U.S. DEPARTMENT OF ENERGY | Office of Science | JSA

Generic R&D for an EIC:
eRD20 - Developing Analysis Tools and Techniques for the EIC

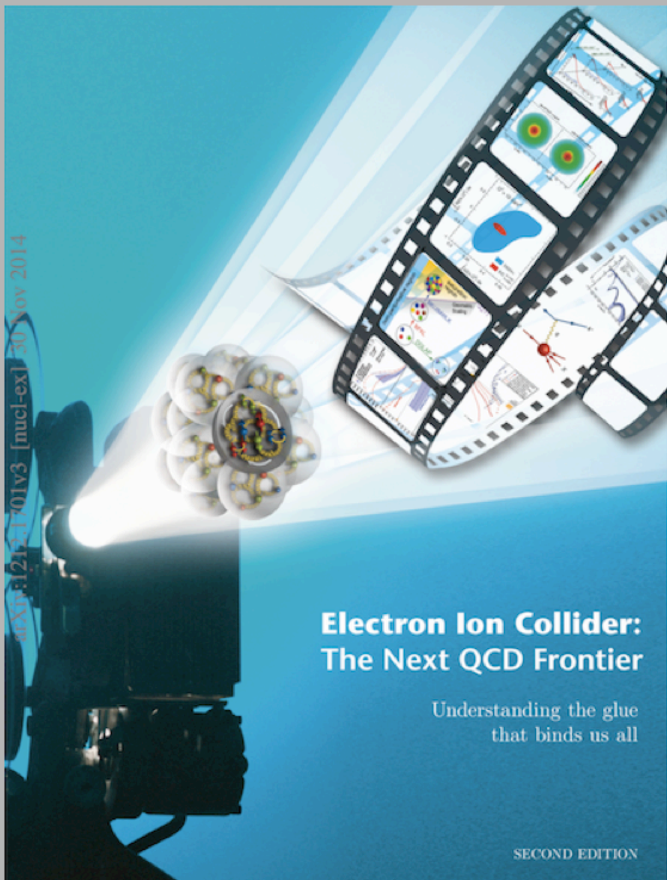
Whitney Armstrong (ANL), Elke-Caroline Aschenauer (BNL), Franco Bradamante (INFN Trieste), Andrea Bressan (INFN Trieste), Andrea Dotti (SLAC), Sergei Chekanov (ANL), **Markus Diefenthaler (Jefferson Lab, co-PI)**, Alexander Kiselev (BNL, co-PI), Anna Martin (INFN Trieste), Christopher Pinkenburg (BNL), Stefan Prestel (SLAC)

EIC Generic R&D Meeting
January 26th – January 27th 2017

Jefferson Lab
Thomas Jefferson National Accelerator Facility

The EIC project

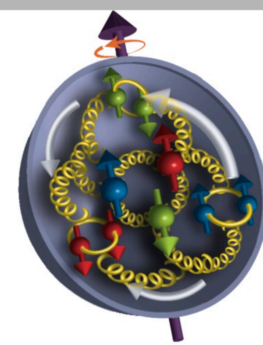
EIC: The Next QCD Frontier

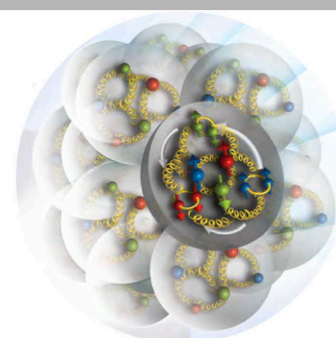


Eur.Phys.J. A52 (2016) no.9, 268

Program aim

- Revolutionize the understanding of nucleon and nuclear structure and associated dynamics.
- For the first time, get (almost?) all relevant information about quark-gluon structure of the nucleon



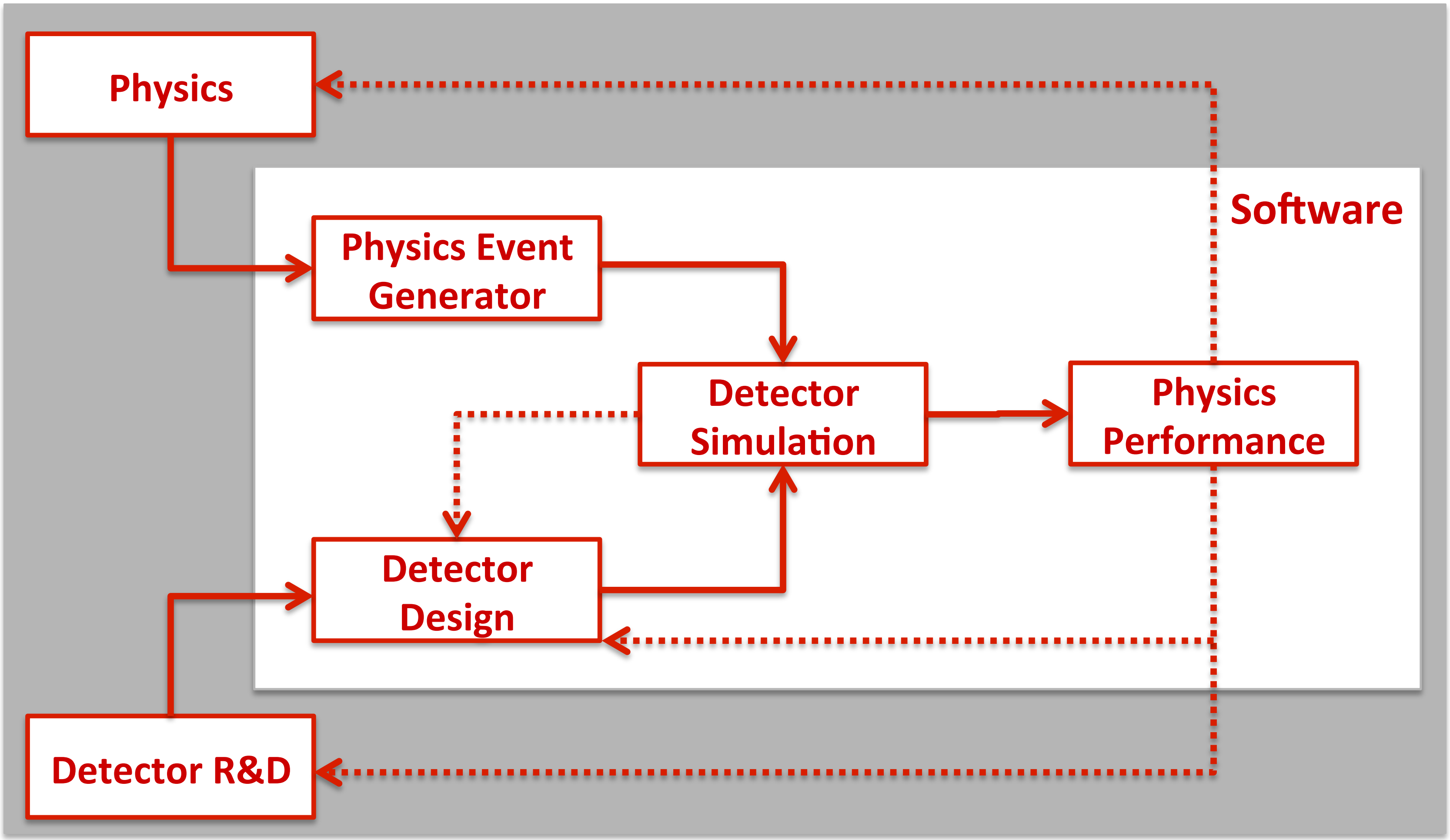


When the EIC is built

it will be *the* new NP machine for many decades. EIC could revolutionize NP — but only if we:

- build the right machine
- build the right detectors
- develop the right analysis environment

EIC R&D and software development



Section

Computing vision

Analysis environments

Developments of analysis environments:

- EIC on the horizon
- think about the **next generation(s) of analysis environments** that will **maximize** the **science output**

LHC experiments: tremendous success in achieving their analysis goals and producing results in timely manners

Lesson learned at LHC experiments:

- as the complexity and size of the experiments grew
- the complexity of analysis environment grew
- time dealing with the analysis infrastructure grew

Anecdote from LHC
a typical LHC student or post-doc spends up to 50% of her/his time dealing with computing issues

New analysis environments

User centered design

- understand the user requirements first and foremost
- engage wider community of physicists in design whose primary interest is not computing
- make design decisions solely based on user requirements
- web-based user interfaces, e.g. interactive analysis in Jupyter Notebook

Future compatibility (both hardware and software)

- most powerful future computers will likely be very different from the kind of computers currently used in NP (Exascale Computing)
- structures robust against likely changes in computing environment
- apply modular design: changes in underlying code can be handled without an entire overhaul of the structure

Think out of the box

- the way analysis is done has been largely shaped by kinds of computing that has been available
- computing begins to grow in very different ways in the future, driven by very different forces than in the past (Exascale Computing)
- think about new possibilities and paradigms that can and should arise

Section

Global objectives and FY17 development

Global objectives

Interfaces and integration

- connect existing frameworks / toolkits
- identify the key pieces for a future EIC toolkit
- collaborate with other R&D consortia

Planning for the future with future compatibility

- workshop to discuss new scientific computing developments and trends
- incorporating new standards
- validating our tools on new computing infrastructure

Organizational efforts with an emphasis on communication

- build an active working group and foster collaboration
- documentation about available software
- maintaining a software repository
- workshop organization

Immediate development in FY17

FY17

Interfaces and integration

- start the development of a library for simulating radiate effects
- work towards a common geometry and detector interface
- work towards an unified track reconstruction
- collaborate with **TMD MC** and **DPMJetHybrid** (eRD17) and other software projects that are essential for an EIC

FY17

Planning for the future with future compatibility

- validation of critical Geant4 physics in the energy regime of the EIC
- start the development of an universal event display for MC events
- promote open-data developments for efficient data-MC comparison from the beginning
- build interfaces to forward compatible, self-descriptive file formats

FY17

Organizational efforts with an emphasis on communication

- build a community website
- organize software repositories dedicated to the EIC
- organize a workshop

Section

Progress in FY17

Interfaces and integration

FY17

Interfaces and integration

- start the development of a library for simulating radiate effects
- work towards a common geometry and detector interface
- work towards an unified track reconstruction
- collaborate with **TMD MC** and **DPMJetHybrid** (eRD17) and other software projects that are essential for an EIC

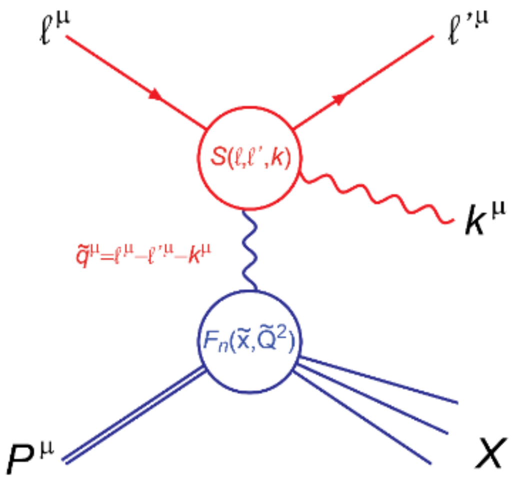
Planning for the future with future compatibility

- validation of critical Geant4 physics in the energy regime of the EIC
- start the development of an universal event display for MC events
- promote open-data developments for efficient data-MC comparison from the beginning
- build interfaces to forward compatible, self-descriptive file formats

Organizational efforts with an emphasis on communication

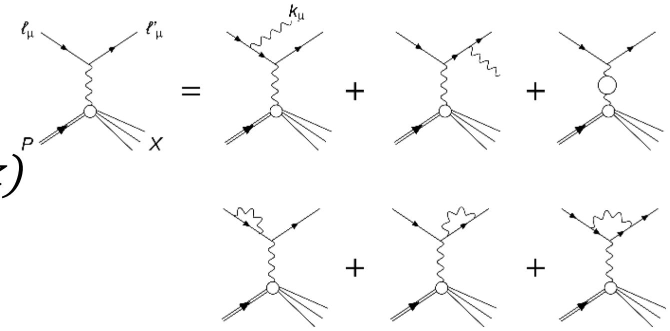
- build a community website
- organize software repositories dedicated to the EIC
- organize a workshop

Radiative corrections



$$Q^2 = -(\ell - \ell' - k)^2$$

$$x = Q^2 / 2P\cdot(\ell - \ell' - k)$$



- Photon radiation from the leptons modify the one boson cross-section and change the DIS kinematics on the event by event basis
- The direction of the virtual photon is different from the one reconstructed from the leptons, giving rise to:
  - False asymmetries in the azimuthal distribution of hadrons calculated with respect to the virtual photon direction
  - Smearing of the kinematic distributions (e.g. z and P_{hT})
- To take into account correctly this effect in the SIDIS cross-section we need both the correct weights for every event and an unfolding procedure for the smearing. THIS can ONLY be done by using a Monte Carlo code for RC

Radiative corrections: Deliverables

Deliverables achieved at the end of the project:

- Calculate radiative corrections for transverse polarized observables to measure TMDs and polarized exclusive observables.
- Provide proof that the MC phase space constrains on the hadronic final state is equal to calculating radiative corrections for each polarized and unpolarized semi-inclusive hadronic final state independently.
- Define a software framework and develop a library based on this framework, which integrates the radiative corrections depending on polarization and other determining factors in a wrapper-software.

Status:

- Radiative correction code HERACLES embedded into DJANGOH identified as the best starting point to develop standalone radiative correction code. **Strategy:**
  - extract HERACLES from DJANGOH
  - establish collaboration with H. Spiessberger, the author of HERACLES
  - fortran and C++ wrapper to make the library flexible to be integrated into different MCEG codes
- Working on test to verify the MC-radiative correction approach gives the same results as dedicated standalone process dependent MC calculations (test inclusive DIS, semi-inclusive DIS and exclusive reactions first).

First meeting in October

EIC Software Consortium Meeting

Monday, 17 October 2016 from **09:00** to **21:00** (UTC)
at **Brookhaven National Laboratory (510 / 2-160)**
Brookhaven National Lab Physics Department Building 510 / 2-160 20 Pennsylvania Avenue Upton, NY 11973

Material: Minutes

Monday, 17 October 2016

09:00 - 09:30 Introduction *30'*
Speaker: Dr. Markus Diefenthaler (Jefferson Lab)
Material: Slides

09:30 - 10:00 Review: SLIC *30'*
Speaker: Dr. Sergei Chekanov (ANL)
Material: Slides

10:00 - 10:45 Discussion *45'*

10:45 - 11:15 Review: GEMC *30'*
Speaker: Dr. Maurizio Ungaro (Jefferson Lab)
Material: Slides

11:15 - 12:15 Review: Fun4All *1h0'*
Speaker: Dr. Christopher Pinkenburg (BNL)
Material: Slides

14:00 - 15:10 Review: GENFIT / Rave *1h10'*
Speaker: Dr. Haiwang Yu (New Mexico State University)
Material: Slides

15:10 - 16:00 Review: EicRoot *50'*
Speaker: Dr. Alexander Kiselev (BNL)
Material: Slides

16:00 - 16:40 Discussion *40'*

16:40 - 17:10 Data model *30'*
Speaker: Whitney Armstrong (Argonne National Laboratory)
Material: Slides

17:10 - 18:30 Discussion *1h20'*

19:00 - 21:00 Dinner *2h0'* (Oceans 5 Seafood Market & Eatery)

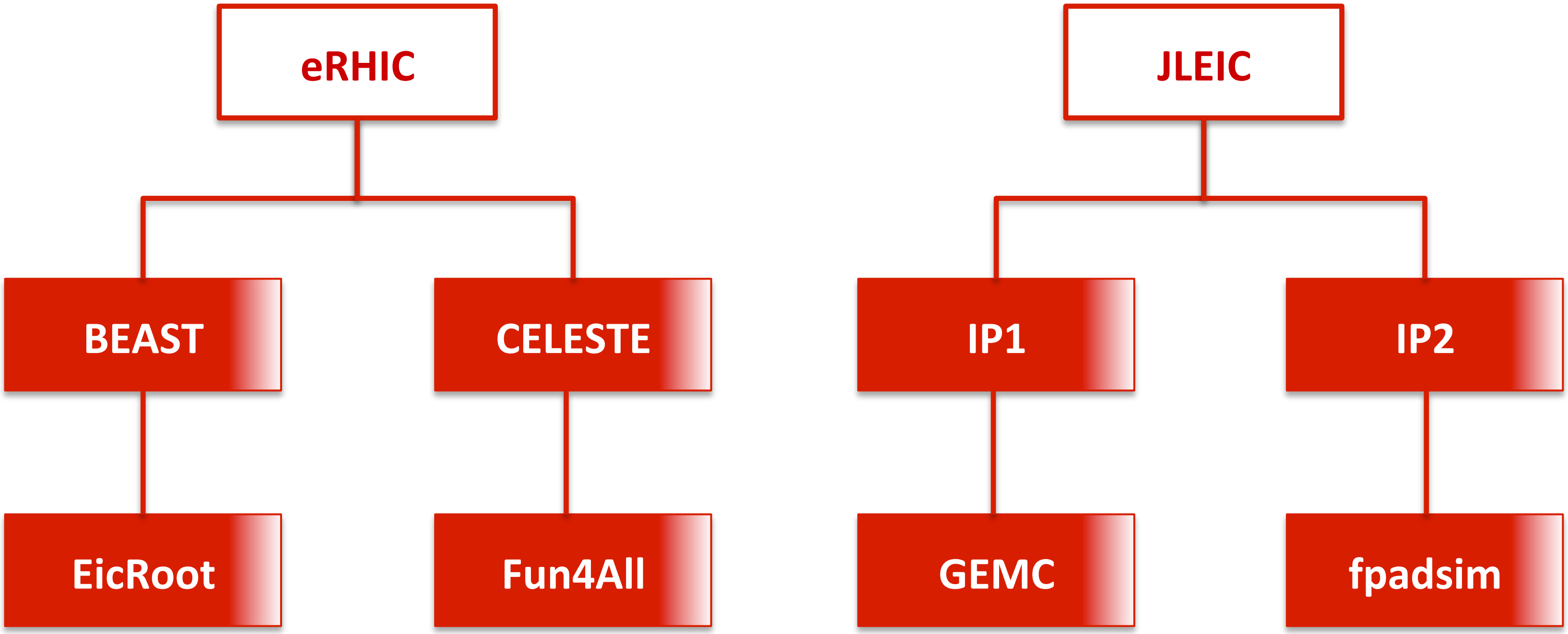
Agenda

- **Review of existing software** (including examples!): What technology is used? What is available? How flexible?
- **Discussion about requirements**: What does the community need? What is urgently required? What long-term goals do we have?
- **Discussion about our common goals and work plan**

Next meeting
February 8th – 10th at BNL

https://www.jlab.org/indico/event/187/

Existing software frameworks for the EIC



Review

- EicRoot, fpadsim, Fun4All, and GEMC, are actively maintained.
- 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.

Discussion of requirements

Use case 1

Requirements for studying physics processes at EIC:

- interface to MCEG
- open access to accelerator specifications
- open access to detector information and simulation
- documentation

Use case 2

Requirements for studying detectors at EIC:

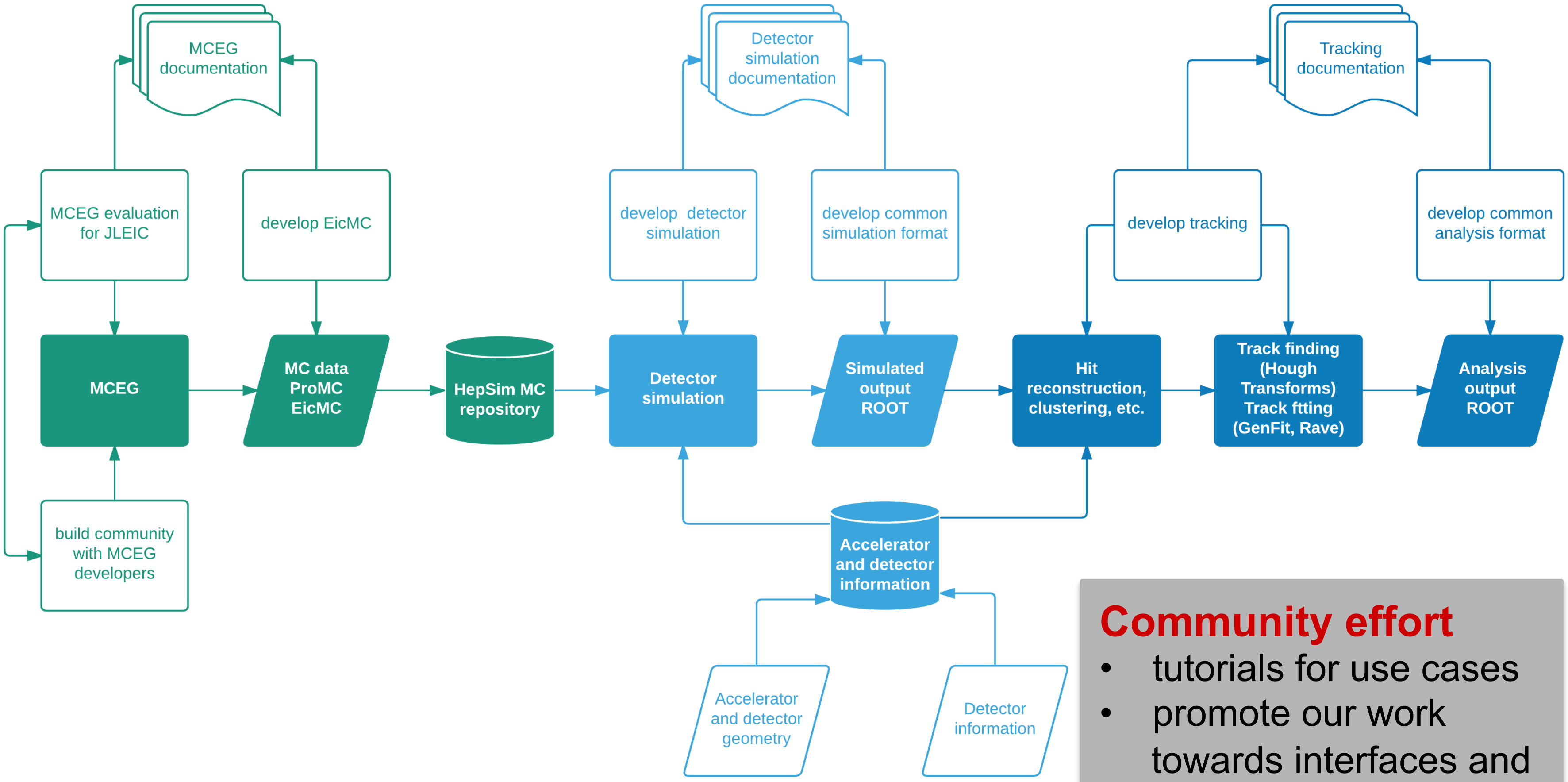
- open access to physics simulations or interface to MCEG
- open access to accelerator specifications
- open access to detector information and simulation
- documentation

Use cases 1 and 2 might involve comparison of eRHIC and JLEIC:

- eRHIC setup might be used in JLEIC software
- JLEIC setup might be used in eRHIC software

Common goals

initial focus on fast simulations



Community effort

- tutorials for use cases
- promote our work towards interfaces and integrations and the tutorials within EIC UG

Planning for the future

Interfaces and integration

- start the development of a library for simulating radiate effects
- work towards a common geometry and detector interface
- work towards an unified track reconstruction
- collaborate with **TMD MC** and **DPMJetHybrid** (eRD17) and other software projects that are essential for an EIC

FY17

Planning for the future with future compatibility

- validation of critical Geant4 physics in the energy regime of the EIC
- start the development of an universal event display for MC events
- promote open-data developments for efficient data-MC comparison from the beginning
- build interfaces to forward compatible, self-descriptive file formats

Organizational efforts with an emphasis on communication

- build a community website
- organize software repositories dedicated to the EIC
- organize a workshop

Validation of Geant4 for the EIC

Collaboration with Geant4/SLAC team (Andrea Dotti)

Makoto Asai (Geant4 spokesperson) will join

Goal:

- Validation and tuning of critical Geant4 physics in the energy regime of the EIC

Progress:

- applications needed to perform physics validations have been completed
- *single interaction application* has been created and regression-testing macros prepared for relevant physics processes, to be released in the first half of 2017
- existing application tailored to CPU-performance measurements has been generalized beyond the initial HEP domain
- application can read GDML files and can be used to test EIC geometries
- a new requirement has emerged: creation of a Geant4 standalone example to read the Google Protobuf based data-files, example should define the strategy on how to read the data files in a multi-threaded application

Interfaces to self-descriptive file formats

New data format for EVGEN: ProMC
baseline in addition to ROOT

S.C., E.May, K. Strand, P. Van Gemmeren, Comp. Physics Comm. 185 (2014), 2629

- "Archive" self-described format to keep MC events:
  - Event records, NLO,original logfiles, PDG tables etc.
- 30% smaller files than existing formats after compression

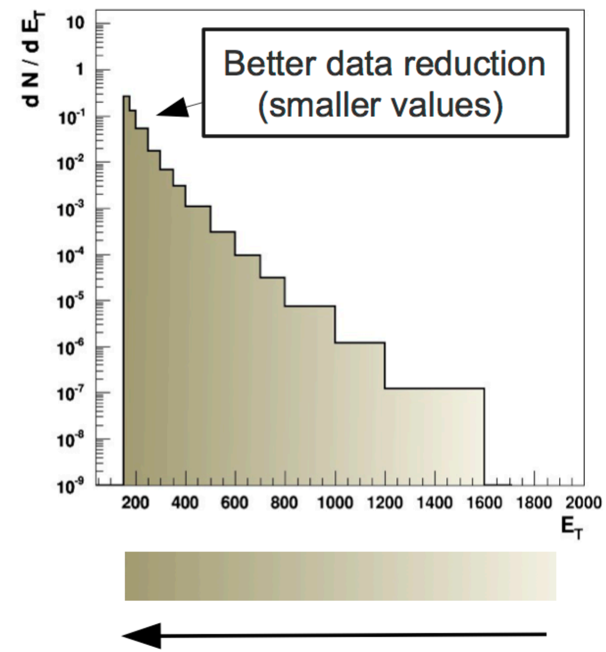
> Number of used bytes depends on values.
> Small values use small number of bytes

Google's Protocol buffers — protobuf (Protocol Buffers - Google's data interchange format)

- Effective file size reduction for pile-up events
  - Particles with small momenta → less bytes used
- Installed on Mira (BlueGene/Q).
- Supports C++/Java/Python
- Separate events can be streamed over the Internet:
  - similar to avi frames (video streaming)

http://atlaswww.hep.anl.gov/asc/promc/

8-bytes (int64) → varint



compression strength keeping precision of constant

EicMC – self-descriptive file format for MCEG

by-product of ProMC import incorporation into EicRoot framework

➔ compact and portable C++ code (~3k lines) developed by Alexander Kiselev

Main objectives

- unified format for all MCEG for EIC (and beyond)
- avoid using either ROOT or ASCII file formats in the future MC repository
- overcome few intrinsic ProMC format limitations

EicMC features

in final stage of validation

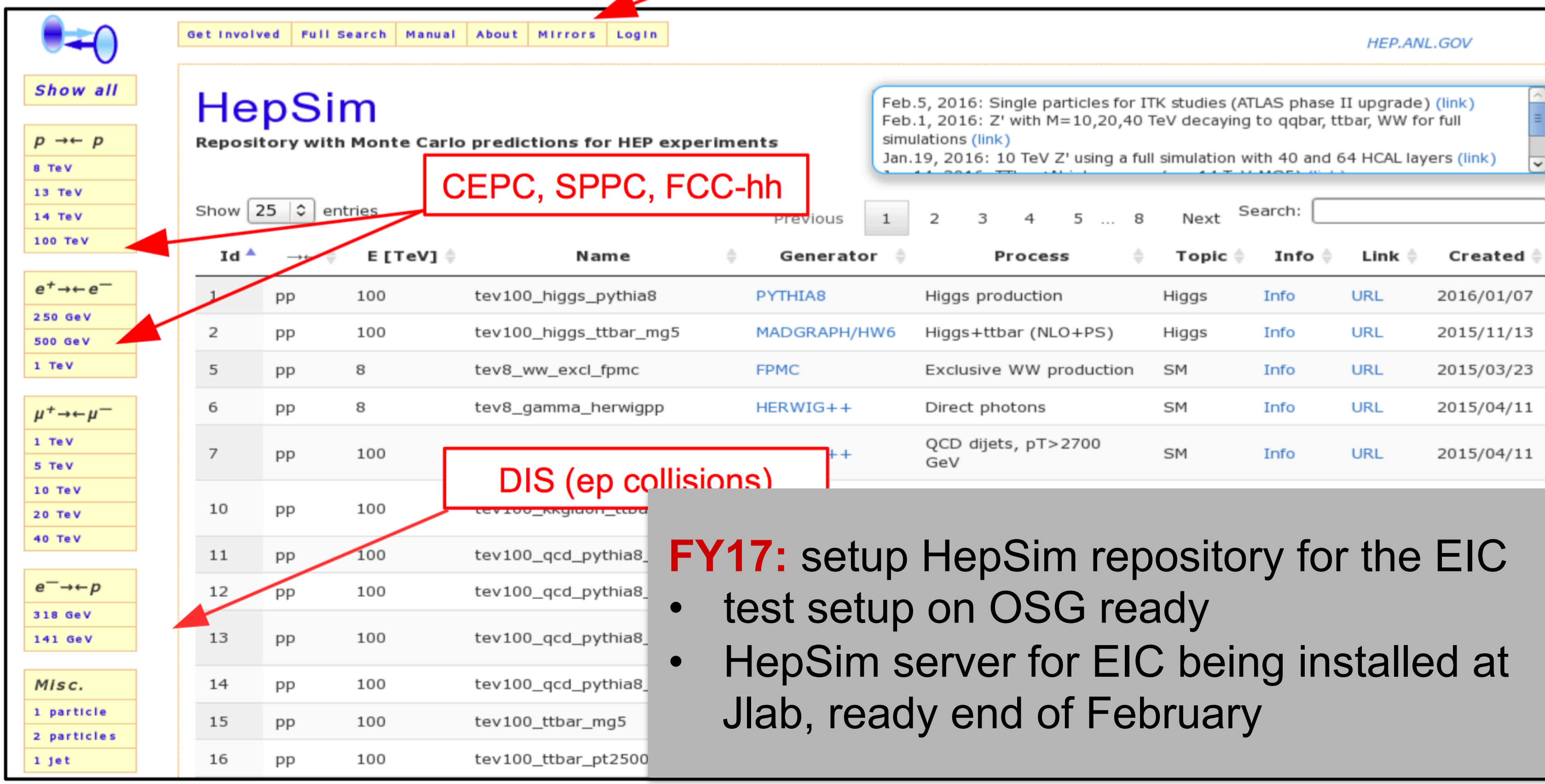
- unified extendable binary MCEG format (all generators from eic-smear are supported)
- true automatic self-description built into the format core .proto file and the user library
- 64-bit implementation ➔ no 16-/32-bit limitations on file size, record count
- flexible set of compression schemes (for file size, input speed, floating point precision)
- both sequential and direct access to event records (with scalable multi-level catalogues)
- no external dependencies on the user (input) side other than Google Protobuf libraries
- convenient user interface
- performance (file size, speed) similar to or better than ProMC and ROOT equivalents

HepSim repository for the EIC

HepSim: Repository of generated events (MC) and detector reconstructed events

uses ProMC

EicMC support to be added



FY17: setup HepSim repository for the EIC

- test setup on OSG ready
- HepSim server for EIC being installed at Jlab, ready end of February

HepSim stores EVGEN files (LO,NLO, etc), fast simulations, full Geant4 simulations

Organizational effort

Interfaces and integration

- start the development of a library for simulating radiate effects
- work towards a common geometry and detector interface
- work towards an unified track reconstruction
- collaborate with **TMD MC** and **DPMJetHybrid** (eRD17) and other software projects that are essential for an EIC

Planning for the future with future compatibility

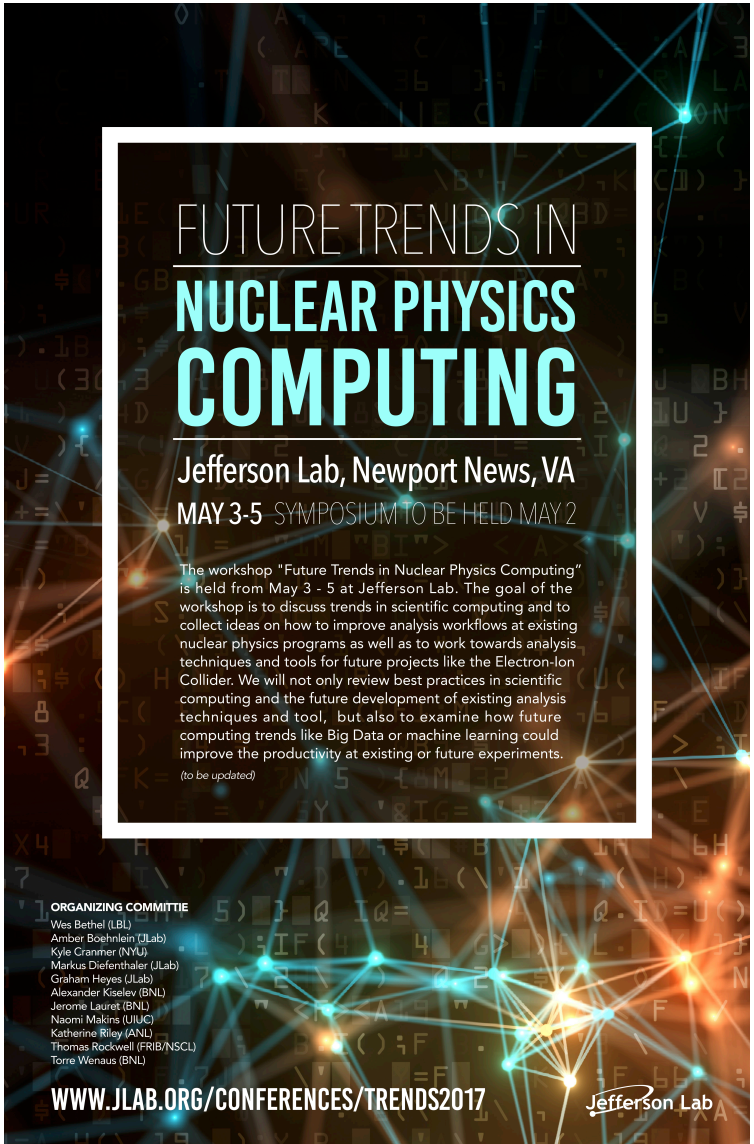
- validation of critical Geant4 physics in the energy regime of the EIC
- start the development of an universal event display for MC events
- promote open-data developments for efficient data-MC comparison from the beginning
- build interfaces to forward compatible, self-descriptive file formats

FY17

Organizational efforts with an emphasis on communication

- build a community website
- organize software repositories dedicated to the EIC
- organize a workshop

Software and workshop organization



- eRD20 involved in **workshop** on "**Future Trends in NP computing**". Topics will include:
  - How can data be used (e.g., combining data or long-lived data repositories)?
  - Visions for future data analysis environments
  - API / methodologies for combining data of various experiments
- **GitLab repository** setup for eRD20 software and documentation

GitLab

Computing for the Electron-Ion Collider

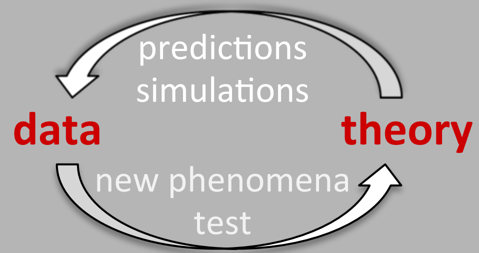
- **EIC Computing**: think about the **next generation(s) of analysis environments** that will **maximize** the **science output**
- **eRD20 – working towards global objectives**:
  - interfaces and integration
  - planning for the future with future compatibility
  - organizational efforts with an emphasis on communication
- eRD20 has started in FY17 and grown by 25% since start, expect to add more members in the next months
- community effort has started, plan on working with EICUG

Section

Backup

Interplay of data and theory

Feedback loop between data and theory



Comparison to:

- analytical calculations
- **Monte Carlo (MC) simulations**
- **Lattice-QCD calculations**

Data-theory comparison: relies on

- open access to data-theory tools
- standardization of data-theory tools
- comparison tools for quick turnaround

MC event generator:

- faithful representation of QCD dynamics
- based on QCD factorization and evolution equations

Usage by experimentalists:

- detector corrections
- analysis prototyping
- comparing to theory

Usage by theoreticians:

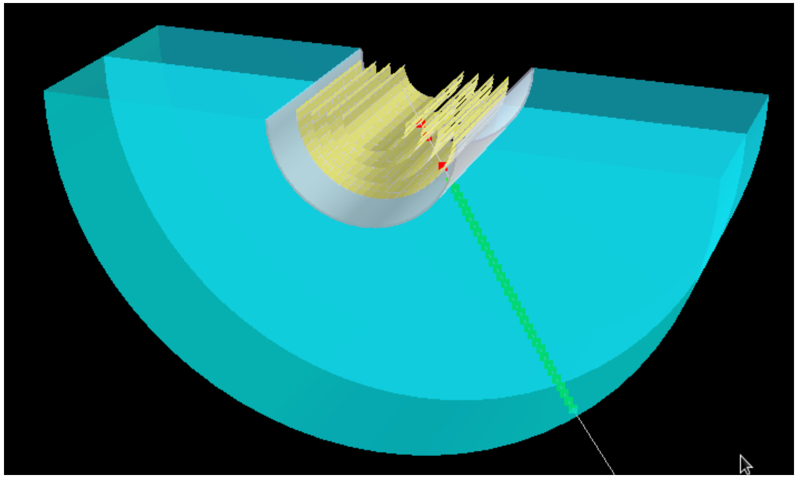
- easy off-the-shelf state-of-the-art tool that looks like data
- validate against and investigate theoretical improvements

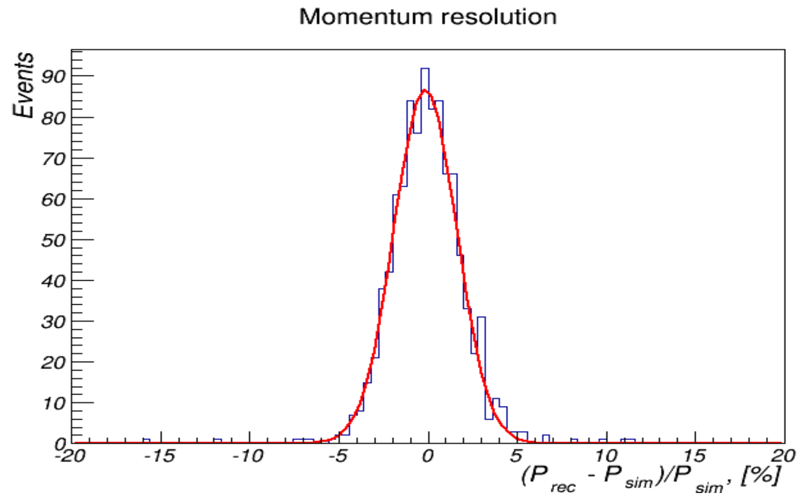
Unified track reconstruction library

Pre-conditions

- Similar requirements for and similar tracker outline of all proposed EIC detectors
- Similar analysis dataflow from simulation to event reconstruction
- Existence of powerful generic libraries for track and vertex fitting (genfit, rave)
- Expertise in the EIC community
- Well-advanced EIC-related set of tracking R&D tools exists already (EicRoot):

Consider a basic example: a vertex tracker + a TPC in a realistic ~3T magnetic field; what is the momentum resolution for pions at p=10 GeV/c and $\theta=75^{o}$?





Distance between the above question and the momentum resolution plot is only ~200 lines of trivial ROOT scripts

But: the tool is at present software-framework-bound!

Unified track reconstruction library

The proposal

- Pull the relevant fraction of tracking-related tools out of the EicRoot framework
- Complement and/or upgrade them with up-to-date libraries (genfit2, rave, etc)
- Provide a suitable unified track finder code for the EIC tracker geometry

- Make use of EIC-specific and framework-independent geometry definition format
- Decide on flexible detector hit formats (raw; digitized; suitable for reconstruction)

Possible first year deliverables

- Perform a detailed feasibility study of the above plan
- Should the task look doable, start code development with a universal standalone library of track *fitting* tools for a typical EIC tracker geometry

Potential benefits

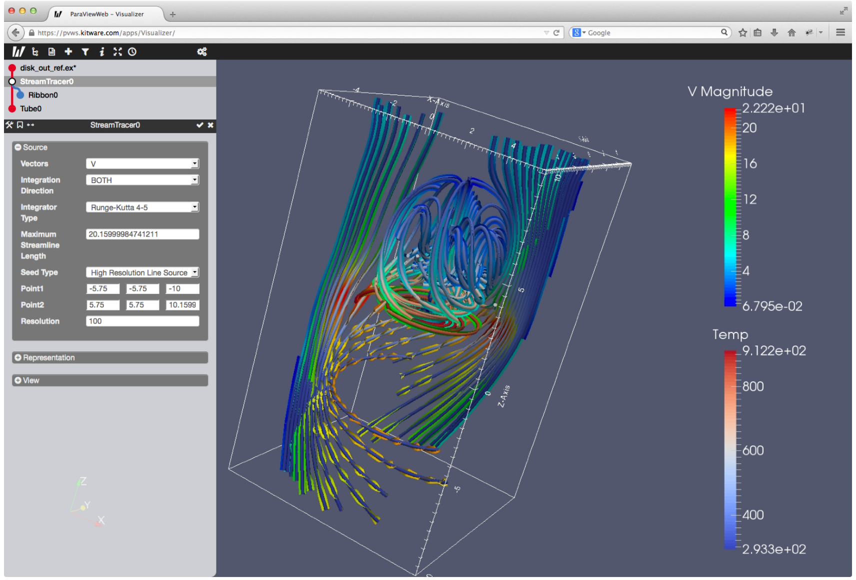
- Provide a unified track reconstruction library which can be used in *any* EIC framework

- Leverage proposed geometry exchange procedure between different implementations
- Simplify detector performance comparisons between site-specific implementations

Development of an universal event display

Motivation:

- outreach
- validation of the EIC simulations
- comparison of different detector designs using an unified approach
- **web interfaces:** forward compatible, universal user interfaces



Goal:

- generic event display for viewing generated (and detector reconstructed) events on web browsers

FY17 Deliverables:

- evaluate how the **CMS** and **ParaViewWeb** event displays can be used for the existing software frameworks and the HepSim MC repository