Date: 01/06/2017

EIC Detector R&D Progress Report

Project ID: eRD20
Project Name: Developing Analysis Tools and Techniques for the EIC
Period Reported: 10/01/2016 - 01/06/2017
Project Leader: Markus Diefenthaler (JLab) and Alexander Kiselev (BNL)
Contact Person: Markus Diefenthaler (mdiefent@jlab.org)

Abstract

We develop analysis tools and techniques for the EIC, and facilitate communication and collaboration among current and future developers and users. We help coordinate the EIC software effort, providing organization and guidance to help seed growth of a software community that will exist for well over a decade. While our localized efforts are typically focused on completing specific tasks or developing certain tools, the consortium focuses also on achieving the following forward-looking goals:

1. **Organizational efforts with an emphasis on communication:** We help with the organization of the software effort for the EIC by providing documentation about the available EIC software and by maintaining a software repository. This function will be eventually taken over by an official EIC software group. We encourage participation in our consortium and schedule regular meetings to build an active working group and foster collaboration.
2. **Planning for the future with forward compatibility:** We incorporate new standards and validate our tools on new computing infrastructures.
3. **Interfaces and integration:** Given the current stage of the EIC project, it is too early to define the analysis tools of the EIC. However, it is important to connect the existing frameworks / toolkits and to identify the key pieces for a future EIC toolkit. We work on interfaces between the existing frameworks / toolkits and aim to collaborate with other R&D consortia and projects in general to integrate their tools into existing frameworks / toolkits. By doing so, we have started to define the key pieces of the EIC toolkit and are identifying the high-priority R&D projects.

Past

The development of Monte Carlo Event Generators is a priority for our project. In the "EIC Software Meeting" in September 2015, we have reviewed the MCEGs that are available for the EIC and identified MCEGs and other Monte Carlo tools that need to be developed. For the development of the EIC analysis, a MCEG for TMDs is urgently required. There is a separate project for this task funded by a JLab LDRD project. Thus, we have not included this important project in our funding request for FY17. There is also work required on MCEG for eA processes. The development of DPM-JetHybrid, a tool to refine detector requirements for eA collisions in the nuclear shadowing / saturation regime has been funded by the Generic R&D program for the EIC (eRD17). Elke-Caroline Aschenauer is part of this project. Within our R&D consortium, we make fundamental contributions to the development of the following Monte Carlo tools:

- Consistent approach to integrate radiative corrections into MCEGs (page 5 of our proposal)
- Validation of critical Geant4 physics in the energy regime of the EIC (page 10 of our proposal)
- Start the development of an universal event display for MC events (page 12 of our proposal)
- Promote open-data developments for efficient data-MC comparisons (page 13 of our proposal)

We also work on these tasks:

- Work towards a common geometry and detector interface (page 15 of our proposal)
- Work towards an unified track reconstruction (page 17 of our proposal)
- Develop interfaces to forward compatible, self-descriptive file formats (page 19 of our proposal)
- Build a community website and organize software repositories dedicated to the EIC

The work on each of task listed above is coordinated separately. In our monthly meetings, we review the status of each project and discuss the future direction of our work.

Our project started on October 1st 2016. Updates can be reported on the following projects:

Consistent approach to integrate radiative corrections into MCEGs (E.C. Aschenauer and A. Bressan)

We have identified the radiative correction code HERACLES embedded into DJANGOH as the best starting point to develop and standalone radiative correction code library. We are currently working on the following points:

- extract HERACLES from DJANGOH
- write a fortran and C++ wrapper to make the library flexible to be integrated

into different MC codes
- but most importantly we are currently establishing a testing procedure to verify the MC-radiative correction approach gives the same results as dedicated standalone process dependent MC calculations. To execute this test for several different processes, i.e. inclusive DIS, semi-inclusive DIS and exclusive reactions is our next milestone.

We are in contact with Dr. H. Spiessberger, the author of HERACLES, aiming to establish a collaboration for this project.

Validation of critical Geant4 physics in the energy regime of the EIC (A. Dotti)

The technical aspects of the applications needed to perform physics validations have been completed. A "single interaction application" has been created and regression-testing macros prepared for relevant physics processes, it is expected that the application will be made somehow publicly available in the first half of 2017. An existing second application tailored to CPU-performance measurements has been generalized beyond the initial HEP domain **[1]**. The application can read GDML files and thus could be used to test EIC geometries.

A new requirement has emerged: creation of a Geant4 standalone example to read the ProMC data-files. The example should also define the strategy on how to read the data files in a multi-threaded application.

[1] https://twiki.cern.ch/twiki/bin/view/Geant4/Geant4HepExpMTBenchmark).

Work towards a common geometry and detector interface (W. Armstrong)

The common geometry and detector interface is challenging due to a coupling to both event data models and input geometry formats. We have started to explore existing or under-development solutions in nuclear and high-energy physics. The DD4hep project, which is under active development, appears to be a good choice and we will examine in our February 8th - 10th meeting how it can be integrated with the existing simulation frameworks at ANL, BNL, and JLab. We will work towards defining an event data model, which in turn, allows for flexible, modular and independent development of the digitization and reconstruction tools. Here, many lessons can be learned from the LCIO data model for which a plethora of tools are available. As a first step towards an EIC data model, we aim for a common data format for the simulation input and output.

Develop interfaces to forward compatible, self-descriptive file formats (S. Chekanov)

We have created a software program to convert Pythia8 event records to the LCIO data format used for full simulations by HepSim. This provides the opportunity to generate ep/eA events using the Pythia8 event generator and to process such events using the SLIC software package for Geant4 simulation and full event reconstruction. The conversion tool uses ProMC file format created by Pythia8 and convert it to the LCIO files. Previously, such conversion was only possible for Pythia6 event records.

Develop interfaces to forward compatible, self-descriptive file formats (A. Kiselev)

During the in-person meeting on October 17th, it was suggested, that a non-ASCII and non-ROOT file format is preferred to maintain the repository of Monte-Carlo event files produced by various DIS event generators. The suggested primary solution for these purposes was the ProMC library, using Google Protocol Buffers format. Since this library was originally developed for Pythia pp-generator events, it appeared to be not flexible enough to be adapted easily to the EIC simulation environment. Some effort was invested to add the missing features and to get rid of unwanted dependencies. As a by-product of this tuning work a small standalone package (EicMC) emerged with its own custom file format, based on the same Google Protobuf library, which is potentially better suited for EIC community needs.

To this moment the developed codes allow to convert either ASCII or ROOT files produced by all known DIS generators from the so-called eic-smear package to a unified extendable file format, which is at least as efficient as ROOT and ProMC implementations in terms of file size and file reading performance. For file import operations EicMC does not have any external dependencies other than the Google Protobuf libraries themselves. File format is oriented to the modern 64-bit platforms, in particular it does not have any built in 16(32)-bit limitations on record count, file size, etc. Both sequential (streaming) and direct access to event records with scalable record catalogues is provided. EicMC converter behavior can be configured in order to produce the smallest file size (gzipped format), the fastest import (unzipped version), or fastest direct access (short catalogue records), all options using native Google Protobuf library streaming classes as a backbone.

The format is truly self-descriptive in a sense that the complete Google Protobuf message structure is automatically encoded in the file header and can be retrieved during import if needed. The software potentially may be better suited for e.g. multi-threaded GEANT4 environment, than other presently available solutions. Validation work is still in progress.

Build a community website and organize software repositories dedicated to the EIC (W. Armstrong and M. Diefenthaler)

We have created a GitLab repository for our project and will use this repository also to link to existing documentation and provide additional documentation. We will link our repository to the website of the EIC UG.

Interfaces and Integration

On October 17th 2016, we have held a workshop at BNL **[2]** where we discussed the following questions:

- Review of existing simulation software: What technology is used? What features are available? How flexible is the simulations software?
- Software requirements: What software does the community need? What software is urgently required? What long-term goals do we have?

- Discussion about our common goals and work plan (with focus on the geometry and detector interface and the unified tracking projects)

In our workshop, we have reviewed the following frameworks:

- EicRoot, the framework used by the EIC task force at BNL
- fpadsim, the collection of analysis tools used by ANL.
- Fun4All, the framework used for PHENIX and sPHENIX
- GEMC, the framework used by the JLEIC project

EicRoot, fpadsim, Fun4All, and GEMC, are actively maintained. SLIC is part of fpadsim. It was developed for the International Linear Collider (ILC) and is now maintained by the community on a best effort basis.

The analysis environments for the EIC will be chosen when the EIC collaborations will form. Until then, we will examine the requirements for the EIC analysis environment and work on the R&D aspects of the EIC analysis environment. We discussed the requirements for an EIC analysis environment and have identified two main use cases:

- Use Case 1: Requirements for studying a physics process at the EIC:
    - Interface to MCEG
    - Open access to accelerator specifications
    - Open access to detector geometry and detector simulations
    - Documentation
- Use Case 2: Requirements for studying a detector at the EIC
    - Open access to physics simulations and / or interface to MCEG
    - Open access to accelerator specifications
    - Open access to geometry and detector simulations
    - Documentation

Use cases 1 and 2 might involve comparison of eRHIC and JLEIC. The eRHIC settings / geometry might be used in JLEIC software (and vice versa). Based on the use case and the resulting requirements, we have made the following plan:

- We will put our initial focus on fast simulations.
- We will use HepSim as a repository for EIC MC simulations. The format for the generated MC information will be based on Google protocol buffers.
- We will maintain a repository of accelerator settings and detector geometries (GDML and a list of sensitive detectors).
- We will use store the results from the physics and detector simulations as ROOT trees in a fixed format.
- We will work with the EicRoot, Fun4All, GEMC, and SLIC developers so that the existing frameworks meet the following requirements:
    - Readin EicMC files
    - Readin GDML files and the files containing the list of sensitive detectors from our repository
    - Create those files from within the framework or a converter
    - Write the results as flat (POD) ROOT trees in our format.

- We will provide for each framework tutorials for use cases 1 and 2.
- We will promote our work towards interfaces and integrations and the tutorials within the EIC UG.

In addition, we have identified the mid-term requirement for an EIC data model:

- We will develop a data model for the EIC analysis environment.

[2] https://www.jlab.org/indico/event/187/

Future

We will continue our monthly status and coordination meetings.

We will meet on February 8th -10th at BNL to follow up on the discussions from the October meeting and to get feedback from the framework / tool kit developers on the requirements we have agreed on. We will also discuss in detail the requirements for fast simulation packages and review the existing fast simulation packages. The meeting will be joined by Makoto Asai, the spokesperson of the Geant4 collaboration, where will discuss about a closer collaboration between the Geant4 collaboration and the EIC / nuclear physics community. Last but not least, we will decide on the date and topics of the next in –person meeting in FY17.

In March 2016, Amber Boehnlein, Markus Diefenthaler, and Graham Heyes held a workshop at Jefferson Lab to discuss trends in scientific computing and to collect ideas on how to improve analysis workflows at existing nuclear physics projects as well as to work towards analysis techniques and tools for future projects like the Electron-Ion Collider **[3]**. In the successful workshop with 76 participants from 9 national laboratories, 12 universities, and one company, best practices in scientific computing and the future development of existing analysis techniques and tools were reviewed, and it was discussed how future computing trends like Big Data or machine learning can improve the productivity at existing or future experiments.

We would like to help with the "Future Trends in Nuclear Physics Computing" workshop in 2017, which will focus on a data perspective and the question how a data analysis will evolve in the next decade.

[3] https://www.jlab.org/conferences/trends2016/

Manpower

The members of the EIC software consortium have spent in average 10% of their time on eRD20.

- Whitney Armstrong (ANL)
- Elke-Caroline Aschenauer (BNL)
- Andrea Bressan (INFN Trieste)
- Andrea Dotti (SLAC)
- Sergei Chekanov (ANL)

- Markus Diefenthaler (Jefferson Lab, co-PI)
- Alexander Kiselev (BNL, co-PI)
- Christopher Pinkenburg (BNL)
- Stefan Prestel (SLAC)

External Funding

There has been no external funding was obtained.

Publications

There are no publications yet.