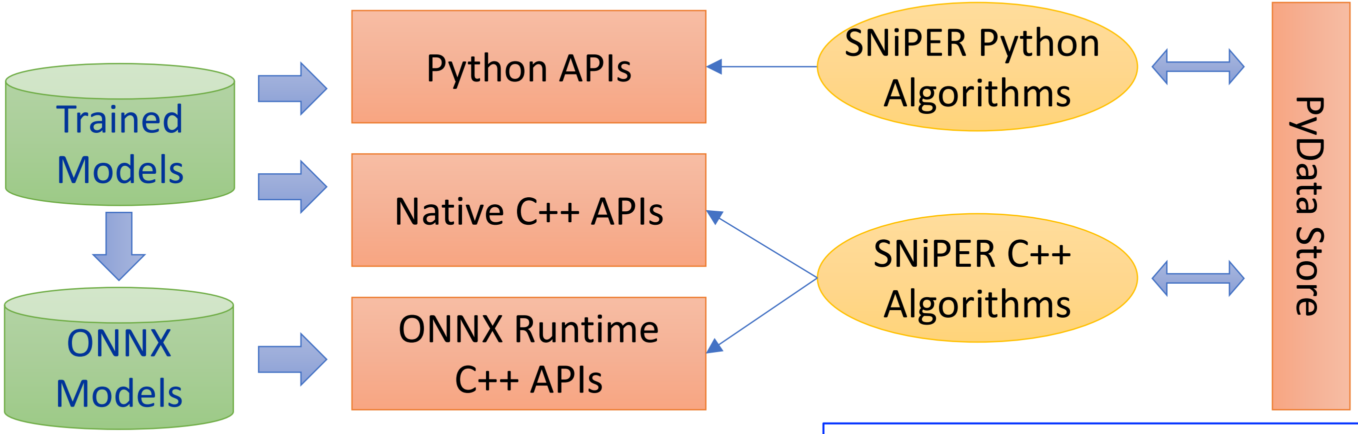
- IAD (Indexed Analysis Data) (Auxiliary data)

  Contains the address to the RAW/ESD and the necessary physics variables

CHEP2023: X.T. Huang et al : JUNO Offline Software for Data Processing and Analysis

Machine learning interface

- A python based datastore is developed for data sharing between C++ and Python
- Popular ML libraries are integrated into JUNOSW in C++ for inference with three solutions:
  - Python API for SNiPER Algorithm in python
  - Native C/C++ APIs: TensorFlow, PyTorch
  - Open standard for ML: ONNX + ONNX Runtime (C++)

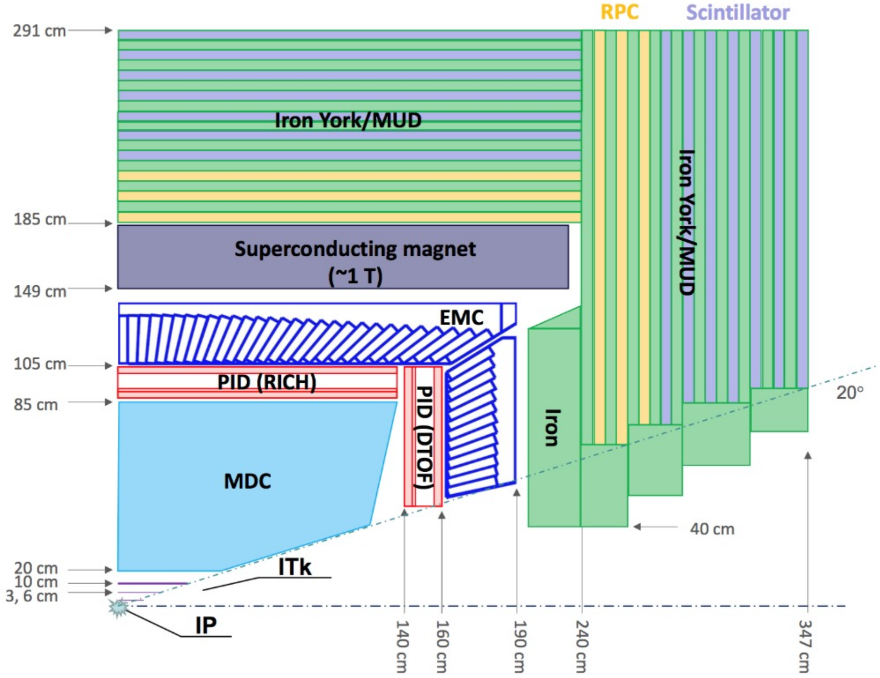


T. Lin et al., *Integration of machine learning-trained models into JUNO's offline software, ACAT2022*

STCF Offline Software System

STCF Detector

立体覆盖角: 94%•4π (θ~20^0)



辐照本底水平

- ~3.5 kGy/y, ~2×10^{11} 1MeV n-eq/cm^2/y, ~1 MHz/cm^2

内径迹探测器 (ITK)

- material<~0.25%X_0 /layer
- σ_{xy}<~100 mm

主漂移室 (MDC)

- σ_{xy}<~120 mm
- σ_p/p ~ 0.5% @ 1 GeV
- dE/dx~6%

粒子鉴别探测器 (PID=RICH+DTOF)

- π/K (and K/p) 4σ separation up to 2GeV/c

电磁量能器 (EMC)

- E：0.025-3.5GeV
- σ_E/E ~ 2.5%@1 GeV
- σ_{xy} ~ 5 mm

缪子探测器 (MUC)

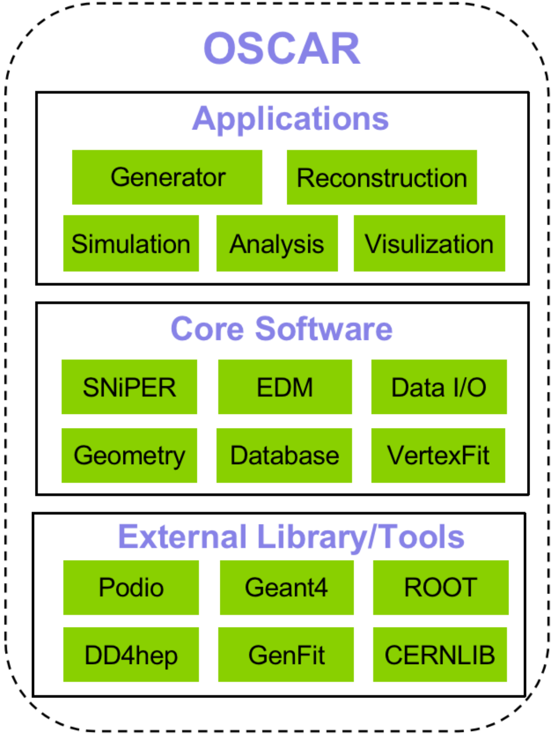
- p：0.5 - 2 GeV/c
- π suppression >30

事例触发、数据传输与获取

- 400 kHz @J/ψ.   200 GB/s

STCF Offline Software

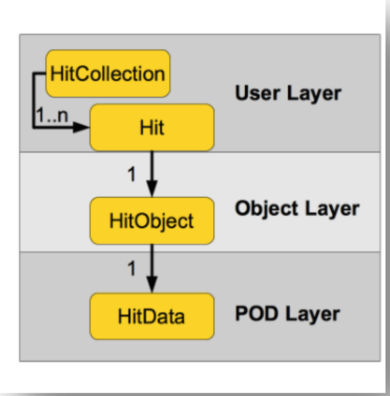
- The Offline Software of Super Tau-Charm Facility (OSCAR), designed for detector design, MC data production and physics analysis since 2018
- OSCAR is partially based on Key4hep
  - Common software stack for future collider experiments
  - Reuse some components. Extend others for STCF
- Core components are developed for common functionalities
  - Event loop control (sequently or concurrently)
  - Detector data and event data management
  - Common tools for data analysis
  - Other common services
- Some applications are migrated from BESIII



Event Data Model Based on Podio (1)

- As the core of the offline software, EDM greatly influences the function and performance of OSCAR
  - EDM defines the structure of event data in memory and in data files
  - Implement relationship between data objects (hit-track-MC particle)
  - Handle schema evolution

- Podio defines the common EDM (EDM4hep adopted by FCC, CEPC, ILC, ...)
  - Generate C++ code based on YAML definition
  - Support both C++ and Python
  - Good multithreading support
  - Powerful and flexible relationship between data objects
  - Support multiple data file format

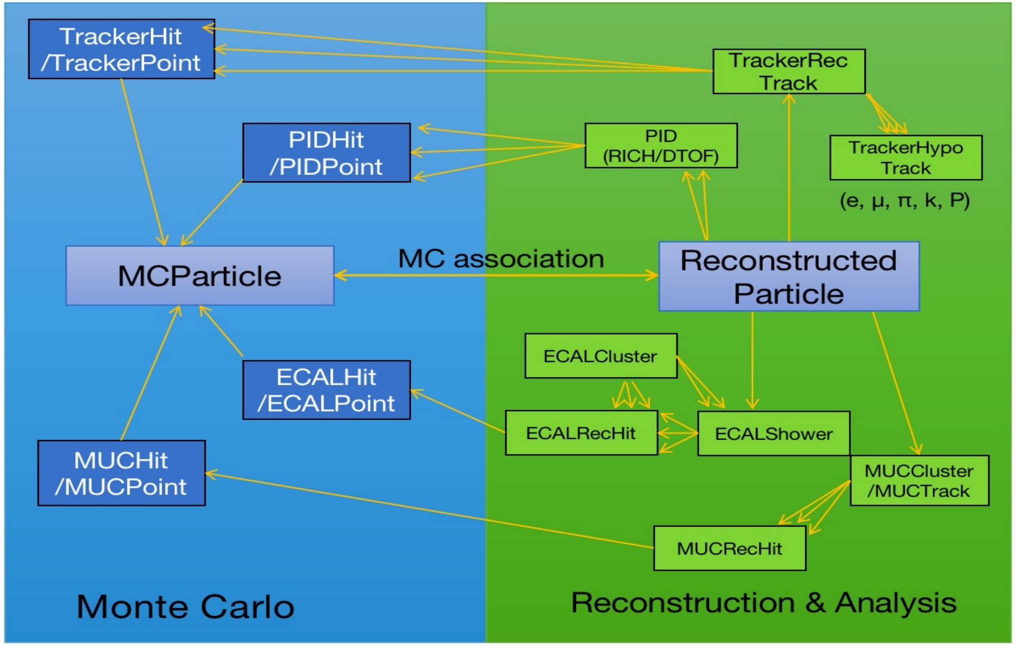


F. Gaede, etc. , CHEP2019

https://github.com/AIDASoft/podio

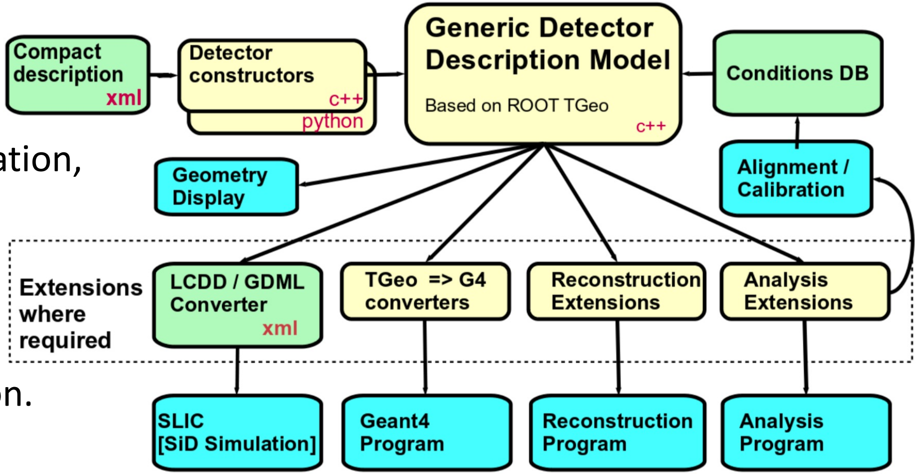
Event Data Model Based on Podio (2)

- Due to the specific requirements of STCF, EDM4hep is not directly used
- Re-design EDM classes based on Podio
  - Simulation Level
    - Point, Hit, MCParticle
  - Reconstruction Level:
    - HypoTrack, RecTrack
    - Hit, Cluster, Shower
    - ReconstructedParticle
  - Relationship between objects
  - Correlated ReconstructedParticle with MCParticle based on track matching algorithm,



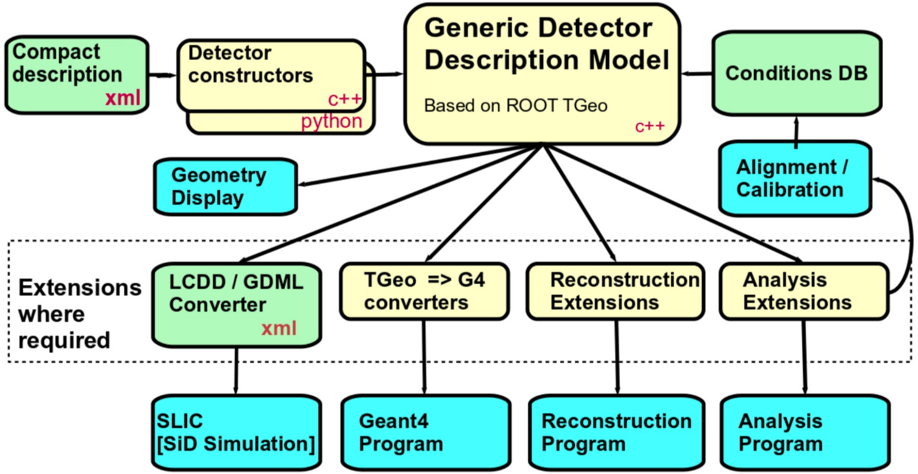
DD4hep

- A general detector (geometry) description toolkit for HEP
- Developed in AIDA and AIDA2020 and used by ILC, CLIC, FCC , many are under evaluation. E.g. EIC, CMS and LHCb for upgrade
- Key functionalities
  - Full Detector Description
    - Includes geometry, materials, visualization, readout, alignment, calibration, etc.
  - Full Experiment life cycle
    - From concept development, detector optimization, construction to operation.
  - Consistent Description with single source
    - for simulation, reconstruction, analysis, etc.
  - Ease of Use



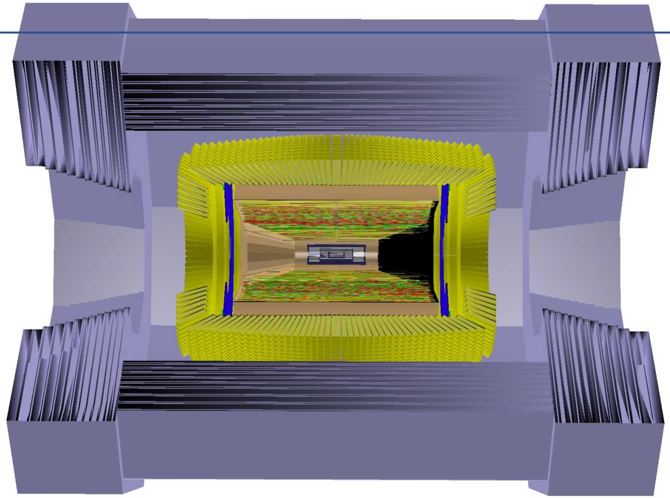
DD4hep

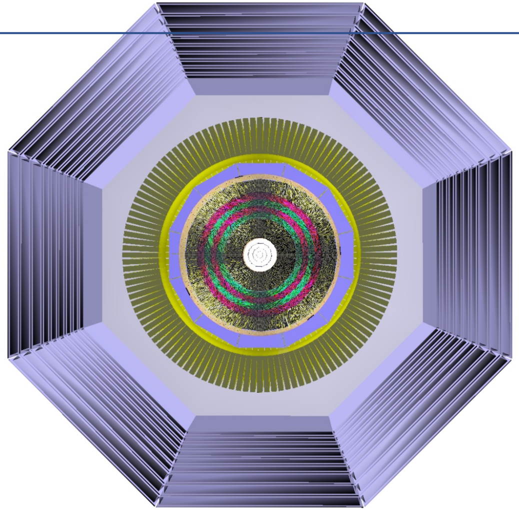
- A general detector (geometry) description toolkit for HEP
- Developed in AIDA and AIDA2020 and used by ILC, CLIC, FCC , many are under evaluation. E.g. EIC, CMS and LHCb for upgrade
- Implementations
  - TGeo : underlying geometry description
    - Open GL viewer for geometry
    - Debugging of geometry
  - Default geometry description
    - Compact xml-files and C++ drivers
  - Output formats/interface
    - GDML
    - Geant4 geometry



STCF Detector with DD4hep

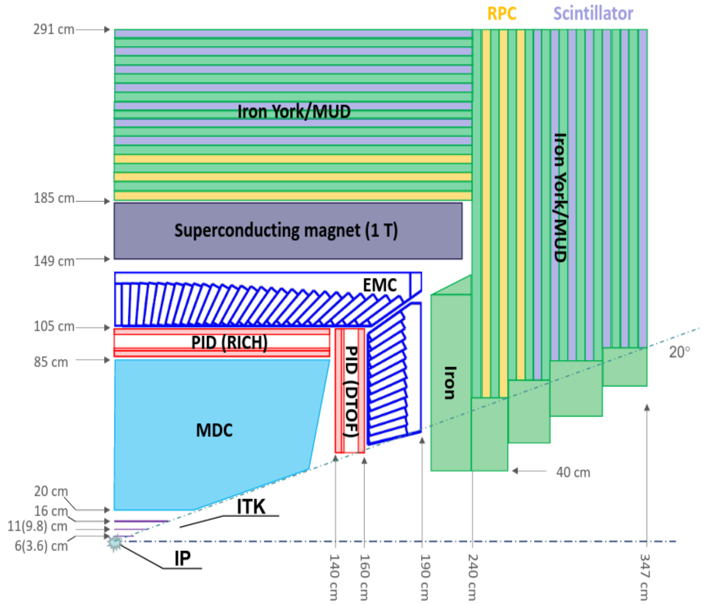
```xml
<detectors>
      <comment>Trackers</comment>

       <detector name="AirTube" type="AirTube" vis="VXDVis" id="42"
insideTrackingVolume="true">
        <dimensions rmin="10.*mm" rmax="11.*mm" zhalf="6.250000000e+01*mm"/>
      </detector>
...
 </detectors>
```

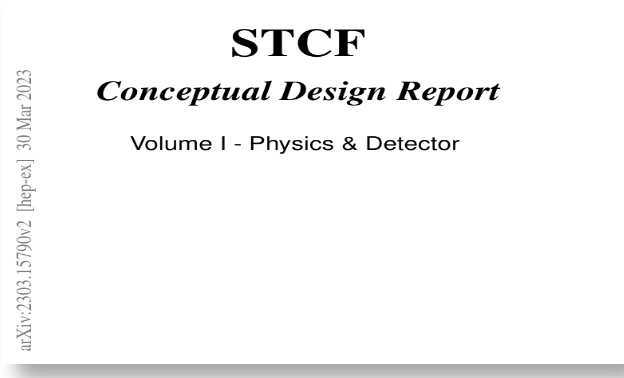




STCF Detector

- ❖ Charged particles: [0.05, 1.6] GeV ,
  - Tracking eff. > 50/90/99 % with pt > 50/100/300 MeV
  - $\sigma(p)/p$ = 0.5% at 1 GeV, dE/dx resolution: < 6%
- ❖ Good e, g detection eff. and energy res.: [0.02,2.5] GeV
  - Energy resolution > 2.5% @ 1 GeV
  - Position resolution > 5 mm @ 1 GeV
  - Time resolution > 300 ps@1 GeV
- ❖ Good PID: [0.05, 2] GeV
  - π/K identification efficiency:>97%, false positive rate: < 2%@2GeV
  - μ identification efficiency:>95%@ [0.4-1.8GeV/c], π suppression power ~ 33
- ❖ Good vertex detection: 50 mm

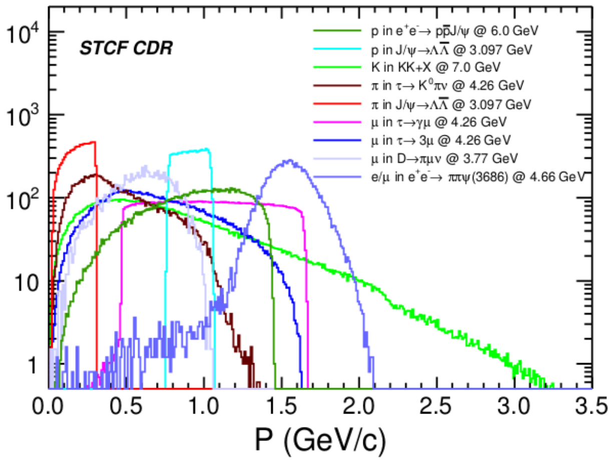


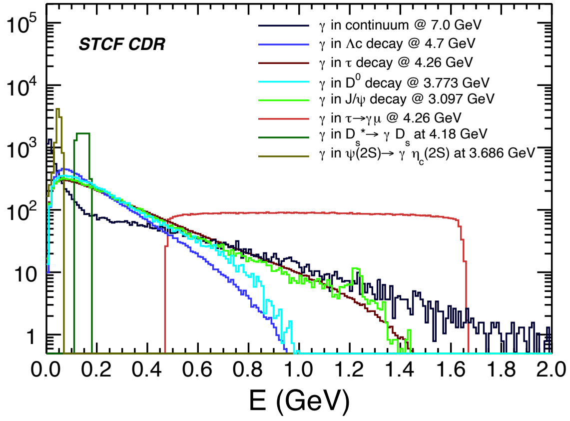


arXiv: 2303.15790

STCF Detector

- ❖ Charged particles: [0.05, 1.6] GeV ,
  - Tracking eff. > 50/90/99 % with pt > 50/100/300 MeV
  - σ(p)/p = 0.5% at 1 GeV, dE/dx resolution: < 6%
- ❖ Good e, g detection eff. and energy res.: [0.02,2.5] GeV
  - Energy resolution > 2.5% @ 1 GeV
  - Position resolution > 5 mm @ 1 GeV
  - Time resolution > 300 ps@1 GeV
- ❖ Good PID: [0.05, 2] GeV
  - π/K identification efficiency:>97%, false positive rate: < 2%@2GeV
  - **μ** identification efficiency:>95%@ [0.4-1.8GeV/c]， π suppression power ~ 33
- ❖ Good vertex det.: 50 mm

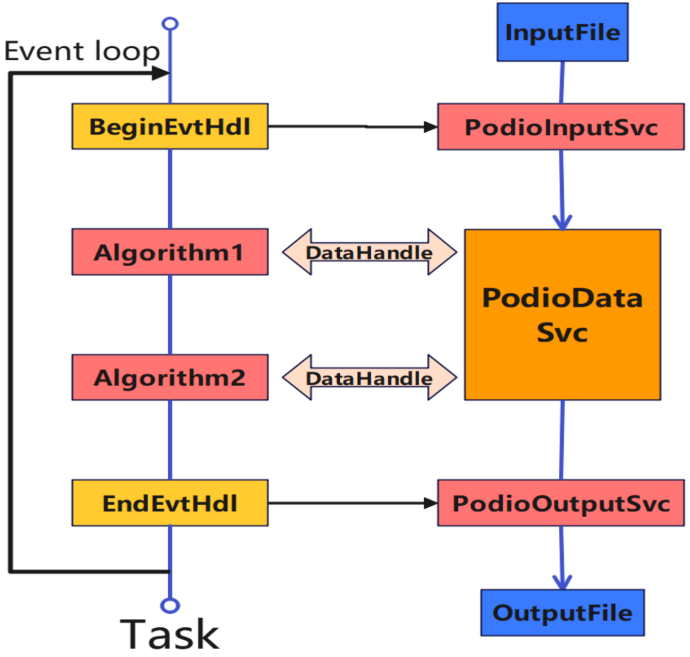




arXiv: 2303.15790

Event Data Management

- Event data management system manages event data in memory, provides interfaces for user applications and handles data I/O
- OSCAR DM system is developed based on Podio
  - PodioDataSvc: memory management
  - PodioInputSvc: data input
  - PodioOutputSvc: data output
  - DataHandle: interface
- OSCAR DM system is integrated with SNiPER based on the Incident mechanism, to integrate the DM system and event loop
- Event data and user application are completely decoupled

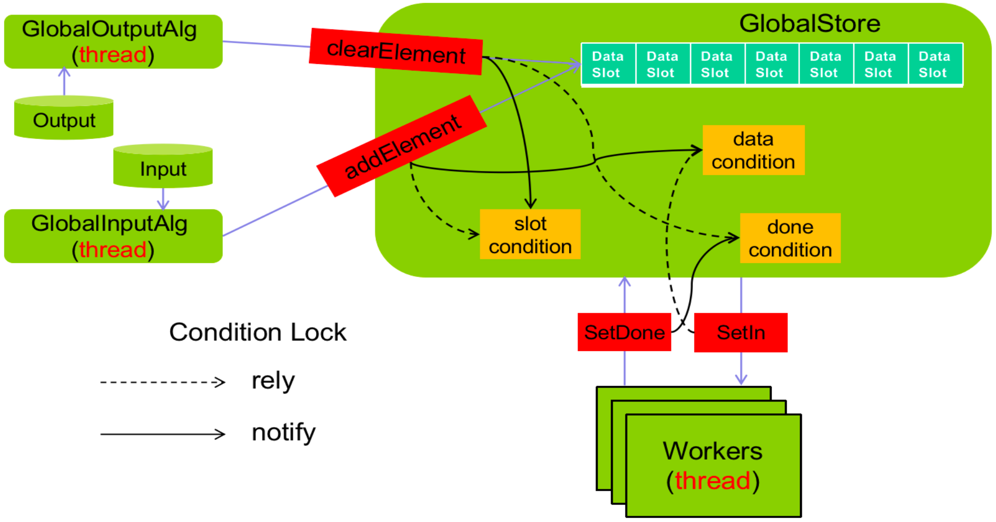


Parallelized Event Data Management

- To enable parallelized data processing, a GlobalStore is developed based on Podio
  - Re-implement podio::EventStore to cache multiple events (each within one data slot)
  - Use several condition lock to enable safety exchanging data between threads
  - I/O services are binded to dedicated I/O threads, to ensure performance and flexible post- or pre-processing

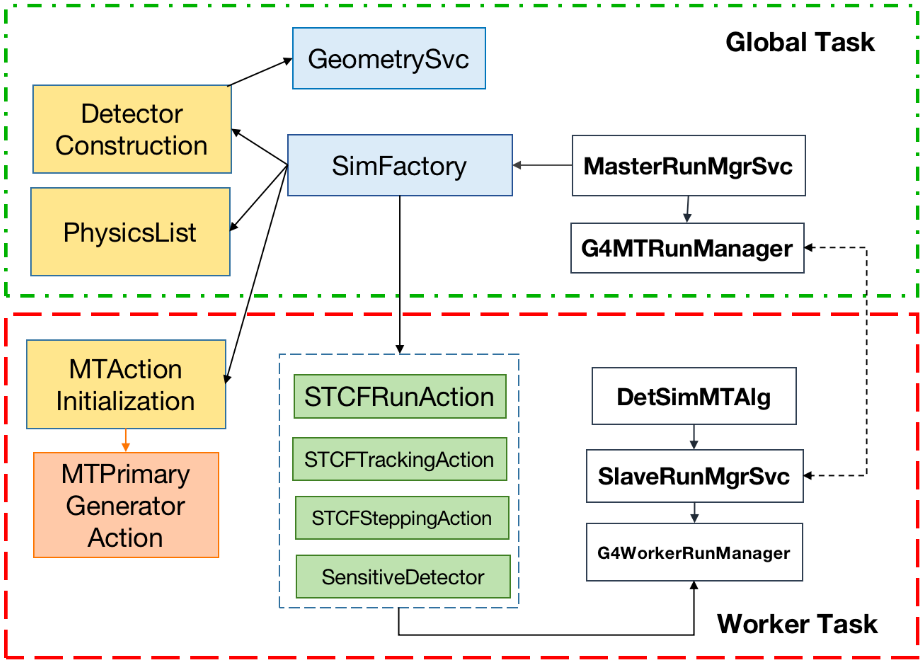
- Based on parallelized DM system, parallel detector simulation and reconstruction works well

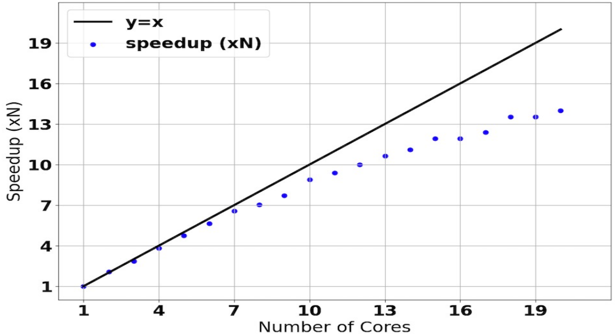
- Users could switch serial/ parallel by just changing job configuration

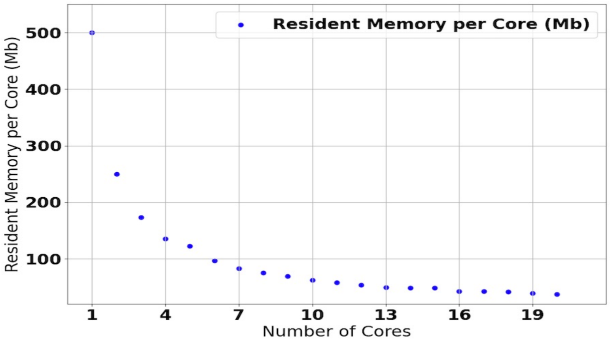


Parallelized Detector Simulation

- Based on the MT-SNiPER and parallelized DM system, parallelized detector simulation applications are developed
  - Basic performance tests show promising scalability



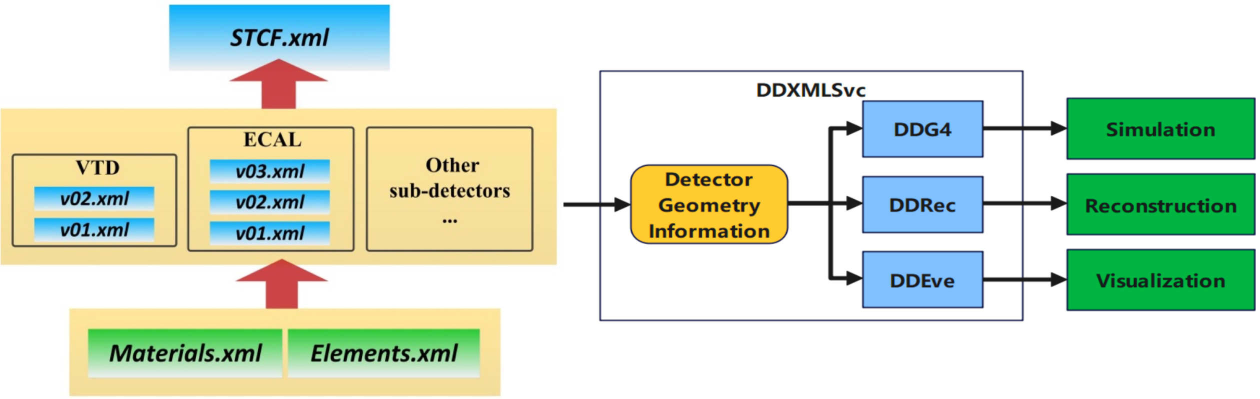




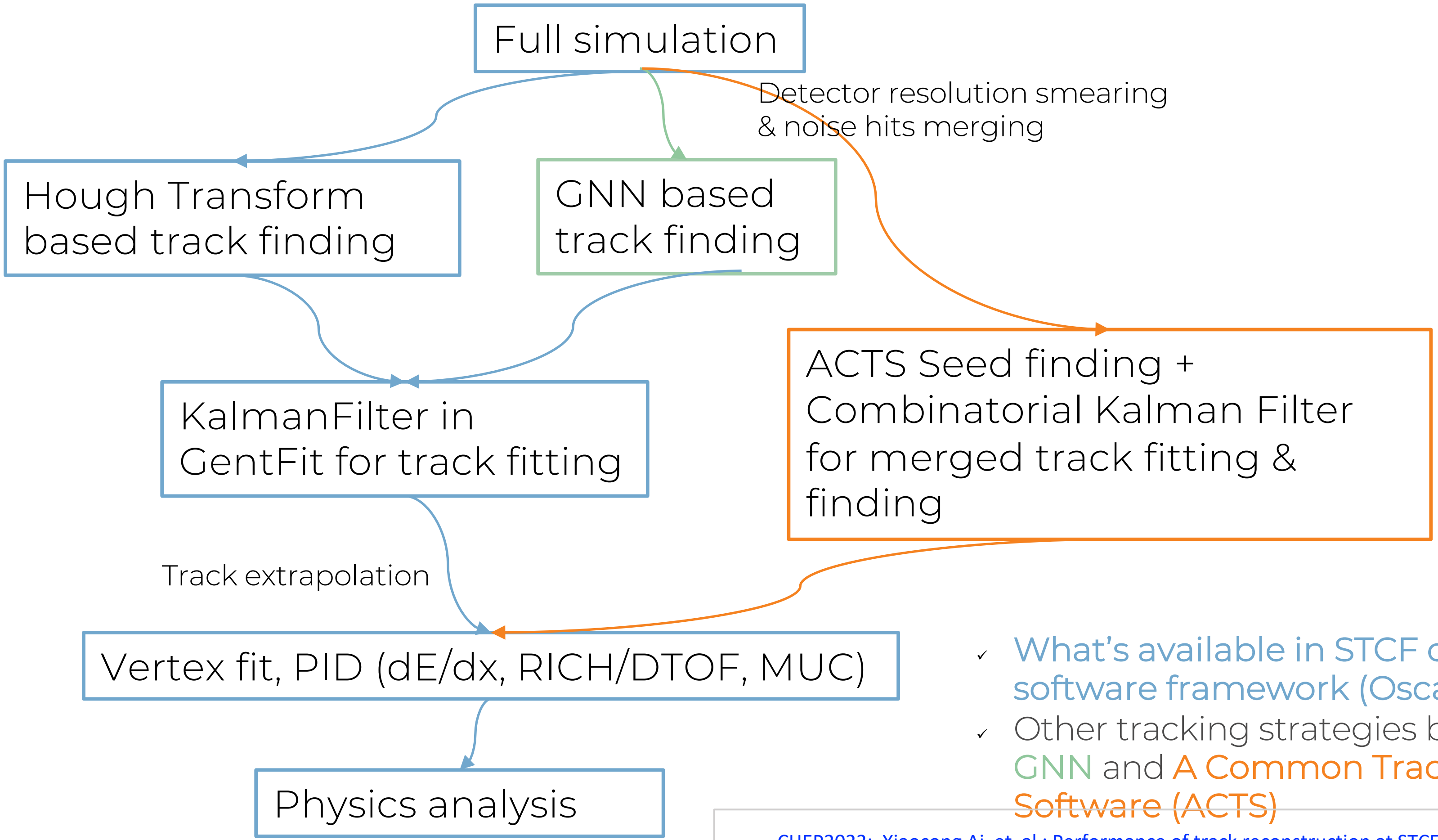
Geometry Management System

- ❖ Detector description in OSCAR is developed based on DD4hep
- ❖ Single source of detector information for detector description, simulation and reconstruction
  - DDG4 for delivering detector geometry to Geant4
  - DDRec for delivering detector geometry to reconstruction algorithms
  - DDXMLSvc: the unified interface to DD4hep, including DDG4 and DDRec

- ❖ Flexible combinations of different versions of detector design, and combinations of sub-systems



STCF tracking landscape

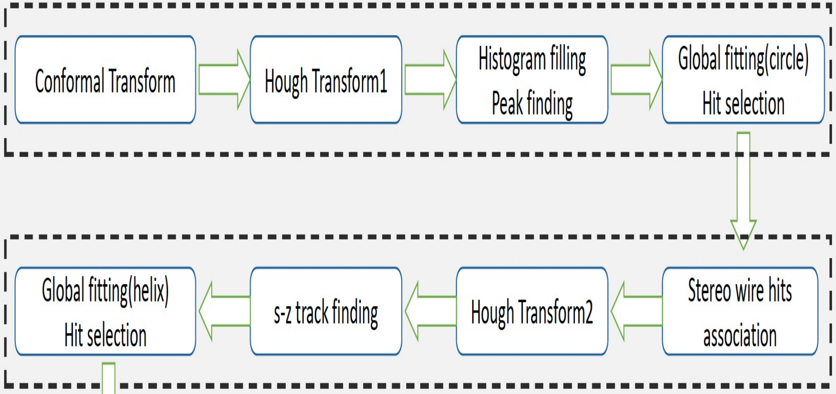


- ✓ What's available in STCF offline software framework (Oscar)
- ✓ Other tracking strategies based on GNN and A Common Tracking Software (ACTS)

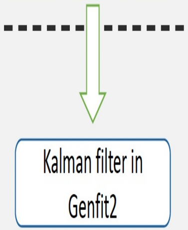
Tracking based on Hough transform

Hang Zhou

2-D track finding



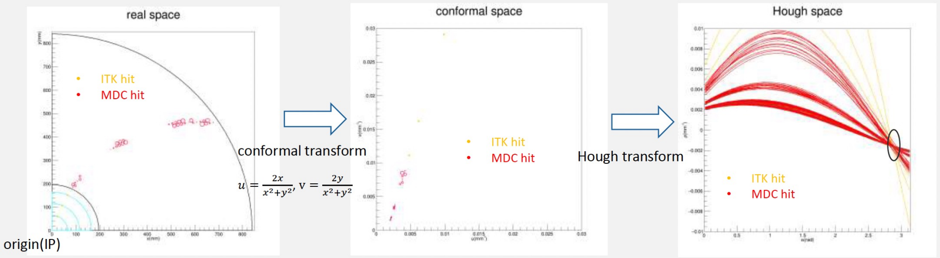
3-D track finding

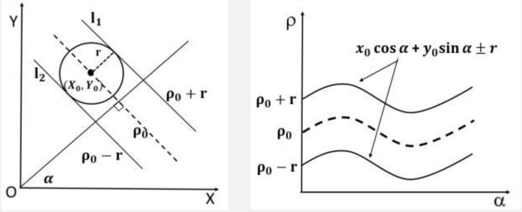


全局的寻迹算法同时使用ITK和MDC击中

两个主要步骤

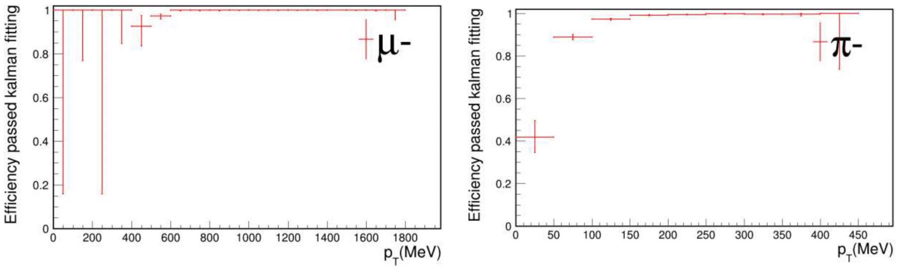
- 使用漂移室直丝和内径迹探测器测量位置寻找圆径迹
- 匹配漂移室斜丝，确定三维螺旋线参数





- 共形变换
- 考虑漂移圆的霍夫变换 (勒让德变换)

ψ(3686)→ π+π- J/ψ(μ+μ-) 20° <theta<160° 至少5个击中



A Common Tracking Software (ACTS)

- A modern open-source detector-independent tracking toolkit for current&future HEP experiments (ATLAS, ALICE, sPHENIX, FASER, MUC, CEPC, STCF…) based on LHC tracking experience
- A R&D platform for innovative tracking techniques (ML) & computing architectures

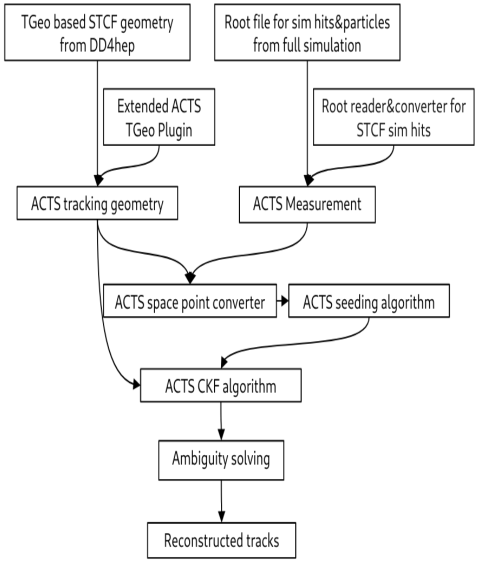
- Modern C++ 17 (→20) concepts
- Detector and magnetic field agnostic
- Strict thread-safety to facilitate concurrency
- Minimal dependency (Eigen)
- Highly configurable, well documented and maintained

Github:
https://github.com/acts-project/acts
Readthedocs:
https://acts.readthedocs.io/en/latest/

See other talks related with ACTS:

- "Machine learning for ambiguity resolution in ACTS" by Corentin ALLAIRE
- "Kiwaku, a C++20 library for multidimensional arrays, applied to ACTS tracking" by Sylvain Joube
- "Flexible, robust and minimal-overhead Event Data Model for track reconstruction in ACTS" by Paul Gessinger
- "Potentiality of automatic parameter tuning suite available in ACTS track reconstruction software framework" by Rocky Garg

ACTS application to STCF



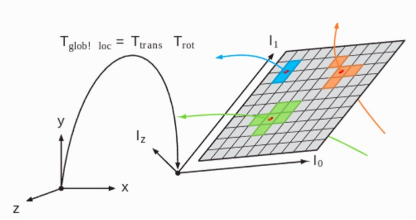
space point making
(ITk local coordinates→
3D global coordinates)

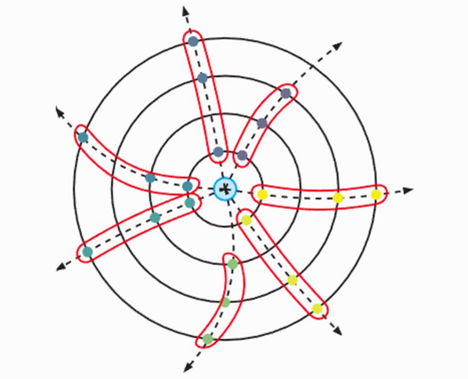
→

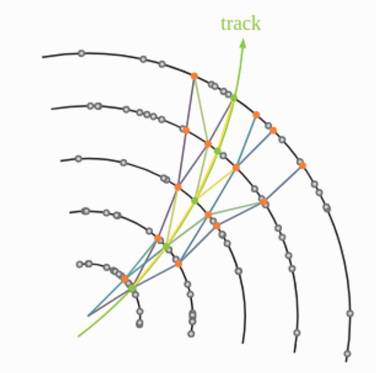
Seeding
(find seeds using hits
on ITk layers)

→

Combinatorial
Kalman Filter (CKF)
(simultaneous track
fitting + finding)





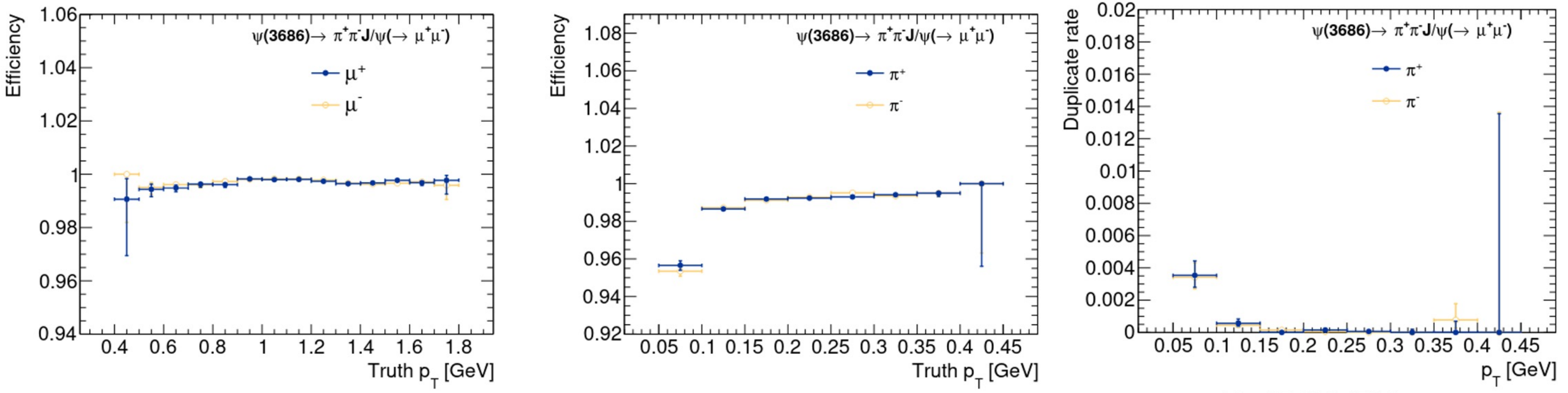


Figures from ACTS readthedocs

Automatic tuning of seeding parameters
(See talk by Rocky Garg)

- ❖ First application and validation of ACTS for a drift chamber

Tracking performance with ACTS



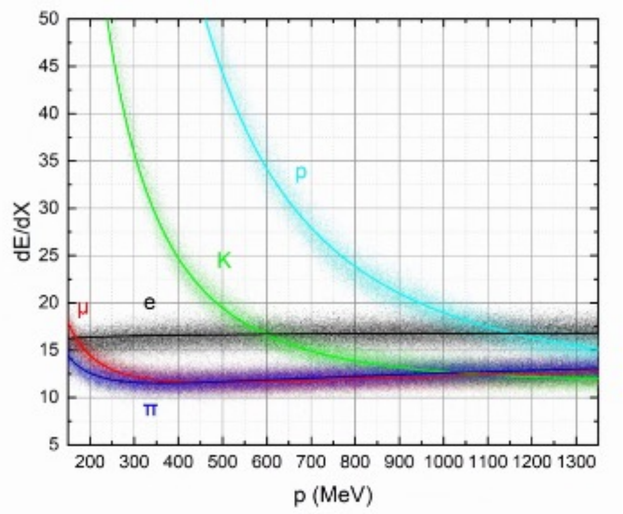
arXiv: 2301.04306

- Above 99% efficiency for $p_T > 400$ MeV
- 95% efficiency for pion with p_T in [50, 100] MeV
- <0.5% duplicate tracks for $p_T < 130$ MeV due to duplicate seeds for looping tracks
- Negligible fake tracks (<0.01%)

CHEP2023: Xiaocong Ai et al : Performance of track reconstruction at STCF using ACTS

Global PID algorithm based on the ML techniques

- The PID for the full momentum range is essential for charm physics studies and fragmentation function studies.
  - The hadrons with low momentum identified with measurements of dE/dx in the MDC.
  - The leptons and neutral particles identified with measurements of the EMC and the MUD.
- Taking BDT (based on XGBoost) as a base model, further exploration of its physical potential
- Integrating all sub-detector information and exploring the PID performance of the detector

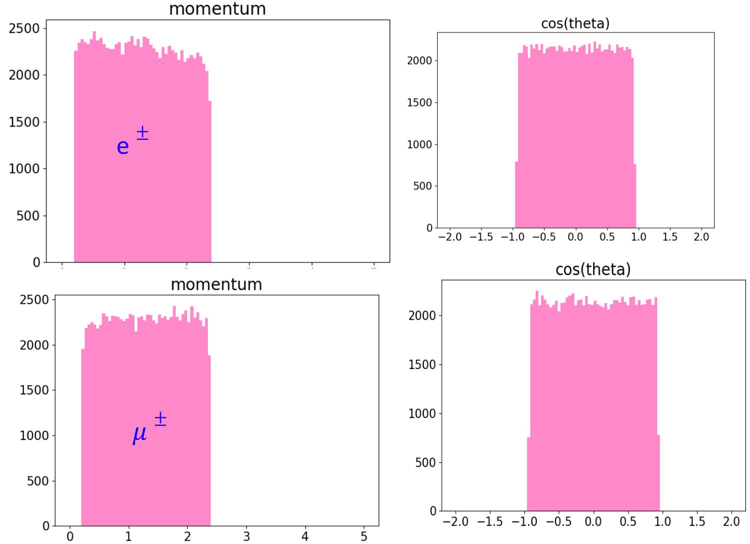


- Data sample produced by OSCAR
  - 50000 tracks for each type ($e^{\pm}$, $\mu^{\pm}$, $\pi^{\pm}$, $K^{\pm}$, $p^{\pm}$)
  - $p \in (0.2, 2.4)$ Gev/c, $\theta \in (20°, 160°)$, phi = 0°
  - Train:Validation:Test = 8:1:1 after pre-processing

Global PID algorithm based on the ML techniques

- The PID for the full momentum range is essential for charm physics studies and fragmentation function studies.
  - The hadrons with low momentum identified with measurements of dE/dx in the MDC.
  - The leptons and neutral particles identified with measurements of the EMC and the MUD.
- Taking BDT (based on XGBoost) as a base model, further exploration of its physical potential
- Integrating all sub-detector information and exploring the PID performance of the detector

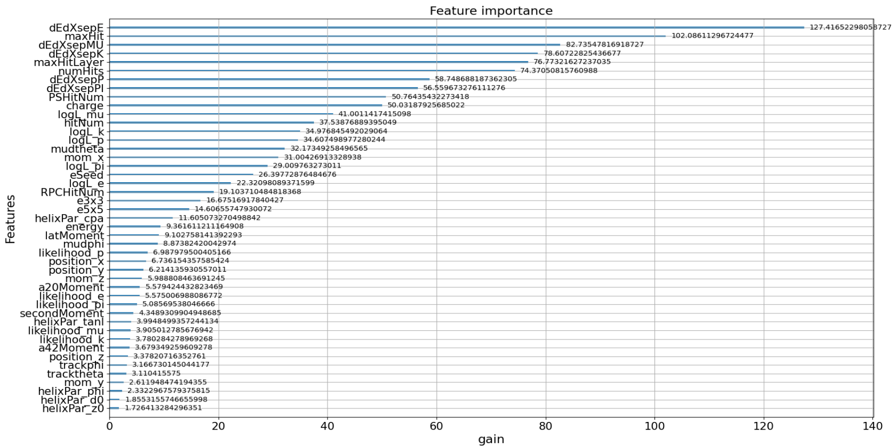
- Data sample produced by OSCAR
  - 50000 tracks for each type ($e^{\pm}$, $\mu^{\pm}$, $\pi^{\pm}$, $K^{\pm}$, $p^{\pm}$)
  - $p \in (0.2, 2.4)$ Gev/c, $\theta \in (20°, 160°)$, phi = 0°
  - Train:Validation:Test = 8:1:1 after pre-processing



Global PID algorithm based on the ML techniques

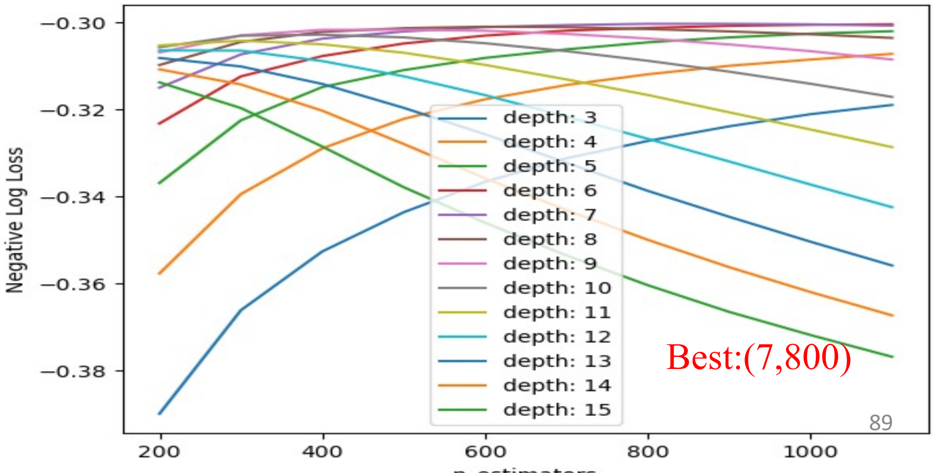
❖ Feature Selection

- Tracker/dEdx/RICH/DTOF/ECAL/MUD reconstructed variables
- 45 features kept according to feature importance distribution



❖ Optimal Hyperparameters

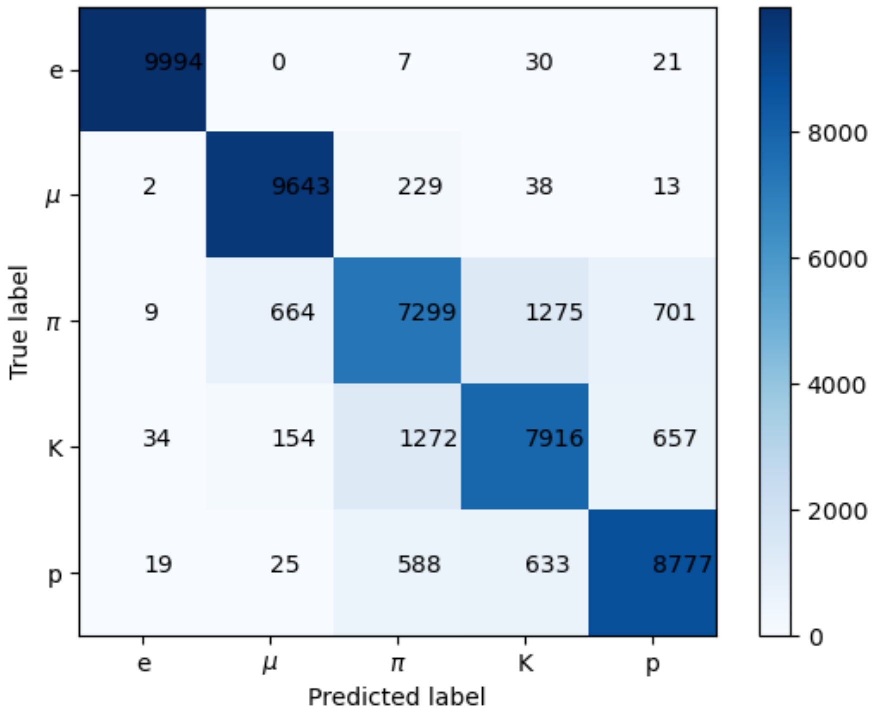
- Search range of max_depth: [200,1200]
- Search range of n_estimators: [3,15]



Global PID algorithm based on the ML techniques

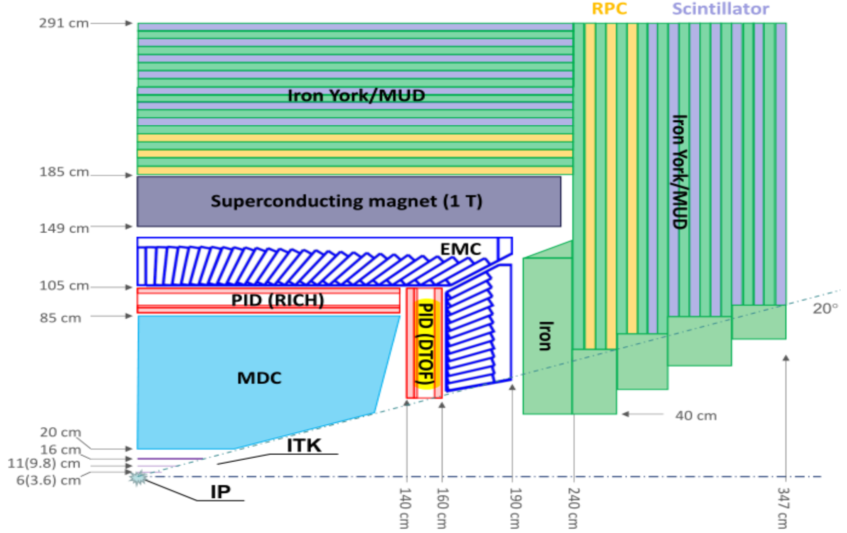
- Performance Analysis
  - BDT model(based on XGBoost) is trained and optimized to discriminate (e, μ, π, k, P)
  - Signal efficiency : *number of* signal *selected correctly/total number of* signal
  - Good performance for leptons
  - Hadron performance is sub-optimal due to the optimization of PID systems
  - Performance needs to be further validated !!

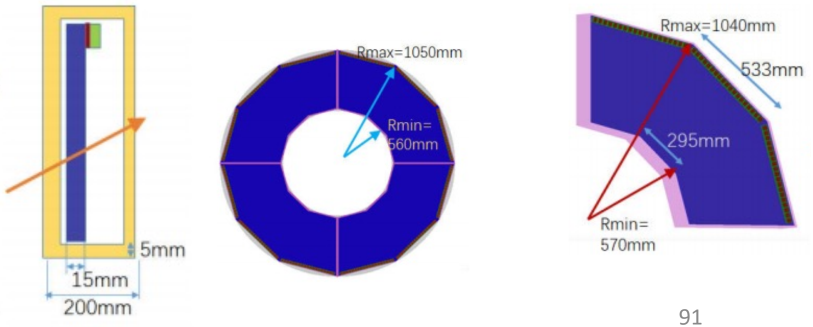
- **Confusion matrix**



STCF Particle Identification System

- To achieve a 3σ separation between kaons and pions with a momentum up to 2 GeV/c.
- the PID system of the STCF is designed
  - A ringing imaging Cherenkov detector (**RICH**) in the barrel
  - A time-of-flight detector based on the detection of the internal total-reflected Cherenkov light (**DTOF**) in the endcap



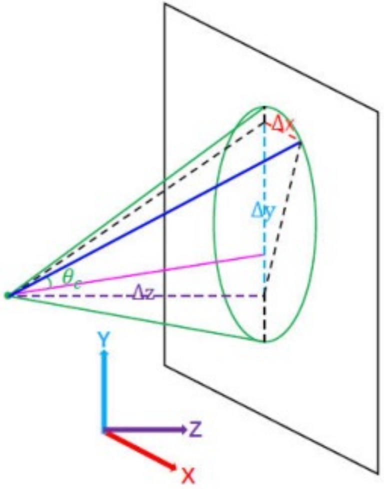


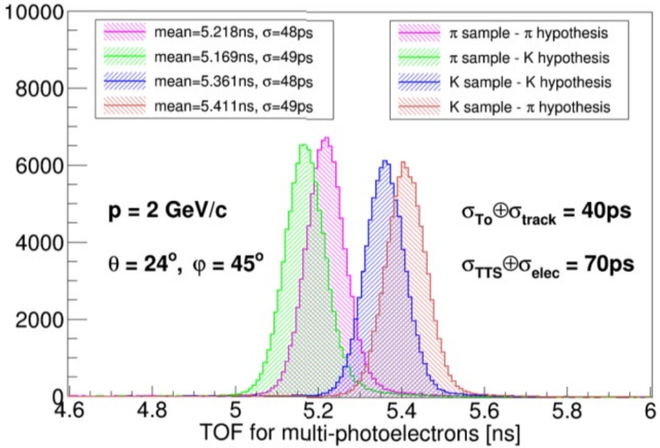
The likelihood method for DOF PID

- Building likelihood probability density function based on reconstructed TOF distribution

- $cos(\overline{\theta_c}) = \frac{1}{n_p \beta} = \frac{\overrightarrow{v_t} \cdot \overrightarrow{v_p}}{|\overrightarrow{v_t}| \cdot |\overrightarrow{v_p}|}$
  - $\overrightarrow{v_t} = (a, b, c), \quad |\overrightarrow{v_t}| = 1$
  - $\overrightarrow{v_p} = (\Delta x, \Delta y, \Delta z)$
- $LOP = H \cdot \sqrt{\frac{\Delta x^2 + \Delta y^2 + \Delta z^2}{\Delta y^2}}$
- $TOF = T - TOP - T_0 = T - \frac{LOP \overline{n_g}}{c} - T_0$

$$\mathcal{L}_h = p_h(N_{p.e.}) \prod_{i=0}^{N_{p.e.}} f_h(TOF_i)$$





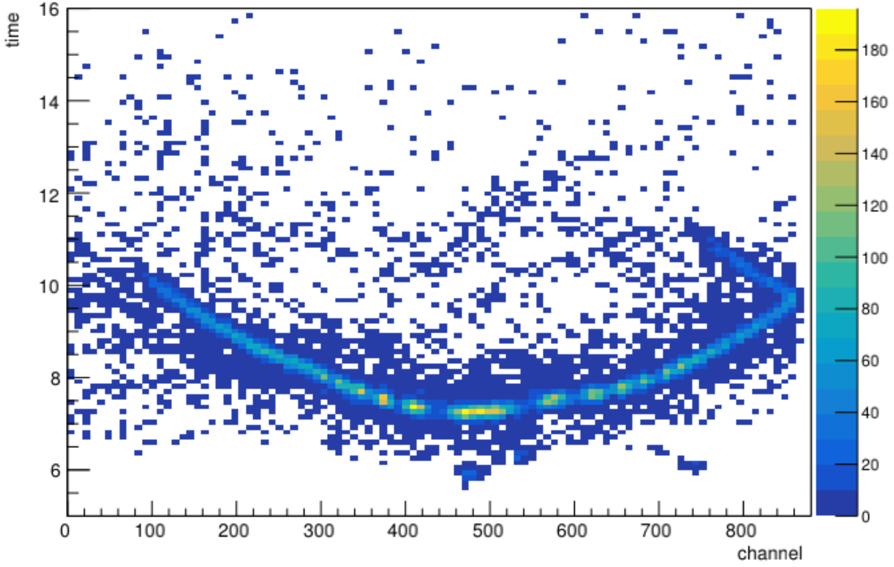
Qi, B et al., DIRC-like time-of-flight detector for the experiment at the Super Tau-Charm Facility. JINST, 16(08), P08021.

The CNN method for DOF PID

- A convolutional neural network (CNN) for pions/kaons identification is developed to utilize both timing and spatial information of the hits and takes two-dimensional pixel maps as the input

- The pixel map of photons
  - X-axis: the hit position of Cherenkov photon collected by PMT
  - Y-axis: the arrival time of Cherenkov photon collected by PMT
  - Value: the number of photons within in this bin

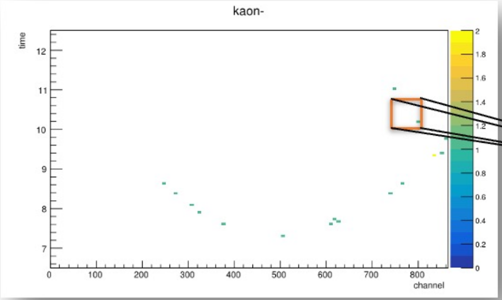
- The image-like data represents the topologies of Cherenkov photons generated by different particles



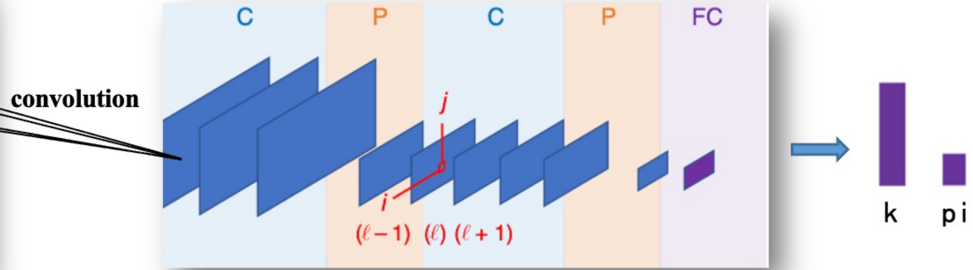
CHEP2023: Zhipeng Yao et al : Pion/Kaon Identification at STCF DTOF Based on CNN/QCNN

The Structure of CNN

- CNN consists of interlaced convolutional layers and pooling layers, and ends with a fully connected layer.
  - The primary purpose of the convolution layer is to extract new hidden features using convolution kernels
  - The pooling layer is used to reduce the dimension of data, reducing the resources required for learning and avoiding overfitting
  - The full connection layer adopts softmax full connection, and the activation value obtained is the picture feature extracted by convolutional neural network.



single event



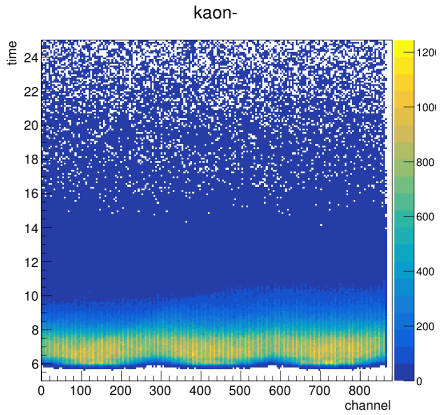
CHEP2023: Zhipeng Yao et al : Pion/Kaon Identification at STCF DTOF Based on CNN/QCNN

Data Sample

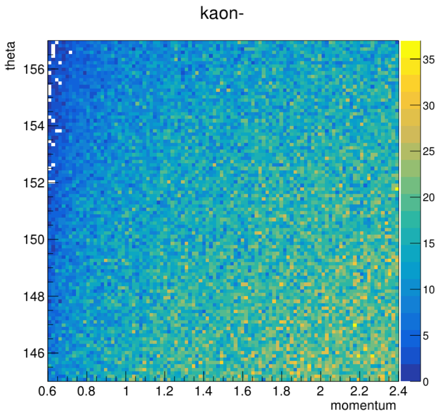
- MC sample is produced with OSCAR
  - pi+ : pi- : k+ : k- = 1 : 1 :1: 1
  - 0.6 Gev/c < p < 2.4 Gev/c
  - 23° < theta < 35°
  - N_photons > 10

- $0 \leq$ channel ≤ 868
- $5.5 \leq$ time ≤ 15.5 ns
- Bin number: channel * time = 217 * 200

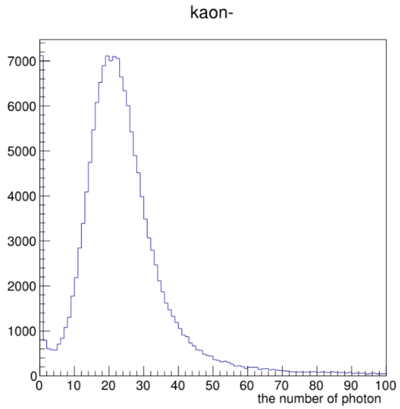
- Several distributions of all same Kaon$^-$ are as follows



2D channel-time



momentum(Gev/c)-theta



the number of photons

The Performance of CNN

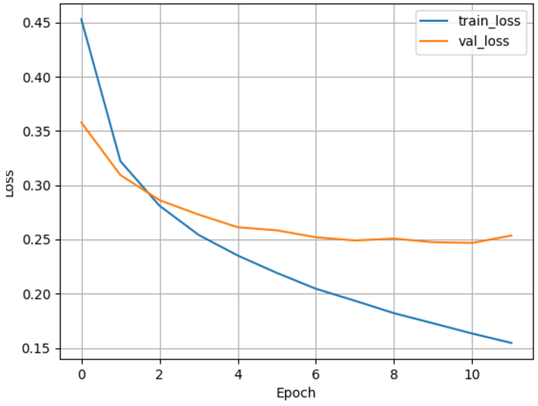
❖ The structure and parameters of CNN:

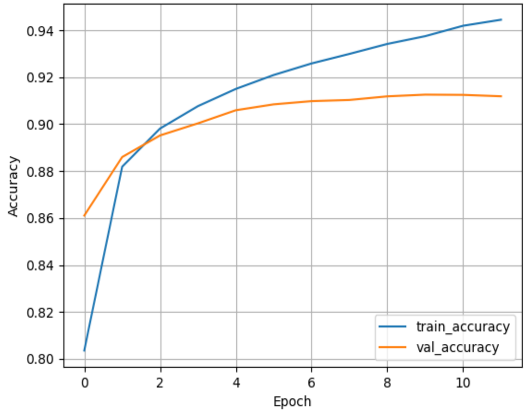
- Conv2D (32, (5, 5), activation='relu'), MaxPooling2D ((2, 2))
- Conv2D (32, (5, 5), activation='relu'), MaxPooling2D ((2, 2))
- Flatten(), Dense(1024,activation='relu'), Dense(2)
- learning_rate = 1e-5, batch_size = 64

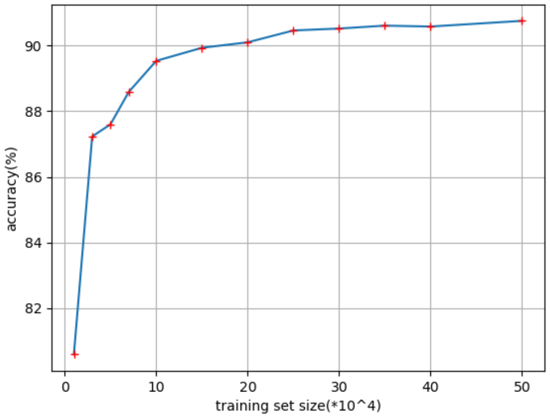
❖ Data set:

- training set: 200k
- validation set : 70k
- test set : 70k

❖ Test set accuracy : 91.76%

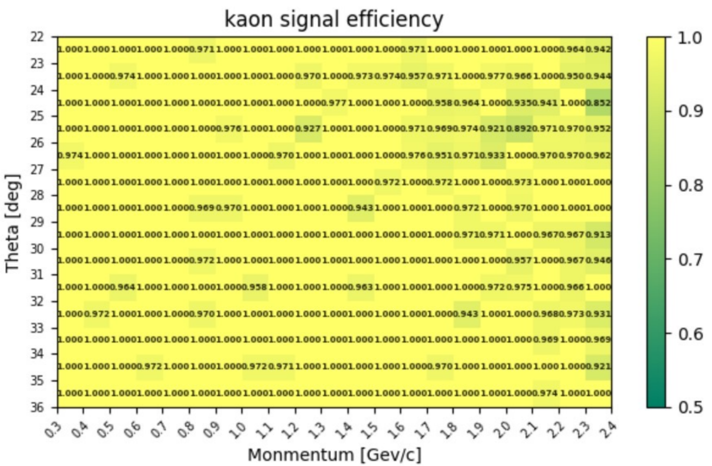


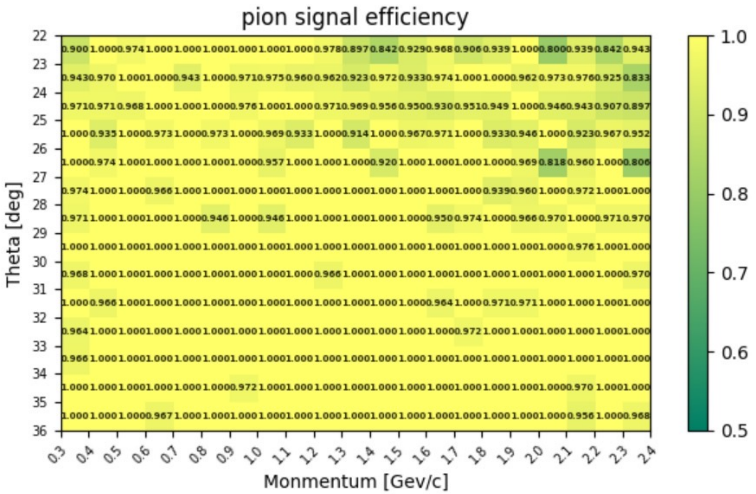




The preliminary results show the CNN model has a promising performance against the pion/kaon identification

Performance of CNN method

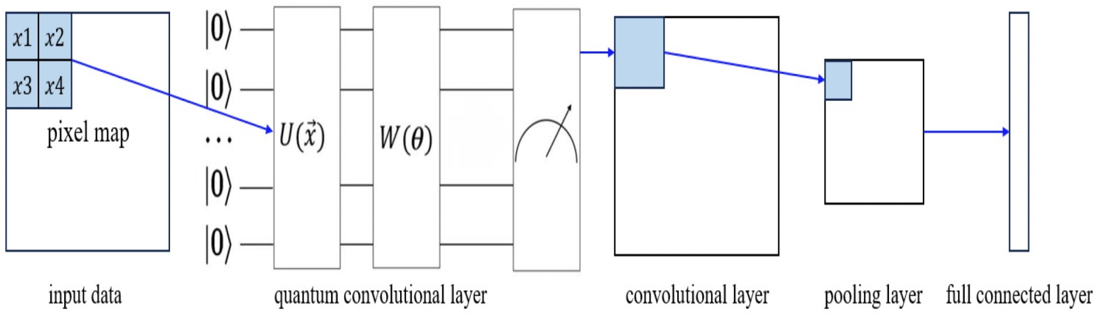




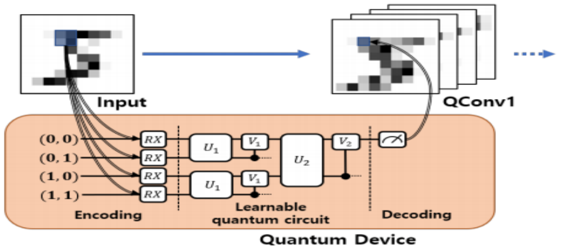
- The results show the CNN model has a promising performance

The QCNN method for DOF PID

- A quantum convolution neural network (QCNN) is developed as a proof-of-concept work exploring possible quantum advantages provided by quantum machine learning methods.
- Under TensorFlow Quantum Simulator platforms, we have developed a trainable quantum convolution layer that can replace the traditional convolution layer in CNN.
- Data Encoding Circuit
  - A small region of the input image, a 2×2 square, is embedded into a quantum circuit. This is achieved with RX rotation gate applied to the qubits initialized in the |0> state.



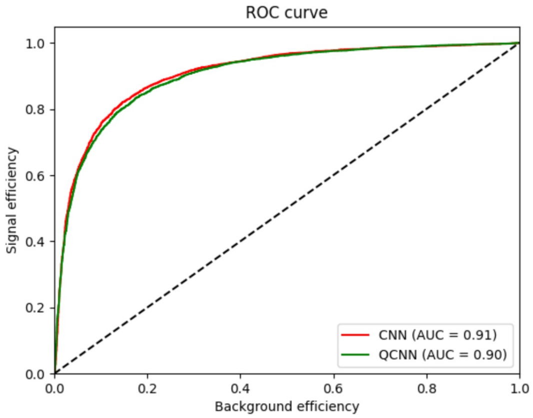
$$RX(\theta) = \exp\left(-i\frac{\theta}{2}X\right) = \begin{pmatrix} \cos\frac{\theta}{2} & -i\sin\frac{\theta}{2} \\ -i\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{pmatrix}$$



http://doi.org/10.1109/ICTC49870.2020.9289439

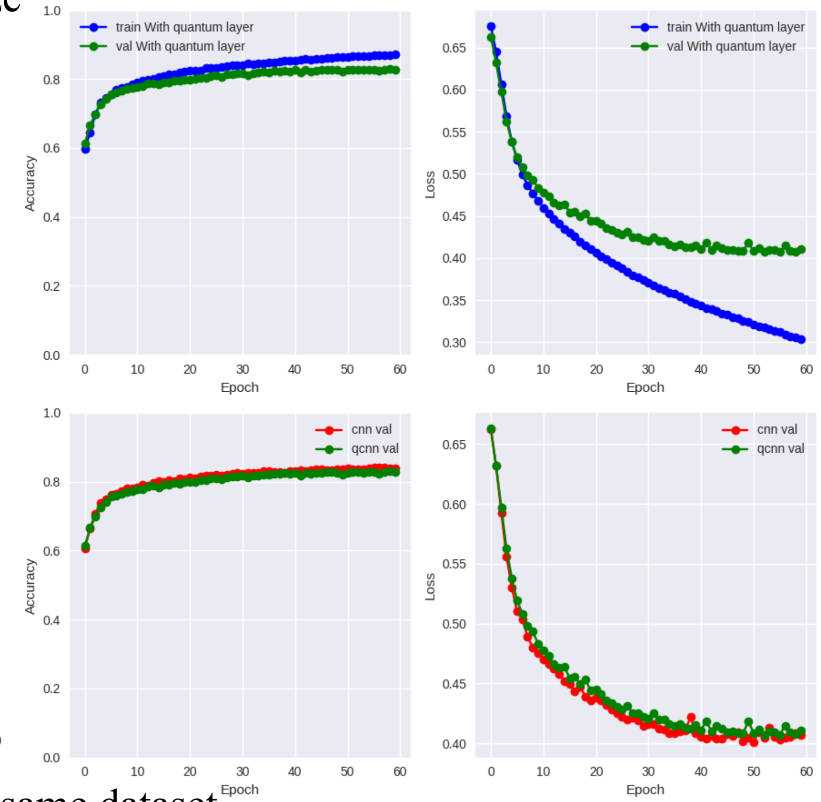
Performance of QCNN

- The 217×200 size dataset was downscaled to 32×32 size
- The structure and parameters of QCNN:
  - **QCONV**(1, (2, 2), activation='relu'), MaxPooling2D ((2, 2))
  - Conv2D (16, (2, 2), activation='relu'), MaxPooling2D ((2, 2))
  - Flatten(), Dense(128, activation='relu'), Dense(2)
  - learning_rate = 0.0001, batch_size = 16



- Data set:

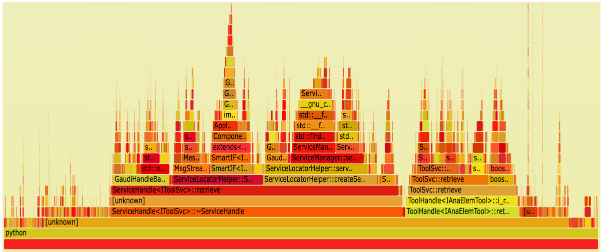
  training set: 20000
  validation set : 10000
  test set : 10000

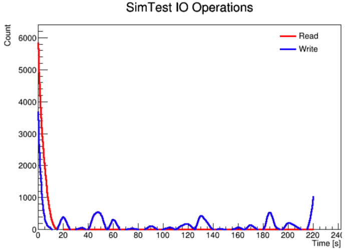


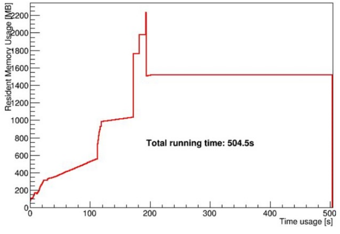
- QCNN achieved similar performance with CNN on the same dataset

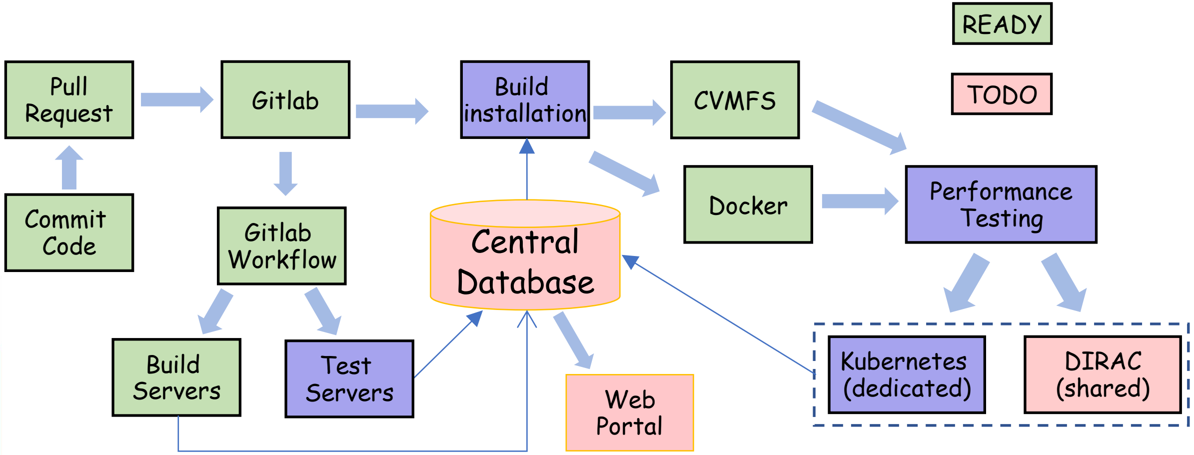
Automated Software Validation

- A software validation toolkit is developed, to support building software validation on different levels
  - Unit test, integrated test, software performance profiling and physics result validation
- Integrated with Gitlab Action system for automated validation
  - Trigger validation jobs on different levels on schedule/commits
- Adopted by CEPC and Key4hep







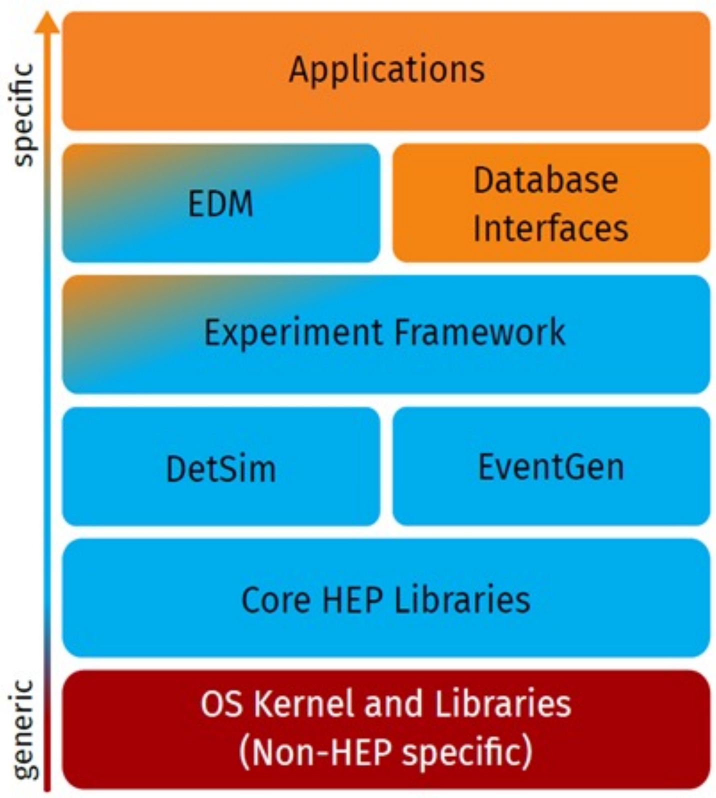


CEPC Offline Software System

History of CEPC software

- The development of CEPC software first started with the iLCSoft
  - Reused software iLCSoft modules: Marlin, LCIO, MokkaC, Gear
  - Developed CEPC components for simulation and reconstruction
  - Produced M.C. data for detector design and physics potential studies
  - CDR was released in Nov, 2018, based on results from the iLCSoft
- New CEPC software (CEPCSW) prototype was proposed at the Oxford workshop in April 2019
- The consensus among CEPC, CLIC, FCC, ILC and other future experiments was reached at the Bologna workshop in June, 2019
  - Develop a Common Turnkey Software Stack (Key4hep) for future collider experiments
  - Maximize the sharing of software components among different experiments

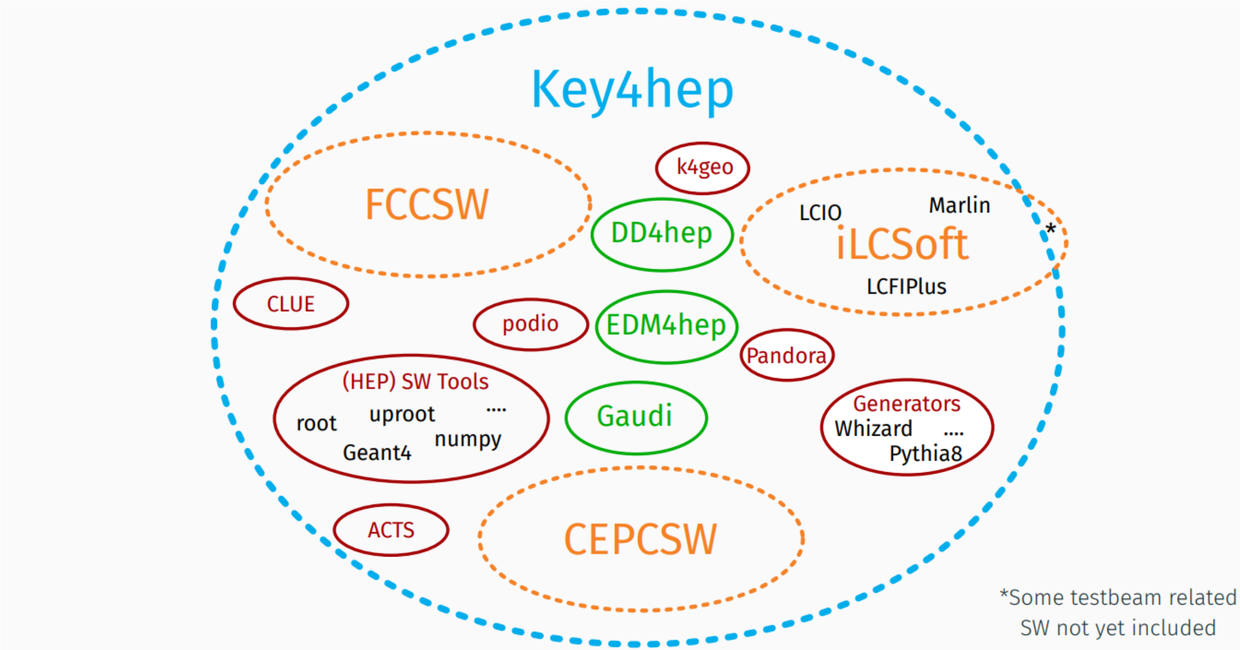
Key4hep



From Thomas Madlener, Epiphany Conference 2021

- The Key4hep is being developed to provide a common software stack for CEPC, CLIC, FCC and ILC experiments:
  - Application layer of modules/algorithms/processors performing physics task
  - Data access and representation layer
  - Experiment core orchestration layer
    - (Marlin, Gaudi, CMSSW, ...)
  - Specific components reused by many experiments
    - (DD4hep, Delphes, Pythia, ...)
  - Core HEP libraries
    - (ROOT, Geant4, CLHEP, ...)
  - Commonly used tools and libraries
    - (Python, CMake, boost, ...)
- IHEP and SDU are involved in Key4hep development as non-EU members.

Common Software Stack: Key4hep



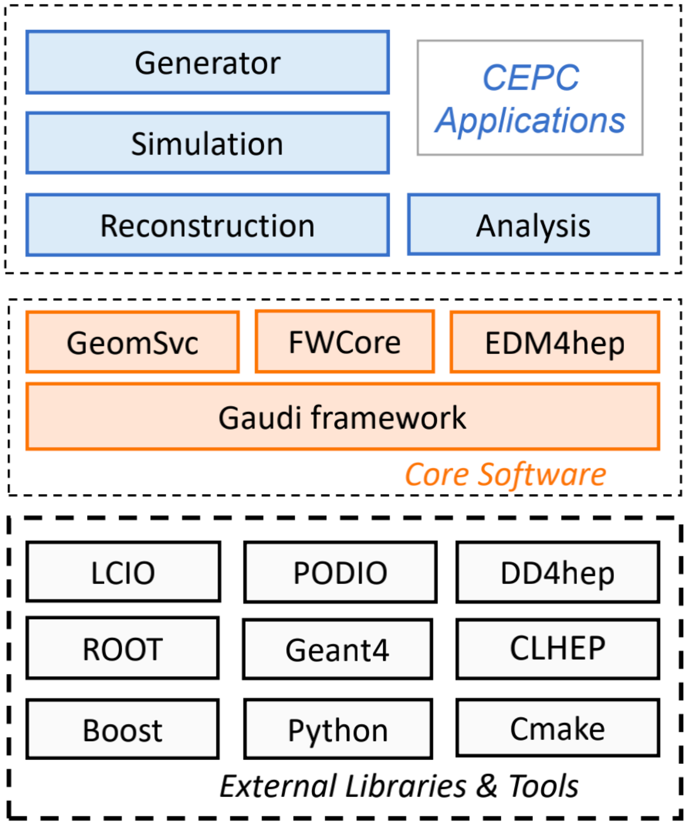
- Develop and maintain project specific software and workflows

  (T.Madlener | Key4hep & EDM4hep, CEPC workshop, Edinburgh)

CEPCSW Core software

https://github.com/cepc/CEPCSW

- CEPCSW software structure
  - Core software
  - Applications: simulation, reconstruction and analysis
  - External libraries
- Core software
  - Gaudi/Gaudi Hive: defines interfaces to all software components and controls their execution
  - EDM4hep: generic event data model
  - K4FWCore: manages the event data
  - DD4hep: geometry description
  - CEPC-specific components : generator, Geant4 simulation, beam background mixing, fast simulation, machine learning interface, etc.



Status of CEPCSW

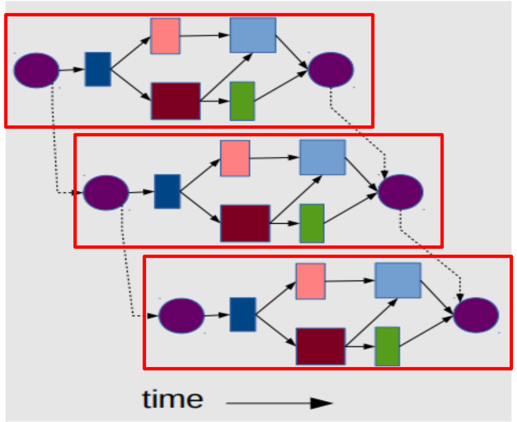
- Detector concepts
  - CDR (baseline design)
  - The 4th concept
- MC Generators
  - Multiple formats supported: HepMC, HepEvt, StdHep, LCIO
  - GuineaPig++ for MDI
  - Particle Gun
- Simulation
  - G4 simulation framework
  - Fast simulation algoritrhms e.g.ML-based dE/dx simulation
  - Digitization 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
  - Usage of EDM4hep, Identifier, etc.

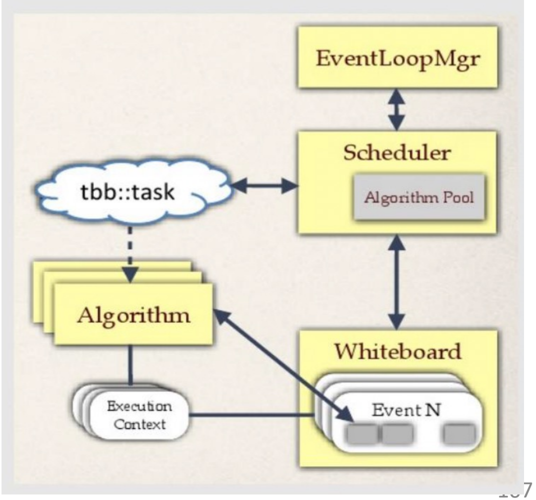
50 packages in total

Multi-threading with Gaudi Hive

- Gaudi Hive is a Gaudi extension supporting multi-threading and concurrent computing
  - Configuration, Initialization, Finalization are performed serially in "master" thread
    - only Algorithm: execute is concurrent
  - Algorithms must declare their inputs at initialization or dynamically with DataHandles
  - tbb::task wraps the pair (Algorithm*, EventContext)

- Multiple algorithms and events can be executed simultaneously using the data flow driven mechanism
  - Algorithms declare their data dependencies
  - Scheduler automatically executes Algorithms as the data becomes available

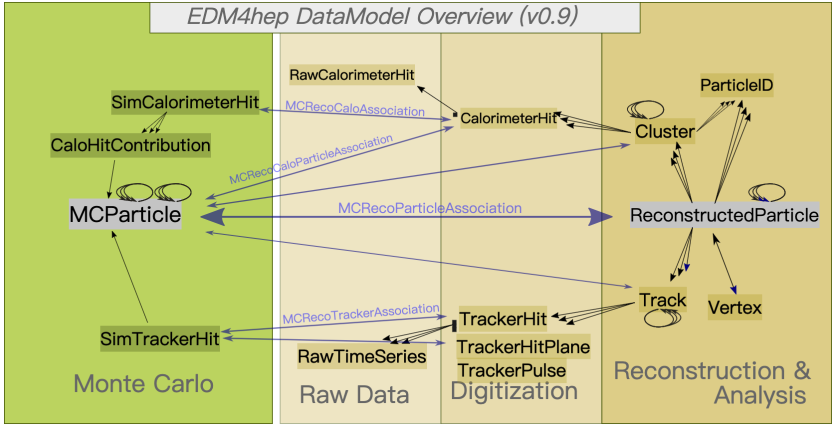


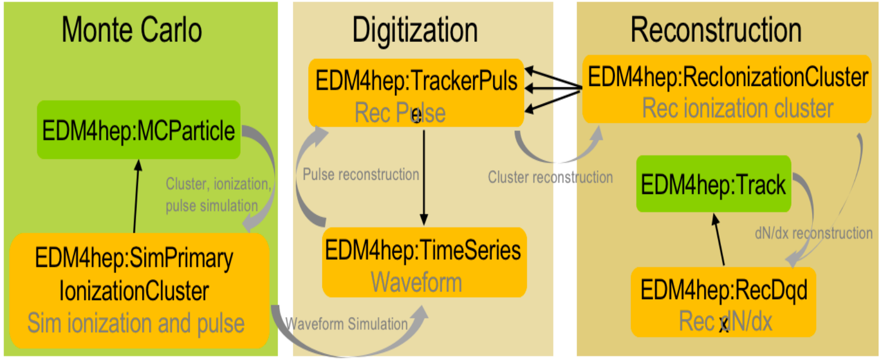


From Charles Leggettt

Event Data Model

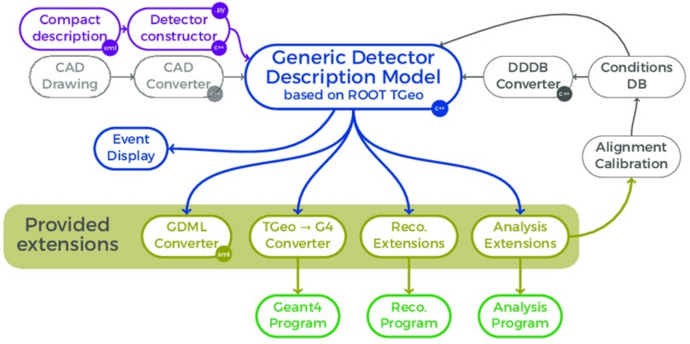
- EDM4hep is the common event data model (EDM) being developed for the future experiments like CEPC, CLIC, FCC, ILC, etc.
  - describing event objects created at different data processing stages and also reflecting the relationship between them.
- Due to the strong flexibility of EDM4hep, TPCHit was extended to accommodate the new needs:
  - By using the upstream mechanism of PODIO, a common EDM was implemented for both TPC and drift chamber

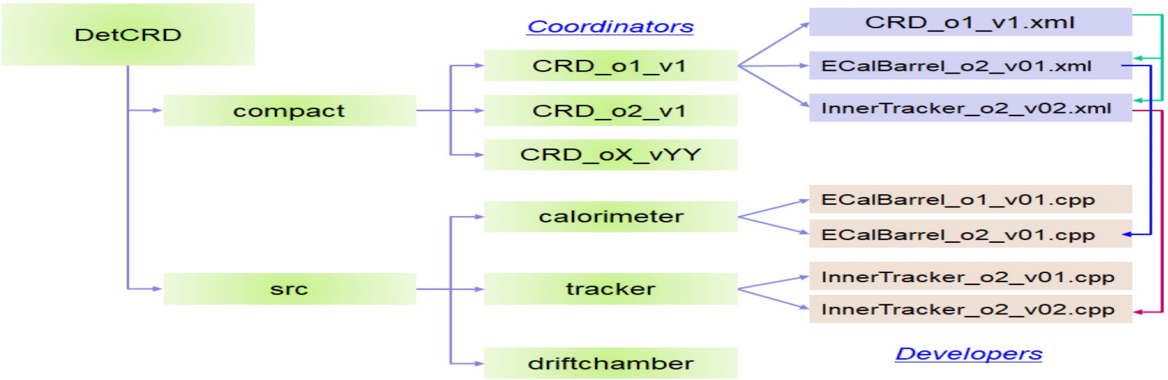




Detector Description

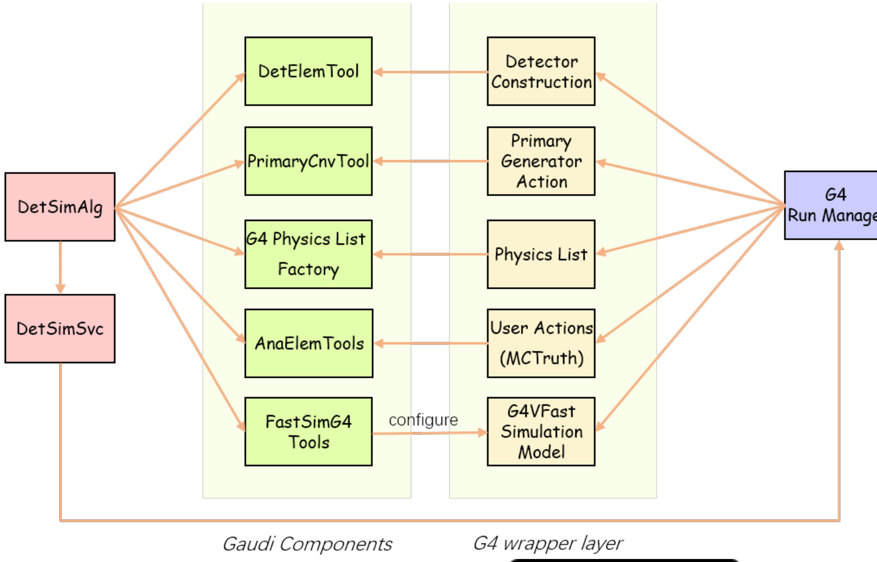
- DD4hep was adopted to provide a full detector description, which was generated from a single source
- Different detector design options are managed in the Git repository and a simulation job can be easily configured in runtime
- The non-uniform magnetic field was also implemented in CEPCSW

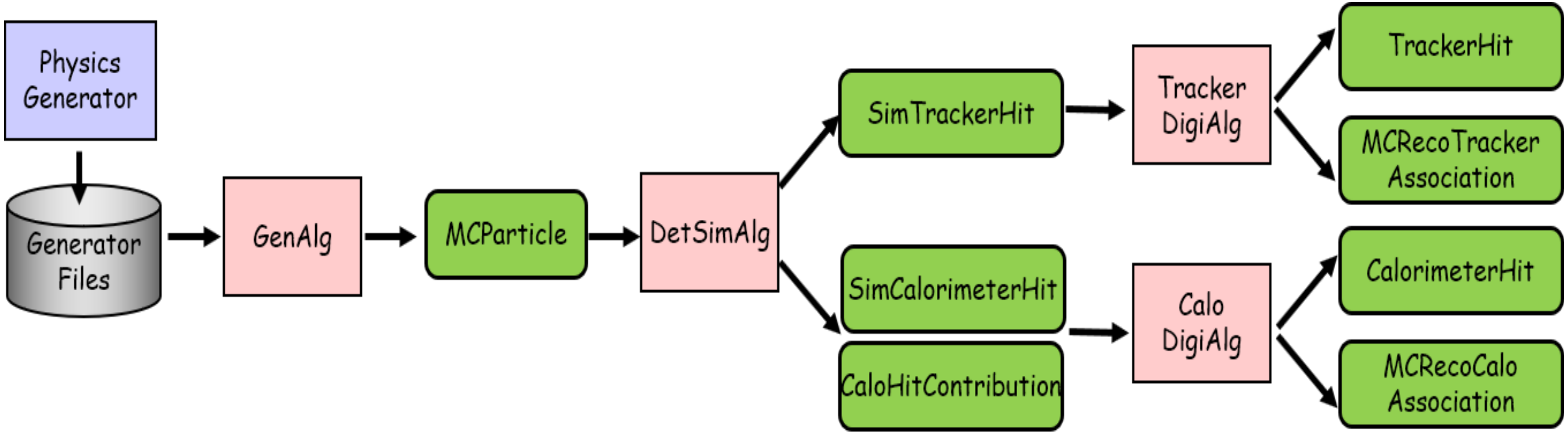




Simulation framework

- The detector simulation framework has been developed in CEPCSW.
  - A thin layer is developed to connect Geant4 and Gaudi.
  - The event loop is controlled by Gaudi with a customized G4RunManager.
  - The geometry conversion from DD4hep to Geant4 is done by DDG4.

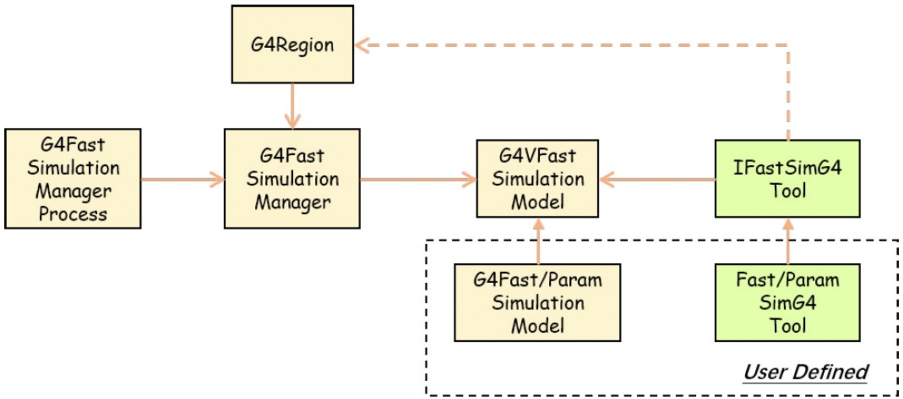


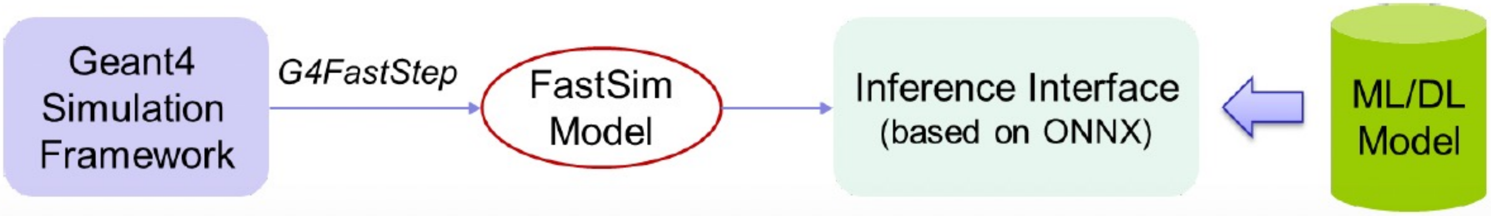


Fast Simulation

❖ To speed up the simulation , two fast simulation interfaces are developed to integrate different fast simulation models into Geant4 in CEPCSW

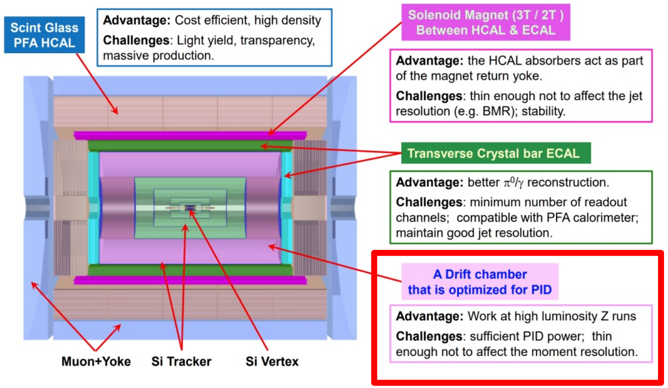
- Region based: when a particle enter a region, fast simulation will be triggered by Geant4.
- Support ML methods via ONNX inference interface.



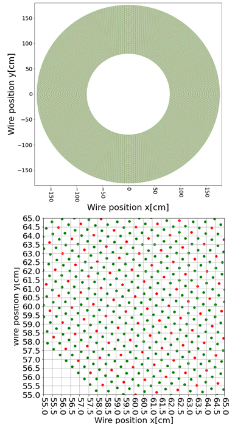


Drift chamber

- The CEPC experiment mainly aims to precisely measure the property of the Higgs boson.
- Physics requirements: high track efficiency (~100%), momentum resolution (<0.1%), PID (2σ p/K separation at P < ~ 20 GeV/c), etc.
- For the 4th conceptual detector, silicon detector and drift chamber (DC) are designed to provide both tracking and PID for charged particles.
- Both detector design and physics potential studies needs strong support of simulation and reconstruction software.



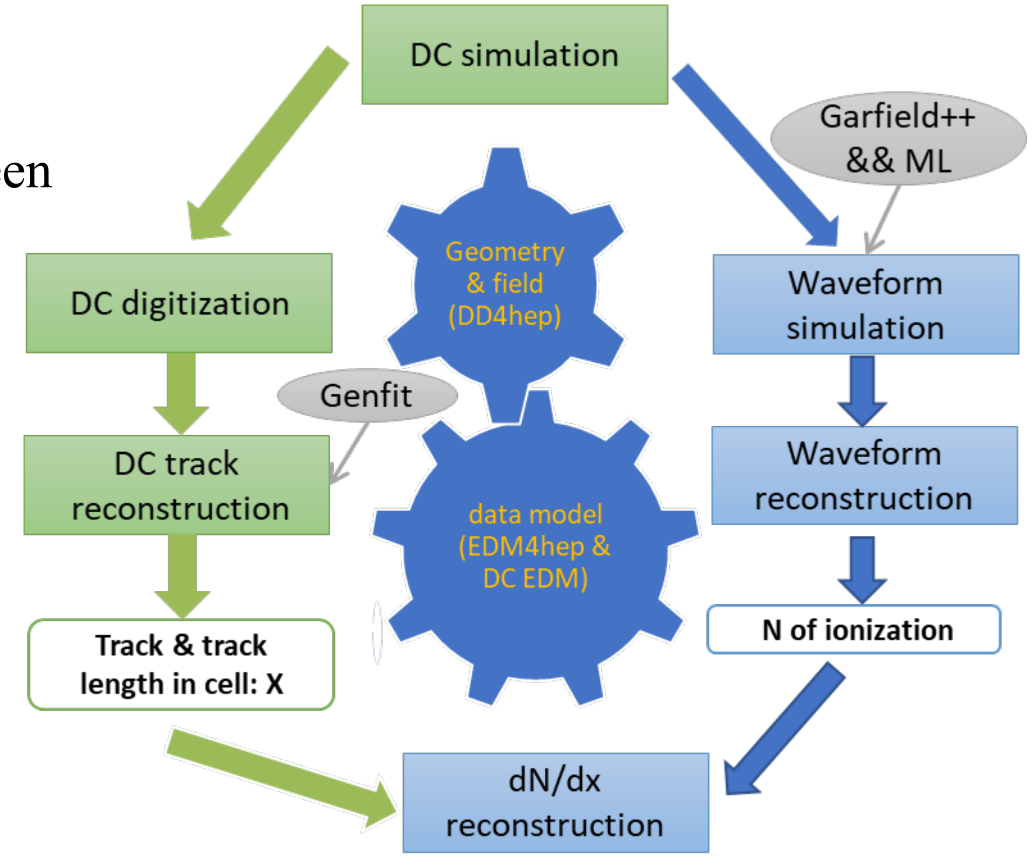
Half length	2980 *mm*
Inner and outer radius	800mm to 1800 *mm*
# of Layers	100/55
Cell size	~10mmx10mm/18mmx18mm
Gas	$He:iC_4H_{10}$=90:10
Single cell resolution	0.11 *mm*
Sense to field wire ratio	1:3
Total # of sense wire	81631/24931
Stereo angle	1.64~3.64 *deg*
Sense wire	Gold plated Tungsten ϕ=0.02*mm*
Field wire	Silver plated Aluminum ϕ=0.04*mm*
Walls	Carbon fiber 0.2 *mm*(inner) and 2.8 *mm*(outer)



Cell structure

DC software

- ❖ The drift chamber software has been developed from scratch
- ❖ Geometry and field map
  - DD4hep
  - Non-uniform magnetic field
- ❖ Data model
  - EDM4hep and dN/dx event model
- ❖ Drift chamber
  - DC simulation
  - DC digitization
  - Track finding
  - Track fitting
  - Multi track reconstruction

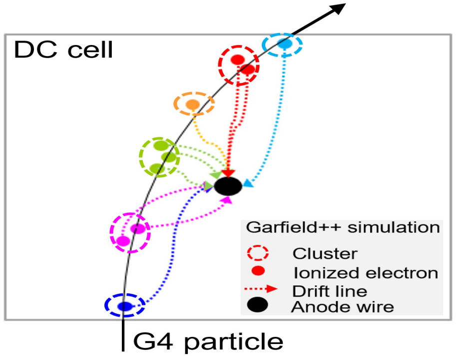


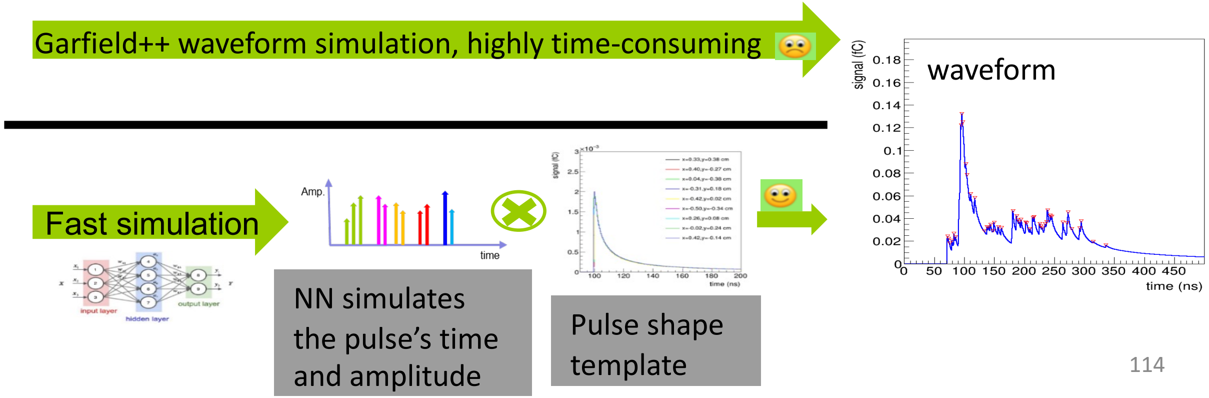
Drift chamber simulation and reconstruction flow

Simulation of drift chamber

❖ TrackHeedSimTool (Gaudi tool) combining Geant4 and Garfield++ to simulate the complete response of the gaseous detector

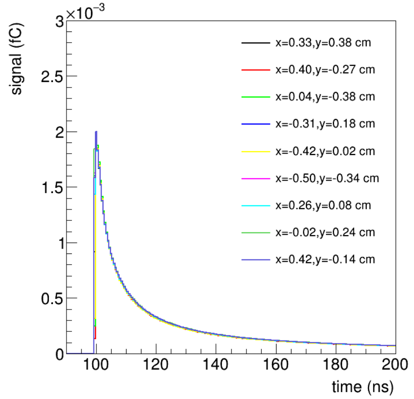
- For each G4Step, information (particle type, initial position, momenta, and step length) will pass to Heed to simulate ionization process. The kinetics of G4Track will be updated according to its energy loss.
- Using TrackHeed (from Garfield++) to create the ionization electron-ion pairs (for both primary and secondary ionizations), the deposited energy will be used to update the energy of the G4Particle
- Using NN to simulate the time and amplitude of each pulse for each ionized electron (for fast waveform sim.)
- Output: primary, total ionization, and pulse information, saved in EDM

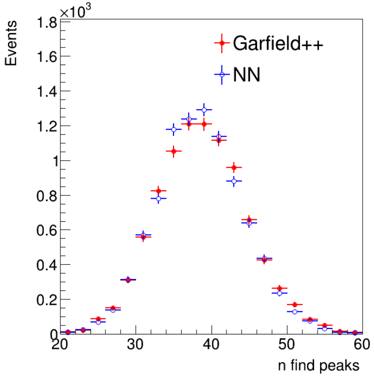




Fast waveform simulation

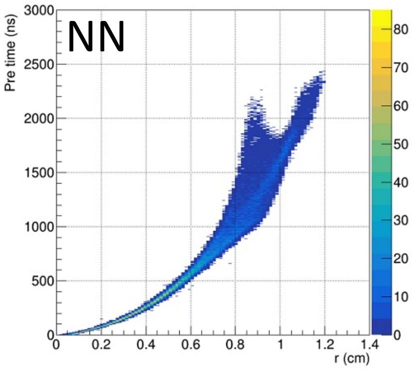
- ❖ Extremely time consuming to use Garfield++ to simulate
  - Drift of ions and electrons, amplification via electron avalanches and final signal generation
- ❖ Studies show that the waveform shape of each ionized electron in Garfield++ is similar. Main difference is the beginning time and amplitude
- ❖ Using machine learning technique to learn the distributions of beginning time and amplitude for each ionized electron
  - Training sample is produced by Garfield++

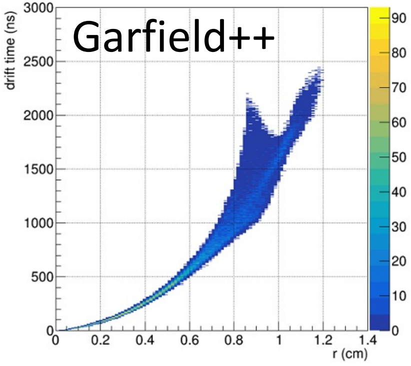




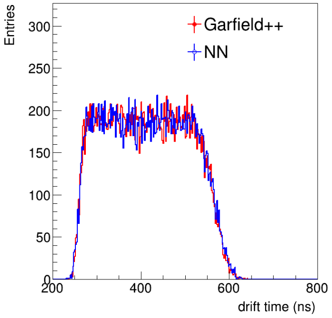
- ❑ Good agreement between NN and Garfield++
- ❑ ~200 times speed up

ML-based simulation

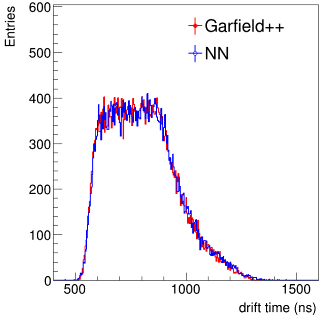




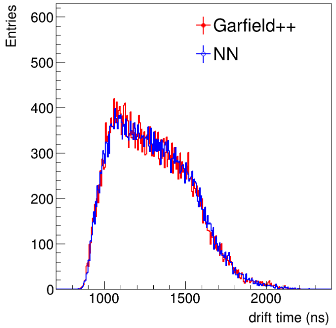
r: 0.4-0.6 cm



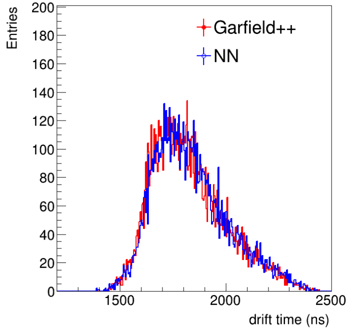
r: 0.6-0.8 cm



r: 0.8-1.0 cm



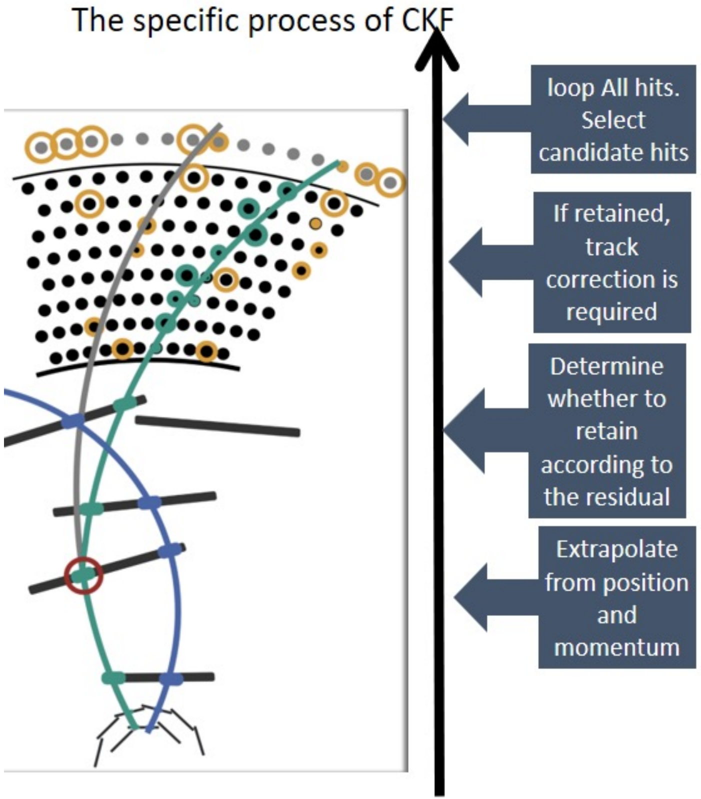
r: 1.0-1.2 cm



- Good agreement between the NN and Garfield++ simulation

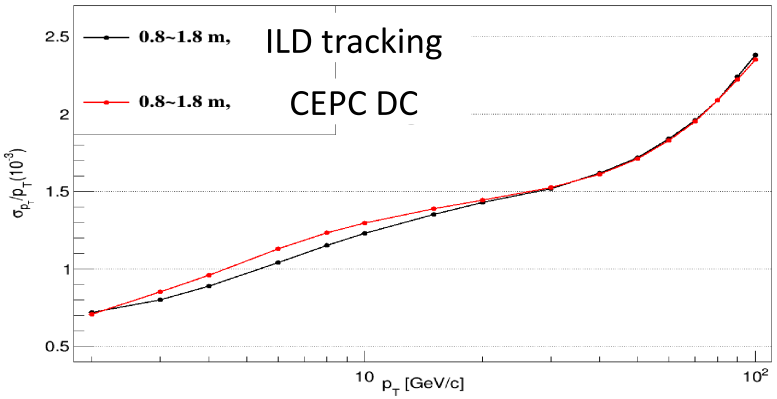
Track reconstruction (1)

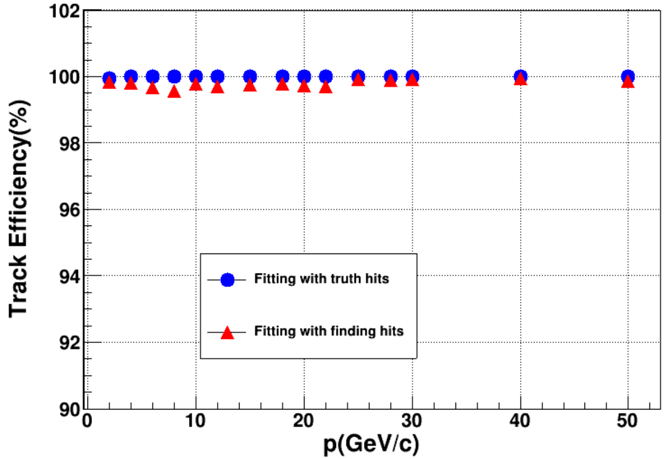
- Tracking with Combinatorial Kalman Filter (CKF) method
  - Used by many high energy physics experiments
- Track finding with CKF in drift chamber
  - Migrate from Belle2
  - Track segments reconstructed in the silicon detector, called seeds, are extrapolated to the DC and all the DC hits belonging to the track are collected
- Track fitting tool: Genfit
  - Experiment-independent generic track fitting toolkit
  - Official track fitting for BelleII, also used by PANDA, COMET, GEM-TPC etc.
  - Using DAF kalman filter

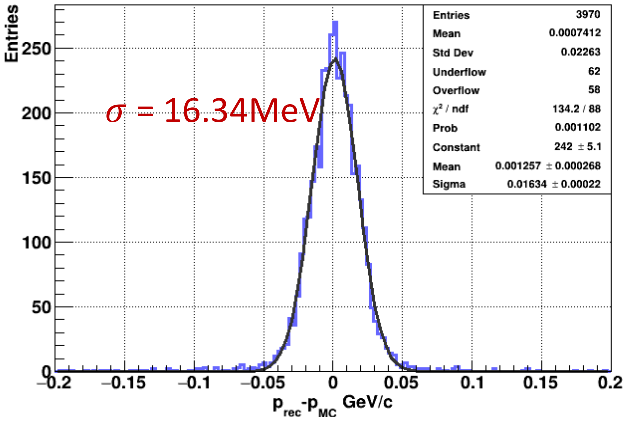


Tracking performance

- Data sample: Single particle $\mu^{-}, \theta = 50°$
- Track efficiency $\boldsymbol{\varepsilon} = N_1/N_2$
  - N1 is the number of track satisfying:
    - $\chi 2 < 400$
    - $N\_(DC\ hits\ on\ track)$>50
  - N2 is the numbre of silicon track
- Combined measurements of Silicon and Drift Chamber



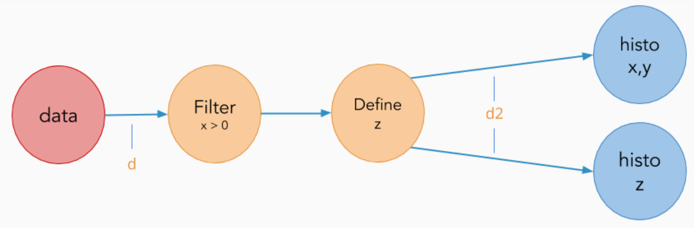




Momentum distribution of 10 GeV/c μ^{-}

Analysis toolkit based on RDataFrame (1)

- RDataFrame is a powerful tool for data analysis
  - Program language: Python and C++
  - Declarative programming and parallel processing are supported
  - EDM4hep data can be read directly
  - Being used by many experiments such as FCC-ee

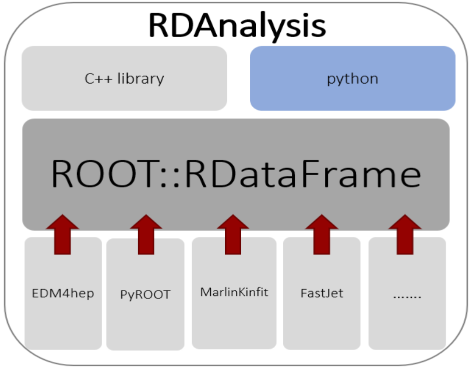


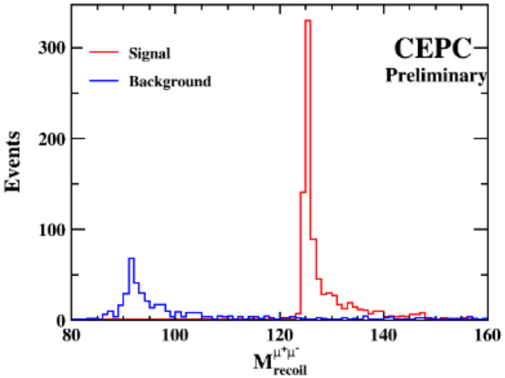

```cpp
// d2 is a new data-frame, a transformed version of d
auto d2 = d.Filter("x > 0")
           .Define("z","x*x + y*y");
// make histograms out of it
auto hz = d2.Histo1D("z");
auto hxy = d2.Histo2D({"hxy","hxy",16,-1,1,64,-1,1},"x","y");
```

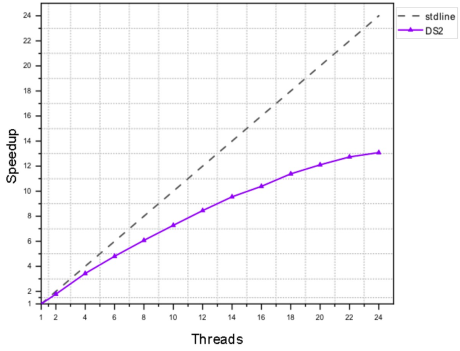
- Development of analysis tool for CEPCSW
  - Development of common components (functions)
    - Analysis functions in C++: event selection, filtering, Jet clustering, vertex fitting
    - Python for configuration: define analysis functions, input samples, output variables
  - Performance test

Analysis toolkit based on RDataFrame (2)

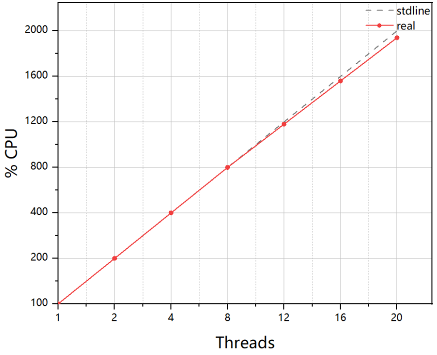
- Several packages are ported from FCC analysis, more are being implemented
  - FastJet, MarlinKinfit
  - Vertex fit, jet tag, PID etc.
- Functionalities and performance test with two analysis channels
  - e+e- -> Z(mumu)H
  - e+e- ->H(2jet) mumu







Identical results with Marlin



performance test

Machine Learning Integration

- ONNX/ONNX Runtime have been integrated with CEPCSW
- Provided an example, OrtInferenceAlg,
  - In initialize()
    - Create a session object of ONNX runtime
    - Load and run an ONNX model
  - In execute()
    - Compute output for an input data
- Fast pulse simulation in the drift chamber provided as an example (MLP)

```
Ort::MemoryInfo info("Cpu", OrtDeviceAllocator, 0, OrtMemTypeDefault);

auto input_tensor = Ort::Value::CreateTensor(info,
                                             inputs.data(),
                                             inputs.size(),
                                             dims.data(),
                                             dims.size());
std::vector<Ort::Value> input_tensors;
input_tensors.push_back(std::move(input_tensor));

auto output_tensors = m_session->Run(Ort::RunOptions{ nullptr },
                                     m_input_node_names.data(),
                                     input_tensors.data(),
                                     input_tensors.size(),
                                     m_output_node_names.data(),
                                     m_output_node_names.size());

for (int i = 0; i < output_tensors.size(); ++i) {
    LogInfo << "[" << i << "]"
            << " output name: " << m_output_node_names[i]
            << " results (first 10 elements): "
            << std::endl;
    const auto& output_tensor = output_tensors[i];
    const float* v_output = output_tensor.GetTensorData<float>();

    for (int j = 0; j < 10; ++j) {
        LogInfo << "[" << i << "]" << "[" << j << "] "
                << v_output[j]
                << std::endl;
    }
}
```

```
bool OrtInferenceAlg::initialize() {

    m_env = std::make_shared<Ort::Env>(ORT_LOGGING_LEVEL_WARNING, "ENV");
    m_seesion_options = std::make_shared<Ort::SessionOptions>();
    m_seesion_options->SetIntraOpNumThreads(m_intra_op_nthreads);
    m_seesion_options->SetInterOpNumThreads(m_inter_op_nthreads);

    m_session = std::make_shared<Ort::Session>(*m_env, m_model_file.c_str(), *m_seesion_options);
```

Introduction to Particle Physics at SDU

山东大学物理学科——历史沿革

1930年6月
国立青岛大学物理系成立

1946年3月
国立山东大学物理系在青岛复系

1958年秋
物理系随山东大学迁往济南

1960年
电子系

1970年
光学系

2000年7月
新山东大学成立，物理与微电子学院组成

2003年
空间科学与物理学院（威海）

2008年1月
物理与微电子学院更名为物理学院

2016年
微电子学院

2018年5月
粒子科学技术中心（青岛）

- 青岛：粒子物理与原子核物理为主
- 济南：凝聚态物理为主
- 威海：空间物理为主

粒子物理与原子核物理-发展历史

1930	1956	1981	1984	2005	2007	2008	2011	2017	2018
青岛 物理系 王普 核物理 导师 Lisa Meitner	济南	高能物理 博士点 (高校唯一)	全国 第一位 高能物理 学博士	王克明等 国家 自然科学 二等奖	国家 重点学科 王克明 当选院士	物理学院 粒子物理 与核物理 研究所	教育部 重点 实验室	双一流 学科 建设	青岛 粒子科 学技术 研究中心

王普教授(1902-1969)
物理系第一位教师
核裂变中缓发中子的发现者

山大粒子物理与核物理
学科奠基人

1958年，王普先生与基本粒子与原子核理论讲习班成员合影留念。（二排左起：第四人为冼鼎昌、第五戴元本、第六余寿绵、第八王普、第九张宗燧、第十谢毓章、第十一吴富恒、第十二朱洪元、第十六何祚庥，四排右一于良）（冼鼎昌院士提供照片）

讲师：朱洪元、张宗燧
主题：量子场论，色散关系等
学员：冼鼎昌、戴元本、何祚庥、……

“量子场论在我国第一次普及”
“标志着我国粒子物理理论研究的开始”

王克明院士(1939-2013)
2005年获国家自然科学二等奖
2007年当选中国科学院院士

山大粒子与固体相互作用
研究方向的开创者

粒子物理与粒子辐照教育部重点实验室

实验室主任梁作堂教授：山东大学讲席教授，1964年出生，杰青（2005），长江特聘（2009）

山东大学本科，德国柏林自由大学博士，导师孟大中教授

国务院学位委员会学科评议组成员（2008- ），中国物理学会常务理事（2011- ）
物理学院院长（2008-2012），威海空间科学与物理学院首任院长（2003-2012）
山东大学前沿交叉科学青岛研究院首任院长（2017-2019）
山东物理学会理事长（2010-2014），教育部物理学类专业教指委委员（2012-2018）

研究方向为粒子物理理论。从事强相互作用自旋物理理论研究。发表论文104篇，总引用2730余次

粒子物理理论

司宗国 教授
物理学院院长

1969年出生
杰青（2013）

霍英东青年教师基金
山东省有突出贡献的中青年专家
《中国物理C》编委

粒子物理软件平台

黄性涛 教授
实验室副主任

1976年出生
杰青 (2020)

BESIII实验IB 主席，STCF、Daya Bay、JUNO、LHAASO等离线软件平台负责人

粒子探测技术

王萌 教授
实验室副主任

1970年出生
中科院百人（2007）

中法粒子物理联合实验室(FCPPL)副主任

粒子与物质相互作用

王雪林 教授

1967年出生
全国优博（2004）

中国核学会辐照效应分会理事、
兰州重离子加速器国家实验室束流评审材料科学组专家

空间粒子与辐射探测

陈耀 教授
实验室副主任

1975年出生
杰青（2008）

赵九章优秀中青年科学奖及空间天气科学青年创新奖获得者
中国地球物理学会空间天气学委员会副主任

粒子科学技术中心（青岛）：师资队伍

固定人员：53人，
研究人员47人：教授 35人，副教授/副研究员13 人
技术支撑辅助人员5人

长江/杰青/千人/万人 ：4人

XZ、梁作堂、黄性涛、徐庆华

国家“四青”人才：17人

优青（2）：徐庆华、刘鹏
青千（7）：马连良、周剑、刘智青、易立、 曹杉杉、陈震宇、周启东
海外优青（5）：刘天博、张洋、熊伟志、李夏卿
青年拔尖（3）：高宁、 杨驰、张金龙

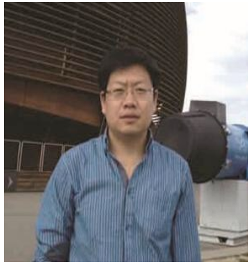
山大杰出人才体系：11人

特聘教授（1）：王萌
杰青学者（1）：祝成光
齐鲁学者（9）：李海峰、张宏、张金龙、杜杨洪岳、李冰、胡坤、魏树一、林挺、刘彦麟

加速器粒子物理实验与软件研究团队

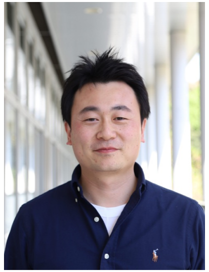
张学尧 教授
二级教授、博导
BES物理、软件

黄性涛 教授
国家杰青、博导
BES物理、软件



马连良 教授
青年千人、博导
BES物理、ATLAS物理

刘智青 教授
青年千人、博导
BES物理、Belle 物理

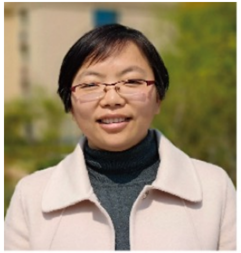


周启东 教授
海外优青、博导
Belle 物理, DAQ

李冰 研究员
齐鲁学者
ATLAS物理

刘彦麟 研究员
齐鲁学者
ATLAS物理

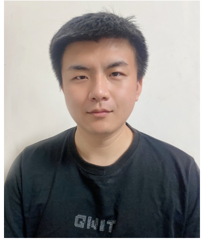
焦健斌 教授、博导
BES物理、Belle 物理



李晓玲 副教授、硕导
BES物理、Belle 物理

孙振田 副研究员
软件

秦小帅 副研究员
软件



李腾 副研究员
软件

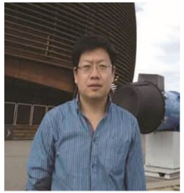
尹娜 正高级实验师
软件

左志宏
助理实验师

加速器粒子物理实验与软件研究团队

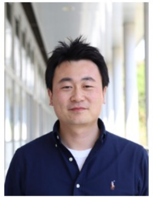
张学尧 教授
二级教授、博导
BES物理、软件

黄性涛 教授
国家杰青、博导
BES物理、软件



马连良 教授
青年千人、博导
BES物理、ATLAS物理

刘智青 教授
青年千人、博导
BES物理、Belle 物理

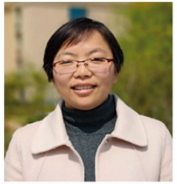


周启东 教授
海外优青、博导
Belle 物理, DAQ

李冰 研究员
齐鲁学者
ATLAS物理

刘彦麟 研究员
齐鲁学者
ATLAS物理

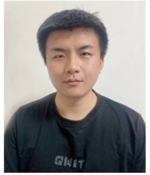
焦健斌 教授、博导
BES物理、Belle 物理



李晓玲 副教授、硕导
BES物理、Belle 物理

孙振田 副研究员
软件

秦小帅 副研究员
软件



李腾 副研究员
软件

尹娜 正高级实验师
软件

左志宏
助理实验师

- 正负电子对撞机实验（BESIII， Belle II)
- 大型强子对撞机实验（ATLAS)
- 粒子物理实验大数据处理软件与技术（BESIII、LHAASO、DayaBay、JUNO、HERD、STCF、CEPC)

国际合作

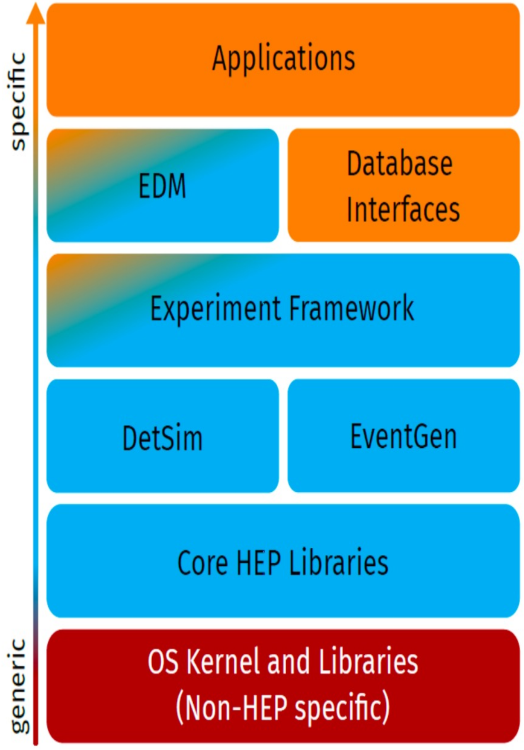
- 参与key4hep国际合作，共同开发未来对撞机软件

AIDAINNOVA COLLABORATION AGREEMENT
(THE "AGREEMENT")

BETWEEN: THE EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH ("CERN"), an Intergovernmental Organization having its seat at Geneva, Switzerland, acting as Coordinator in the names and on behalf of the Parties of the AIDAinnova Consortium (hereinafter collectively referred to as the "Consortium" and individually as "Consortium Members"),

AND: SHANDONG UNIVERSITY ("SDU"), established at Jinan, China,

Hereinafter each individually referred to as a "Party" and collectively as the "Parties",

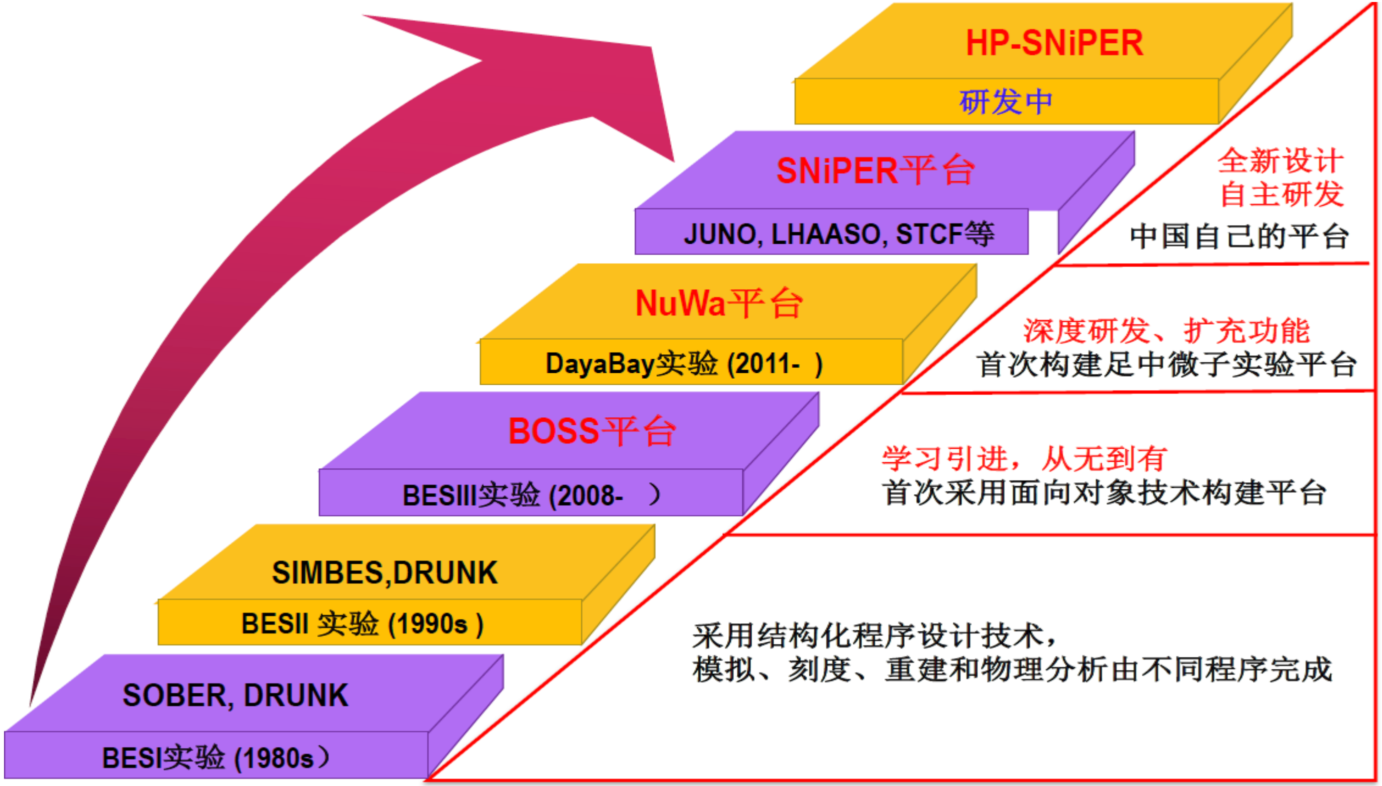


and STCF

国际合作

SDU @ CHEP2023： 6 talks, 2 posters

我国高能物理软件平台的研发历程



总结

- 离线软件是粒子物理实验研究的重要组成部分，主要包括框架、模拟、刻度、重建以及分析等软件系统，为分析海量实验数据获得物理成果提供关键技术
- 粒子物理高精度、高能量两个前沿物理研究对离线数据处理以及离线软件提出了新需求和新挑战
- 介绍我国牵头的BESIII、JUNO、STCF，CEPC等实验（或预研项目）离线软件系统研发的现状，很好地支撑了这些实验的预研、建设、运行和科学产出。
  - LHCb实验软件框架Gaudi
  - 我国自主研发的软件框架SNiPER
- 在发展和优化传统软件技术的同时，深入开展并行计算、异构计算、机器学习、量子计算等技术在离线数据处理中的应用研究和软件开发是离线软件发展的重要研究方向。

欢迎更多的优秀青年参加我国BESIII、JUNO、STCF、CEPC等实验的离线软件研发中!