BROOKHAVEN NATIONAL LABORATORY HSF Jefferson Lab

FUTURE TRENDS IN NUCLEAR PHYSICS COMPUTING

Hosted by Center for Frontiers in Nuclear Science (CFNS)

Organized by Amber Boehnlein (JLab), Benedikt Hegner (CERN, HSF), Brad Sawatzky (JLab), Graeme Stewart (CERN, HSF), Jan Bernauer (SBU, CFNS), Kolja Kauder (BNL), Markus Diefenthaler (JLab), Nathan Baltzell (JLab), Ofer Rind (BNL), Paul Laycock (BNL), Torre Wenaus (BNL)

Report from previous workshops by Markus Diefenthaler (Jefferson Lab)

History

Future Trends in Nuclear Physics Computing in 2016

76 participants

Motivation

- **New experiments coming up** (CEBAF 12 GeV, EIC, FRIB, sPHENIX)
- **Advance nuclear science** *“The purpose of computing is insight, not numbers.”* Richard Hamming (1962)

Goals

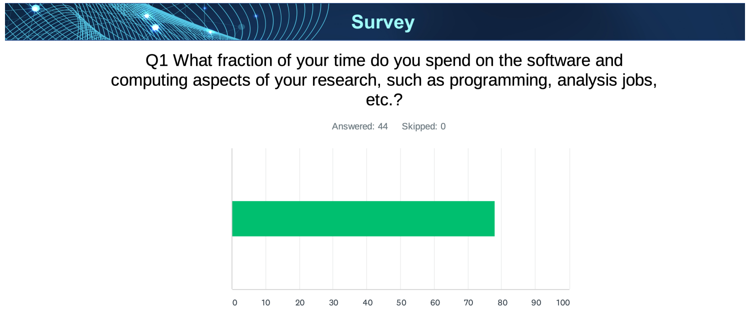
- **Examined computing strategy at a time horizon of ten years**
  - E.g.: Concurrent and parallel computing, heterogenous computing
- **Defined common vision for NP computing**
  - E.g.: Streaming Readout
- **Recommended future directions for development**
  - E.g.: Data science, User-centered design

Website

- https://www.jlab.org/conferences/trends2016/

An Emerging Vision

Software & computing are an integral part of our research



Survey among Nuclear Physics Ph.D. students and postdocs in preparation of "Future Trends in NP Computing" in 2020

- **Goal** We would like to ensure that scientists of all levels worldwide can participate in EIC analysis actively.
- **User-Centered Design**: To achieve this goal, we must develop simulation and analysis software using modern and advanced technologies while hiding that complexity (turnkey) and engage the wider community in the development.

Rapid turnaround of data for the physics analysis and to start the work on publications

- **Goal**: Analysis-ready data from the DAQ system.
- **Compute-detector integration** with AI at the DAQ and analysis level.

History

Future Trends in Nuclear Physics Computing in 2017

74 participants

Themes

- **Resource management and the interplay of I/O, compute and storage**
- **Machine learning for enhancing scientific productivity and appropriate task based approaches**
- **Software portability, reusability and common infrastructure components**
  - Including: The HEP Software Foundation and NP

Website

- **https://www.jlab.org/conferences/trends2017/**

History

Future Trends in Nuclear Physics Computing in 2020

Focus on the Nuclear Physics Software & Computing community

- Identify what is unique about our community
- Discuss how we could strengthen common efforts
- Chart a path for **Nuclear Physics Software & Computing** for the next ten years

Website

- https://indico.bnl.gov/event/9023/

Workshop Report

Summary by Paul Laycock and Torre Wenaus

Common Scientific Software – The keys to success

- **The team is the most important** Do not separate development and operations, both ACTS and Rucio benefited from experience with developing and operating a worse software package, crucial experience. Developers keen to use modern software paradigms, open-source and open-minded, proactively searching out best practice and adopting it.
- **The project** Clear, well-focused short-term goals are important, grounded in real-world deliverables. Aligned with the long-term plan of building something sustainable and designed to be used by outside collaborators.
- **The management** Accept that the long-view takes longer to deliver the short-term product, manage expectations of the collaboration and funders to ensure the team have sufficient time and space to succeed.

Scientific software careers need support

- Recognition, encouragement and reward: Need to make software citations a priority
- Career paths of Research Software Engineers (RSE) need to be supported and not only at the labs

NP Software - should NP participate in HSF or build its own organization?

- Pros and cons, the balance of opinion favored NP participation in HSF. HSF is a do-ocracy, active participation will yield the biggest rewards.
- NP often has small groups developing solutions in-house, work with this reality.

The Role of Data Centers in Scientific Discovery

Summary by Graham Heyes and Ofer Rind

- **Evolution of Data Center Infrastructure**
  - Use of disruptive technologies - AI/ML and coprocessors (GPU, FPGA)
    - Tension between HTC and HPC in resource planning
  - Role of tape and cloud storage
  - Integration of distributed computing resources, where do LCFs fit in?
    - Federated authentication/authorization
  - Evolution of high bandwidth networking
- **Long term Data and Analysis Preservation (DAP)**
  - Importance of preserving metadata and code alongside raw data
  - Importance of documentation and proper choice of tools
  - Importance of building DAP into the infrastructure and policies at an early stage
- **Containerization as mechanism for DAP as well as for operating on distributed resources**
- **Challenges in developing common, relevant computing benchmarks for NP**
- **Interactions between facilities and stakeholders**
  - Need for improved two-way communication
    - Educating users on how to use the data center; communicating stakeholder needs to the facility
  - Embedding facility personnel within experiments and vice versa
  - Connecting ops personnel with CS researchers (at labs, at universities, etc.)
- **Promoting computing within the NP community**
  - Support for education and training in software development
  - Provide career paths and funding that allow for and value software development

Unique Software Challenges for Nuclear Physics

Summary by Alexander Kiselev, MD, and Mark Ito

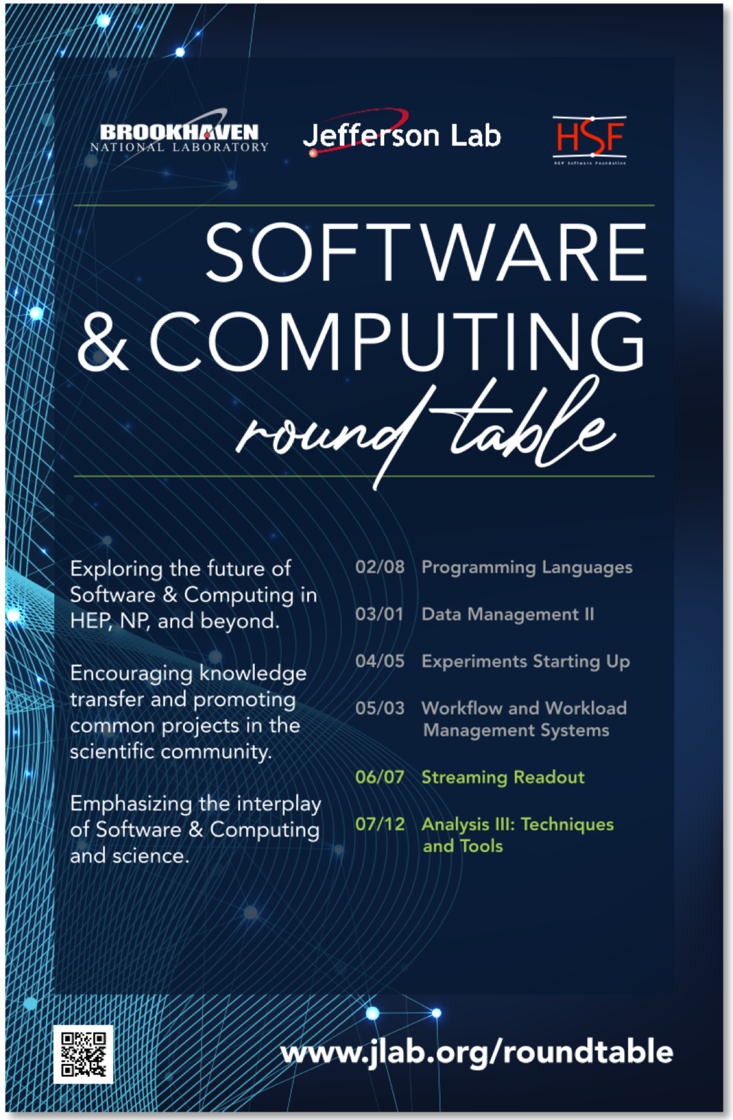
Scientific Problem Space

- Focus on non-perturbative QCD phenomena
- MC event generators for spin-dependent measurement, including novel QCD phenomena (GPDs, TMDs)
- **Analyses considering large number of signal events simultaneously** (or multiple times)
  - **Contrary** to separating a few events from a large number of background events
  - **Example** complexity of multi-dimensional, strongly correlated relationships among data opposed to search of rare events with novel topologies
  - **Example** high-precision results which require complex analyses to control systematic uncertainties
  - Require unique software and computing strategies
- Relatively smaller size of experiments goes along with shorter experimental life cycles and faster changes in scientific goals

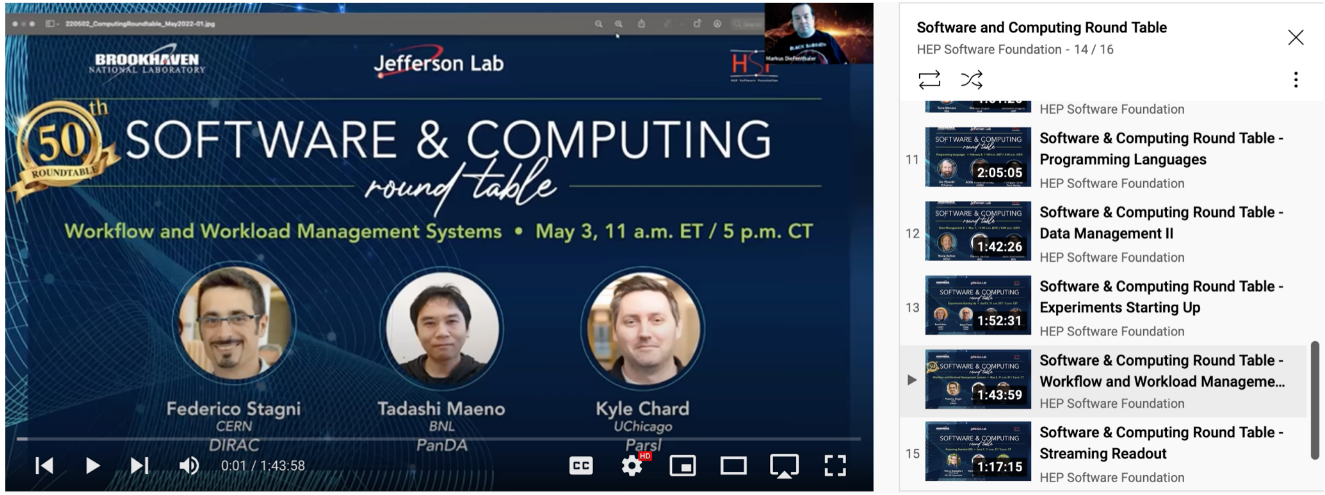
Small Group Size

- Collaboration size in average smaller in NP than in HEP
- Tendency for everyone *"doing their own thing"*
  - Larger experiments, individual analyses can be numerous and quite different from another, with a small team on each top.
- Non-unified approach has inhibited progress in the field in the past.
- Transition to experiments with larger data size and more complex analyses
- Old culture cannot effectively address problems of scale of future experiments
- Relatively smaller group size asks for careful planning and design of the software effort: mix of in-house development, adoption of outside packages, and the choice of appropriate scale throughout.
- Challenge in finding the right balance.

Community Forum



- Seminar series on the **interplay of computing and science**
- With O(50) participants per month
- Initiated at **Jefferson Lab** after the first "Future Trends in NP Computing" workshop in 2016 with two main goals:
  - **Knowledge transfer**
  - **Encourage common projects**
- Since 2020 **jointly organized with BNL and the HSF** with software & computing topics from the wider NP and HEP community.
- Recordings available on YouTube:



HEP Software Foundation and NP

MD appointed as

Convener of Physics Generators WG

Common forum for:

- **Discussion** on the physics event generators used by **N**HEP experiments
- **Technical work** on these physics event generators

Promotes collaboration among experiments and experiment and theory.

Jin Huang appointed as

Convener of Reconstruction and Software Triggers WG

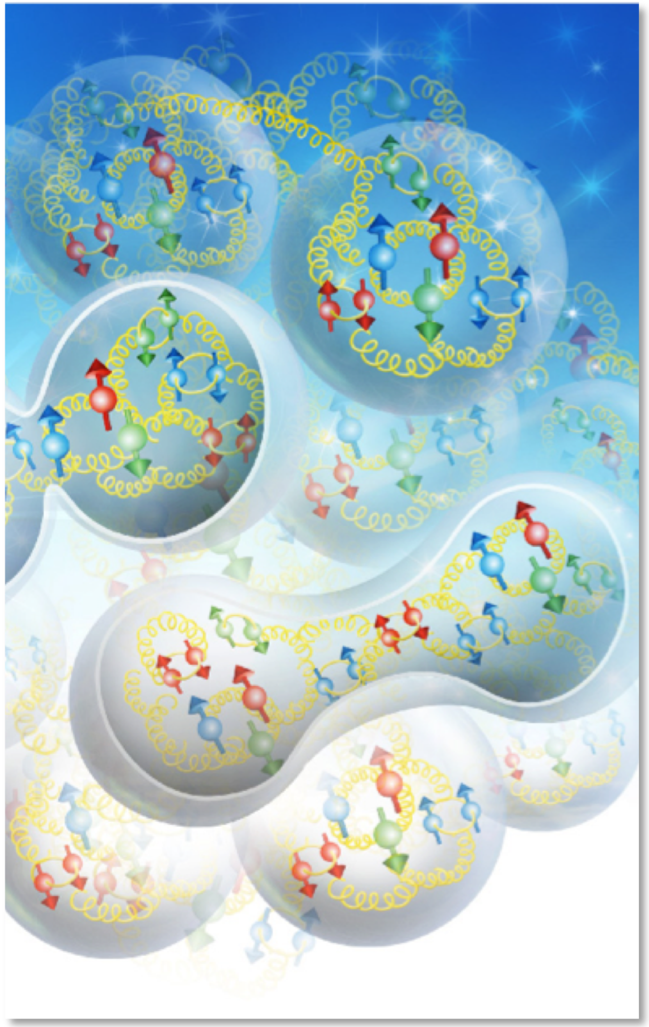
- Considers approaches and solutions to common challenges in the area of event reconstruction and software triggering
- Algorithms and data structures that perform the heavy lifting data processing linking real/simulated data → analysis

Wouter Deconinck appointed as

Convener of Training WG

- Provide training in the computing skills needed for researchers to produce high quality and sustainable software.
- Offer and expand a training program, ranging from basic core software skills needed by everyone to the advanced training required by specialists in software and computing.
- Organize periodic software carpentry workshops.

Working with the Electron-Ion Collider Community



Lessons Learned

2016 – now **Future Trends in Nuclear Physics Computing**

2016 – now **Software & Computing Round Table**

2016 – 2020 **EIC Software Consortium (ESC)**

2018 – now **EICUG Software Working Group (SWG)**

2019 – 2021 **Yellow Report Initiative**

- **"State of Software"** survey to understand and quantify software usage throughout the EIC community.

2021 – 2022 **Detector Collaboration Proposals**

- Repeated **"State of Software"** survey to understand and quantify software usage throughout the EIC community.

Work Towards one Software Stack

- **May 11 – now** Software decision process
- Requirements based on "**EIC Software: Statement of Principles**"
  - Endorsed by EIC Community

EIC Software: Statement of Principles

EIC SOFTWARE:
Statement of Principles

1 We aim to develop a diverse workforce, while also cultivating an environment of equity and inclusivity as well as a culture of belonging.

2 We will have an unprecedented compute-detector integration:

- We will have a common software stack for online and offline software, including the processing of streamed data and its time-ordered structure.
- We aim for autonomous alignment and calibration.
- We aim for a rapid, near-real-time turnaround of the raw data to online and offline productions.

3 We will leverage heterogeneous computing:

- We will enable distributed workflows on the computing resources of the worldwide EIC community, leveraging not only HTC but also HPC systems.
- EIC software should be able to run on as many systems as possible, while supporting specific system characteristics, e.g., accelerators such as GPUs, where beneficial.
- We will have a modular software design with structures robust against changes in the computing environment so that changes in underlying code can be handled without an entire overhaul of the structure.

4 We will aim for user-centered design:

- We will enable scientists of all levels worldwide to actively participate in the science program of the EIC, keeping the barriers low for smaller teams.
- EIC software will run on the systems used by the community, easily.
- We aim for a modular development paradigm for algorithms and tools without the need for users to interface with the entire software environment.

5 Our data formats are open, simple and self-descriptive:

- We will favor simple flat data structures and formats to encourage collaboration with computer, data, and other scientists outside of NP and HEP.
- We aim for access to the EIC data to be simple and straightforward.

6 We will have reproducible software:

- Data and analysis preservation will be an integral part of EIC software and the workflows of the community.
- We aim for fully reproducible analyses that are based on reusable software and are amenable to adjustments and new interpretations.

7 We will embrace our community:

- EIC software will be open source with attribution to its contributors.
- We will use publicly available productivity tools.
- EIC software will be accessible by the whole community.
- We will ensure that mission critical software components are not dependent on the expertise of a single developer, but managed and maintained by a core group.
- We will not reinvent the wheel but rather aim to build on and extend existing efforts in the wider scientific community.
- We will support the community with active training and support sessions where experienced software developers and users interact with new users.
- We will support the careers of scientists who dedicate their time and effort towards software development.

8 We will provide a production-ready software stack throughout the development:

- We will not separate software development from software use and support.
- We are committed to providing a software stack for EIC science that continuously evolves and can be used to achieve all EIC milestones.
- We will deploy metrics to evaluate and improve the quality of our software.
- We aim to continuously evaluate, adapt/develop, validate, and integrate new software, workflow, and computing practices.

The "Statement of Principles" represent guiding principles for EIC Software. They have been endorsed by the international EIC community. For a list of endorses, see https://eic.github.io/activities/principles.html

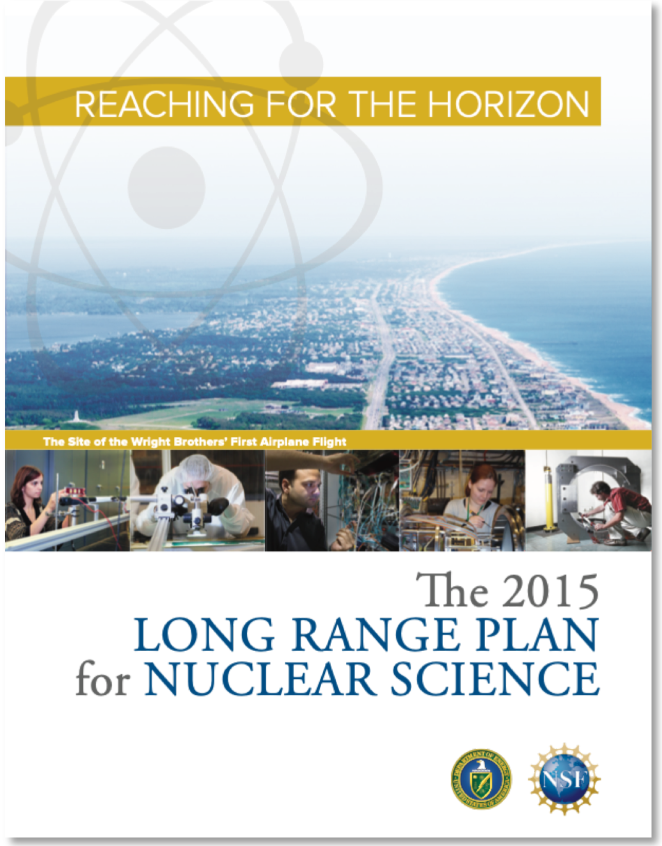
A Core Principle: The Software Community

7 We will embrace our community:

- EIC software will be open source with attribution to its contributors.
- We will use publicly available productivity tools.
- EIC software will be accessible by the whole community.
- We will ensure that mission critical software components are not dependent on the expertise of a single developer, but managed and maintained by a core group.
- We will not reinvent the wheel but rather aim to build on and extend existing efforts in the wider scientific community.
- We will support the community with active training and support sessions where experienced software developers and users interact with new users.
- We will support the careers of scientists who dedicate their time and effort towards software development.

Next Steps

The design and development of software & computing constitute a cornerstone for the future success of the Nuclear Physics program. **Future Trends in Nuclear Physics Computing** will serve as a **forum to discuss priorities for design and development** as **input for a community white paper to inform the next Long Range Plan for Nuclear Science.** Each day of the workshop will have a theme to frame the discussion:



Where are we as a community?

- Wednesday, September 28

How can we make analysis easier?

- Thursday, September 29

Careers and DE&I

- Thursday, September 29

How can we scale up computing?

- Thursday, October 1