Generic R&D Proposal submitted on June 21st 2019 for FY20:

eRD20: Developing Simulation and Analysis Tools for the EIC

Proponents

ANL (2) Whitney Armstrong, Sergei Chekanov
BNL (4) Elke-Caroline Aschenauer, **Alexander Kiselev (co-PI)**, Jerome Lauret, Christopher Pinkenburg
JLAB (4) **Markus Diefenthaler (co-PI)**, David Lawrence, Dmitry Romanov, Maurizio Ungaro
Lund University (1) Stefan Prestel
SLAC (2) Makoto Asai, Dennis Wright
University of Trieste (1) Andrea Bressan
University of Manitoba (1) Wouter Deconinck

Contact Markus Diefenthaler, Alexander Kiselev

Abstract

Developing the physics program for the EIC, and designing the detectors needed to realize it, requires a plethora of software tools and multifaceted analysis efforts. Many of these tools have yet to be developed or need to be expanded and tuned for the physics reach of the EIC. Currently, various groups use disparate sets of software tools to achieve the same or similar analysis tasks such as Monte Carlo event generation, interaction region and detector simulations, track reconstruction, and event visualization to name a few examples. With a long-range goal of the successful execution of the EIC scientific program in mind, it is clear that early investment in the development of well-defined interfaces for communicating, sharing, and collaborating, will facilitate a timely completion of not just the planning and design of an EIC but ultimate delivery of the physics capable with an EIC.

Overview

The EIC Software Consortium (ESC) is part of the EIC Generic Detector R&D program and works on common interfaces among the simulation tools for the EIC (e.g., geometry, file formats, tracking) and explores new avenues of software development (e.g., artificial intelligence). It engages with the EIC User Group (EICUG) to communicate the status of EIC software, bring existing EIC software to end users, produce publicly available consensus-based documents on critical subjects, and provide a vision for the future. The ESC works in liaison with the Geant4 Collaboration Meetings, the HEP Software Foundation, MCnet, and the ROOT team at CERN.

The EIC Detector Advisory Committee recommended for FY18 that we *"take a more active role in working with the detector consortia to help with the simulations and set up a process to easily implement new detector configurations to optimize the detector design."* We have followed up on the recommendation by providing containers for various simulation tools in the EIC community and putting our main focus on common interfaces among the simulation tools. This work resulted in "EIC Software Sandbox" containers based on a geometry interface in GDML and a common set of IO standards, and a community reference reconstruction that are available for the community for the EIC.

In FY18, we have started to engage with the EICUG:

- ❑ We have worked with the EICUG Steering Committee on a computing vision for the EIC that has been presented to the National Academy of Sciences (NAS) and has been included in the NAS report on the EIC science case.

- ❑ The EICUG steering committee has also asked us to summarize the status and plans for the EIC software at the recent EICUG Detector Discussion meeting at Temple University. Our review covered the whole palette of existing EIC-related GEANT-based software frameworks (ANL software suite, EicRoot and fun4all from BNL, GEMC from JLab) as well as the legacy fast smearing generator (eic-smear code developed at BNL), EIC R&D PID Consortium software, existing tools to model EIC Interaction Regions and few advanced standalone simulation applications. Reflecting the ESC philosophy, it was stated that in the present diverse environment and given a very limited dedicated manpower one should not expect various software efforts to merge with each other but we should rather focus on developing common interfaces and shared tools, which in particular would allow one to exchange the detector geometries between different frameworks and to design the software algorithms, libraries and tools in a way they can later be used in various environments rather than be strictly bound to a particular framework and/or toolkit. It was also anticipated that in these early days of EIC the users should not expect to find a *"perfect EIC simulation environment"* under any of the existing frameworks, but rather be ready to contribute to the common development by either providing feedback to the software maintainers or, even better, by contributing to the code base directly.

In June 2018, the EICUG formed a software working group, with currently Andrea Bressan (ESC), Markus Diefenthaler (ESC), and Torre Wenaus (BNL HEP) as conveners. In the announcement of the working group, the EICUG steering committee mentioned the *"considerable progress made within the EIC Software Consortium"* and presented a charge that is very similar to the ESC objectives:

> *"The EICUG Software Working Group's initial focus will be on simulations of physics processes and detector response to enable quantitative assessment of measurement capabilities and their physics impact. This will be pursued in a manner that is accessible, consistent, and reproducible to the EICUG as a whole. It will embody simulations of all processes that make up the EIC science case as articulated in the White-paper. The Software Working Group is to engage with new major initiatives that aim to further develop the EIC science case, including for example the upcoming INT program(s), and is anticipated to play key roles also in the preparations for the EIC project(s) and its critical decisions. The Working Group will build on the considerable progress made within the EIC Software Consortium (eRD20) and other efforts. The evaluation or development of experiment-specific technologies, e.g. mass storage, clusters or other, are outside the initial scope of this working group until the actual experiment collaborations are formed. The working group will be open to all members of the EICUG to work on EICUG related software tasks. It will communicate via a new mailing list and organize regular online and in-person meetings that enable broad and active participation from within the EICUG as a whole."*

We are working with the EICUG Software Working Group on common simulation tools for the EICUG:

- ❏ In cooperation with the MCnet we organized in February the second workshop on "MCEGs for future ep and eA facilities". In the three-day workshop at DESY we have reviewed the R&D on Monte Carlo Event Generators (MCEGs) for polarized lepton-proton (ep) and lepton-nucleus (eA) collisions and discussed in particular the merging of QED and QCD effects (radiative corrections) and the ongoing validation of DIS simulations with the HERA data.
- ❏ In March, we have had a plenary talk at the "Joint HEP Software Foundation Workshop, Open Science Grid All-Hands Meeting, and the Worldwide LHC Computing Grid Workshop" at Jefferson Lab. There has been high interest in the software initiatives for the EIC with various suggestions from the ILC, HL-LHC and other related efforts in HEP.

- In May, we have organized the “EIC Software Meeting” in Trieste, Italy, to reach out to the EIC community in Europe, to coordinate our work with the HEP Software Foundation, MCnet, and the ROOT team at CERN, and to review the progress on the simulation tools we are preparing for the EICUG.
- During the upcoming EICUG Meeting in Paris, France, we will give a tutorial on how to use simulation tools for the EIC. More tutorials are being planned for the rest of the year.
- We help organizing the “Geant4 Collaboration Meeting“ which will be held on September 23-27 at Jefferson Lab. The meeting will include on September 24 a Technical Forum on EIC detectors which will be coordinated with the EIC Generic Detector R&D program.

This is in agreement with your recommendation from January 30, 2019: *“The effort of providing common interfaces is strongly supported and the ESC is strongly encouraged to interact with the user community at the next EIC user meeting”*.

Prior to the EIC Generic Detector R&D meeting, we will meet at BNL to work out the requirements for simulation tools including geometry exchange between the eRHIC and JLEIC concepts and plan next steps for common tools for the EICUG. The work will be also relevant for TOPSiDE and other detector concepts for the EIC as well as the EIC Generic Detector R&D program. The goal of the meeting is to clarify the strategy for EIC detector and physics simulations and to meet the requirement for common tools and documentation in the EICUG.

The timely implementation of the common tools for physics and detector simulations would benefit enormously from dedicated manpower that will be not involved in the CD1 preparations for the EIC that are expected to ramp up in FY20. Thus, we are requesting funds for hiring a postdoctoral researcher who works solely on the community simulation tools and closely with the EICUG on user requests.

Plan for FY20

In FY20, we will be placing our emphasis on simulation of physics and simulations with a focus on Geant4 Simulations (p. 5), Interfaces and integration (p. 6), and Monte Carlo Event Generators (p. 8).

Development of Detector Simulations

Geant4 Simulations

Makoto Asai and Dennis Wright are serving as points of contact between the EIC Software Consortium and the Geant4 International Collaboration, representing the EIC community needs to the Geant4 collaboration, monitoring progress, and making sure feedback is delivered to ESC in a timely manner. One of the most important aspects for the success of the Geant4 simulation of EIC detectors is the correctness of the physics simulation modeling. While efforts made for the simulation of LHC detectors are a good starting point for the EIC, it is however expected that tuning of the physics models will be needed to fully address the energy range of the EIC physics and peculiarities of EIC simulations.

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. We have made the application publicly available in our repository. An existing second application tailored to CPU-performance measurements has been generalized beyond the initial HEP domain. The application can read GDML files and thus could be used to test EIC geometries. We have created a Geant4 standalone example for reading EicMC and ProMC data files. The example code gives an idea of our strategy on how to read the data files in a multithreaded application.

Together with Chris Pinkenburg, Dennis and Makoto are working on a physics list optimized and tuned for the EIC needs. The choice of physics models and transition regions between them require validation against published data and against alternative MC codes. Some of the detector components for future EIC detectors (calorimeter prototypes in particular) have been tested at Fermilab's Test Beam Facility. These tests cover the relevant energy ranges for the EIC and will be used to develop and verify the physics lists used in the accompanying Geant4 simulations. These test beam setups have been incorporated into the Geant4 simulations and comparisons have been published. The test beam data are analyzed within the same framework as the Geant4 based simulation which will result in a fast turnaround time when testing even the newest Geant4 beta versions. The ESC containers provide an easy access for Geant4 developers to the test beam results.

In FY20, we plan to continue our work on a physics list tuned for EIC needs. Available Geant4 hadronic models will be further reviewed and selected. The energy transition regions between the models will also be tuned to fit EIC physics. This work will benefit greatly from the Geant4 Technical Forum planned on September 24 that will be coordinated with the EIC Generic Detector R&D program.

Interfaces and Integration

We are setting up a "software sandbox" environment, which is supposed to unify to a large extent presently disconnected event simulation and reconstruction pieces of code used by our various communities (in particular at ANL, BNL and JLAB). This unified EIC Software Sandbox environment should greatly improve the coherence of our efforts in physics simulations, detector modeling as well as the machine IR background studies, in particular in terms of comparing the performance of the proposed accelerator and EIC detector configurations. The working proposal, which is based on the noticeable progress we made over the last two years in terms of defining EIC event I/O model, interchange file formats and container infrastructure was finally shaped up during an ESC meeting at William & Mary in May 2018. The sandbox will represent itself as a Docker and singularity container image with a well-defined set of "checkpoints", namely the I/O standards describing:

1) Monte-Carlo input events,
2) digitized hits after transporting these events through a particular detector geometry in a particular software framework and/or toolkit,
3) reconstructed physics events as output.

It is anticipated that any EIC detector simulation which we have available at the moment (TOPSiDE at ANL, EicRoot and fun4all at BNL, GEMC at JLAB) should be able to perform a "transition" step from the checkpoint #1 to the checkpoint #2 according to its own way of implementing the particle transport and generating hits in the sensitive detectors.

We have started with the tracking implementation, the EIC Community Reference Reconstruction, and use JANA (a mature and well-maintained piece of JLAB Hall D in-house software) as a workflow tool. Once the detector hits are represented in a unified way at the checkpoint #2, it is assumed that a very generic package based on a straightforward GENFIT implementation will do the rest of the track reconstruction independently of the internal details of a particular software framework which produced the file with the digitized hits and will represent the output events in a unified standard at checkpoint #3. This way the internal machinery of a particular software framework is hidden completely from the end user, who is typically only interested in comparing of reconstructed events to the simulated ones while (as mentioned earlier) both will be available in a framework-independent way. We believe that since in this approach the digitized hit information is effectively decoupled from the geometry details (material distribution in particular), it should be sufficient for the track reconstruction code to be able to import just a standard GDML or ROOT TGeo file and a magnetic field map of a given setup in order to perform the track fitting. The unified format for the magnetic field maps needs to be worked out, but does not look like a task of overwhelming complexity to us. Track finding can be implemented at a later stage. The singularity flavor of this sandbox container image should allow easy portability of the modeling infrastructure between the desktop environment and the production grade computing resources.

D.Romanov with D. Lawrence started the implementation of the Community Reference Reconstruction. The software is based on the JANA workflow tool. The source code is available at https://gitlab.com/eic/ejana (the name "ejana" comes from EIC JANA). One of the strong advantages of the framework is a plugin system which allows to achieve several important goals:

- First, hiding the internal complexity, it allows to automatically activate different parts and utilize various data processing paths depending on execution context.
- Which in turns allows to build easily extensible software that could be used in different scenarios, analysis, testing, calibration, etc., and environments, ordinary PCs, computing farms, cloud services or even real time. The latter is important in the context of the R&D by the EIC Streaming Readout Consortium.
- The plugin system also helps users to organize a workflow in such a way, that independent groups of experts could work in parallel with minimal code interference with each other during development cycles.

The first goal to achieve was to get GENFIT tracking and RAVE vertexing to work in terms of JANA framework (and its plugins), utilizing standard ROOT TGeo or GDML (through TGeo) formats for geometry input, TTree and TFile for data input/output and a magnetic field given in a text file. The overall architecture and plugin development takes into account the final goal to have the framework which is easily extendable and is agnostic to source of data input.

At this point the framework major dependencies are ROOT, JANA, GENFIT and RAVE. The build system takes into account that each of dependent packages may be optionally excluded in the future. In reference to EIC Container Initiative, eJANA is to be released on several docker images: with minimal dependencies, on top of JLAB environment and later on top of other community images.

D. Romanov introduced "EJPM" packet which stands for EIC JANA Packet Manager. Located https://gitlab.com/eic/ejpm . The software main goal is to unify eJANA application deployment workflow in different environments: containers (in reference to Container Initiative), ordinary user machines and production (computing) environments. Secondary goal, is to enhance user experience of working on eJANA (and its plugins) development cycles.

D. Romanov coordinated within the PID consortium (eRD14) a series of meetings about the current status of reconstruction software of PID subsystems and future possibilities of integration into the Community Reference Reconstruction. A detailed review of each PID subsystem has been done and a plan for the integration has been proposed.

In FY20, we will continue the development of the EIC Community Reference Reconstruction and deploy via JupyterLab a next-generation web-based user interface that will facilitate accessible and reproducible simulations for the EICUG.

Monte Carlo Event Generators

Elke-Caroline Aschenauer, Andrea Bressan and Markus Diefenthaler are initiating a project with the Monte Carlo communities in the US and Europe (MCnet) to work on the Monte Carlo Event Generators (MCEGs) for the EIC. The MCEG initiative will connect the MCEG efforts in NP and HEP and should encourage a strong interplay between experiment and theory also at an early stage of the EIC.

As an initial step, they have started a workshop series on "MCEGs for future ep and eA facilities" with the next workshop taking place in November 2019 at the Erwin Schrödinger International Institute for Mathematics and Physics in Vienna, Austria.

The MCnet community is interested in working on ep and eA processes and would like to understand the detailed requirements for MCEGs for the EIC. We are working with the EICUG on a community document on the MCEG requirements for an EIC. The requirements will include R&D on MCEGs for GPDs and TMDs including the simulation of target remnants and spin-orbit correlations and MCEGs for eA processes including the simulations of various nuclear effects and saturation models.

Validating MCEG results requires easy access to analysis results, e.g., from HERA, and the detailed description of the analysis. In HEP, there is an existing workflow for the MCEG R&D using tools such as HZTool, Rivet, and Professor. There is work needed to leverage the HEP analysis tools in NP and to make the HERA analysis results and analysis descriptions available. The validation of MCEG for ep processes is coordinated on the rivet-ep-l@lists.bnl.gov mailing list.

Precision measurements require a good understanding of radiative corrections. Elke-Caroline Aschenauer and Andrea Bressan will continue their development of a library for simulating radiative effects for both polarized and unpolarized observables. They have established a collaboration with H. Spiesberger on our project for a radiative correction library.

Funding request for FY20

We are requesting USD 120,000 for hiring a postdoctoral researcher who works solely on the community tools for physics and detector simulations and interacts closely with the EIC community to not only help with using the tools but in particular assist with integrating existing detector designs and reconstruction algorithms into the community tools. We understand that these funds might not be available. However, our request is timely: With CD0 expected in FY19 and CD1 expected in FY21, the interest in EIC Science in general and EIC Detector R&D in particular will increase considerably and the community would benefit enormously from dedicated manpower that can provide support for the community tools. The funds would be not only a service for the EIC community but would of course also allow proponents of the EIC Generic Detector R&D program to easier test their detector designs in the BeAST, JLEIC or other detector concepts. While postdoctoral researcher would be supervised by the ESC, (s)he would work for EIC Generic Detector R&D program and the EIC community at large.

In addition to that, we request USD 20,000 in travel funds. This budget would be used by the ESC members meet in person and work together on key tasks and be to invite visiting scientists that are essential to the R&D effort. As the ESC consists not only from members of ANL, BNL, and JLAB we cannot rely solely on the respective Lab travel budgets for our meetings.

Nominal budget In total, we request a FY20 budget of **USD 140,000**.

Nominal budget minus 80% If the funds for the EIC Generic Detector R&D program do not allow to hire a postdoctoral researcher for the ESC, then we limit our FY20 request to USD 20,000 for travel.