Proposal to Develop a Generic Tracking Tool

Submission: 21 June, 2019

Principal Investigator: Sergei Chekanov and José Repond[1] (Argonne National Laboratory)

Investigators: Jessica Metcalfe, Manoj Jadhav, Paul Reimer (Argonne National Laboratory)

Abstract: In the context of the TOPSiDE detector concept, we propose to develop a generic package for tracking charged particles. The development will start with the adaptation of a tracking algorithm based on conventional techniques, such as the Kalman filter, to the geometry of TOPSiDE. In a second step we propose to develop a tracking algorithm based on machine-learning techniques. The latter will provide a) some independence from the geometrical layout of the tracker, and b) circumvent the need to retune the tracking algorithm by hand after geometrical changes during the detector optimization process.

[1] Contact person: repond@anl.gov

Introduction: TOPSiDE

TOPSiDE [1] is a modern EIC detector concept based on novel detector paradigms and technologies recently developed by the High-Energy and Nuclear physics community. The tracking of charged particles is achieved with a precision vertex detector and a five-layer silicon tracker in the barrel region, supplemented with the appropriate number of forward disks. The calorimeter is imaging, i.e. it features extremely fine granularity, both laterally and longitudinally [2]. The particle identification (pion-kaon-proton separation) is obtained through precision timing in the calorimeter. In the very forward (hadron) region, the detector features a Ring Imaging Čerenkov counter, a dipole magnet, additional tracking disks, and again imaging calorimetry. In the backward (electron) direction, the required energy resolution is achieved by a crystal calorimeter, again with fine granularity. The central solenoid is placed outside the calorimeter.

There are many advantages to this approach: Each particle produced in ep/eA collisions is measured individually with the subdetector providing the best momentum/energy resolution (Particle Flow paradigm) [3]; the energy resolution for (neutral particles) can be improved through software compensation techniques [4]; due to the imaging capability of the calorimeter, a muon identification system becomes redundant; the material in front of the calorimeter is minimized, as there is no need for additional time-of-flight counters, Čerenkov counters or Transition Radiation Detectors; last but not least, the output of the detector is a list of identified particles with their momenta, similar to the hadron level output of Monte Carlo simulations.

The advantages for physics are many as well: higher photon detection efficiency (due to the minimal amount of 'dead' material in front of the calorimeter and hence an improved material budget. This is particularly important for the measurement of the DVCS process); improved kinematic reconstruction, in particular when using the double angle or Jacquet-Blondel method, jet reconstruction, and background rejection (due to the excellent reconstruction (energy and angle) of the hadronic system), etc. The superior reconstruction capability of TOPSiDE will result in an optimal use of the luminosity and provide precision measurements for all physics processes of interest at the EIC.

Introduction: The simulation of TOPSiDE

We have assembled a complete tool chain for the simulation of the TOPSiDE detector concept and any other detector concept proposed for the EIC, for that matter. Table I lists the various simulation and event reconstruction steps and the tools available for their execution.

The data model, ProIO [5], deserves special mention, as it was recently developed (at Argonne) and has been adopted by the EIC software community.

The geometry implementation is based on DD4HEP, a recent development from CERN. The geometry of TOPSiDE has been coded, but has not yet been optimized. For instance, for most of the solid angle the energy range of photons is below 10 GeV which needs to be taken into account when defining the thickness of the absorption layers in the calorimeter. Also, details of the Ring Imaging Čerenkov, such as the light collection system, are entirely missing.

Task	Tool
Generate collision events	Lepto, PYTHIA8, Milou….
Transport of particles through matter	GEANT4
Digitizing the response (making it look like real data)	Digitizer, e.g. RPC_sim
Data Model	ProIO [5]
Reconstruction of tracks	LCSim [6]
Reconstruction of particles	Pandora: Particle Flow Algorithm
Depository for events	HepSim
Analysis of events	Root
Event display	Root based

Table I. List of links in the simulation tool chain.

The track finding and fitting algorithms are out-of-date and difficult to maintain. In particular, part of the code is hard wired which makes changes to the geometry very tedious to implement. In the meantime, we are using a cheater taking advantage of the true information from the event generator to identify hits belonging to a given charged particle.

Research program

In a first step we propose to replace the outdated track finding with maintainable algorithms, such as one's based on the Hough transform. However, these new algorithms still require some tuning or adaption after each change to the tracking detector layout. To avoid this in the future, we propose, in a second step, to develop a novel track finding tool based on machine learning ML techniques. The advantage will be that the track finding task can be optimized in a straightforward way without the need to reprogram the track reconstruction algorithms after changes made to the tracking geometry during the detector optimization process. We will take advantage of existing, highly developed ML freeware packages, such as TensorFlow [7].

In the longer term, the same approach will be applied to the Particle Flow Algorithm (PFA) stage of the event reconstruction. Again, current implementations of PFAs require manual intervention after each change in detector geometry. A ML-based approach to PFAs will be used, for instance, for estimating the probability that a given cluster of calorimeter hits was created by one or several particles and will find the optimal way to split the cluster if necessary. The same approach will be utilized to perform particle identification (PI) based on the precision timing information provided by the electromagnetic calorimeter. We believe that the ML implementation of the PFA and PI stages will require less effort once the track finding task has been successfully implemented with ML techniques.

Multi-dimensional problems, such as track finding, PFA, and PI are excellent candidates where ML techniques are expected to provide a superior performance compared to what is possible with the application of hard coded algorithms. However, to date no prove to this effect exists yet.

The proposed development, if successful, will constitute a breakthrough in the way we approach the optimization of detector geometries. Once available, any detector concept will benefit from this tool. In other words, this algorithm can easily be adopted by other detector concepts being developed by the EIC community. And, indeed, the leaders of the EIC Software Consortium have expressed support for this development [8].

The proposed work will lead to improved tracking, PFA, and PI performances on one hand and a detector design optimized through a controlled process involving minimal re-programming of the event reconstruction algorithms on the other hand. We believe that these improvements will benefit all physics topics to be studied at the EIC.

Deliverables

The main deliverable of this work is the development of a track finding algorithm based on ML. This algorithm will be modular and independent of a particular detector geometry. Thus, it can be used by the TOPSiDE detector concept as well as by other detector concepts currently being developed by the EIC community [9]. The development of this software will simplify the detector optimization studies and will provide enhanced control over the detector development process geared towards meeting the physics goals of the EIC experiment.

Funding request

We request support for ½ a postdoctoral research assistant in the amount of $62.5k. The major task of the assistant will be the adaptation of a conventional track finding algorithm to the geometry of TOPSiDE, followed by a novel development of track finding algorithms based on ML techniques, leading eventually to the optimization of the geometry of TOPSiDE. In addition, a small amount of funds for travel is requested.

The postdoctoral research assistant will be supervised by Sergei Chekanov and José Repond. Once the funding is established a search for a suitable candidate with interest/expertise in ML techniques will be initiated. As in the past, with David Blyth, we will encourage our postdoc to work closely with the EIC software consortium.

	Postdoc	**Travel**	**Sum**
Argonne National Laboratory	$62.5k	$5k	$62.5k

Table II. Funding requests.

References

[1] J. Repond, *TOPSiDE: Concept of an EIC Detector,* Proc. of the XXVI International Workshop on Deep-Inelastic Scattering and Related Subjects, Kobe, Japan (2018). **PoS DIS2018 (2018) 179**

[2] F. Sefkow et al., *Experimental tests of particle flow calorimetry*, Rev. of Modern Physics **88**, 15003 (2016).

[3] M. A. Thomson, *Particle flow calorimetry and the PandoraPFA algorithm*, Nucl. Instr. Meth. **A611** (2009) 25-40.

[4] C. Adloff et al., *Hadronic energy resolution of a highly granular scintillator-steel hadron calorimeter using software compensation techniques*, JINST 7 (2012) P09017.

[5] D. Blyth et al., *ProIO: An event-based I/O stream format for protobuf messages*, Computer Physics Communications, Vol. 241 (2019), 98 – 112.

[6] http://www.lcsim.org/

[7] https://www.tensorflow.org/tutorials/

[8] M. Diefenthaler, private communication.

[9] J. Repond, *Detector Concepts for the Electron-Ion Collider*, Proc. of 13th International Workshop on High pT Physics in the RHIC and LHC Era (high-pT2019), Knoxville, TN.