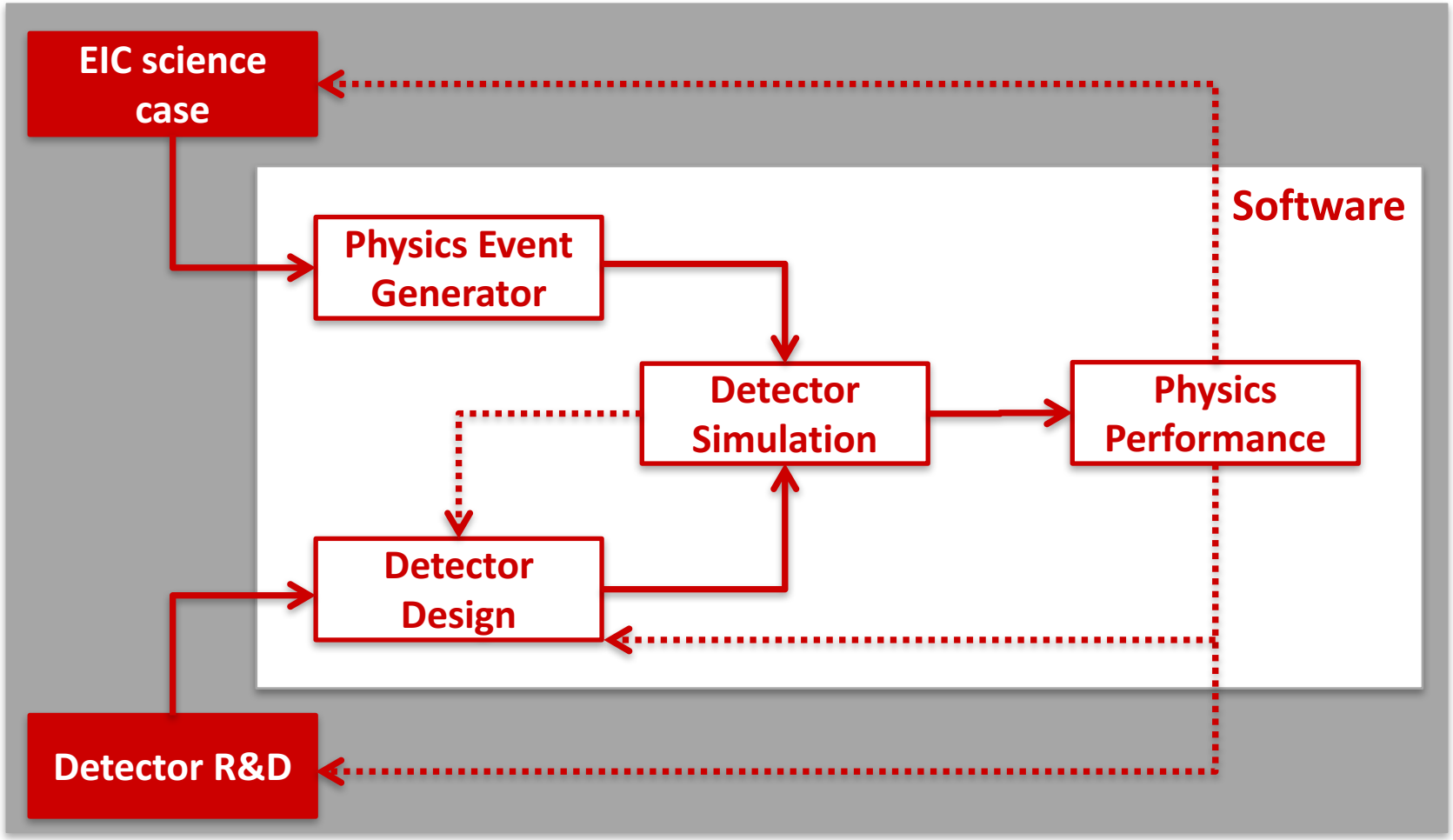
Generic R&D for an EIC: **eRD20 - Developing Simulation and Analysis Tools for the EIC**

Markus Diefenthaler (mdiefent@jlab.org)

on behalf of the EIC Software Consortium (ESC)

EIC R&D and software development



EIC Software Consortium (ESC)

ANL (2+**1**) W. Armstrong, S. Chekanov, **D. Blyth**

BNL (2+**2**) E.-C. Aschenauer, A. Kiselev (**co-PI**) , **J. Lauret**, C. Pinkenburg

Fermilab (1) S. Prestel

INFN Trieste (1) A. Bressan

Jefferson Lab (1+**3**) MD (**co-PI**) , **D. Lawrence**, **D. Romanov**, **M. Ungaro**

SLAC (1+**1**) **M. Asai**, A. Dotti

William & Mary (1) **W. Deconinck**

Joined during FY17 or for FY18

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

New analysis environments

User centered design

- understand the user requirements first and foremost
- engage wider community of physicists, whose primary interest is not computing, in software design
- make design decisions largely based on user requirements

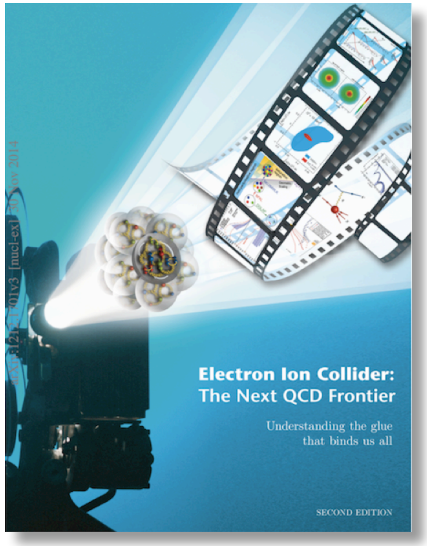
Future compatibility (both hardware and software)

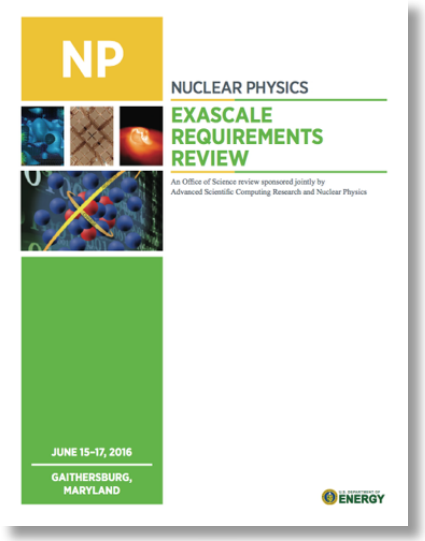
- most powerful future computers will likely be very different from the kind of computers currently used in NP (Exascale Computing)
- structures robust against likely changes in computing environment
- apply modular design: changes in underlying code can be handled without an entire overhaul of the structure

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

Towards a computing vision for the EIC





EIC

- extremely broad science program
- strong **interplay theory – experiment**
- **EIC Software Consortium** (EIC Generic R&D Program)

Computing requirements

- flexible, modular analysis ecosystem
- interplay of data analysis, MC calculations, and Lattice QCD effort

Lessons learned from LHC

- computing central to success of scientific goals
- complexity of analysis ecosystem limits time on physics analysis
- strong role of machine learning

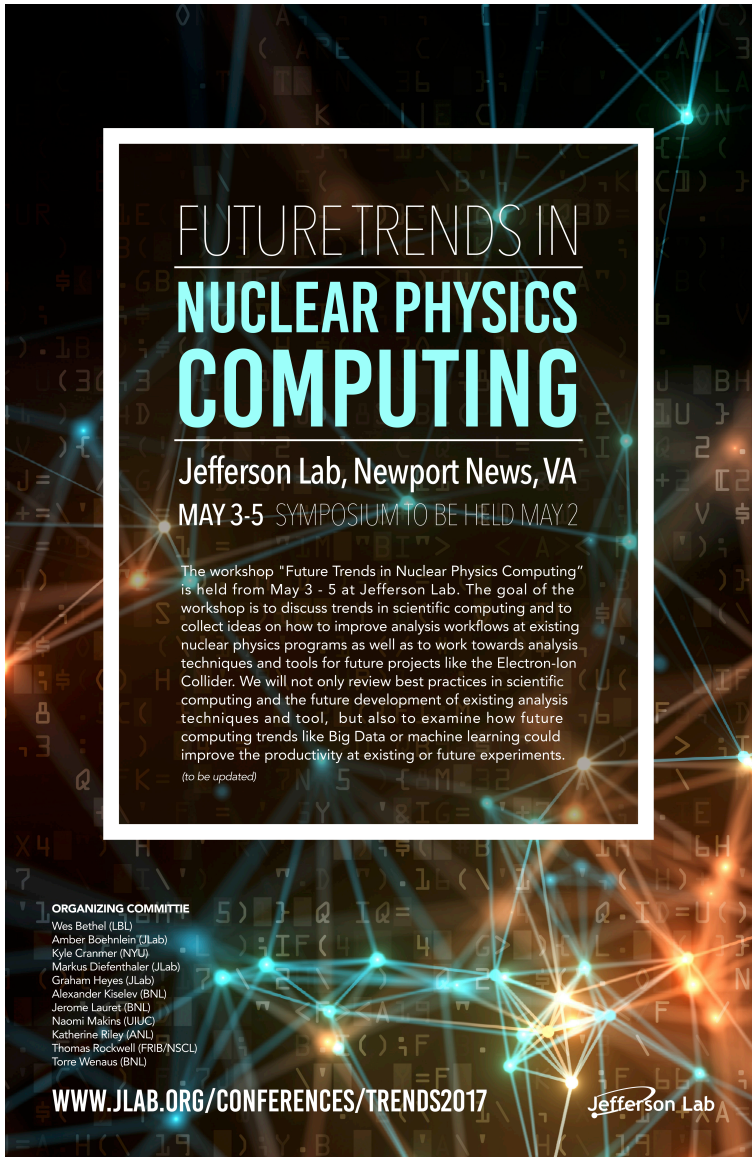
Era of Exascale Computing

- changing the paradigm for I/O, storage and compute
- high-precision non-perturbative QCD measurements (MC, Lattice QCD)

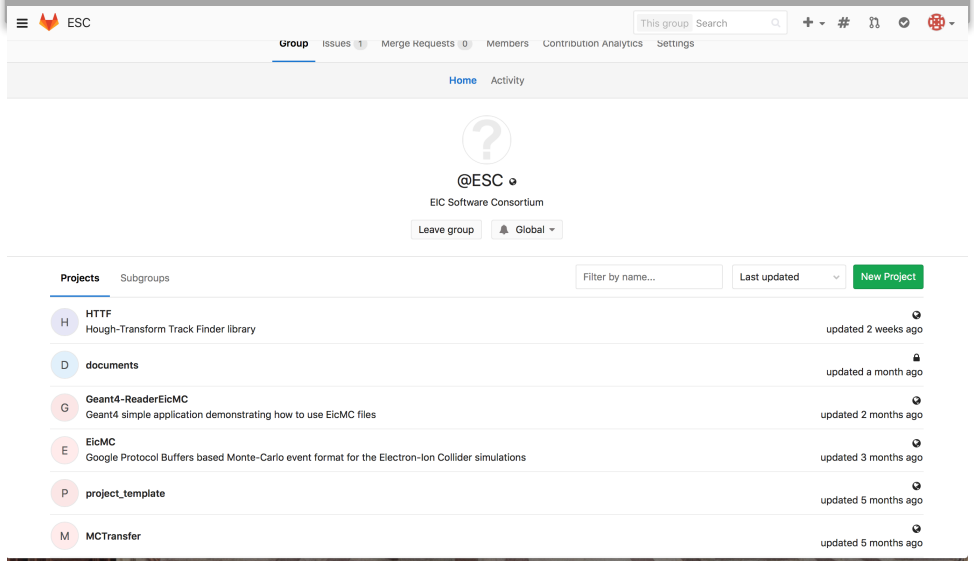
Section

ESC projects (FY17 and FY18)

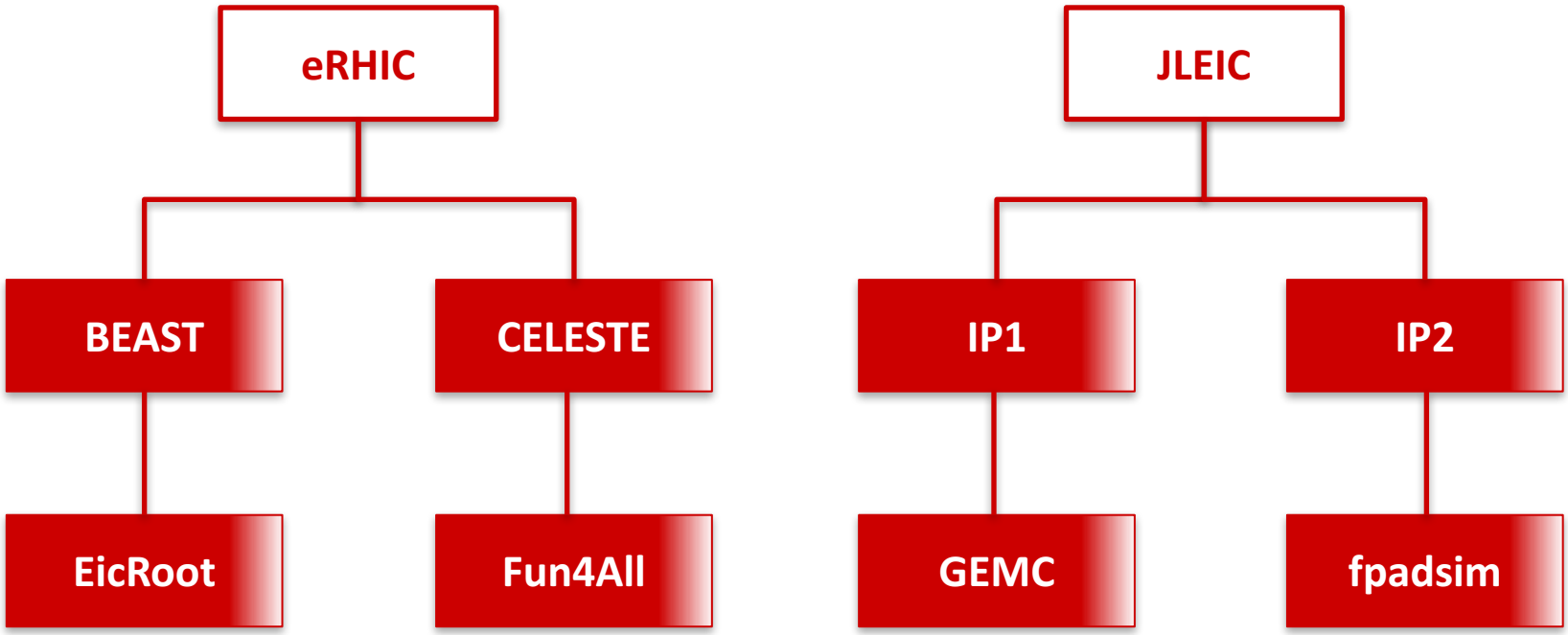
Organizational efforts



- ESC involved in **workshop** on "**Future Trends in NP computing**"
  - 76 participants
  - symposium with forward looking topics (D. Geesaman (ANL), M. Savage (INT), S. Hoeche (SLAC), R. Ent (JLab)
  - workshop report with recommendations for moving forward in progress
- **GitLab repository** setup for ESC software and documentation: **https://gitlab.com/ESC/**



Existing software frameworks



ESC Review

- EicRoot, Fun4All , GEMC, and fpadsim 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:

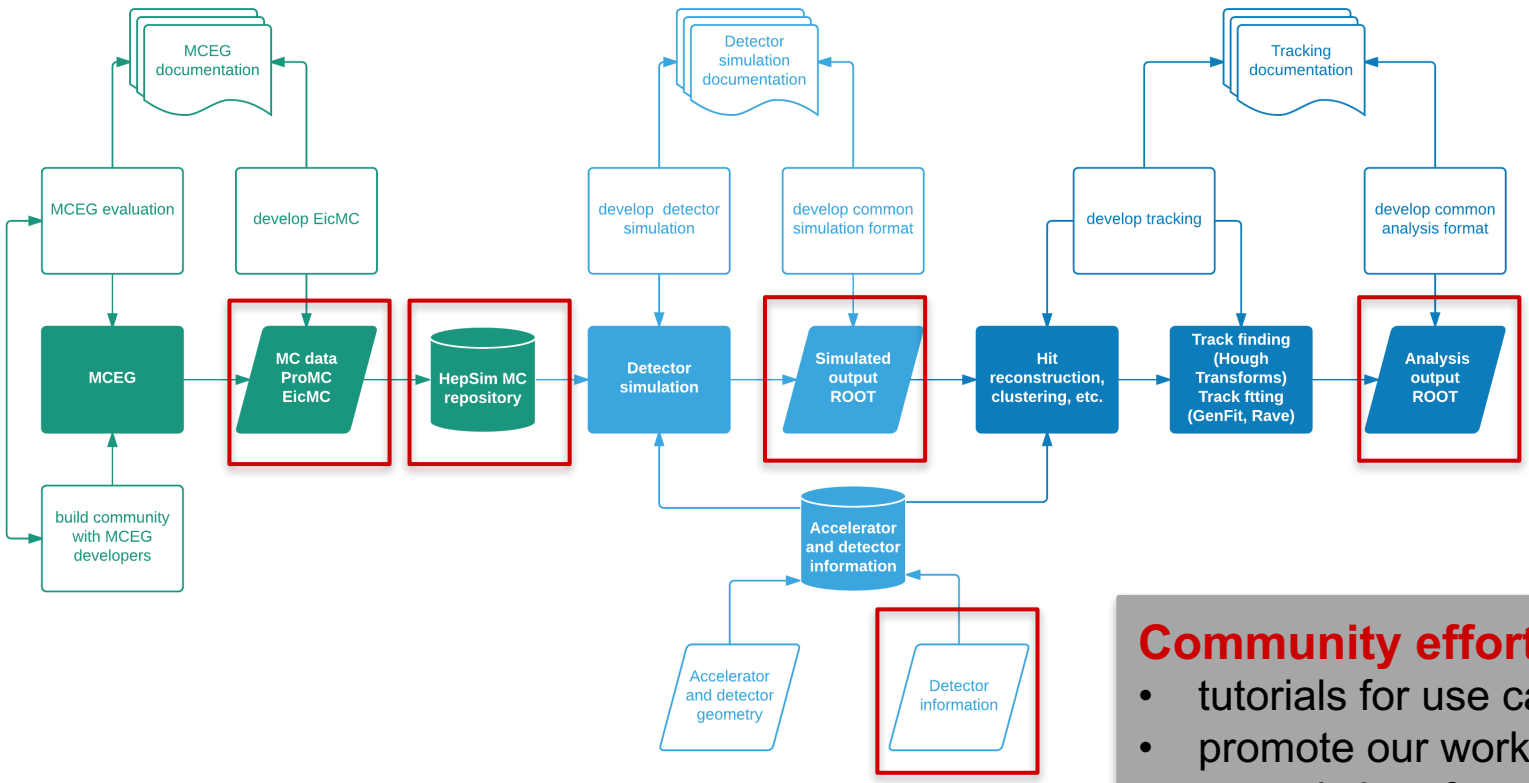
- open access to physics simulations or interface to MCEG
- open access to accelerator specifications
- open access to detector information and simulation
- documentation

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

ESC contributions

focus on **common interfaces** between frameworks



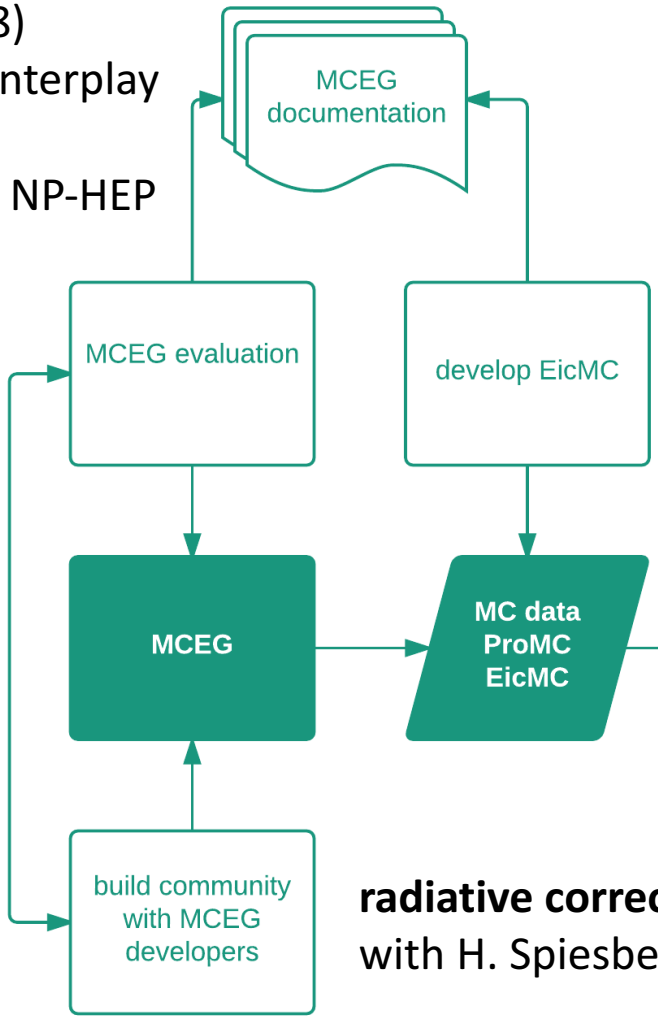
Community effort

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

Monte Carlo simulations

EIC MCEG initiative (FY18)

- emphasize on strong interplay experiment –theory
- connect MCEG efforts NP-HEP



ESC MC format (FY17, FY18)

- self-descriptive data format
- HPC compatible

eicsim01.jlab.org

radiative correction library (FY17, FY18)
with H. Spiesberger (HERA)

EicMC – self-descriptive MC files

by-product of ProMC import incorporation into EicRoot framework

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

tested with Geant4

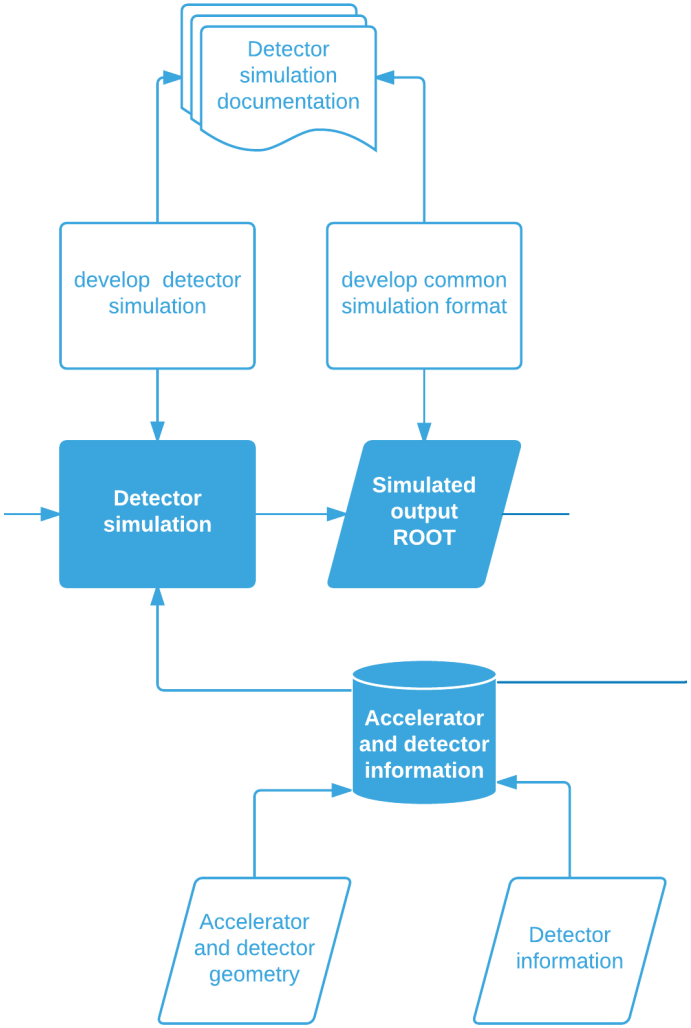
EicMC: feature-rich, flexible, self-descriptive

- no external dependencies on the user (input) side other than **Google Protobuf** library
- 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)
- convenient user interface
- performance (file size, speed) similar to or better than ProMC and ROOT equivalents

FY18

- comparative review of EicMC and ProMC, ROOT, and **HDF5**
- decide on ESC MC format
- adaptation to HepSim repository and simulation tools

Detector simulations



Geant4 - ESC Collaboration (FY17, FY18)

- Makoto Asai: liaison between Geant4 International Collaboration and ESC
- EIC physics list and validation, tuning and extension (energy range is different from LHC)
- library to exchange geometry, sensitivity and hits among different frameworks

Validation of Geant4 for 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
- 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

Unified track reconstruction

Modular tracking software (FY17, FY18)

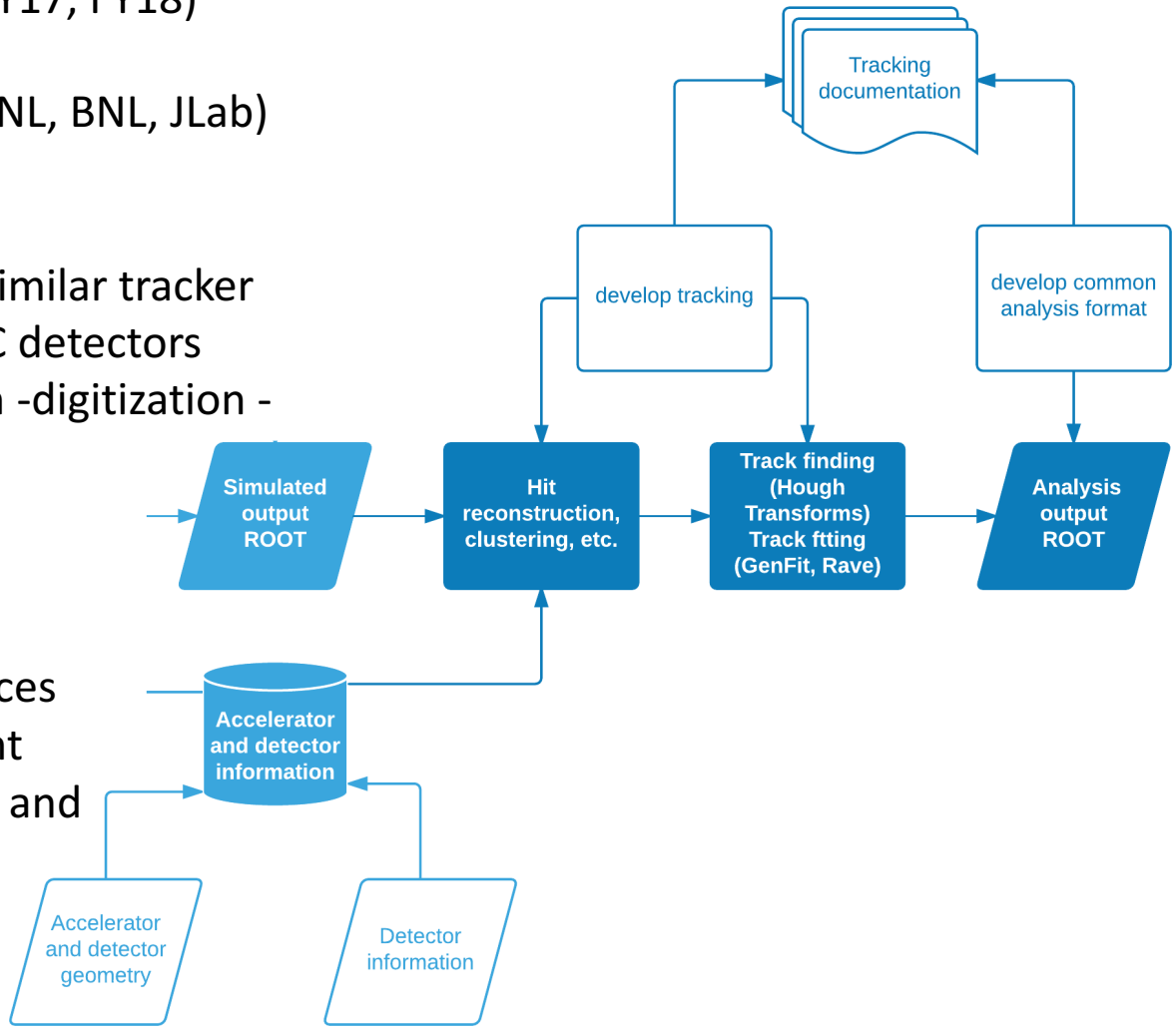
based on EIC tracking tools (ANL, BNL, JLab)

FY17 feasibility study

- similar requirements and similar tracker outline for all proposed EIC detectors
- similar dataflow simulation -digitization - track reconstruction

FY18 start development

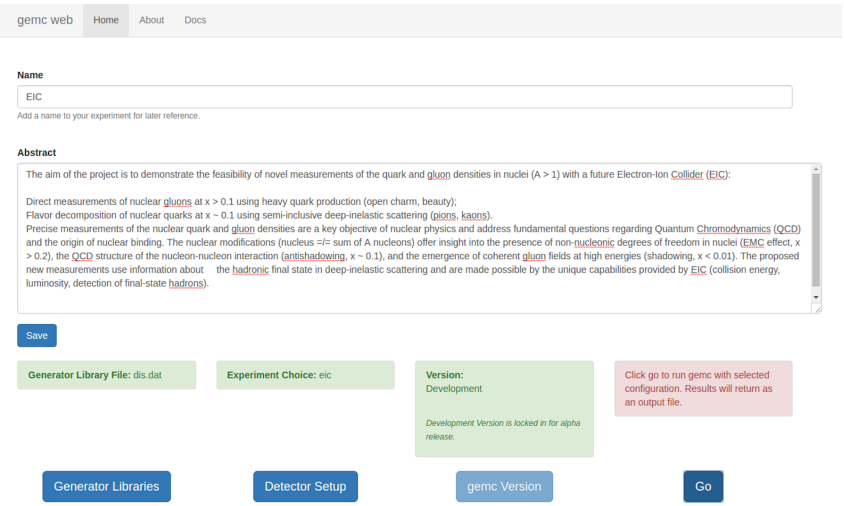
- define libraries and interfaces
- setup sandbox environment compatible with ANL, BNL, and JLab frameworks



Exploring user-centered design

Simulations on the web

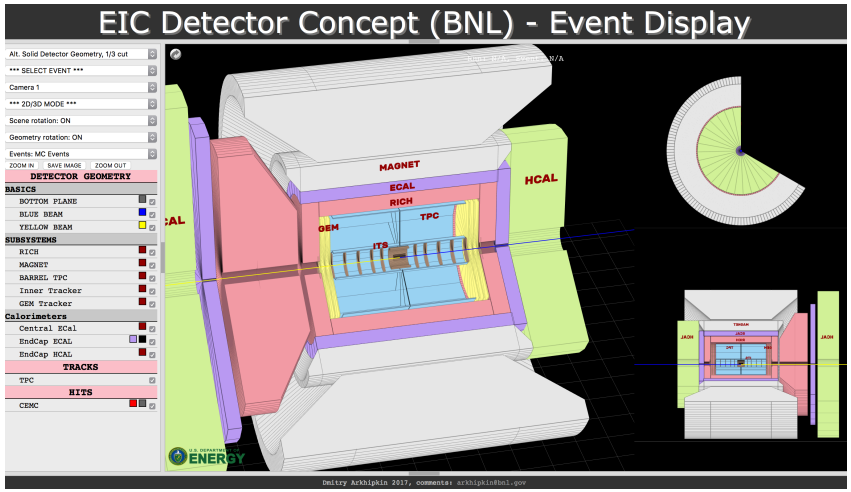
Sam Markelon (UConn)



Universal web-based event display

- validation of the EIC simulations
- comparison of different detector designs using an unified approach
- public outreach

STAR software group effort



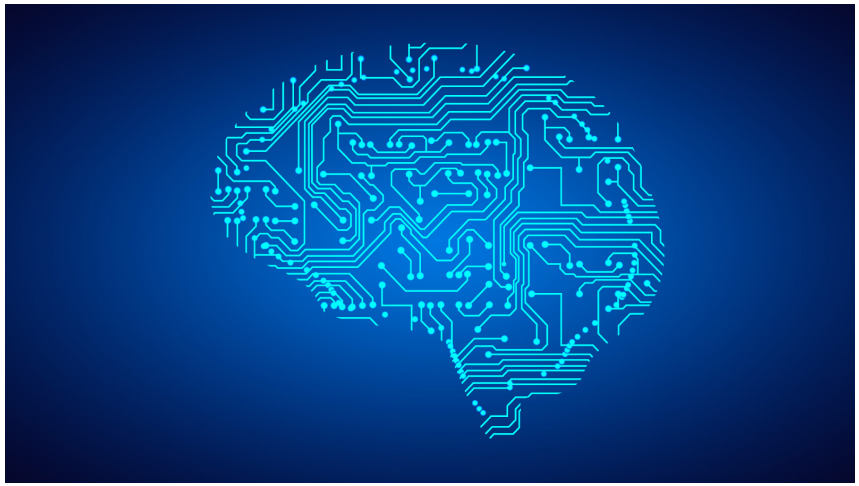
Other initiatives

High-performance computing

FY18

- prepare EIC HPC projects
- begin dialogue with ASCR-operated computational science user facilities

Machine learning



FY18

- Document selected examples for using (D)NN at the EIC, e.g., RICH

Status of FY17 deliverables

List of deliverables in FY17 proposal. Most projects will continue in FY18.

FY17

Interfaces and integration

- start the development of a library for simulating radiative effects ✓
- work towards a common geometry and detector interface ✓
- work towards an unified track reconstruction ✓

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 **(started)**
- build interfaces to forward compatible, self-descriptive file formats ✓

FY17

Organizational efforts with an emphasis on communication

- build a community website **(started)**
- organize software repositories dedicated to the EIC ✓
- organize a workshop ✓

Section

eRD20: Recommendations and funding

Advisory Committee: Recommendations

- **The committee suggest the group investigates the required event size and data size to evaluate the EIC computing needs.**
  - We would use photoproduction as an upper limit for the event rate at the EIC. The photoproduction cross section is 10 microbarn. At a luminosity of 10^34 per cm^2 and s, this corresponds to an event rate of 100KHz.
  - Event size will be 10-100KB.
  - This corresponds to a triggerless data rate of 1-10GB per second and after some trigger 3.6TB per hour or 86.4TB per day.
  - Assuming the EIC would effectively run 20 weeks a year, this would correspond to a 12 PB per year.
  - Comparison to LHCb: Online computing farm processes 250 PB / year. 5 PB/year collected. Triggerless-readout for LHC Run 3 (2021-2023):, process 5 TB/s in real time on the CPU farm.

- **The open, broad approach the collaboration is taking to develop the software infrastructure for the EIC is welcomed.**

- **The group is urged to be cognizant of existing, well-supported efforts that can be deployed for the EIC to ensure optimal use of existing resources.**
  - Geant4-ESC collaboration
  - “Future Trends in NP Computing” workshop
    - Common infrastructure components
    - Software portability and reusability
  - Reached out to HEP Software Foundation (A. Boehnlein, T. Wenaus) and HEP Center of Computing Excellence (S. Habib)
  - Will reach out to ASCR-operated computational science user facilities.

Budget request

- We need to meet and talk in person in order to make progress.
- We should have enough funds to attend relevant conferences …
- … and to hire students to help with real coding work.

	FY17 ($)	FY18 ($)
Travel	~40k	70k
Students	~15k	30k
Total	55k	**100k**

-20% scenario:

- Only 3 face-to-face meetings next year
- Strictly limited computing conference travel

-40% scenario:

- Only 2 face-to-face meetings next year
- No travel to the computing conferences and workshops
- Very limited opportunities to attract (and hire) additional manpower

Summary

- **ESC** initiative from 16 scientists from 7 institutions

- Community effort for EIC User Group
- **Global objectives**:
  - Interfaces and integration for existing EIC software
  - Planning for the future with future compatibility
  - Organizational efforts with an emphasis on communication

- **FY18 focus**
  - MCEG and self-descriptive MC file format (EicMC, HDF5)
  - Geant4 validation
  - Unified tracking with universal geometry and detector interface
  - Exploring user-centered designs via web interfaces
  - Start HPC and (D)NN initiatives