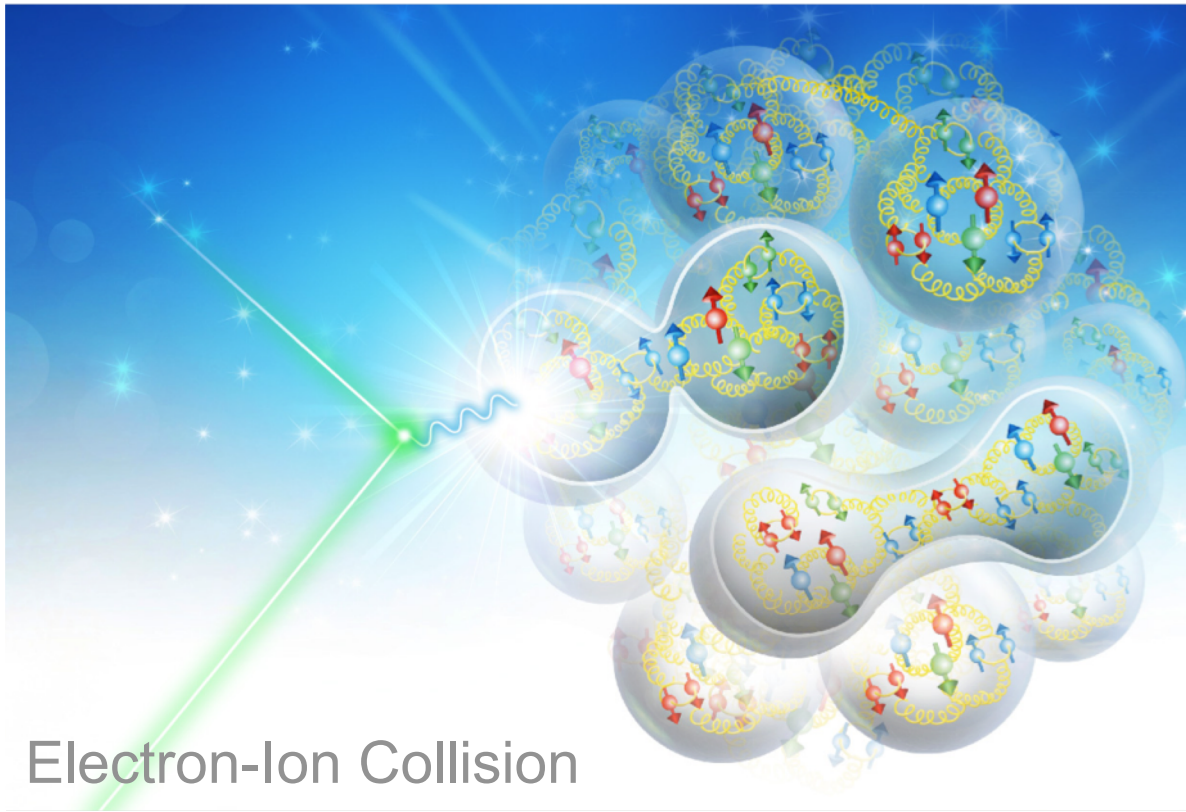
EIC Software Requirements

EICUG Software Working Group



Electron-Ion Collision

Greenfield simulation(framework)

Definition

Greenfield event-processing software := community-wide project on event-processing software freeing us from the legacy of existing options while leveraging everyone's experience. The project will define requirements and build up the simulation toolkit / framework on these requirements. Input by the wider scientific and software & computing communities is encouraged.

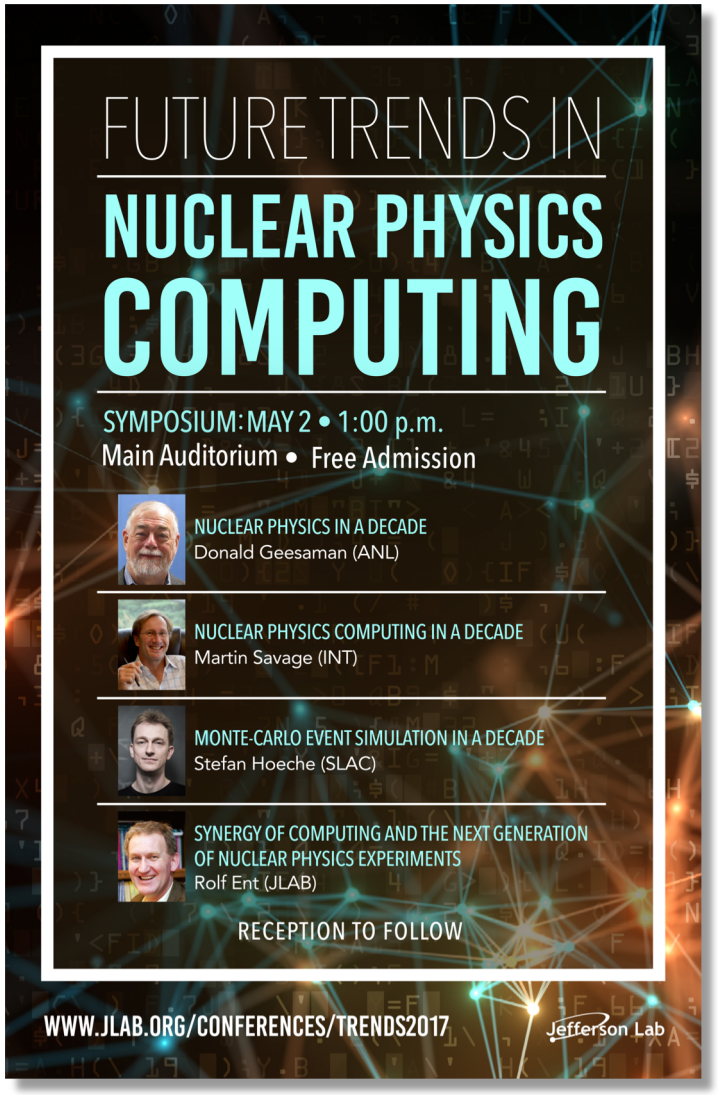
Approach

1. Define **requirements** and write them down.
2. **Study** existing implementations and **consult** with wider scientific community and software developers.
3. **Agree** on design.
4. **Implement** our design.

Initial requirements

EIC offline (and online?) **Software**

Data & Analysis (from TRENDS2017)



Our Goal All scientists of all levels, worldwide, should be enabled to actively participate in the nuclear physics data analysis

- regardless what experiment the data came from
- easy moves from n-dimensional data to (n-1)-dimensional data analysis

To achieve this goal, we must develop analysis toolkits using modern and advanced technologies while hiding that complexity.

- resolving this tension means putting a priority on the user experience and functionality
- user-centered design

We must put stress 'Data' as much as 'Analysis'

- experimental Data must be open access, readily accessible and in a self-describing formats
- data must be trackable and linked to publications

Requirements from EIC User Group

Charge for EICUG Software Working Group

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.

Monte Carlo event generators (MCEGs)

Requirement (charge) MCEG for all physics processes that make up the EIC science case as articulated in the white paper and as further developed by the EICUG.

Related requirements:

- requires organization with theory community in NP and HEP (where most of the MCEG are being developed)
- requires merging of QED and QCD effects in MCEGs
- CPU usage will be dominated by MC (mainly for detector simulation, but the higher MCEG complexity will also impact on resources): Requires organization on how to handle MC, how to keep track of MC data, and how to validate MC.

Detector simulations

Requirement Geant4

Discussion on fast simulations beyond eic-smear

- Are these needed at all?
- these detectors specific, or a common system could be developed?
- What is the needed speedup?
- If is a factor 2-10 technologies already exist (e.g. gflash or biasing in G4), if more one needs to think how to develop these.

Reconstruction

Requirement Flexible. We are still discussing detector concepts.

Related requirements

- well-defined (event) data model
- thread-safe (see also slide 13 on concurrency)

Benchmarks and validation

Requirement (charge) quantitative assessment of measurement capabilities and their physics impact

Related requirements:

- benchmark processes (→ MCEGs)
- validation tools (intrinsic part of simulation tools?)
- benchmark workflow (intrinsic part of simulation workflow?)

Accessible

Requirement (charge) accessible to the EICUG

Related requirements:

- repository
- software distribution: container, cvmfs (also related to user-centered design)
- **User-centered design**:
  - “Experience in HEP has been that as the complexity and size of the experiments grew, the complexity of analysis environment grew. The time spent by physicists in dealing with the analysis infrastructure rather than on doing physics also grew. Currently, the anecdotal data is that a typical LHC student of post-doc spends up to 50 % of his/her time dealing with computing issues. This seems to most of us to be too much. One key idea in beginning to think about the EIC analysis environments is to understand the user requirements of the analysis environment first and foremost. This requires the engagement of the wider community of physicists whose primary (and perhaps not even secondary) interest is not in computing. All design decisions must serve these requirements.” (Future Trends in Nuclear Physics Computing in 2016)
  - “In ATLAS, some postdocs spend even more than 50% of their time on it, on behalf of a larger analysis group. I don’t believe the average is as high as 50%. An important requirement is to ensure the framework and infrastructure support the mechanisms to keep that number low, like managed train production of analysis derivations. I think the new/recent developments in capturing analysis processes in reproducible and portable containers/workflows will also be a factor in keeping this number down.” (Torre)

Consistency

Requirement (charge) consistent to the EICUG

Related requirements:

- simulation tools to be used on all detector concepts
- geometry exchange

Reproducibility

Requirement (charge) reproducible to the EICUG

Related requirements:

- documentation for simulation and validation tools and the related workflow
- unit tests
- self-descriptive data
- self-descriptive analysis
- data / record compatibility after changes
- data preservation build in

Concurrency

Requirement EIC software should be designed with concurrency in mind from the very beginning

Modularity

Requirement "We need to think about structures that are robust against likely changes in computing environment both hardware and software. Modularity is important, and careful thought has to be given to how exactly to modularize so that changes in underlying code, database structure, etc. can be handled without an entire overhaul of the structure. This is important also in that most powerful future computers will likely be very different from the kind of computers that are currently most often used in HEP and NP." (Future Trends in Nuclear Physics Computing in 2016)

Related requirements:

- define interfaces and keep them simple (**discussion** or not?)
- use flat files (**discussion** or not?)
- modular also for simulation of detector effects, digitization, and event reconstruction
- requires geometry exchange between simulation, digitization, and the event reconstruction
- support for heterogenous hardware

Markus Diefenthaler

mdiefent@jlab.org

Discussion

Homework

- write up presentation and share it on our repository for detailed comments
- make it is available on repository