STCF Core Software Status

Teng LI on behalf of the STCF core software development team

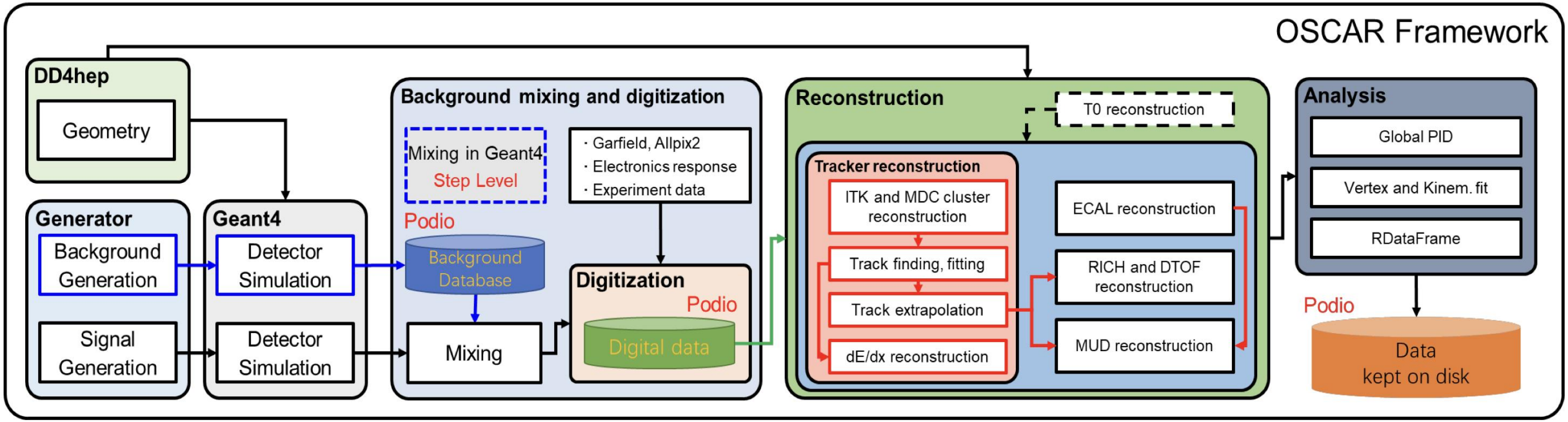
Shandong University

2025-11-25

FTCF 2025

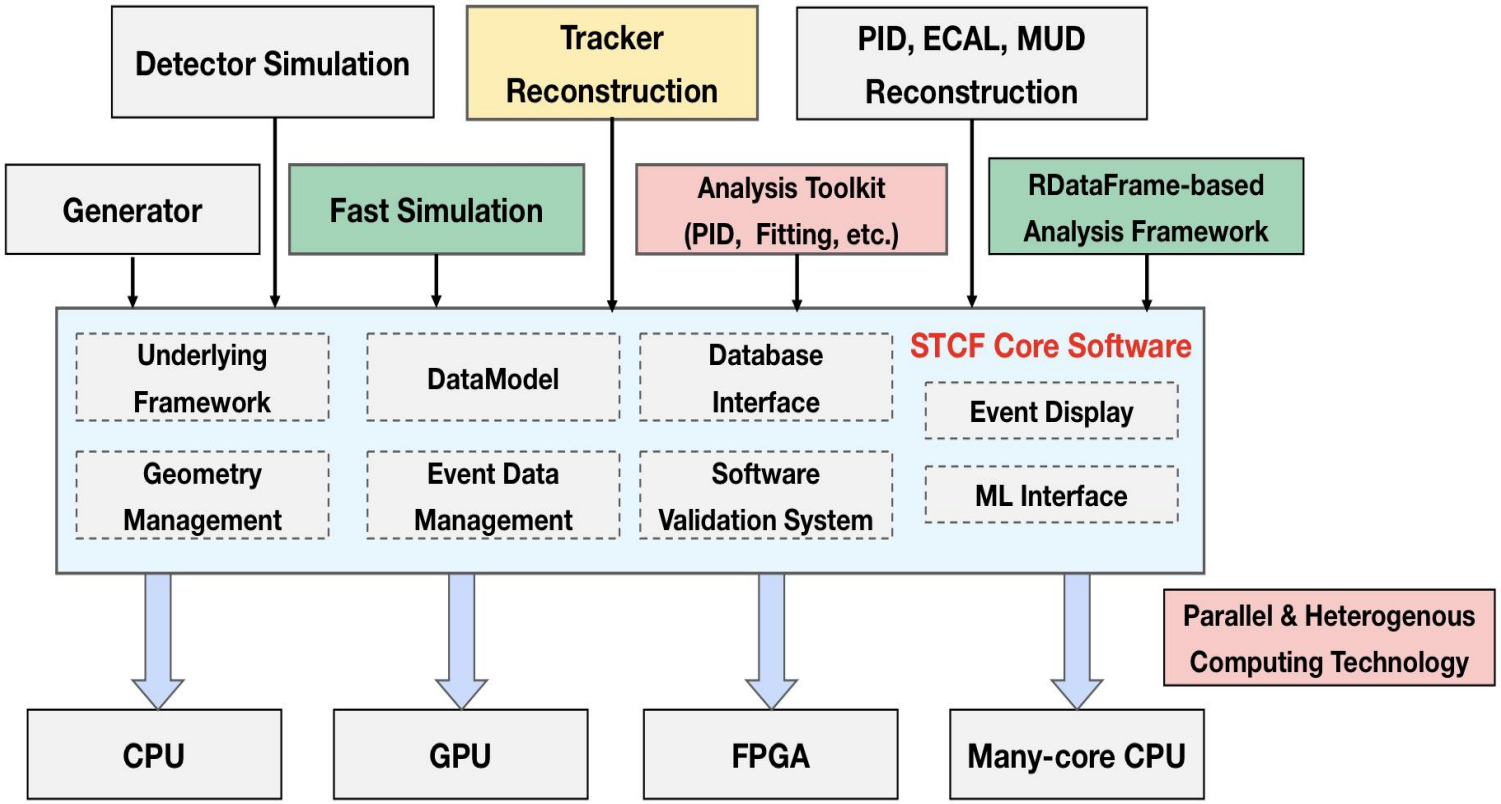
STCF Offline Software (OSCAR)

- OSCAR is develeped to implement **detector simulation**, **digitization, calibration reconstruction**, and provide **physics analysis** platform at STCF
  - Raw data processing chain (~110 PB/year)
  - MC data production (at the similar level of raw data)



Overview of STCF Core Software

- Overview of STCF core software
  - Provide basic functionalities below the applied software
  - Partially based on Key4hep stack

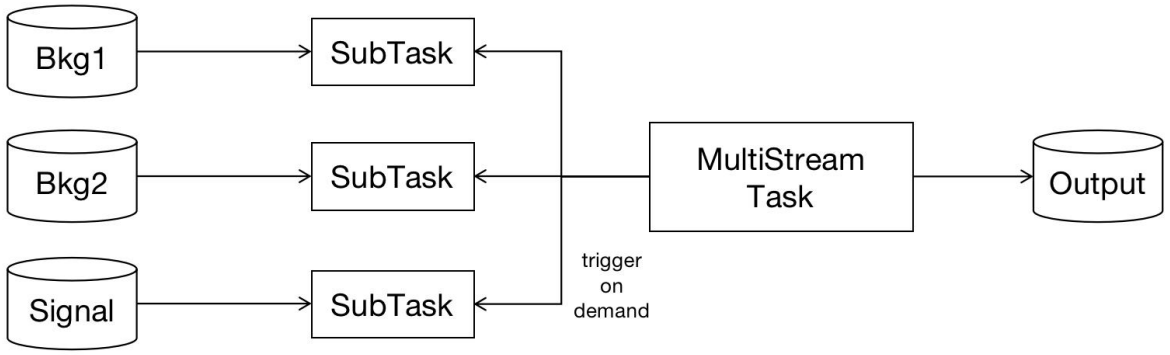


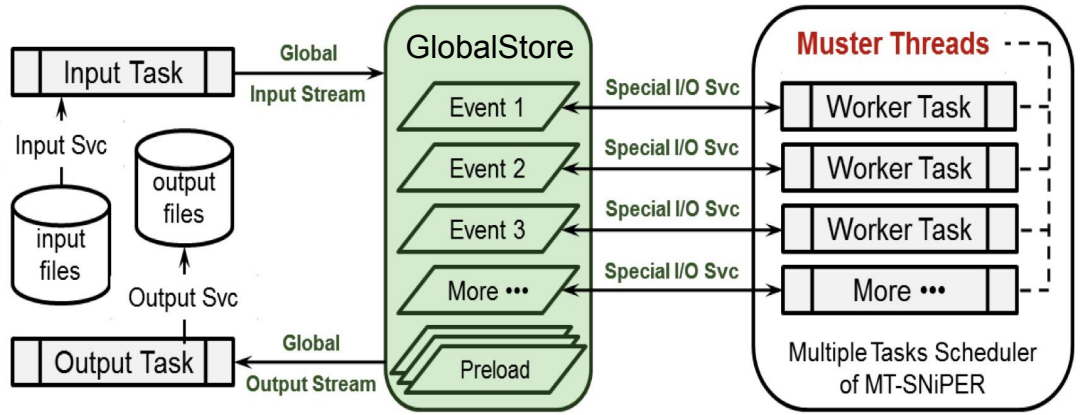
Scope of STCF core software

- Underlying framework
- Event data management
- Detector description and conditions data management
- Event display
- Support of ML, parallel computing, and heterogeneous computing
- Software and physics validation
- Software build, installation and distribution

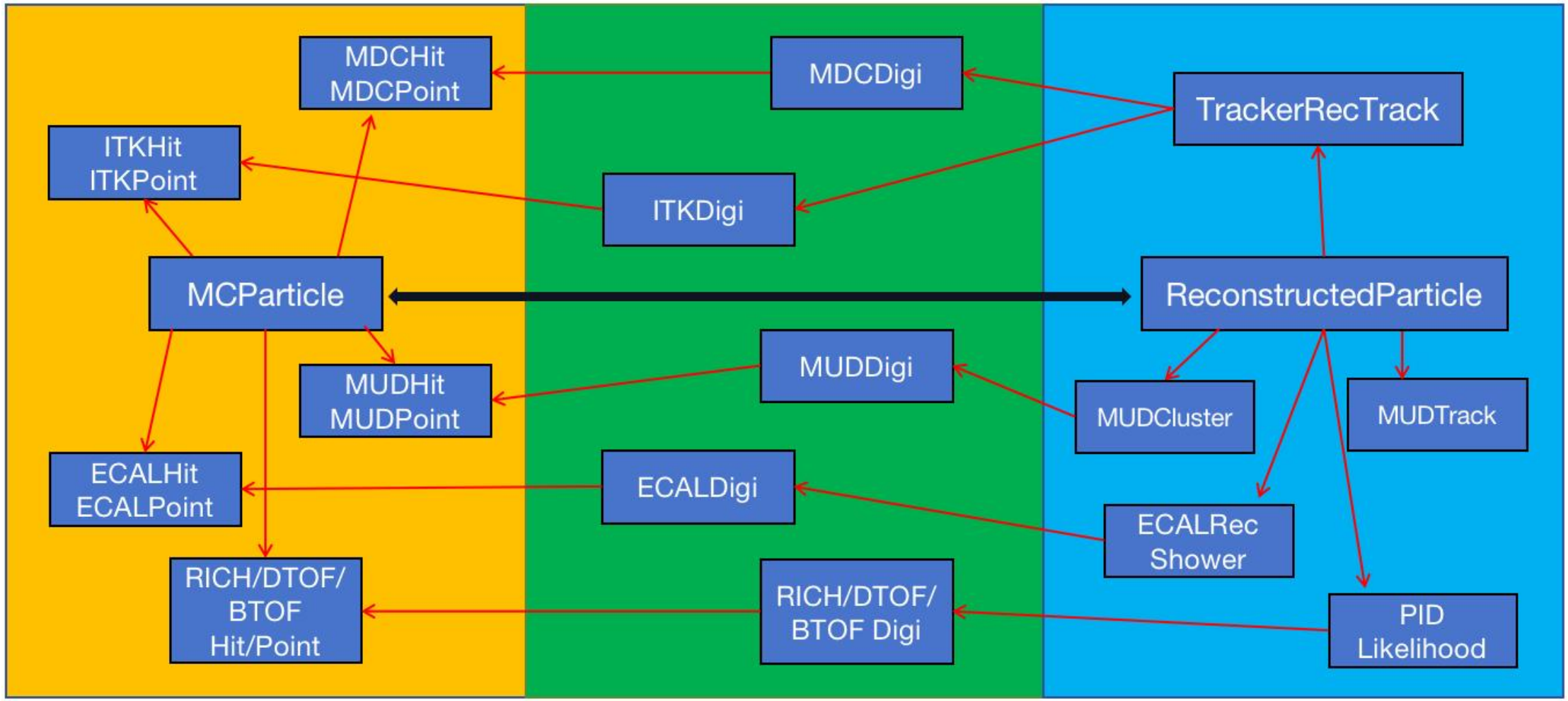
Underlying Framework: SNiPER

- The underlying framework builds the skeleton of OSCAR
  - Provide basic functionalities of **event loop control, algorithm scheduling, thread management, user interface, job configuration, logging** etc. (Like Gaudi for BOSS)
- OSCAR adopts SNiPER as the underlying framework
  - Lightweighted, efficient and highly extendable
  - Developed since 2012, maintained by **10+ developers from IHEP, SDU, etc.**
  - Adopted by JUNO (neutrino), LHAASO (cosmic ray), nEXO (neutrinoless double beta decay) and HERD (dark matter)
- Recent updates
  - Better support for inter- and intra- event level parallism



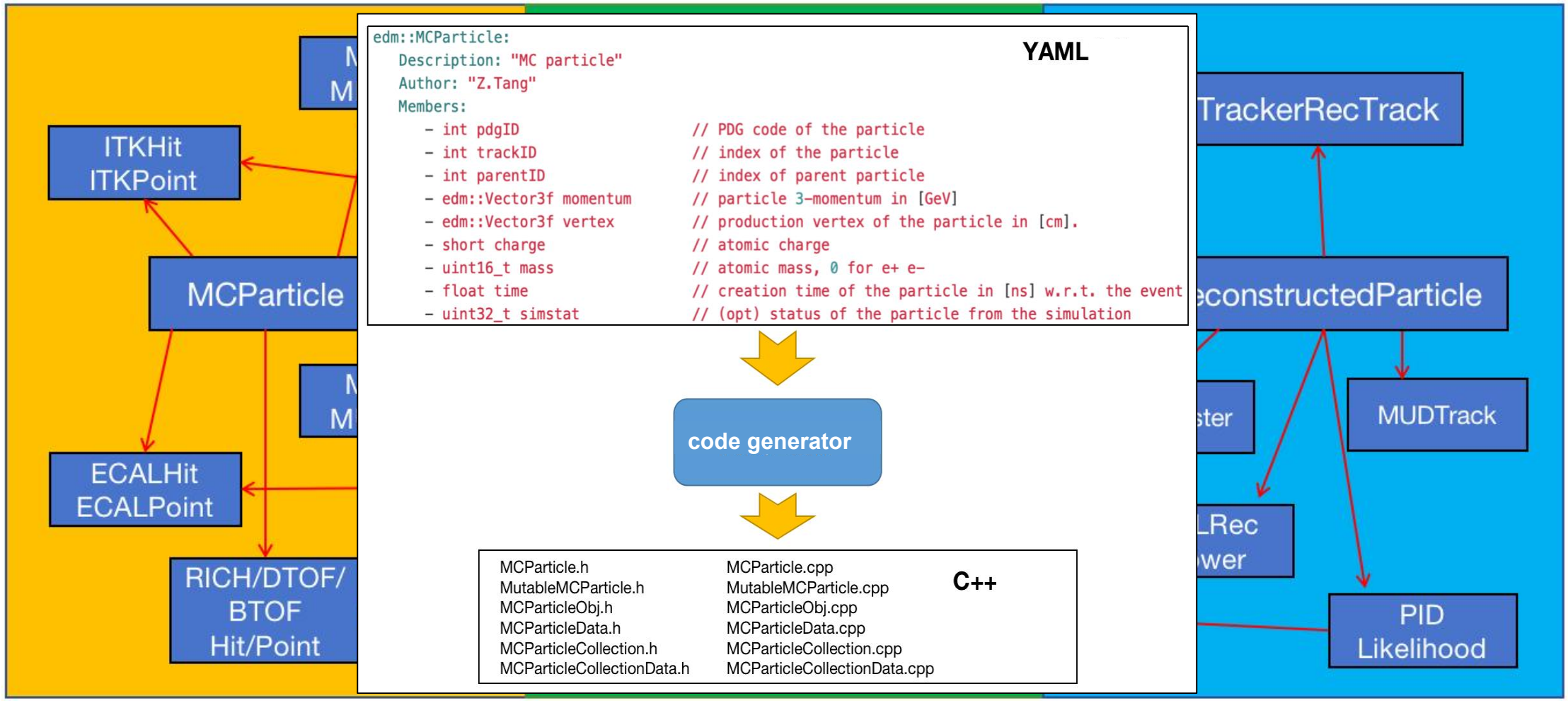


Event Data Model based on podio



Detector Simulation **Digitization** **Reconstruction**

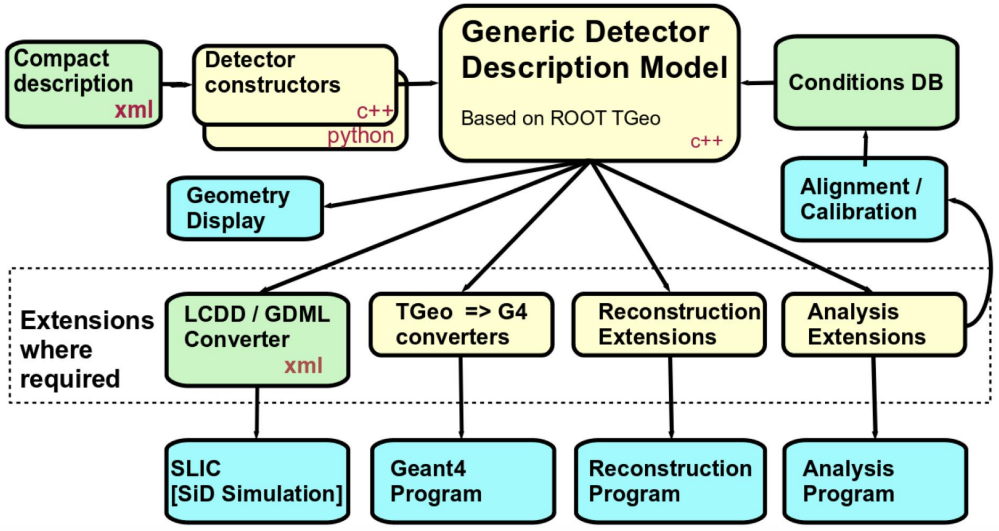
Event Data Model based on podio

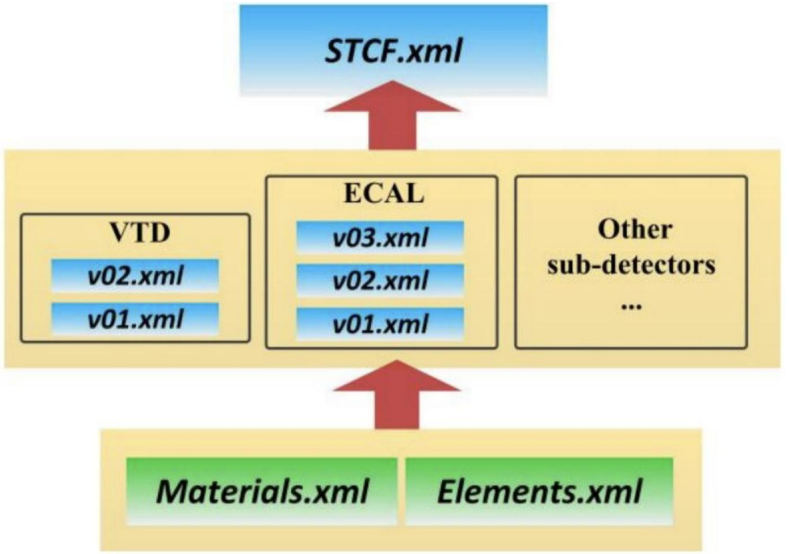


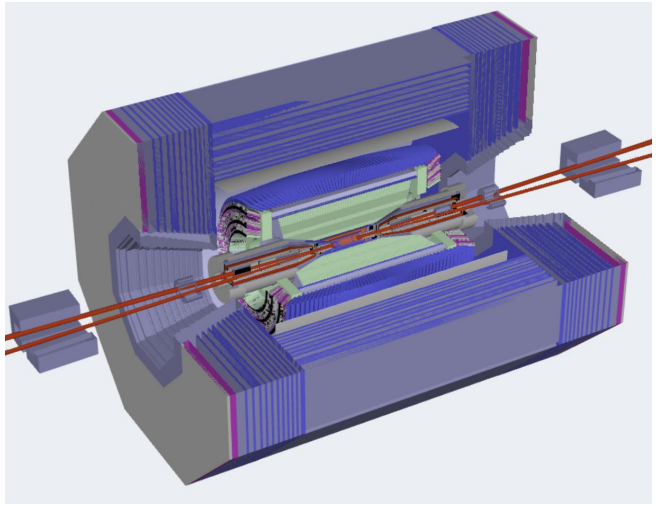
Detector Simulation **Digitization** **Reconstruction**

Geometry Management System

- Geometry Management System (GMS) in OSCAR is based on DD4hep
- Single source of detector information for detector description, simulation reconstruction and event display
  - Complete geometry defined with XML files and C++ parser
  - Various plugins for appied software
  - Interface for alighment and conditions data
- Full detector (and beam-test geometry) defined and stably used, now being optimized

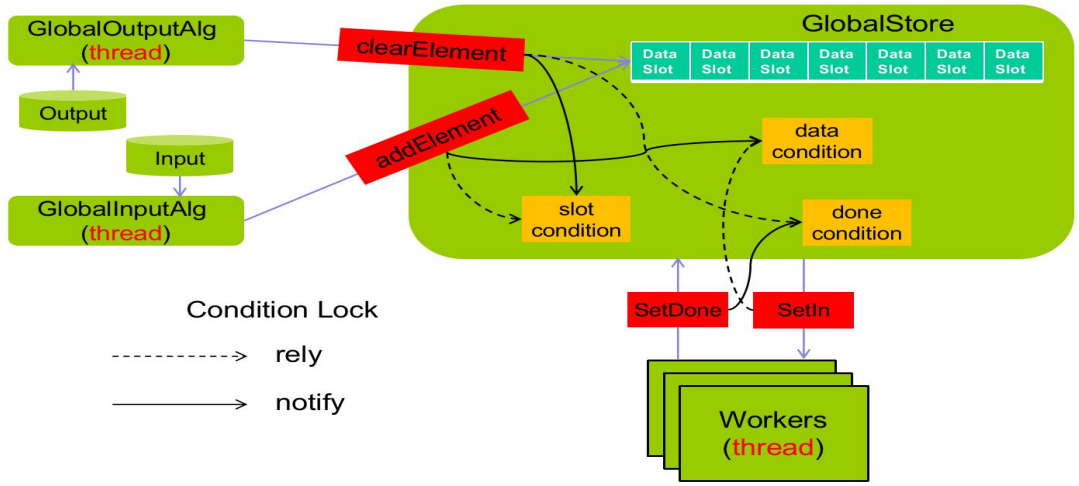


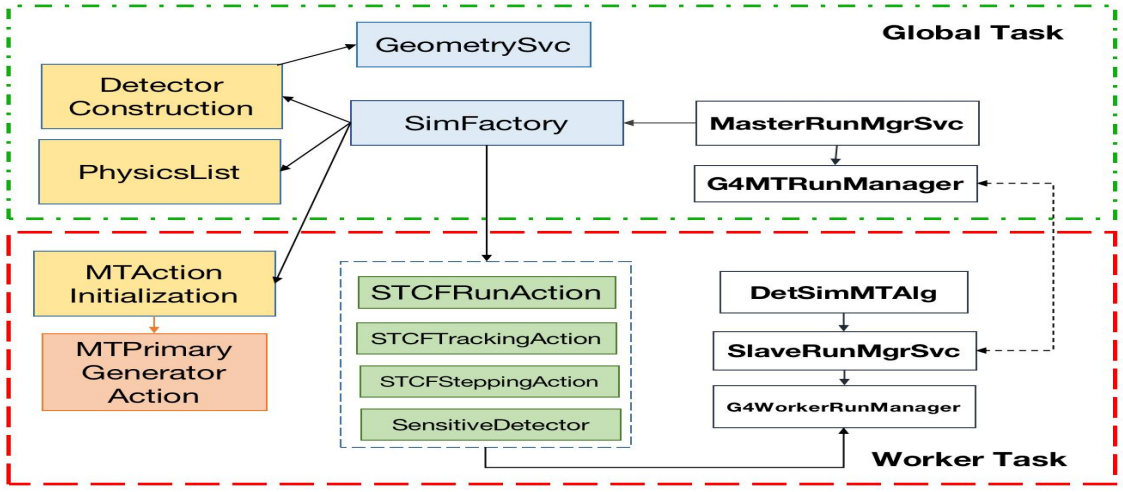


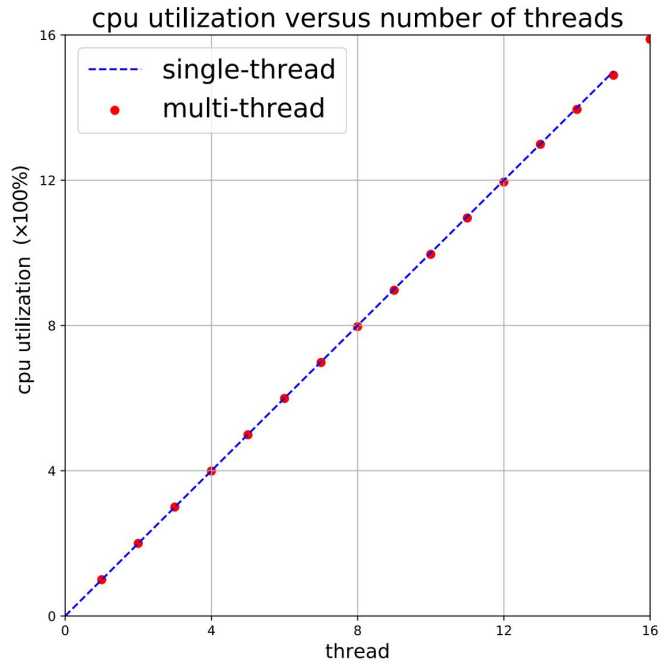


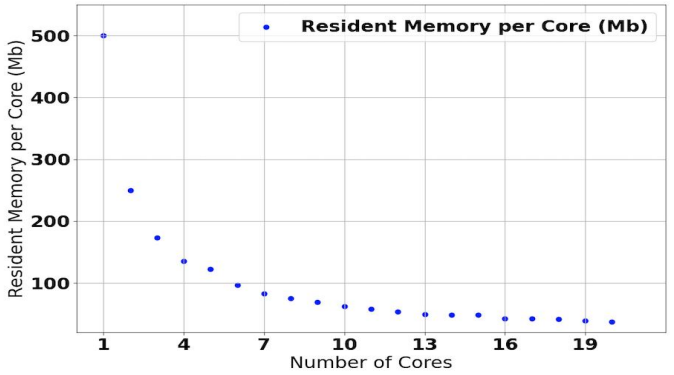
Parallelized Data Processing

- Parallelized detector simulation and reconstruction applications are implemented
  - Basic performance tests show good scalability







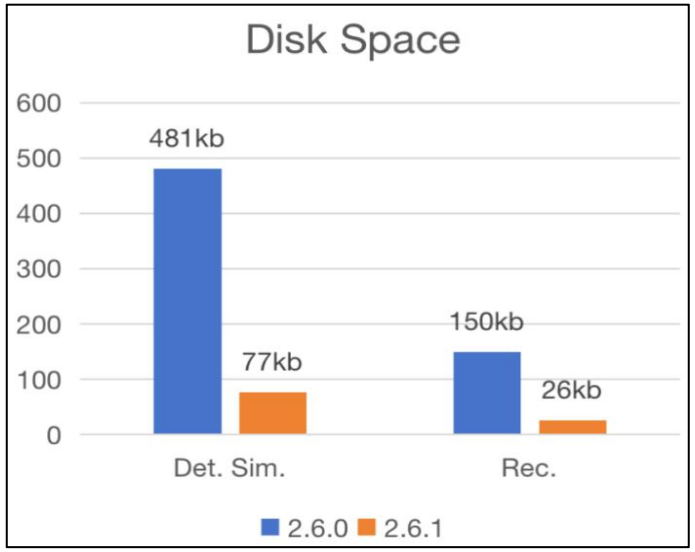


Progress of OSCAR since Xiangtan Workshop

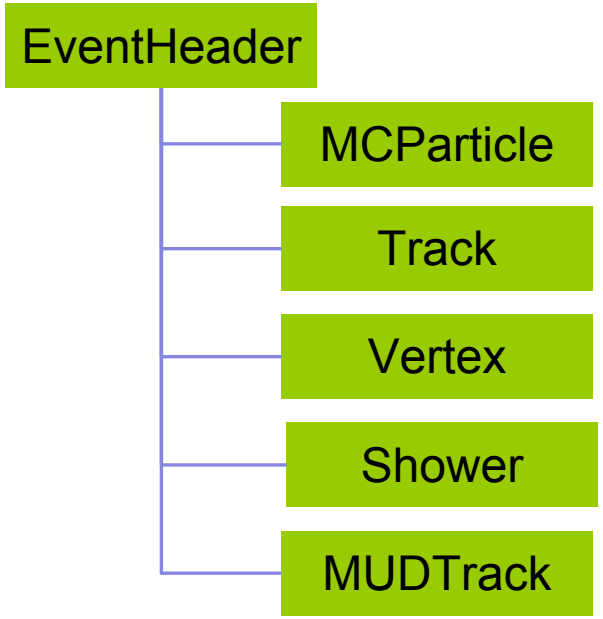
- **Major updates since the Xiangtan STCF workshop:**
  - Development of the analysis event data model
  - Upgrade of modern CI powered by LLM
  - Development and optimization of RDFrame analysis framework
  - Development and optimization of the GlobalVertexFit toolkit
  - Development of software deployment toolkit powered by Spack

Development of the Analysis Data Model

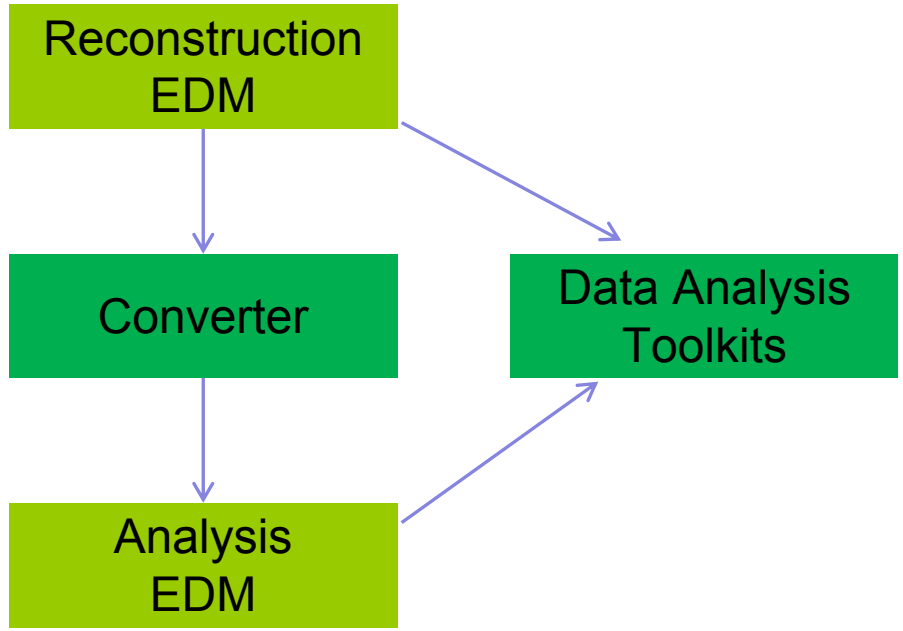
- Motivation: to further reduce disk comsumption of mass data production and to simplify data access interface for analyzers
  - Use flat structure, minimum-required precision type, avoid vector members and relationships
  - Converter from current EDM objects are being developed
  - Analysis service interfaces will be overloaded (e.g. GlobalPID, VertexFit)



Expecting ~50% compression of final data

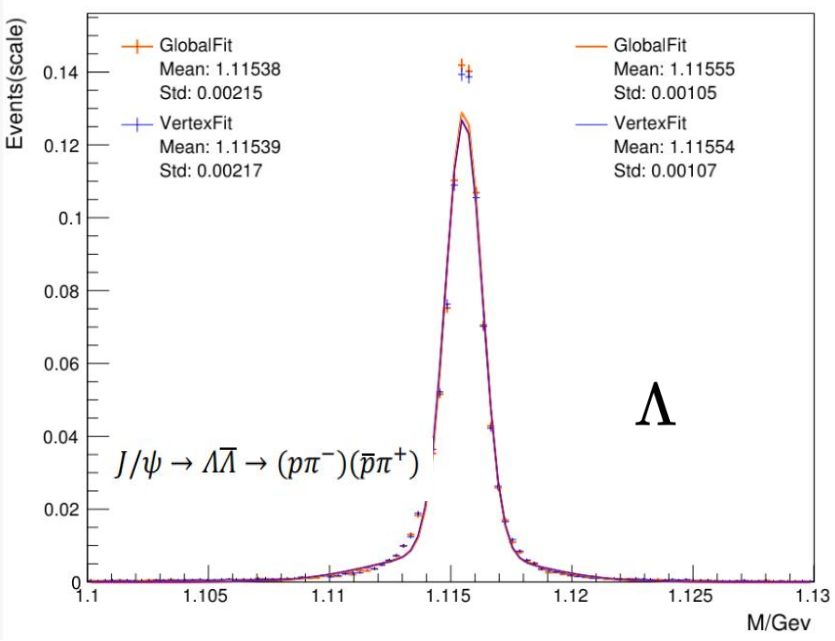


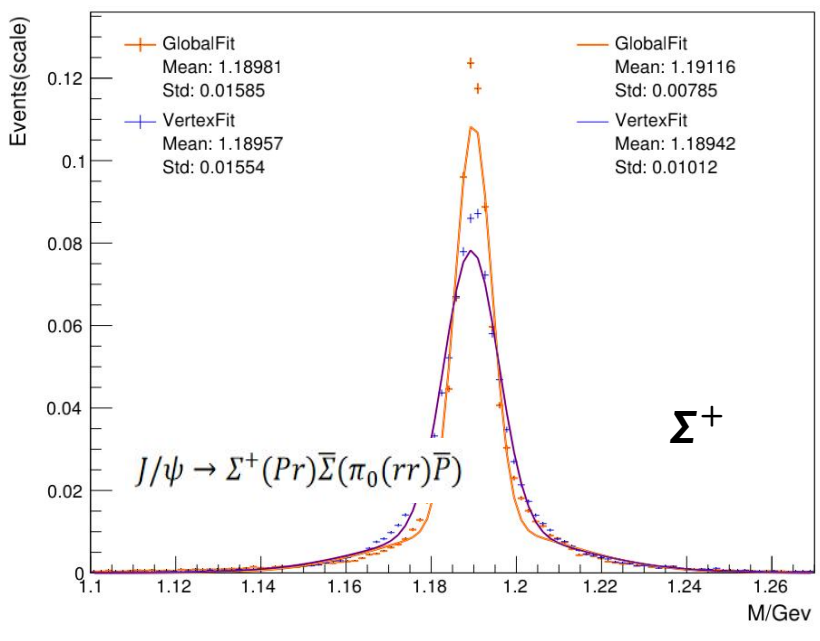
Flat stucture to simpily interface



GlobalVertexFit

- GlobalVertexFit package for STCF is developed based on the global vertex fitting algorithm ported from Belle II
  - Performing kinetic and vertex fit simultaneously to improve fitting performance for long-lived particles
  - Slightly better mass resolution for $\Lambda/\bar{\Lambda}$ and higher reconstruction efficiency reached
  - Much improved $\Sigma/\bar{\Sigma}$ resolution by including photon information





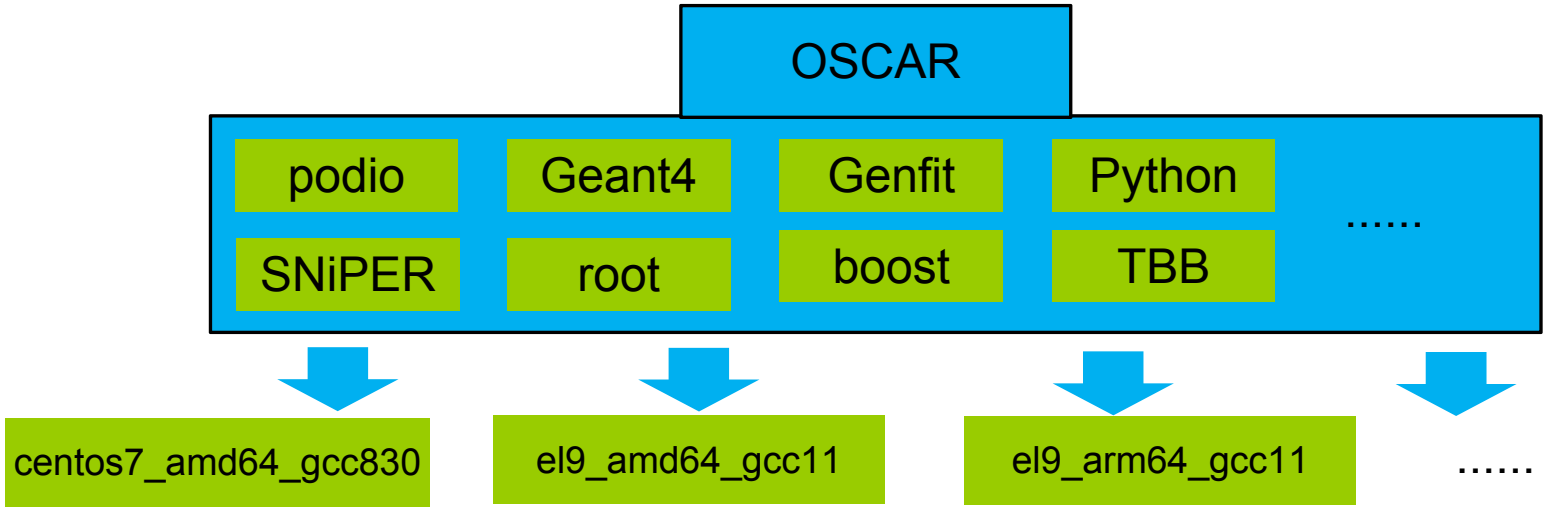
GlobalVertexFit is merged to OSCAR

We encourage analysers to use :)

Software Deployment Powered by Spack

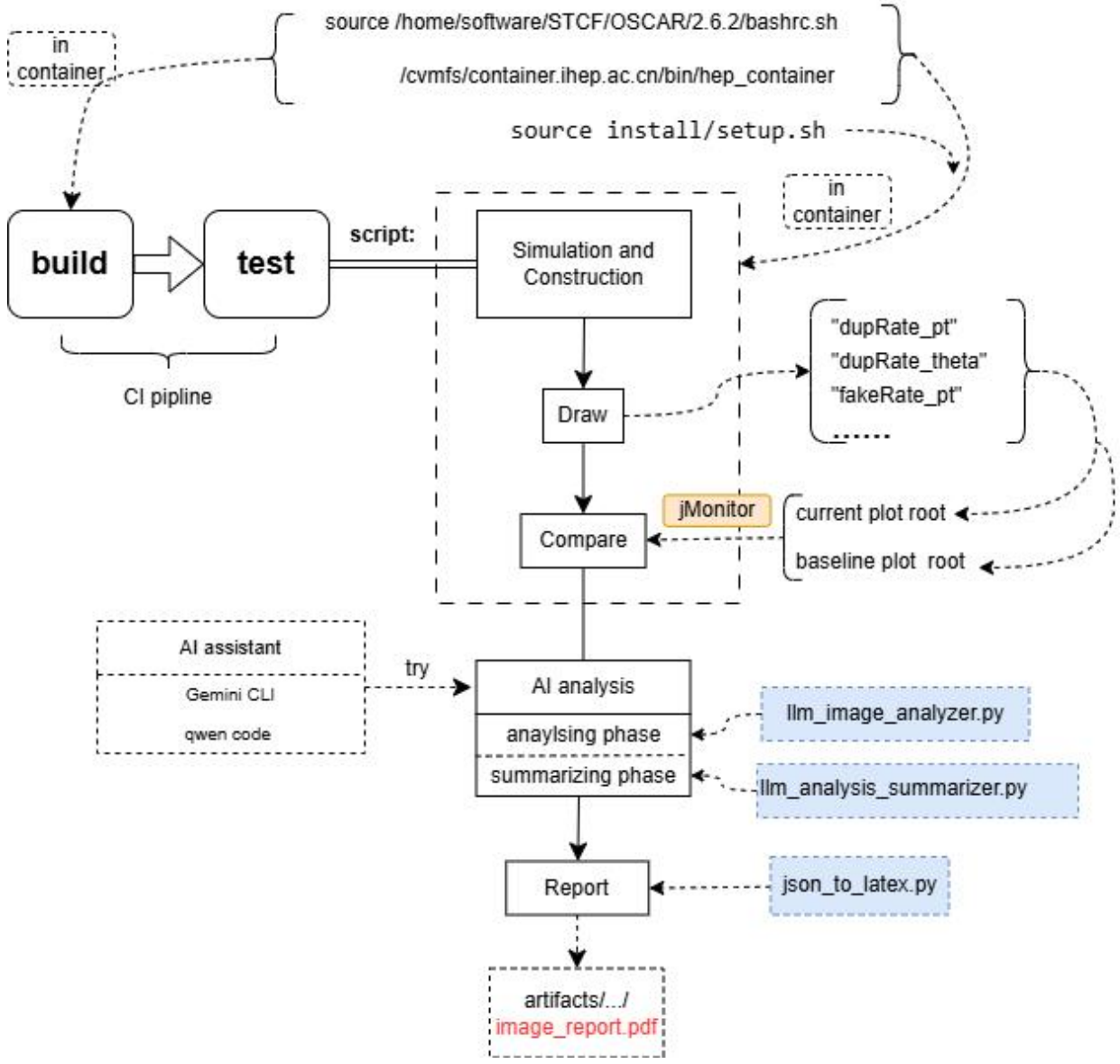
- What is Spack?
  - A package manager that supports multiple software versions and configurations on different platforms and compilers
- Why choose Spack?
  - To replace the outdated bash-script-based OSCAR installation toolkit (nightmare!)
    - All versioning, depending, patching, platform problems have to be handled manually
  - To achieve "One-click installation" regardless of platforms, processor archs, versions ...

We will release this new toolkit together with the el9 updated env



Software Validation

- A modern software validation system powered by LLM is developed
  - Full-scale software validation included:
    - Unit test
    - Integration test
    - Physics validation
    - Performance profiling
- Integrated with Gitlab CI for automation
  - Support both local machine and local cluster (HTCondor)
  - Support of grid will be added
- Integrated with LLM backend to analyze validation results
  - Multiple LLM backend supported



Software Validation

Contents

Tasks Summary

task: analysing_images
total_images: 104
completed_images: 104
success_rate: 100.0
model: qwen-vl-plus
start_time: 2025-11-11 10:22:57
end_time: 2025-11-11 10:24:28
total_time: 0:01:31.174625

residual distributions across all kinematic variables, with the majority of p_T slices for d_0, d_z, ϕ, p, p_T, and θ showing inconsistent shape agreement. These residual inconsistencies are partially offset by systematic improvements in mean residuals and resolutions for several observables, including d_0, d_z, p, p_T, and θ, where the current version demonstrates reduced bias and improved resolution. The overall validation indicates that while the core reconstruction algorithms maintain their efficiency, the software changes have introduced measurable alterations in the detailed shape characteristics of residual distributions while simultaneously improving certain bias and resolution metrics.
model: deepseek-chat
time: 2025-11-11 10:24:40

Inconsistent Plots

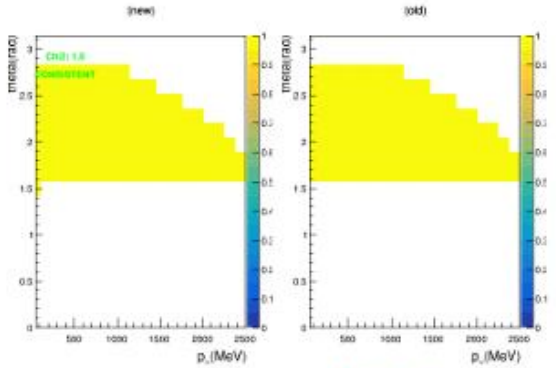
Figure 1.2.1.1, 1.2.2.1, 1.2.4.1, 1.2.5.1, 1.2.6.1, 1.3.1.1, 1.3.2.1, 1.3.6.1, 1.4.1.1, 1.4.1.2, 1.4.1.3, 1.4.1.4, 1.4.1.5, 1.4.1.6, 1.4.1.7, 1.4.1.8, 1.4.1.9, 1.4.1.10, 1.4.1.11, 1.4.1.12, 1.4.1.13, 1.4.1.14, 1.4.1.15, 1.4.2.1, 1.4.2.2, 1.4.2.3, 1.4.2.4, 1.4.2.5, 1.4.2.6, 1.4.2.7, 1.4.2.8, 1.4.2.9, 1.4.2.10, 1.4.2.11, 1.4.2.12, 1.4.2.13, 1.4.2.14, 1.4.2.15, 1.4.3.1, 1.4.3.2, 1.4.3.3, 1.4.3.4, 1.4.3.5, 1.4.3.6, 1.4.3.7, 1.4.3.8, 1.4.3.9, 1.4.3.10, 1.4.3.11, 1.4.3.12, 1.4.3.13, 1.4.3.14, 1.4.3.15, 1.4.4.1, 1.4.4.2, 1.4.4.3, 1.4.4.4, 1.4.4.5, 1.4.4.6, 1.4.4.7, 1.4.4.8, 1.4.4.9, 1.4.4.10, 1.4.4.11, 1.4.4.12, 1.4.4.13, 1.4.4.14, 1.4.4.15, 1.4.5.2, 1.4.5.3, 1.4.5.4, 1.4.5.5, 1.4.5.6, 1.4.5.7, 1.4.5.8, 1.4.5.9, 1.4.5.10, 1.4.5.11, 1.4.5.12, 1.4.5.13, 1.4.5.14, 1.4.5.15, 1.4.6.1, 1.4.6.2, 1.4.6.3, 1.4.6.4, 1.4.6.5, 1.4.6.6, 1.4.6.7, 1.4.6.8, 1.4.6.9, 1.4.6.10, 1.4.6.11, 1.4.6.12, 1.4.6.13, 1.4.6.14, and 1.4.6.15.

Automatic generated PDF summary

1 sim_and_rec_e-

1.1 efficiency

1.1.1 kalman



1.4 residual

1.4.1 d0

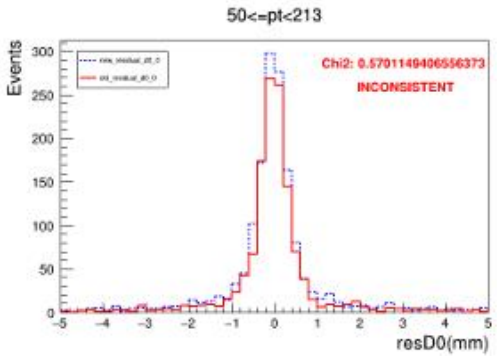


Figure 1.4.1.1: llm_conclusion: Inconsistent: current is better than baseline. The new residual distribution shows a sharper peak around zero and fewer outliers compared to the old distribution, indicating improved performance in the current data.

1.2.2 dz

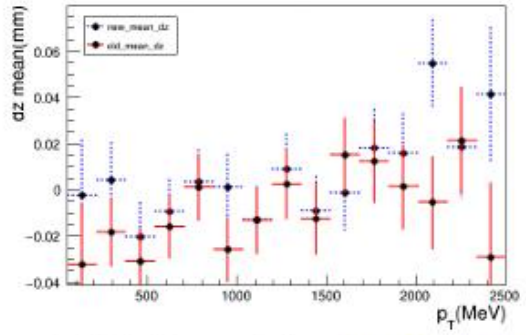


Figure 1.2.2.1: llm_conclusion: Inconsistent: current is better than baseline. The new data distribution shows smaller residuals and tighter error bars compared to the old data, indicating improved accuracy in measuring the longitudinal impact parameter d_z.

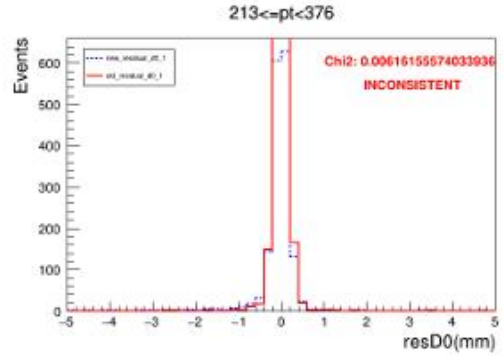


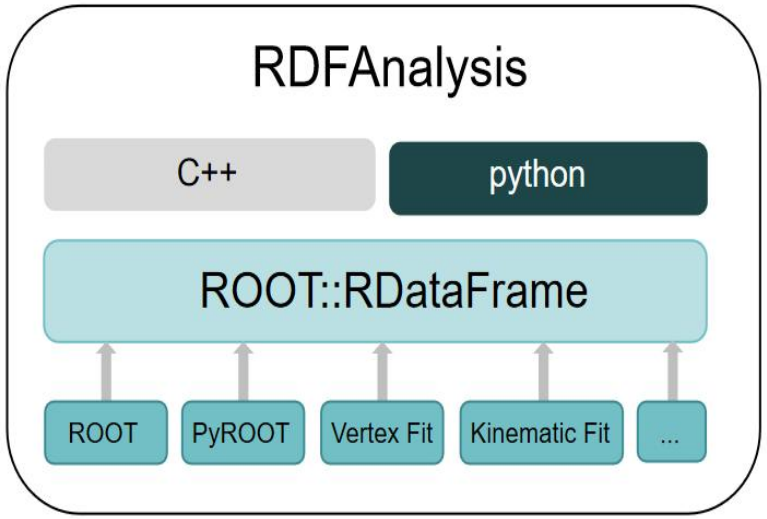
Figure 1.4.1.2: llm_conclusion: Inconsistent. The χ^2 value is very high, indicating a poor fit between the new and old distributions.

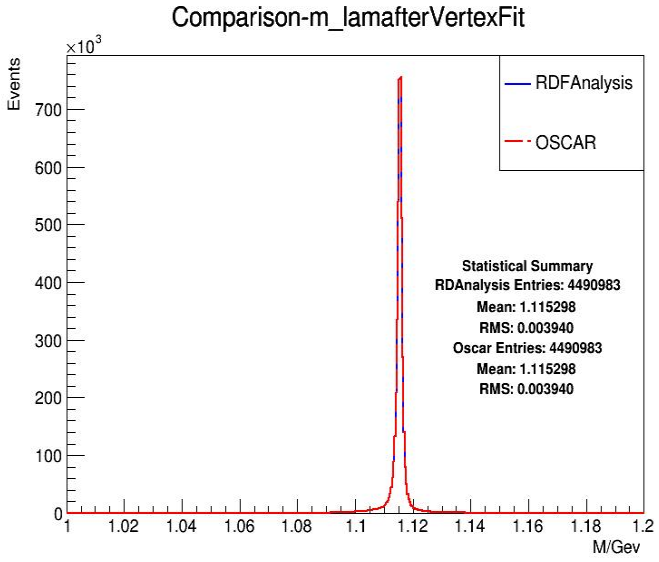
Figures are analyzed by LLM

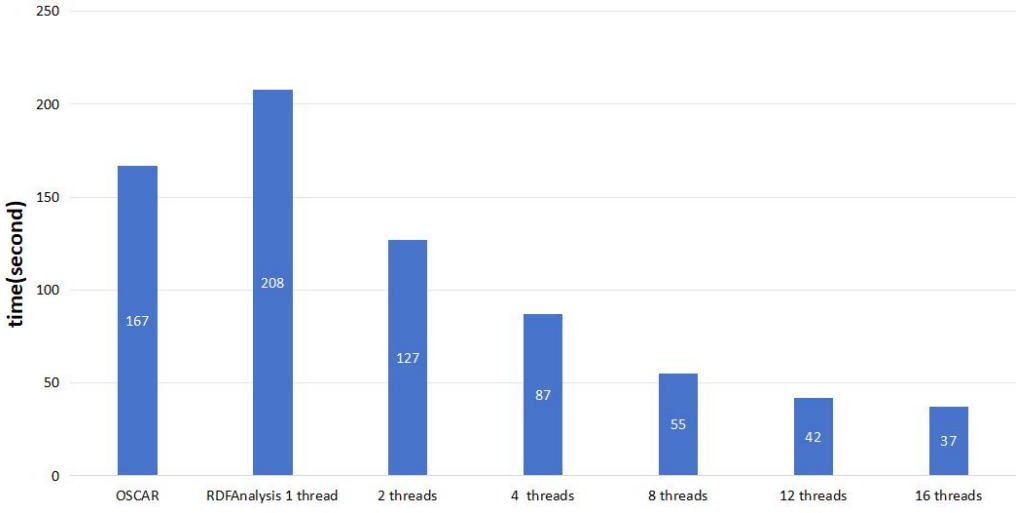
Analysis Framework Powered by RDataFrame

- A Parallelized Analysis Framework is implemented and optimized for STCF
  - Powered by ROOT RDataFrame (support declearitive programing and concurrent processing)
  - Analysis toolkits ported to python (GlobalPID, VertexFit)
  - Validated using multiple decay channels

Details in Ying's talk







We will release the framework once it's ready
Encourage analyzers to use if if you need fit results quickly

Summary

- We have stable release of STCF full simulation and reconstruction chain
- Multiple data analysis studies are (being) performed on top of OSCAR
- We have been continuously improving the core software
  - Improve the data model to further reduce disk burden
  - Upgrade CI for better validation
  - Better data analysis performance
    - Development and testing of the RDFrame framework
    - Optimization of the GlobalVertexFit toolkit (merged)
  - Development of software deployment toolkit powered by Spack