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

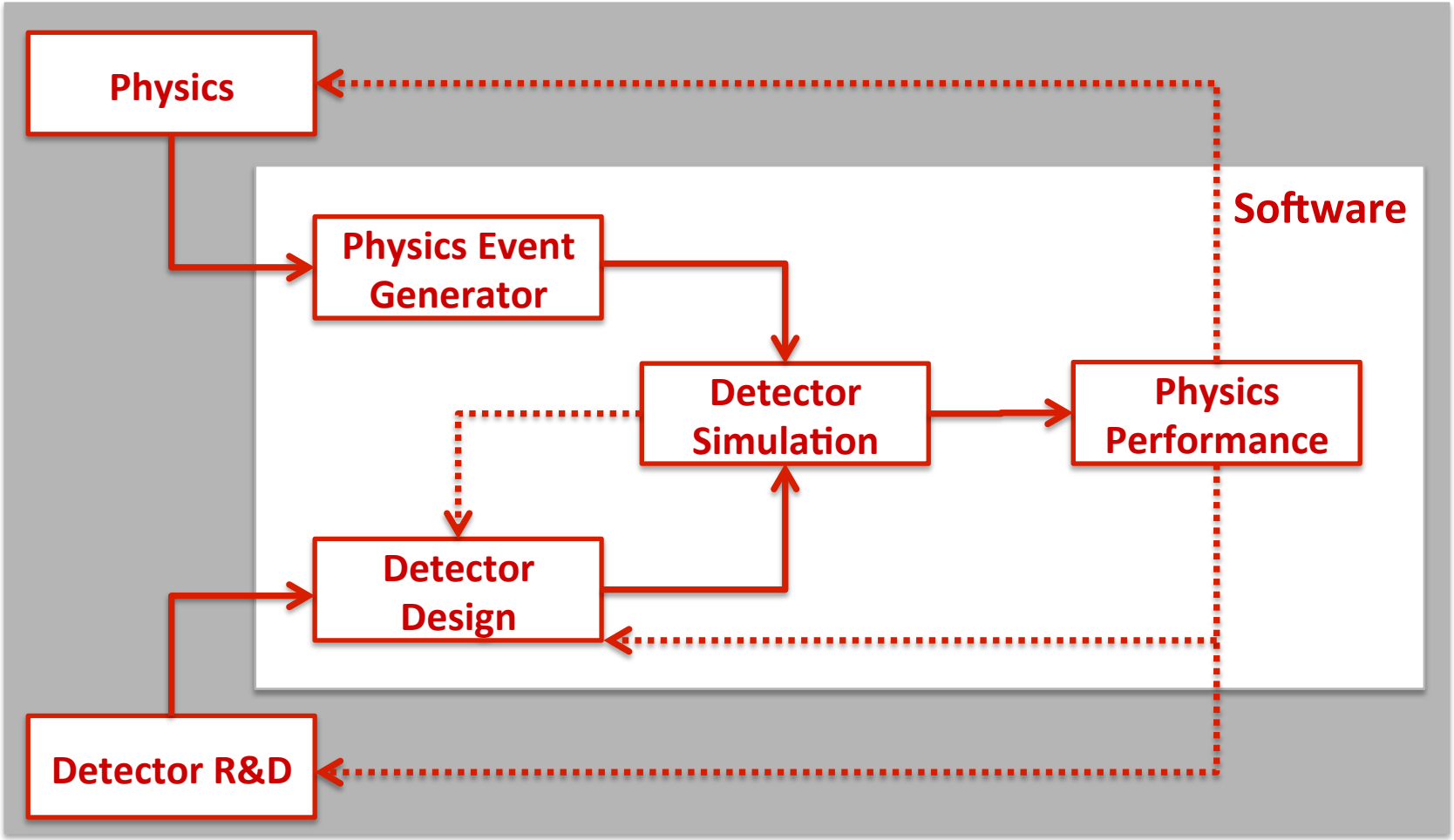
Generic R&D for an EIC:
Developing Analysis Tools and Techniques for the EIC

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)

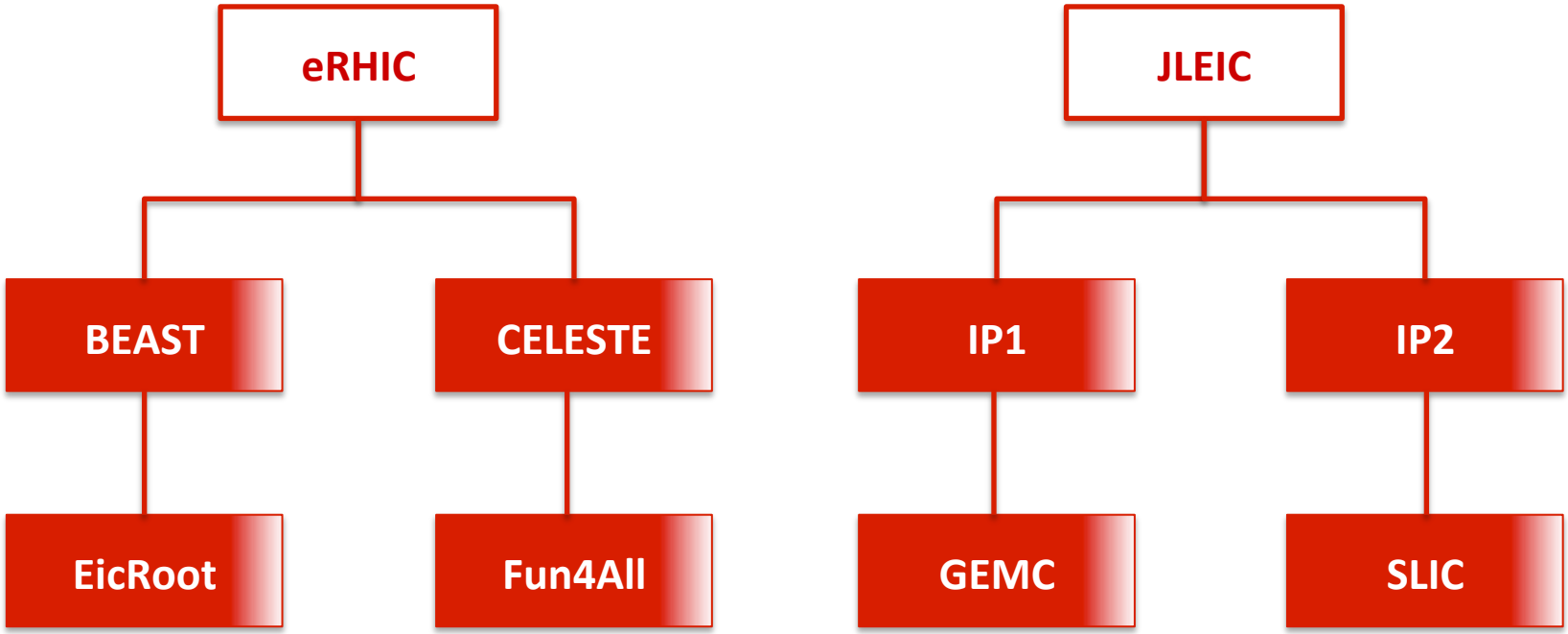
EIC Generic R&D Meeting
July 6th – July 7th 2016

Jefferson Lab
Thomas Jefferson National Accelerator Facility

EIC R&D and software development



Existing software frameworks for the EIC



Building on existing EIC software:

- build forward-compatible interfaces between existing frameworks / tools
- identify common tools and improve them (e.g. MCEG)
- add tools that are forward-compatible with existing frameworks

Forming a software consortium for the EIC

September 2015 EIC Software Meeting
Workshop organized by Elke-Caroline Aschenauer and Markus Diefenthaler
https://www.jlab.org/conferences/eicsw/
review of existing EIC software frameworks and MCEG available for the EIC

January 2016 Generic R&D Meeting: **LOI for Software Consortium**

> **Report** "A robust software environment, compatible with the existing software frameworks, is very important for the development of the physics case for the EIC."

March 2016 Future Trends in NP Computing
Workshop organized by Amber Boehnlein, Graham Heyes, and Markus Diefenthaler
https://www.jlab.org/conferences/trends2016/
discussion of computing trends, e.g., Big Data, machine learning, Exascale Computing
incubator for ideas on how to improve analysis workflows in NP

July 2017 Generic R&D Meeting: **Proposal for Software Consortium**

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

building up on existing documentation:
https://wiki.bnl.gov/eic/index.php/Simulations and related pages

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

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

Workshop review of MC generators for EIC

- **MC generators for ep processes:**
  - several excellent MC generators available
  - but essential pieces are missing:
    - MC generator for (un)-polarized p_T dependent physics
    - radiative corrections not integrated in many generators, required as physics and detector smearing don't factorize

LEPTO (DIS)

PEPSI (polarized DIS)

PYTHIA 6

PYTHIA 8

GMC_TRANS (SIDIS)

CASCADE (ep + pp, p_T)

MILOU (DVCS)

PARTONS (exclusive)

DJANGOH (radiative effects)

+ many more generators

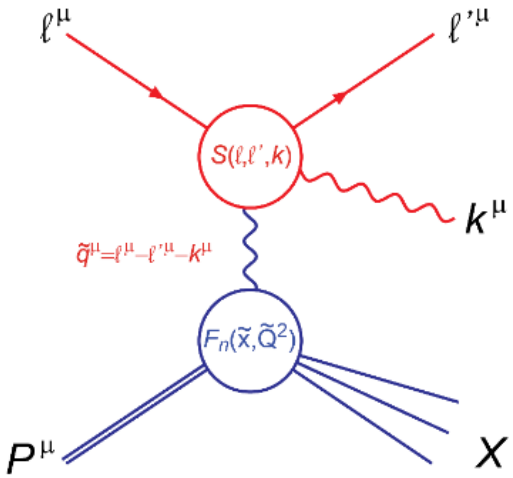
- **MC generators for eA processes:**
  - significantly worse situation than ep
  - need a SIDIS generator w/o saturation
  - need CASCADE like eA generator

PYTHIA + DPMJET

DJANGOH (radiative effects)

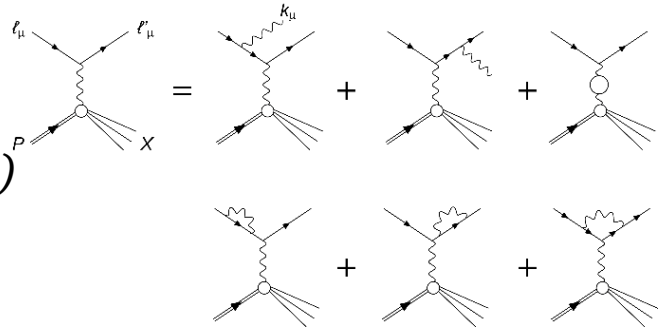
SARTRE (diffractive, DVCS)

Radiative corrections



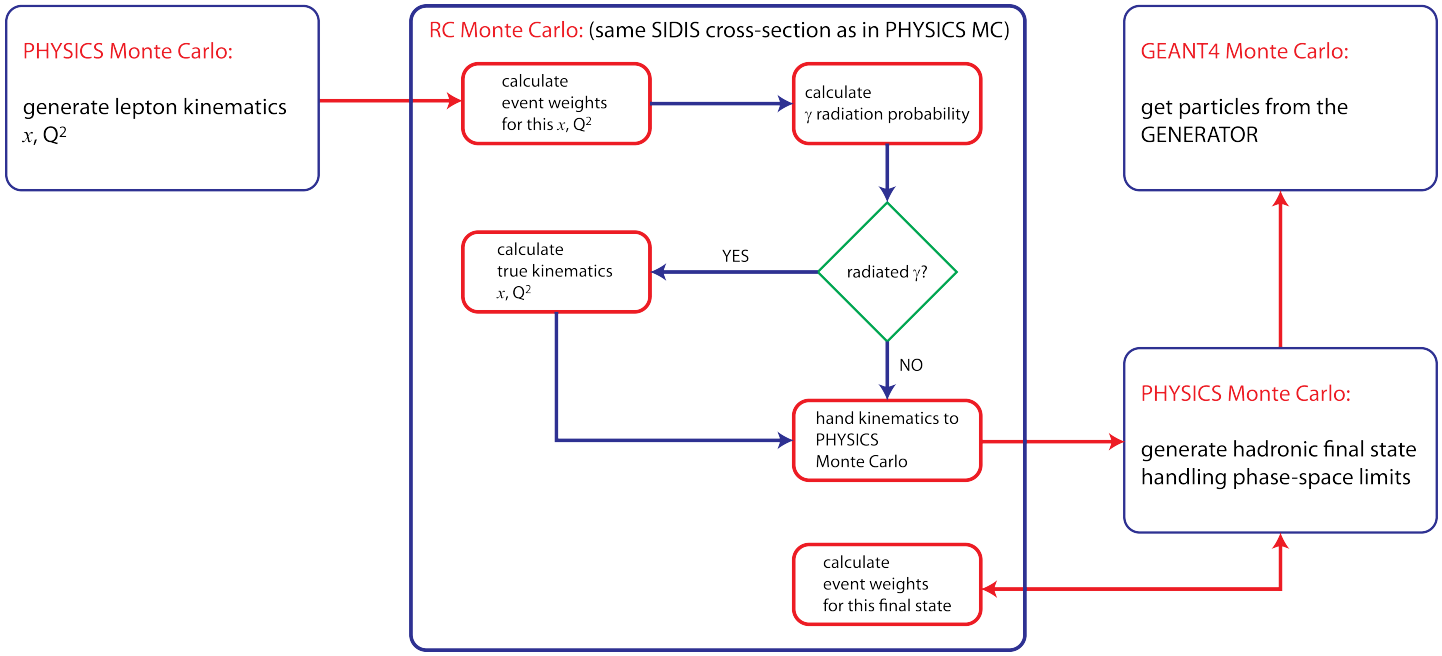
$$Q^2 = -(\ell - \ell' - k)^2$$

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



- 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.

Validation of Geant4 for the EIC

Collaboration with Geant4/SLAC team (Andrea Dotti)

Goal:

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

FY17 Deliverables:

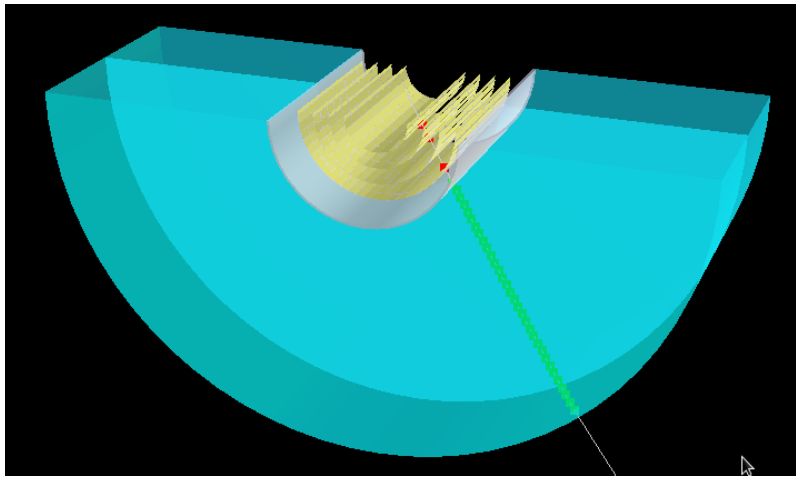
- Review the current validation strategy of Geant4 identifying what are the EIC specific interests that are currently not covered. We will identify which of the data-sets could be used to extend the Geant4 validation test-suite particularly fit to EIC energy/interactions. We plan to feedback these findings to the Geant4 Collaboration and eventually collaborate with experts to address these issues.
- Extend the validation applications to address the EIC specific needs:
  - Develop simulation and analysis macros for *SimplifiedCalorimeter* and *ProcessTest* to generate and study the interactions of most interest for EIC
  - Evaluate a GDML-based simplified setup, to be used with the *HepExpMT* application, to measure CPU time-consumption of alternative physics list

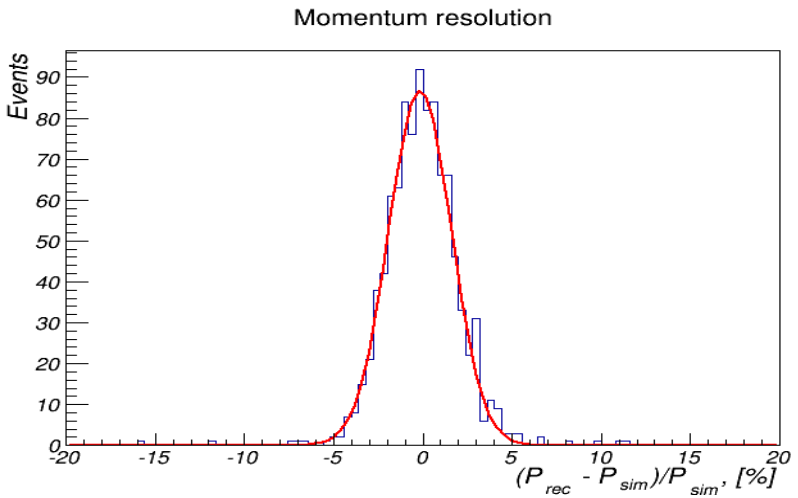
Unified track reconstruction library

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 θ=75°?





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

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

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

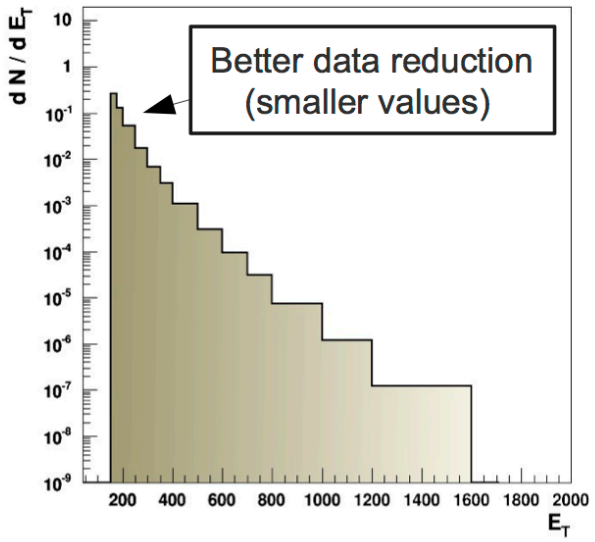
Google's Protocol buffers

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



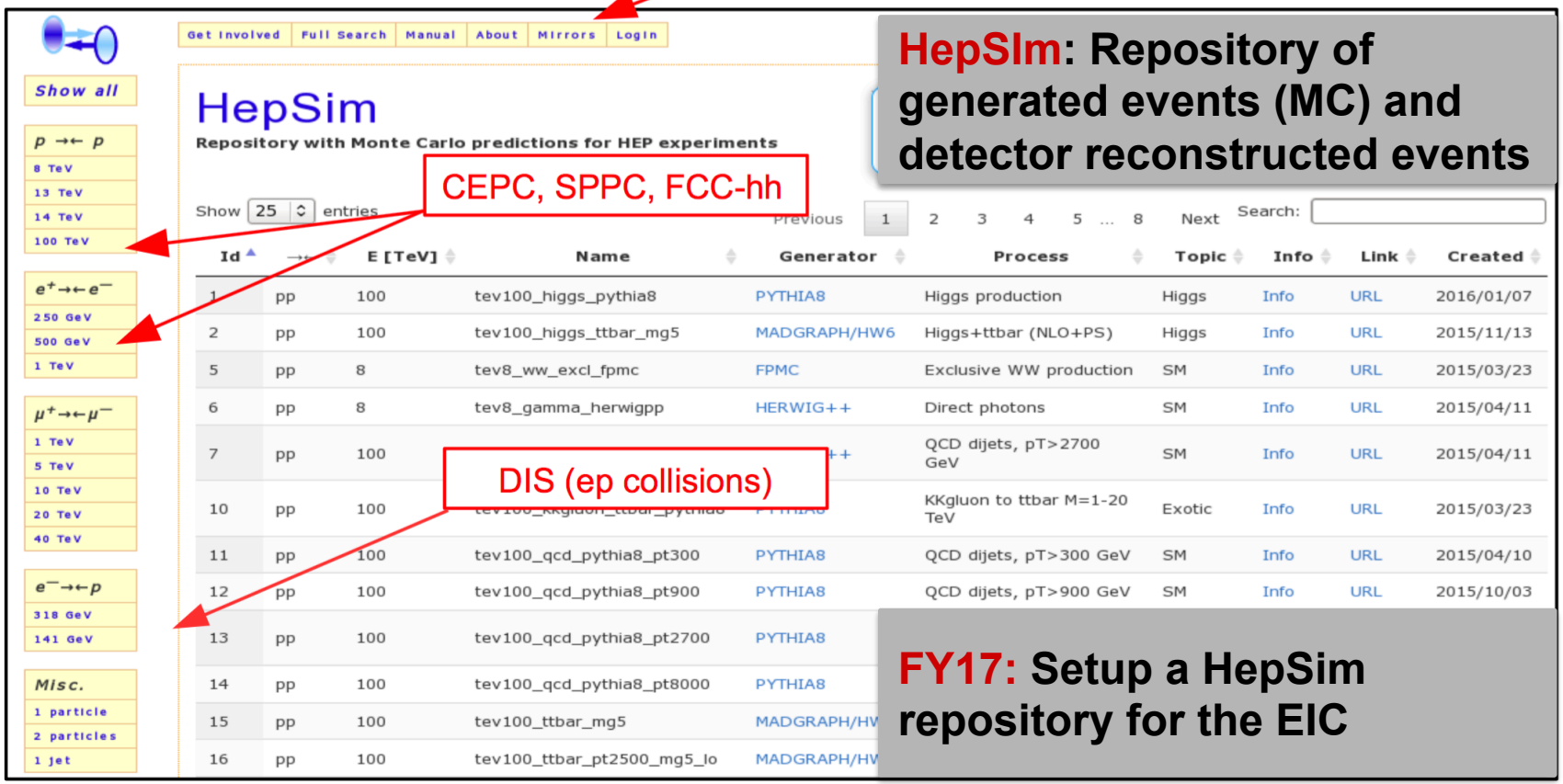
compression strength keeping precision of constant

HepSim repository for the EIC

HepSim simulation uses ProMC

http://atlaswww.hep.anl.gov/hepsim/

NERSC, CERN mirrors



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

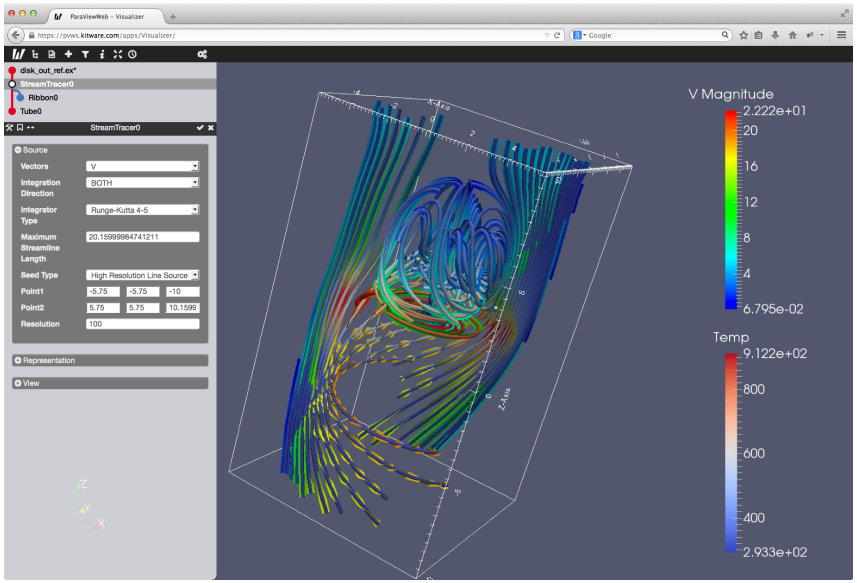
FY17: Setup a HepSim repository for the EIC

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

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

Funding request for FY17

Travel budget allow proponents to meet and to work together on key tasks, invite visiting scientists that are essential to the R&D effort support proponent's travel to annual workshop	USD 30,000
Undergraduate student projects	USD 20,000
Funding request for FY17	**USD 50,000**

Early investment in the development of software tools