

FUTURE TRENDS IN NUCLEAR PHYSICS COMPUTING

Workshop Re-Introduction and Day 2 Recap

BROOKHAVEN & Jefferson Lab
NATIONAL LABORATORY

FUTURE TRENDS IN NUCLEAR PHYSICS COMPUTING

SEPT. 29 - OCT. 1, 2020

The workshop focuses on the Nuclear Physics Software & Computing community. We will identify what is unique about our community and we will discuss how we can strengthen common efforts and chart a path for Software & Computing in Nuclear Physics for the next ten years.

TOPICS:

- Common Scientific Software
- The Role of Data Centers in Scientific Discovery
- Unique Software Challenges for Nuclear Physics

- 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

- We meet for four hours each day in a time window chosen to be as inclusive as possible for participants around the world:

9:00 a.m. – 1:00 p.m. (EDT)

- Substantial discussion time is included in the agenda.

Common Scientific Software

- Tuesday, September 29
- **Moderated by** P. Laycock and T. Wenaus

The Role of Data Centers in Scientific Discovery

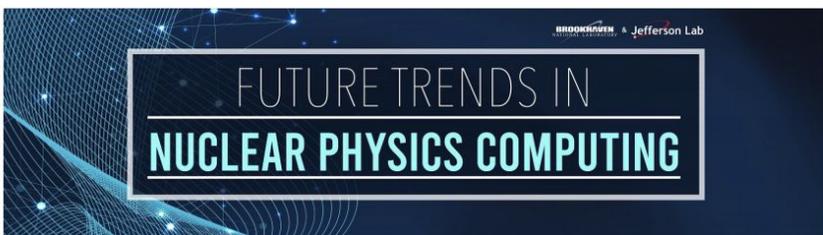
- Wednesday, September 30
- **Moderated by** G. Heyes and O. Rind

Unique Software Challenges for Nuclear Physics

- Thursday, October 1
- **Moderated by** M. Diefenthaler, M. Ito, and A. Kiselev

Attendance ranged from ~30 at beginning to ~66 at peak. All sessions were recorded this time...even if a power outage did knock out one of the moderators for some time.

204 participants registered as of this morning



Future Trends in Nuclear Physics Computing Meeting Notes

[Timetable](#)

This is the live meeting notes document for the [Future Trends in Nuclear Physics Computing Workshop](#) held on September 29 - October 1, 2020. This workshop, the third of the series (previous editions were in [2017](#) and [2016](#)), focuses on the Nuclear Physics Software & Computing community itself. Goals for the workshop are to identify what is unique about our community, find ways to strengthen common efforts, and chart a path for Software & Computing in Nuclear Physics for the next ten years.

We meet for four hours each day in a time window chosen to be as inclusive as possible for participants around the world. Substantial discussion time is included in the agenda, and session conveners will keep speakers to time in order to preserve the discussion time. This google doc will be used in advance to give the discussions structure and focus, as well as during the workshop itself to moderate and record the discussion and gather input from all participants, and after the workshop as the basis for summarizing and report writing. Editing is on, and all participants are encouraged to contribute in all phases.

Each day has a theme. In advance of the workshop, questions and discussion points for each day will be gathered here to guide a moderated common discussion following the talks. A short discussion period will follow each talk to address questions specific to the talk. The content prepared in advance will be augmented during the presentations and discussions.

A brief synopsis of the previous day will be part of an intro talk on days two and three.

The workshop will conclude with a short summary, but summarizing and report writing proper will proceed after the workshop. All participants are welcome and encouraged to join the meeting organizers in this work. The outcome will be a written report, with presentation and discussion of the report in the subsequent meeting of the "Software & Computing Round Table" that is jointly organized

As in the previous two days, there is a short(ish) question/discussion session after each talk, focused on questions arising directly from the talk

A longer discussion slot at the end is for open discussion centered around the topic of the day: **Unique Software Challenges For Nuclear Physics**

We are continuing to use [this google doc](#) to record live notes, and to seed the discussion sessions with questions and discussion points in advance, and during talks....thanks to all who have contributed - this has worked remarkably well so far!

Editing is on, please continue to contribute!

The notes (together with the talks) will be the basis for summarizing the meeting and writing a report in the weeks to follow (more on this in the closing session)

Day 2 Recap

09:00 → 13:00	The Role of Data Centers in Scientific Discovery Conveners: Dr Graham Heyes (Jefferson Lab), Ofer Rind (BNL)	
09:00	Workshop Introduction and Recap Speakers: Paul Laycock (Brookhaven National Laboratory), Torre Wenaus (BNL) Slides	15m
09:15	Data Centers in a Decade Speaker: Eric LANCON (BNL) Data Centers in a D...	30m
09:45	Questions on Talk	10m
10:00	Data and Analysis Preservation and Open Data ↑ Speaker: Maxim Potekhin (NPPS/Physics Department) Data and Analysis ...	30m
10:30	Questions on Talk	10m
10:45	User Perspective and Requirements Speaker: Dr Graham Heyes (Jefferson Lab) Users and Require...	30m
11:15	Questions on Talk	10m
11:30	Discussion of the Role of Data Centers	1h 30m

- Our three speakers did a great job of examining the role of the data center from a number of perspectives:
 - Looking at how this role may evolve over the next decade
 - Looking at its particular role supporting long term data and analysis preservation for experiments
 - Looking at the multifaceted interactions with its user base
- Thanks to all participants for the stimulating discussions, challenging questions, and frank open answers
- Will not attempt to summarize the presentations here as the slides are available
- A summary of the discussion follows, as recorded in the meeting notes
- Please help improve, extend, and correct the summary
 - We ask that you please identify yourself in the text when you add something

Themes in the talks were

- Local infrastructure decisions:
 - The role of tape and cloud storage in the future.
 - The use of disruptive technologies, AI/ML and coprocessors (FPGA and GPU).
- Distributed computing resources.
 - How they are integrated with local resources
 - How users interact with them
 - Access - federated authentication and high bandwidth networking.
- Long term viability of data
 - Metadata and code is almost as important as the raw data from experiments.
 - Documentation of computational resources as well as software.
 - Importance of technology choices for documentation, usability vs long term stability.
- Containerization - as a mechanism for aiding operation on distributed sites as well as preserving code for far future execution.
- Interaction with users or stakeholders.
 - In all three talks communication and overall engagement was emphasized.

Common Discussion

- Common benchmarks that are consistently used by various sites participating in a distributed computing model.
 - Similarly common metrics that allow consistent evaluation of utilization and availability.
- The rapid growth of ML and AI in nuclear physics seen at both BNL and JLab.
 - Need to bring these groups together, at least within labs, to avoid duplication of effort and to share ideas.
- The balance between educating a user on the use of a complex system vs providing a simplified interface.
- The tension between HPC and HTC, where do supercomputers and large data centers fit.
 - Edge computing vs centralized and where the balance lies for NP.
- Collaboration between facility and users.
 - Embedding of people within experiments.
 - Physicists on Sci Comp staff as a bridge between physics and computer science.
 - Connecting ops oriented people with CS research.
 - Collaborating with CS researchers at universities.
- Promoting computing within the community
 - Career paths for joining science and computer science.
 - Pulling in a coherent direction.