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 |
| 2 | 05/07/2020 |  |
| 3 | 01/08/2021 | Update the project and schedules |
|  |  | < 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 |  |
| 8.1.1 | Event Data Model | In Key4HEP, EDM4hep is the common event data model that describes the event data in different processing steps as well as the relationships between them. Customized EDM classes will be added according to CEPC requirements. In order to read M.C. data produced by iLCSoft, we have implemented K4LCIOReader which can convert LCIO data to EDM4Hep objects on the fly. |
| 8.1.2 | Detector Description | A unified detector description will be adopted by the simulation and reconstruction to keep the consistencies. The DD4hep-based geometry service has been implemented to provide a unified way to retrieve the detector geometry data. Standard geometry management policies are released to support the development of Reference Detector, including Drift Chamber and Long bar ECAL. |
| 8.1.3 | Simulation Framework | A simulation framework has been implemented to integrate Gaudi and Geant4. The simulation software will be enhanced by enabling more realistic simulation. The non-uniformity of magnetic field will be considered. The beam-related and electronics noises will be added to physics signals at the MC hit level. Various types of fast simulation tools will be provided such as Delphes and fast ECAL simulation tools etc. Simulated data corresponding to different detector configurations will be produced for optimizing the detector design. |
| 8.1.4 | Analysis Framework | A ROOT based analysis framework will be developed based on the declarative analysis. The performance of the ROOT based analysis software framework will be tuned to achieve good scalability for any analysis running on a multi-core server and a large computing cluster respectively. Event selection for online data reduction will be studied by developing event selection steering software and reusing software components developed in the offline environment. |
| 8.1.5 | Parallelization | Parallelization in different levels will be fully supported, including inter-event level and intra-event level. The thread safety of EDM4hep and data I/O will be tested with GaudiHive and the performance will be optimized. Parallelized simulation of dE/dx or dN/dx in the gaseous detector will be prototyped as a possible application scenario for parallel computing. To take advantage of machines with many cores, both simulation and reconstruction algorithms will be migrated to run as fully multi-threaded applications. In order to use GPU, FPGA etc more efficiently, the framework software will provide a unified and straightforward way to offload time consuming calculations to any kind of accelerator. |
| 8.1.6 | Machine Learning | The interfaces will be developed to integrate TensorFlow/PyTorch tools with the CEPCSW to support various machine learning applications, such as ECAL fast simulation with GANS, Jet/Falvor tagging as well as signal/background discrimination. |
| 8.1.7 | Database | The database service will be developed to manage static metadata of detector geometry files and to provide coherent access to geometry data, and the conditions data management system will be designed to manage conditions data which are varying with time and come from calibration, detector monitoring and running. |
| 8.1.8 | Validation | The validation platform will be built for software validation tasks including automatic compiling, unit testing and physics validation etc. It will be setup based on the Github action system with runners based on Jenkins and Kubernetes, a central database recording the test results and a message queue synchronizing test jobs. To provide user-friendly interface, a web-based validation dashboard will also be developed. |
| 8.2 | Generators(interface) | The integration of generator interface with Gaudi framework has been done by using the Gaudi Tools. The generator interfaces already support different types, such as particle gun, LCIO reader, HepMC reader etc. New generator tools could be implemented when new physics generators are requested. |
| 8.3 | Simulation and Reconstruction |  |
| 8.3.1 | Silicon detector | Continue to optimize the geometry, digitization, tracking and fitting software for silicon tracker in CEPCSW, based on implemented CEPCv4 silicon trackers and migrated reconstruction from Marlin. New type of silicon modules more realistic for CEPC design are considered to be implemented, and then mechanics support, cooling and electronics are also will be implemented step by step. At the same time, detail digitization will be implemented, and effect of background, noise, non-uniform magnetic field and mechanical alignment on performance will be studied. For the tracking algorithm, the tracking packing ConformalTracking will be migrated, and developing a new tracking algorithm using machine learning techniques to look for possible way to improve the tracking performance is in plan. For the fitting, DDKalTest will be implemented to improve geometry interface from Gear to DD4hep. Combining fitting for silicon tracker and drift chamber will be developed. In the developing procedure, Key4hep will be an important station to keep toughing, benefitting and feedback. |
| 8.3.2 | TPC | The TPC geometry, simple digitization and tracking algorithms has been implemented into CEPCSW. The full reconstruction chain for TPC and silicon trackers has been completed, and the full simulation and reconstruction chain in CEPCSW is being validated. Software architecture will be optimized, such as reorganizing the common constants and common tools used in tracking and refining the structure of GaudiAlgorithm ClupatraAlg. More realized TPC digitization and optimization of tracking after mixing background will be performed. Another tracking algorithm ArborTracking will be implemented and compared. Correction on the energy lose and multi-particle scattering for gas in Kalman filter will be investigated. Once the reconstruction has good shape, tracking based alignment will enter the agenda. After ACTS and GenFit are implemented into CEPCSW, they will be considered to apply onto TPC reconstruction processor. Finally, performance study on benchmark physics channels will be performed. Software and design are both optimized as far as possible. |
| 8.3.3 | Drift Chamber | Develop the two drift chamber simulation in CEPCSW. Realize a simple drift chamber digitization by get the POCA and dE/dx of hit in cell. Develop the track fitting algorithm according to MC truth position. Implement a track seeding algorithm based on truth information. Realize a dummy dE/dx reconstruction algorithm by doing a hit level sample from empirical model. Simulation of CRD detector including silicon, drift chamber and ECAL. Do Garfield simulation on the cluster counting method is performed. Develop the track fitting combined DC and silicon measurement in CEPCSW. Integrate a track level dN/dx simulation from Garfield in CEPCSW to study the PID performance. Optimize the tracker through fast simulation. Optimize the tracker by study the resolution of momentum, impact parameter, and dE/dx(dN/dx) in CEPCSW. Optimize the design with benchmark physics channels in CEPCSW. |
| 8.3.4 | Calorimetry | Porting simulation and digitization of ECAL into CEPCSW with EDM4hep based calorimetry hit objects and MC truth information and integrated with DDG4 for detector description. Simulation and digitization algorithms for a long crystal bar solution for ECAL have been developed and a preliminary result has been obtained. Feasibility research on long crystal bar solution for ECAL is in processing, the ghost ambiguity rejection is the key point. Porting simulation and digitization of HCAL in baseline detector design to CEPCSW will be implemented. Optimization of reconstruction of long crystal bar for ECAL to improve the precision of 3D cluster finding, especially with multiple particles incidented into one super module. Extraction of the jet energy resolution, and the limitation of detector and physics reach based on the long crystal bar solution for ECAL will be implemented. Development and upgrade of PFA algorithms is in the plan to improve the performance of long crystal bar solution for ECAL. |
| 8.3.5 | Muon detector | The CEPCv4 coil, yoke and muon detector will be implemented into CEPCSW as comparison at first, and then new design of muon detector for TDR will be described based on DD4hep description. Pandora and Arbor are considered as muon identification using the muon detector as the first attempt. The more developing and optimization will be investigated. |
| 8.3.6 | ACTS | The standalone ACTS implementation of two trackers is going to be integrated into the CEPCSW, then the optimization of silicon tracker will be the driver of the development. A new propagator algorithm to achieve the balance between speed and precision is going to be developed. A new seeding and pattern recogniation algorithms are also in the agenda. |
| 8.3.7 | PFA | In the past year, Pandora from Marlin was ported into CEPCSW. The single γ and electron events were used to check its ECAL and ECAL+Tracker reconstruction performance and the results looks good. Besides, the Pandora was put into Key4HEP project. In the next year, the main work for Pandora will be checking its reconstruction performance for hadrons and jets. Moreover, optimizing the Pandora reconstruction algorithms for different detector designs will also be considered. After that, there are some future plans, such as optimizing the Pandora client to make it more flexible to read geometry information, improving the Pandora reconstruction performance by using machine learning techniques, working with Key4HEP to optimize the PandoraSDK and making it more efficient.  The clustering of Arbor has been ported into CEPCSW. In the next year, the whole reconstruction will be ported and it will be validated using the single particle performance and Boson mass resolution. For future plans, the ArborPFA will be integrated into Key4HEP project, the performance will also be improved. |
| 8.4 | Analysis tools | A a full covariance fast simulation module combined with the analysis framework is being developed. Next plan is migrating the analysis framework from the IlcSoft to CEPCSW, which depends on the EMD development. The vertex finding, jet-clustering, and flavor tagging package will also be interfaced the CEPCSW, more machine learning approaches will be tried to improve the performance.  Implementing the CEPC plot style macros into the analysis framework and the CEPCSW. Integration of the CEPC statistical combination package in the CEPCSW will be realized in near future. |
| 8.5 | Visualization | In the next year, the Druid package will be ported into the CEPCSW and tuned according to the detector design. |
| 8.6 | Computing | The basic Grid Infrastructure has been set up for CEPC: The VOMS(VO Management System) service was deployed for CEPC VO. The IHEP Grid Computing Element and Storage Element prototype was built up and successfully tested using HTCondorCE with more than 500 CPU cores. The CVMFS server was deployed CEPC software in distributed environment. The distributed computing infrastructure (DCI) prototype is built up based on DIRAC. Available CPU and storage resources are being integrated in the infrastructure. The production system prototype will be built up to allow massive production MC jobs to automatically submit, schedule and run in DCI. User job submission and management system will be considered to support user analysis. Also metadata system and data transfer system should be designed and built to share data among sites. Data challenges, tuning, performance study and optimization in real user cases need to be done before production. After in production, a complete monitoring system is needed to ease operations. The issues such as improvements with parallel computing, upgrading with WLCG evolutions needed to be considered. |

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:**

Core software

The Gaudi was adopted as the underlying software framework, based on which CEPCSW was developed. The DD4hep-based geometry service was implemented to provide a unified way to retrieve the detector geometry data. The EDM4Hep was used as the event data model to describe the event data in different processing steps as well as the relationships between them. The K4LCIOReader was implemented to convert LCIO data to EDM4Hep objects, which made it possible to read M.C. data produced by iLCSoft. The Geant4 was integrated to the CEPCSW with the help of newly developed integrating modules. The environment for software developing was established by using tools such as Boost, Python, GitHub, CMake, Spack, and CVMFS etc. in the CentOS operating system.

Detector simulation and reconstruction

Tracking algorithms for Silicon Trackers and TPC detector were ported to the new framework and consistent tracking performances were obtained. Realize the drift chamber geometry and simulation and digitization. Make track seed through MC truth information. Integrate Genfit framework in CEPCSW for the track fitting of tracker. Pandora-based PFA from Marlin was ported into CEPCSW for particle reconstruction. Porting simulation and digitization of ECAL into CEPCSW with EDM4hep based calorimetry hit objects and MC truth information and integrated with DDG4 for detector description. Development of simulation and digitization algorithms for a long crystal bar solution for ECAL.

Computing

The basic Grid Infrastructure was set up for CEPC: The VOMS(VO Management System) service was deployed for CEPC VO. The IHEP Grid Computing Element and Storage Element prototype was built up and successfully tested using HTCondorCE with more than 500 CPU cores. The CVMFS server was set up to successfully deploy CEPC software in distributed environment.

**2021:**

Core software

The simulation software will be enhanced by enabling more realistic simulation. The simulation will use the non-uniformity of magnetic field,and also beam-related and electronics noises will be added to physics signals at the MC hit level. Various types of fast simulation tools will be provided such as Delphes and fast ECAL simulation tools etc. Simulated data corresponding to different detector configurations will be produced for optimizing the detector design. The thread safety of EDM4hep and data I/O will be tested with GaudiHive and performance will be optimized. Parallelized simulation of dE/dx or dN/dx in the gaseous detector will be prototyped as a possible application scenario for parallel computing . The prototype of validation platform will be setup to demonstrate the process of automatic compiling and unit testing triggered by source code commitment.

Detector simulation and reconstruction

Tracking and fitting performance for silicon tracker and TPC will be studied and optimized, through improving the architecture of reconstruction software, implementing more realitic detector modules and comparing more algorithms. The effect of noise and background on silicon and TPC tracker will be implemented into performance study. Besides the CDR baseline, track fitting with the combination of silicon and drift chamber measurements will also be performed. Realize the dN/dx samping with the sampling tool and study the performance of PID with cluster counting method in CEPCSW. Optimize tracker with fast simulation tools. Get the CRD physics performance in CEPCSW. Check particles and jet reconstruction performance using PFA algorithm such as Pandora and optimize the reconstruction algorithms for different detector designs. Feasibility research on long crystal bar solution for ECAL, the ghost ambiguity rejection is the key point. Porting simulation and digitization of HCAL in baseline detector design to CEPCSW.

Interfacing the FastJet package and migrating LCFIplus, the vertex finding, jet-clustering, and flavor-tagging tool in the CEPSW. To support alternative ML-based jet classification algorithms, which will be studied and implemented in CEPCSW. The core software will integrated with ACTS to take advantage of its modern, parallelizable, efficient and fast track and vertexing reconstruction software.

Computing

The distributed computing infrastructure prototype is built up based on DIRAC. Available CPU and storage resources are being integrated in the infrastructure.

**2022:**

Core software

To take advantage of machines with many cores, both simulation and reconstruction algorithms will be migrated to run as fully multithreaded applications. The database service will be developed to manage metadata of detector geometry files and to provide coherent access to geometry data. For machine learning, the interface will be developed to integrate TensorFlow/PyTorch tools with the CEPCSW to support various machine learning applications. A software prototype will be developed based on ROOT RDataFrame to support declarative analysis. According to data format of the beam test, new interfaces will be developed for reading the test beam data and performing analysis on test beam data in the CEPCSW. The software validation platform will be fully functioning by adding database to store unit testing results, synchronizing message queue, and validation algorithms, and a web-based validation dashboard will be provided.

Detector simulation and reconstruction and physics analysis

More silicon tracking optimization will be considered, and try to develop tracking algorithm based on machine learning technique. For TPC, correction on the energy lose and multi-particle scattering for gas in Kalman filter will be optimized.

On ACTS, the new propagator method will be development to balance the speed and precision. Optimization of reconstruction of long crystal bar for ECAL to improve the precision of 3D cluster finding, especially with multiple particles incidented into one super module. Extraction of the jet energy resolution dependence, and the limitation of detector and physics reach. Development and upgrade of PFA algorithms to improve the performance of long crystal bar solution for ECAL.

Computing

The production system prototype will be built up to allow massive production jobs to automatically submit and run.

**2023 and 2024:**

Core software

The complete chain of data production from event generator, detector simulation, digitization, event reconstruction and physics analysis will be established by fully utilizing Key4hep software stack. In order to use GPU, FPGA and TPU etc. more efficiently, the framework software will provide a unified and straightforward way to offload time consuming calculations to any kind of accelerator. The performance of ROOT based analysis software framework will be tuned to achieve good scalability for any analysis running on a multi-core server and a large computing cluster respectively. Event selection for online data reduction will be studied by developing event selection steering software and reusing software components developed in the offline environment.

Computing

The User job submission and management system prototype will be developed to allow users easily used. The data transfer system prototype will be developed to allow data sharing between sites.

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 |