CEPC Detector R&D Project

8. Software and computing

|  |  |
| --- | --- |
| Document Responsible: | Li Weidong, Ruan Manqi, Sun Shengseng, Li Gang |
| Last saved by on |  |
| Revision number: | 1.2 |
|  |  |

Change history

|  |  |  |
| --- | --- | --- |
| **Revision** | **When** | **What changed and why** |
| 1 | 12/22/2019 | First draft |
|  |  |  |
|  |  |  |
|  |  | < Add further lines to table as required > |

Readme first

1. Please do not delete or modify this section or its structure.
2. Only change text enclosed by (and including) angled brackets “< … >”.
3. Don’t change field directly, instead modify the document options, under File🡪 Properties (or similar)
   * Enter name of person that wrote the document in Document:Summary: Author
   * The project ID number, should follow the rules provided to you earlier. The number should be changed in Document:Custom: PBS.
   * The project name should be changed in Document:Summary: Subject.
4. In Section [*Project Objectives*](#ProjectObjectives) provide a brief description of the project goals, i.e. why and what is being produced, for PBS item . If this project includes identifiable sub-projects you can indicate them in the [*Sub-projects Description*](#SubprojectsDescription) Section, otherwise submit a separate document for each of them. The sub-project IDs are free for you to define.
5. Finally, remember to update the [*Change History*](#ChangeHistory).

8. Software and computing: Project Objectives

A software and computing framework with well top design and various key algorithms/software tools are essential for high energy experiments. The CEPC software and computing project aims to provide a complete software chain for detector R&D, physics study, and future operation. Besides maintaining the existing CEPC software developed from the ILCSoft, the project is developing new CEPC software based on the Gaudi and the EDM4HEP, which also includes many new modern software techniques and packages, such as DD4HEP, ACTS, and so on. The software is going to support both full and fast simulation and provide a unified reconstruction and analysis toolkit, which also support flexible switch between fast and full simulation for sub-detectors. The new software is going to support both single thread, multi-thread computing, as well as Heterogeneous computing to make use of various potential computing resources.

8. Software and computing: Sub-projects Description

|  |  |  |
| --- | --- | --- |
| **Project ID** | **Title** | **Description** |
| 8.1 | Core Software | Including software framework and event data model (EDM), Geometry, and so on. The core will be based on Key4HEP, so that we can sharing the efforts with other experiments. |
| 8.1.1 | EDM | We have moved from LCIO to plcio in the CEPCSW prototype. We are going sto move from plcio to EDM4hep soon. |
| 8.1.2 | Framework | We have moved from Marlin to Gaudi in the CEPCSW prototype. It is a good start for the software refactoring based on Key4HEP. |
| 8.1.3 | Geometry | A consistent detector description is important for simulation, reconstruction, calibration and analysis during both the detector R&D and data taken stage. The DD4hep toolkit provides a consistent generic detector description, which could be used by ROOT, Geant4 and so on. In CEPCSW, the DD4hep toolkit is used to describe the materials, geometries and fields. A geometry service is developed to manage the instance of DD4hep, so that other components could access the geometry information directly. At the same time, the geometry service could be configured in the Gaudi steering file, so that the different detector designs could be loaded dynamically and managed easily. |
| 8.2 | Generators(interface) | At present, WHIZARD generator is used to simulate all most all Higgs signal and SM background, and two-photon for gamma-gamma collision processes, both of which are implemented and used via StdHep interface. MadGraph and Pythia8 are very powerful and also used in some cases, in particular for BSM study.  In the plan, a physics generator interface is developed to integrate the external physics generator with the detector simulation software, while the HepMC package is used as the intermediate data format between them. A component called GenTool is responsible to modify or create the HepMC data object from the configurations or from the files in different format. A Gaudi-based algorithm is developed to manage and control the execution of the GenTool components, so that the registered GenTool components will be executed in sequence, which constructs the full event. |
| 8.3 | Simulation | Simulation tool includes fast and full simulation, which is going to support detector design and optimization, as well as physics performance study. It also provides flexible switches for all sub-detectors or different detector region between fast and full simulation to make simulation fast and realistic. |
| 8.3.1 | Full simulation | A Geant4 based simulation framework is developed based on Gaudi. An algorithm called DetSimAlg is in charge of the detector simulation instead of the Geant4 itself. During the initialization of DetSimAlg, the Geant4’s G4RunManager is retrieved from a service called DetSimSvc, then the detector construction, physics lists, primary generation and user actions are all registered into G4RunManager. The geometry information used in Geant4 is converted from the DD4hep detector description using a converter developed in DDG4. The primary generation converts the HepMC objects to Geant4’s vertex and particle objects. The sensitive detectors maintain the hit collections and collect hit information during the simulation. The hit collections are finally saved in the event data model format, which will be used in the digitization stage. |
| 8.3.2 | Fast simulation | A fast, simple simulation detector design and physics study, which has two main components: hit level tracker simulation and calorimeter fast simulation.   1. The tracker fast simulation includes multiple scattering, ionization loss, radiation loss, and the nuclear interaction processes, which could be used to test tracking software, detector and physics study. 2. frozen shower method for calorimeter simulation. During the simulation the low-energy particle are substituted with pre-generated shower from the library. When asked, the library should return the shower with the generation conditions as close to the required, as possible. Library is needed to be performed only ones. 2, Generative Adversarial Networks (GAN). Using GAN from deep learning to do the calorimeter shower fast simulation. Training the GAN using MC samples for different particles (except muon) for different moment. Do the inference in the simulation when a particle comes to ECAL surface. |
| 8.4 | Digitization | Tools convert simulated hits into the ones used in reconstruction, which needs sufficient collaboration with sub-detector groups.  The TPC and silicon detector digitization with simplified parameter model have been migrated to Gaudi framework.  A digitization algorithm for ECAL and HCAL is available and migrating to CEPCSW. |
| 8.5 | Reconstruction | Including various sub-detectors, algorithms, and potentially new software technologies, the baseline is particle flow concept |
| 8.5.1 | Tracker | We have various tracking reconstruction options~~.~~  The first one combines track segments from several reconstruction algorithms optimized for different parts of the silicon tracker (VXD, SIT, SET and FTD) and a mature package Clupatra for TPC to obtain final tracks. Clupatra is based on nearest neighbor clustering seeded pattern recognition using Kalman-Filter for extrapolation and hit search, which could be to be improved as the detector design changed.  The 2nd is ACTS (A Common Tracking Software), a modern, well top-designed, and highly performant tracking finding and fitting package original from the ATLAS. It is going to be integrated into CEPCSW and used for future CEPC tracking.  The 3rd one is A generic Kalman track fitting framework, Genfit, for CEPC tracking detectors, development of TPC track finding and track fitting algorithms. |
| 8.5.2 | Calorimetry | Jet energy resolution is very import for CEPC, which is significantly improved with so-called particle flow algorithm (PFA). We have two PFA options for CEPC, home-made Arbor and the PandoraPFA.  Arbor: the baseline reconstruction PFA of the CEPC.  Pandora: According to ILC experiment’s study, the Pandora can achieve good jet energy resolution. Pandora using dedicated pattern recognition algorithms to decouple the jet energy into different parts including charged particles (~60%), photons (~30%), and neutral hadrons (~10%). In this way, the energy of charged particles can be measured by tracker precisely. The energy of photons can be measured in ECAL with high resolution. The energy from neutral hadrons which is small can be measured by HCAL. Therefore, the jet energy resolution will be improved compared with traditional approach which just separate the jet’s energy into ECAL part and HCAL part. |
| 8.5.3 | Jets | Jet-related topics include vertex finding/fitting, jet-clustering, and jet-flavor-tagging. A well design and almost standalone package LCFIPlus integrate all above functions in a single software tool. Vertexing could be replaced with the one in the ACTS, or as a cross check in future. Jet-clustering has only ee-kt algorithm and more recent developed one could be tested. Jet-flavor-tagging is the most promising one which is expected to be improved with recent artificial intelligence break-through.  Jet-charge is a very challenging mission and could be a very good testbed for PID design. |
| 8.6 | Visualization | Maintain the Druid display package |
| 8.7 | Conditional database | Conditions data are typically heterogeneous and vary with time. A conditions database management system is designed to homogeneously treat all these heterogeneous conditions data, provide easy management and convenient access with both Restful API and web interfaces, support good scalability and maintenance for long time running. |
| 8.8 | Computing | Use Dirac-based distributed computing system to organize heterogeneous and remote resources from members of the collaboration for CEPC detector simulation study and future data processing tasks. Workload Management System and Data Management System are designed and built to manage CEPC workflow and dataflow in distributed computing environment. |

8. Software and computing: CEPC Relationship

Software and computing projects play an essential role, which interplay with detector, accelerator, and physics study. A framework and complete toolkit with good top design, wide collaboration among different experiments and industry will make CEPC a fruitful and successful experiment.

Physics performance. Physics is the ultimate goal of any experiment. CEPC is going to focus on Higgs, electroweak precision measurements to test SM, as well as lots of QCD and flavor physics topics. Software and computing is supporting this kind of study by providing complete simulated samples with the software chain including generation, detector simulation, reconstruction, and analysis. It should be reliable and fast, so the physics performance could be evaluated according the experiment design quickly.

Detector design and optimization. Detector design depends many factors, such as physics performance, background, and other constraints. Full simulation and realistic digitization

New technology. High energy experiment could and should benefits from the development of artificial intelligence and data science.

8. Software and computing: Project Schedule

2020:

1. Software environment
2. Event data model: EDM4Hep
3. Uniform geometry interface of simulation and reconstruction, and DD4Hep for geometry
4. Tracker: silicon and Clupatra
5. PFA: Arbor and Pandaora
6. Jet / Flavor tag
7. Integration of ACTS
8. Visualization

2021:

1. Fast simulation for physics analysis
2. Support data analysis of beam test
3. Simulation of uniformity of magnetic field, noise and background mixing, also optimization of reconstruction

2022:

1. Integration with Key4Hep
2. Optimization of algorithms
3. Online event filter
4. Parallel
5. Machine learning and other new technologies
6. Performance and validation

8. Software and computing: Funding Availability

There is no dedicated funding for software.

8. Software and computing: Leadership Arrangement

IHEP: LI Weidong, Sun Shengseng, Ruan Manqi, Li Gang

Others: Huang Xingtao(SDU)

8. Software and computing: Manpower Resources

|  |  |
| --- | --- |
| **Type** | **Average FTE Expected** |
| Faculty | 7 |
| Postdoc | 2 |
| Students | **3** |
| Engineers | 0 |