BROOKHAVEN NATIONAL LABORATORY & Jefferson Lab

FUTURE TRENDS IN NUCLEAR PHYSICS COMPUTING

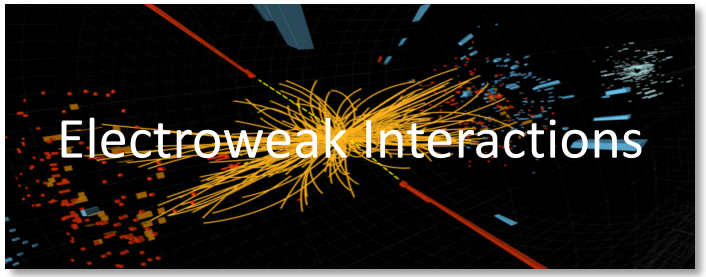
Workshop Summary

Nuclear Physics

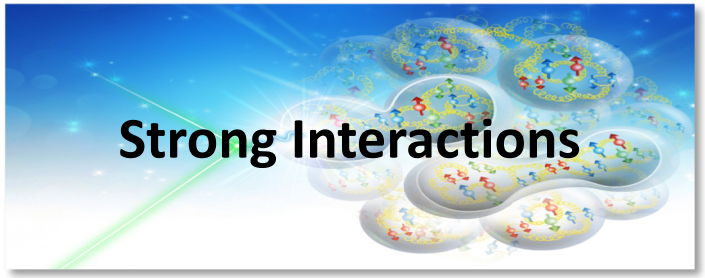
Further exploration of the Standard Model

Dark matter searches

Electroweak symmetry breaking



Deeper understanding of QCD:



The mission of the Nuclear Physics program in the U.S. is to discover, explore, and understand all forms of nuclear matter.

Frontiers in Nuclear Physics

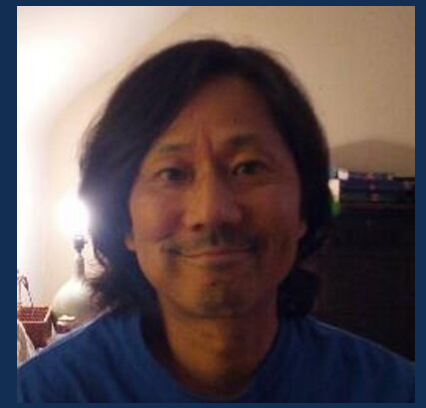
- One of the enduring mysteries of the universe is the nature of matter—what are its basic constituents and how do they interact to form the properties we observe? The largest contribution by far to the mass of the matter we are familiar with comes from protons and heavier nuclei.
- Although the fundamental particles that compose nuclear matter—quarks and gluons—are themselves relatively well understood, exactly how they interact and combine to form the different types of matter observed in the universe today and during its evolution remains largely unknown.

Joint Organization by BNL and JLab

Alexander Kiselev

Amber Boehnlein

Graham Heyes



Mark Ito

Markus Diefenthaler

Ofer Rind

Paul Laycock

Torre Wenaus

History

Future Trends in Nuclear Physics Computing in 2016

Goals

- Examined computing strategy at a time horizon of ten years
- Defined common vision for NP computing
- Recommended future directions for development

Website

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

Future Trends in Nuclear Physics Computing in 2017

Goals

- Discussed trends in scientific computing
- Collected ideas on how to improve analysis
- Worked towards next-generation analysis techniques and tools

Website

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

Goals

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

THANK YOU FOR JOINING US!

207 participants

Thanks to HSF for support and active participation.

Organization

- We met 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 was 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

Discussion

Live notes

Scope

- moderate and record the discussion
- gather questions and discussion points from all participants
- **after the workshop** summarizing and report writing

- https://docs.google.com/document/d/1mug_UB31WngFvvlLv8CFRBd4dSdKZR0iROL3m3lFY5Y/edit
- 26 (!) pages

Common Scientific Software

Common Scientific Software: Password for watching the recording: sF0JQ^t#

Conveners: Paul Laycock (Brookhaven National Laboratory), Torre Wenaus (BNL)

Recording

9:00 AM **Workshop Goals** 15m

Speakers: Alexander Kiselev (BNL), Dr Amber Boehnlein (Jefferson Lab), Dr Graham Heyes (Jefferson Lab), Dr Mark Ito (Jefferson Lab), Markus Diefenthaler (Jefferson Lab), Ofer Rind (BNL), Paul Laycock (Brookhaven National Laboratory), Torre Wenaus (BNL)

TRENDS2020-Welc...

9:15 AM **Developing Common Software: ACTS** 30m

Speaker: Andreas Salzburger (CERN)

2020-ACTS-Future-T...

9:45 AM **Questions on Talk** 10m

10:00 AM **Software Sustainability** 30m

Speaker: Daniel S. Katz (University of Illinois Urbana-Champaign)

nuclear-physics-co...

10:30 AM **Questions on Talk** 10m

10:45 AM **Developing Common Software: Rucio** 30m

Speaker: Mario Lassnig

Rucio @ Future Tren...

11:15 AM **Questions on Talk** 10m

11:30 AM **Discussion on Common Software** 1h 30m

- No attempt to summarize the presentations, the slides stand on their own
- Our excellent speakers **on all days** made the session as interesting and useful as we could have hoped, thank you to them all
- And thanks to all participants for stimulating discussions, challenging questions, frank open answers

Common Scientific Software

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

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

The Role of Data Centers in Scientific Discovery

- **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

09:00 → 13:00 **Unique Software Challenges for Nuclear Physics Community: Password for watching the recording: D!thz2Vu**

Conveners: Alexander Kiselev (BNL), Dr Mark Ito (Jefferson Lab), Markus Diefenthaler (Jefferson Lab)

Recording

09:00	**Workshop Introduction and Recap** **Speakers**: Dr Graham Heyes (Jefferson Lab), Ofer Rind (BNL) Day 2 Summary · TRENDS2020-Day2...	15m
09:15	**Experience from 12 GeV Science Program** **Speaker**: David Lawrence (Jefferson Lab) 2020.10.01.12GeV... · Experience From th...	30m
09:45	**Questions on Talks**	10m
10:00	**Software Challenges in Streaming Readout** ¶ **Speaker**: Jan Bernauer (Stony Brook University and RBRC) futuretrends.pdf	30m
10:30	**Questions on Talks**	10m
10:45	**Survey from Nuclear Physics Students and Young Postdocs** TRENDS2020-Surv...	30m
11:15	**Questions on Survey**	10m
11:30	**Discussion of Unique Software Challenges for Nuclear Physics**	1h 15m
12:45	**Workshop Closing: Summary and Next Steps** **Speakers**: Alexander Kiselev (BNL), Dr Amber Boehnlein (Jefferson Lab), Dr Graham Heyes (Jefferson Lab), Dr Mark Ito (Jefferson Lab), Markus Diefenthaler (Jefferson Lab), Ofer Rind (BNL), Paul Laycock (Brookhaven National Laboratory), Torre Wenaus (BNL) TRENDS2020-Closi...	15m

Identified challenges from three perspectives

- Experience from JLab 12 GeV Science Program
- Software challenges in Streaming Readout
- Survey among NP Students and Young Postdocs

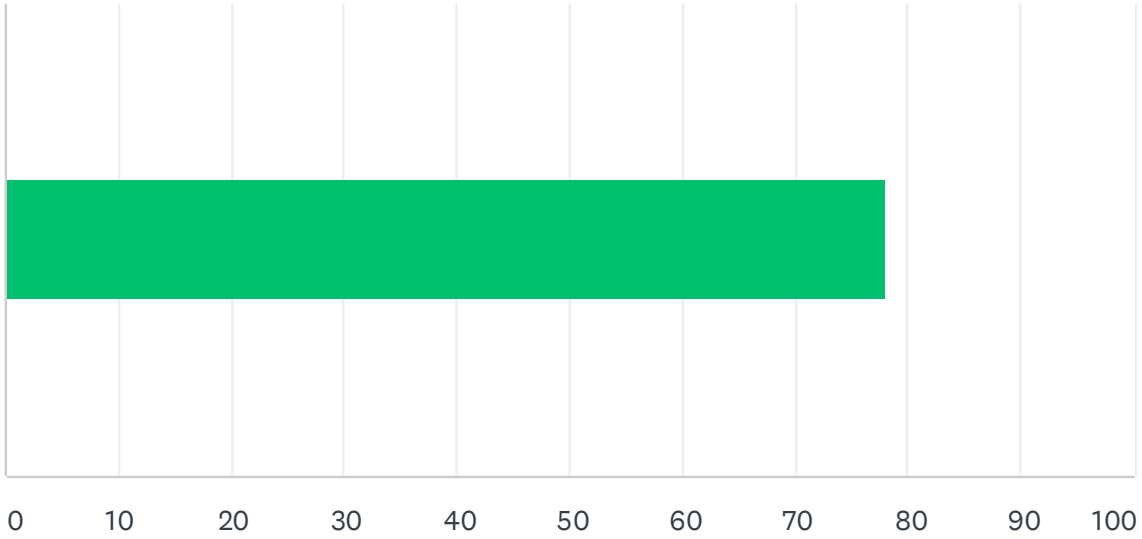
Discussion highlights

- Importance of fast turnaround in data processing and having adequate monitoring and DQ tools
- **DAP issue** Even big enough experiments do not have a working model in place by the time data taking starts.
- Connection between SRO and DAP, since the archived data is available for the analyses which were not anticipated at a time the experiment is taking data. A lot of data mining associated with the new physics topics can happen later in time.
- Quality tutorials are extremely important for teaching young scientists.

Survey: Input from 44 Ph.D students and postdocs

Q1 What fraction of your time do you spend on the software and computing aspects of your research, such as programming, analysis jobs, etc.?

Answered: 44 Skipped: 0

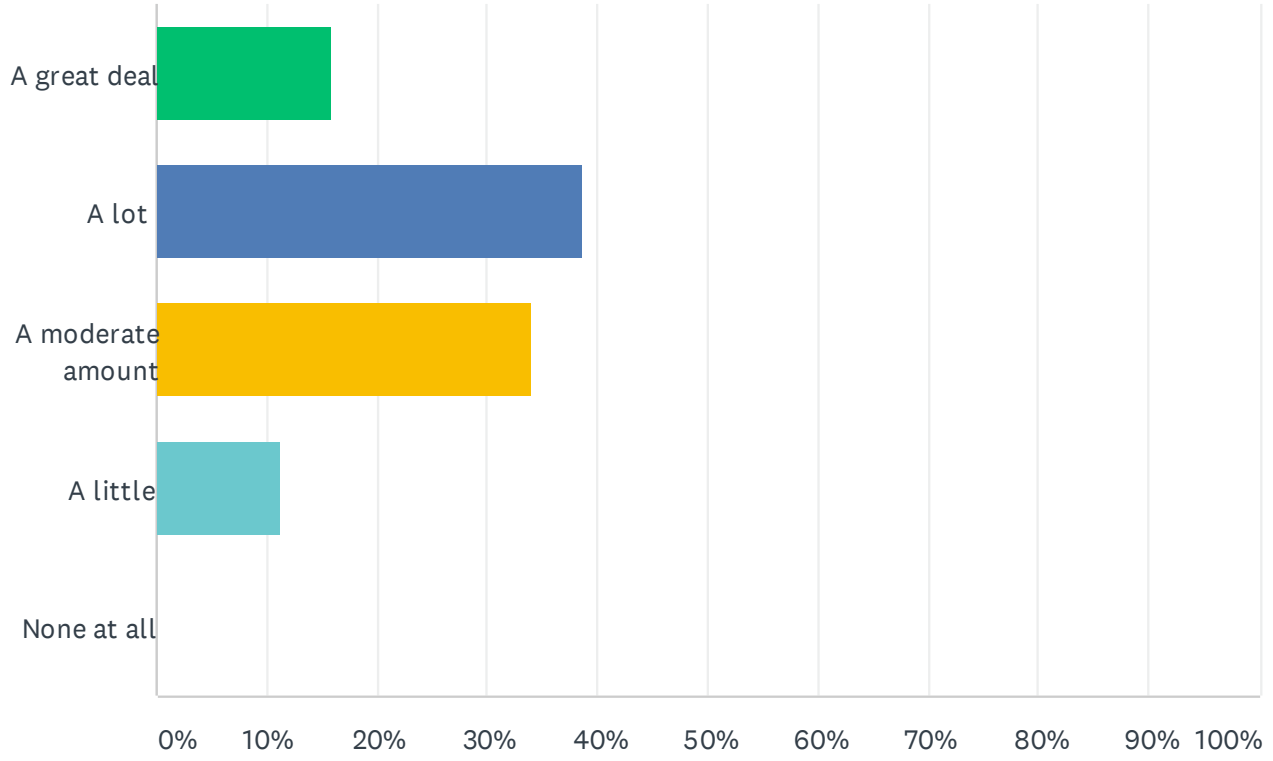


ANSWER CHOICES	AVERAGE NUMBER	TOTAL NUMBER	RESPONSES
	78	3,435	44
Total Respondents: 44			

Interesting Comments from Survey

Q9 Are the computing resources you need readily available?

Answered: 44 Skipped: 0



4	It seems that funds are spent mostly on computing resources, but very little on computing expertise (if we had half the computers and twice the number of software experts we would be in a much better place)

Unique Software Challenges for Nuclear Physics

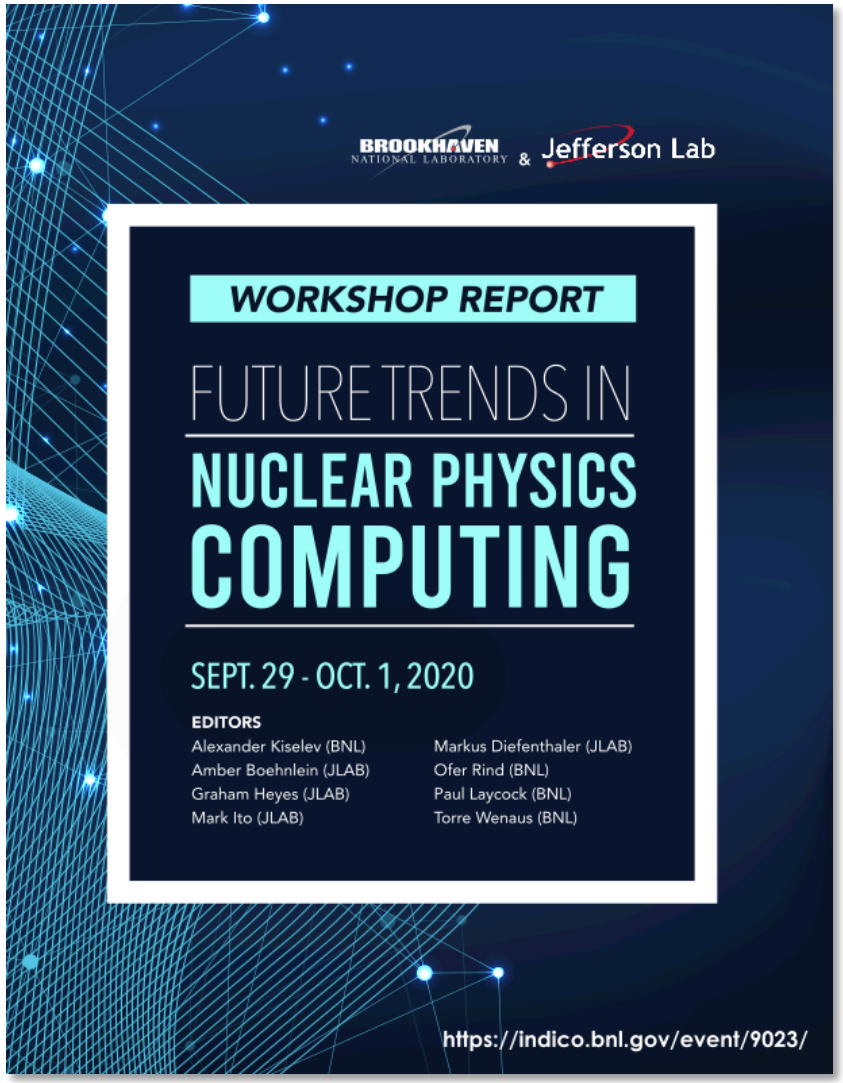
Scientific Problem Space

- 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.

Workshop Summary and Report



Workshop Report

- **Draft** (28 pages and growing)
  - https://docs.google.com/document/d/1HT0dQJwZkUGaxGS0bUhZgkz0l2G5QR-4Dm08-hr8DJc/edit?usp=sharing

- Everyone was welcome and encouraged to **provide comments** between **Nov. 10 – Nov. 17.**

- **Status**
  - Finalizing the report.
  - Will be shared on arXiv.

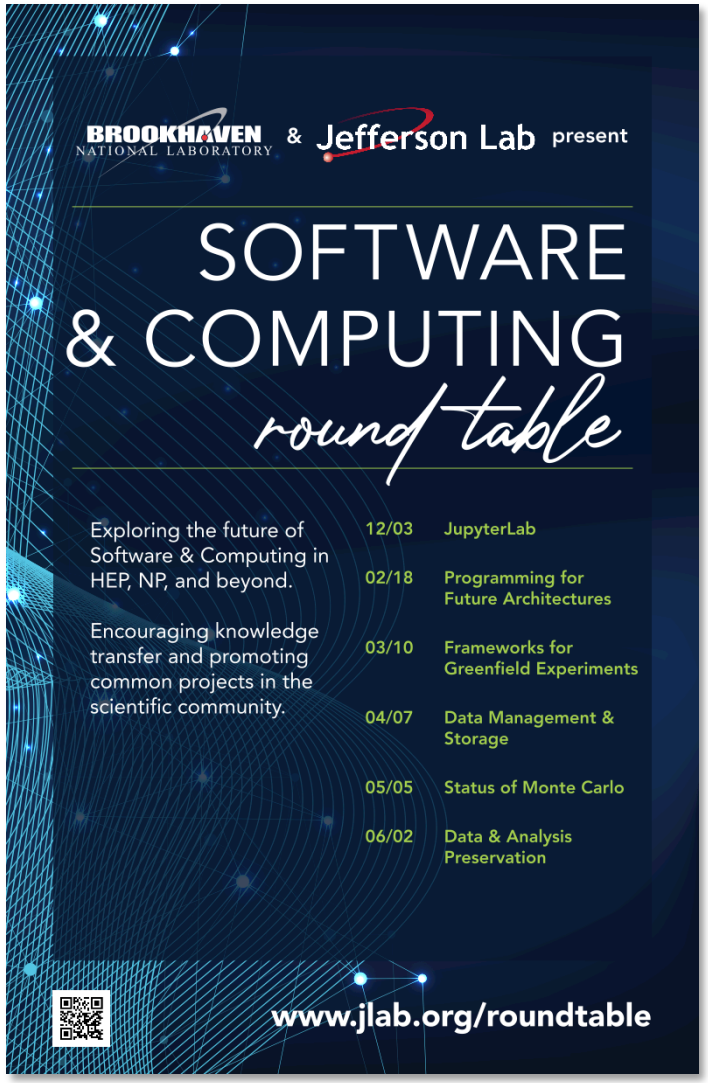
Community Forum

- There was consensus on starting a new **community forum for NP** to discuss common projects, the role of data centers, unique challenges etc.

- Ongoing discussion on details and possible connection to HSF.

- Possible **goals of the community forum**
  - Inform on building successful scientific software projects, taking the unique challenges of NP in consideration
  - Foster collaborative common software projects in NP
  - Promote scientific software career support

Software & Computing Round Table



Jointly organized by BNL and Jlab

- role of Software & Computing Round Table in 2016 – now
  - encourage knowledge transfer
  - promote common projects
- exploring the future of software & computing in HEP, NP, and beyond
- emphasis on the interplay of software & computing and science

- **website** https://www.jlab.org/roundtable

News story on Software & Computing Round Table

NUCLEAR PHYSICS DATA DEMAND MORE POWERFUL PROCESSING

Jefferson Lab and Brookhaven National Lab partner on a Software & Computing Round Table to track the leading edge of computing and foster collaboration

Fans of the popular TV show "The Big Bang Theory" can picture the sitcom's physicists standing at a whiteboard, staring hard at equations.

It's an iconic image. But is that the future — or even the present — of how nuclear physicists do their jobs? Not really. Not when new experiments demand ever-more powerful data processing and thus ever-more-powerful software and computing.

"Scientists being at a blackboard and writing up some equations — that is not always the reality," said Markus Diefenthaler, an experimental nuclear physicist at the U.S. Department of Energy's Thomas Jefferson National Accelerator Facility in Newport News, Virginia.

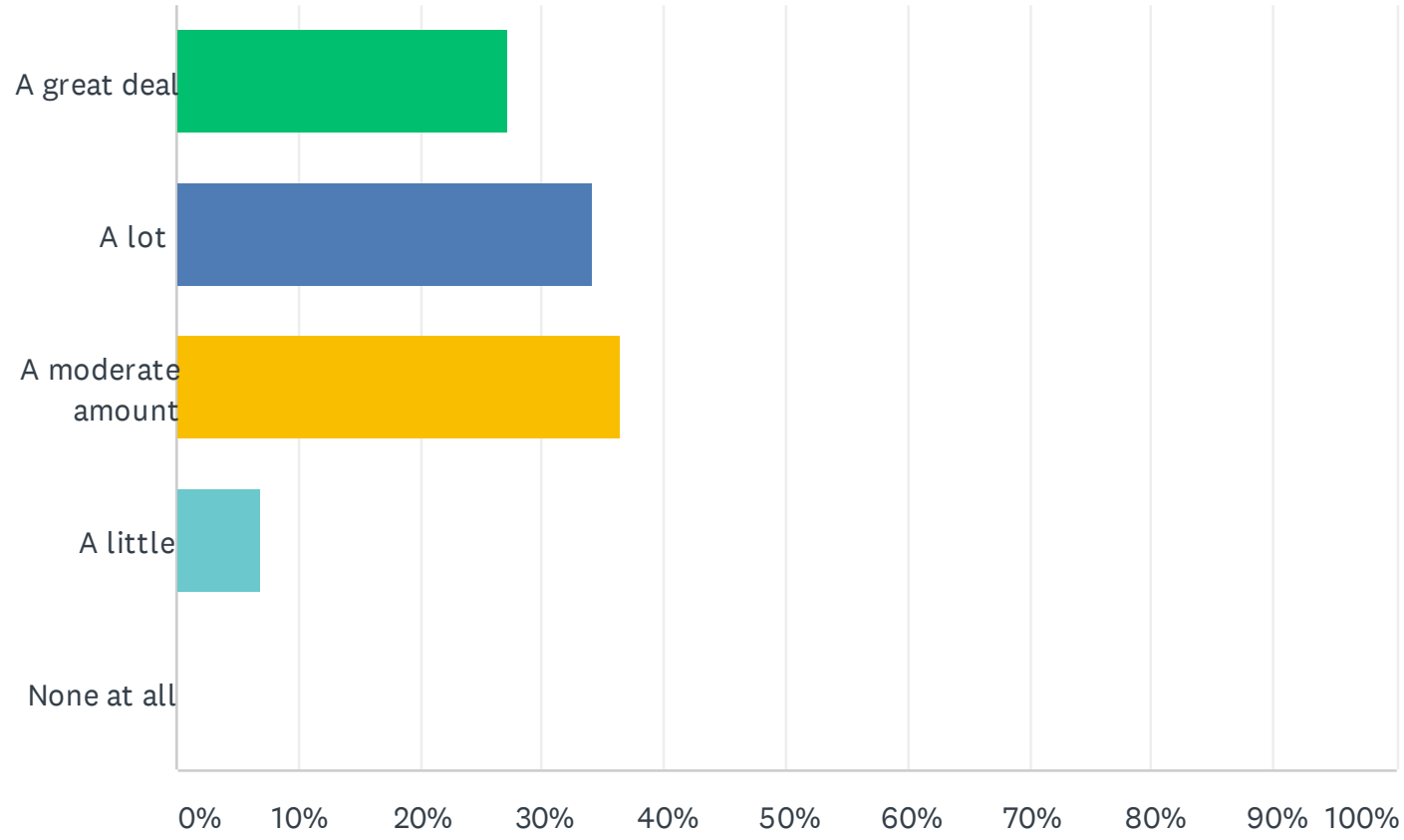
BROOKHAVEN NATIONAL LABORATORY & Jefferson Lab

FUTURE TRENDS IN NUCLEAR PHYSICS COMPUTING

Backup

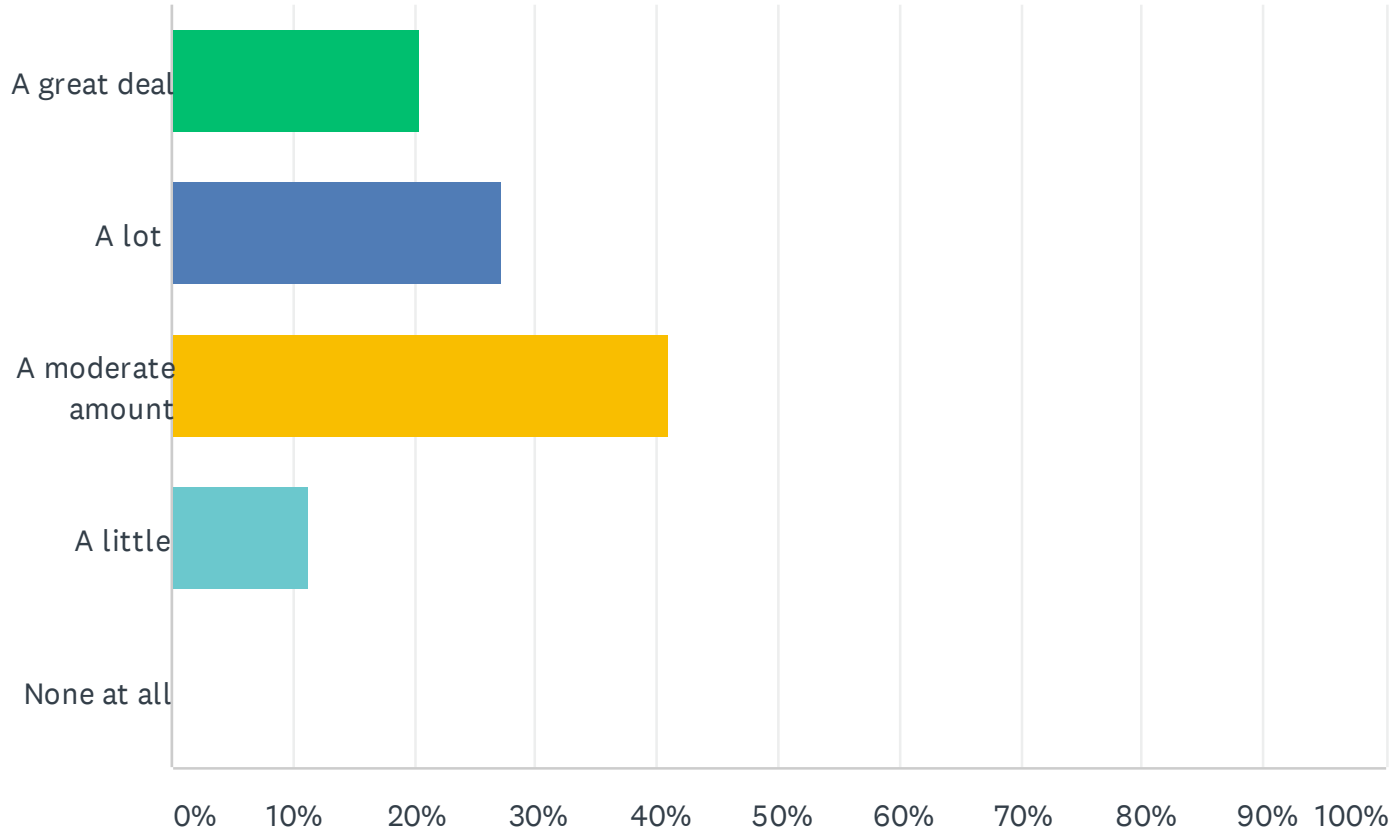
Q6 Are you currently able to perform the software and computing tasks needed for your research?

Answered: 44 Skipped: 0



Q7 Are you confident performing software and computing tasks needed for your research?

Answered: 44 Skipped: 0



Q3 How many computer science, engineering and/or programming courses did you take in college?

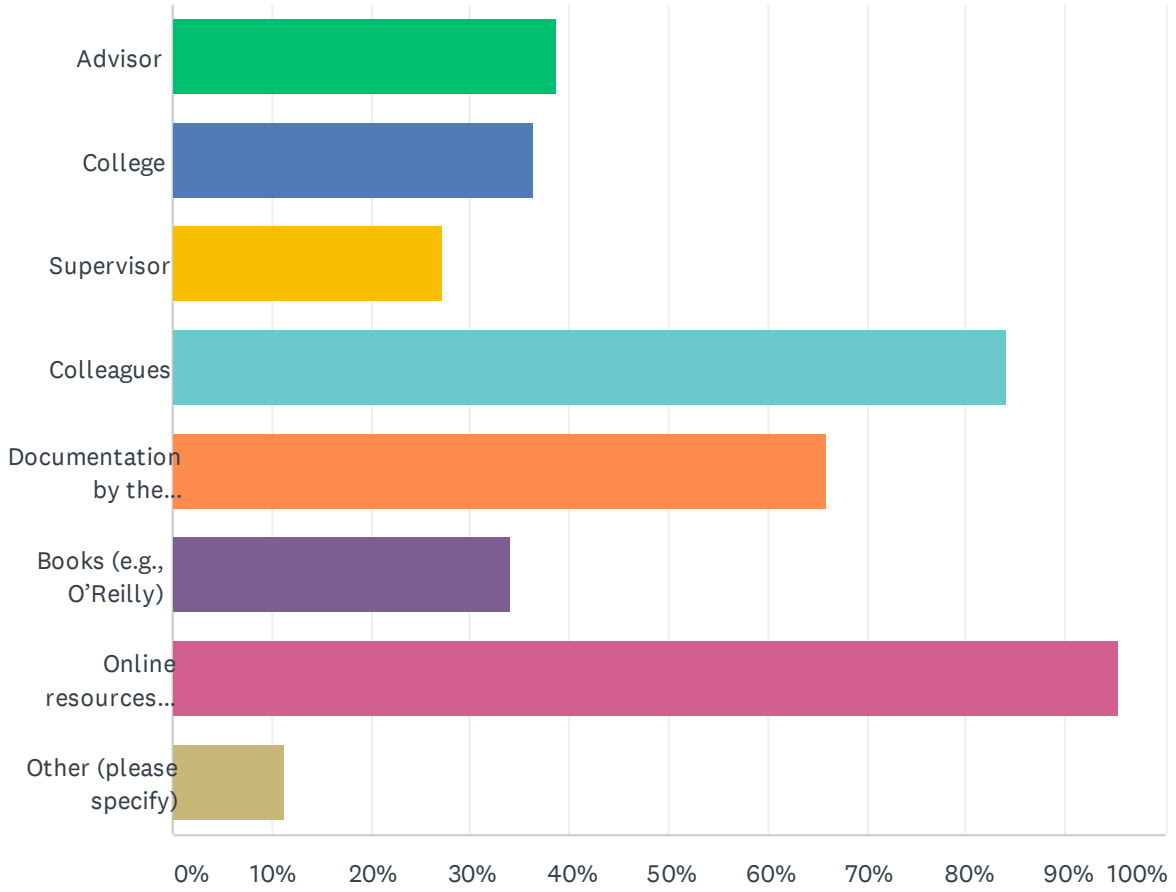
0	9	
1	14	mentioned S&C introduction in python
2	4	
3	7	
4	4	
5	3	
8	1	
10	1	

4	Full-time computer science major.

Survey

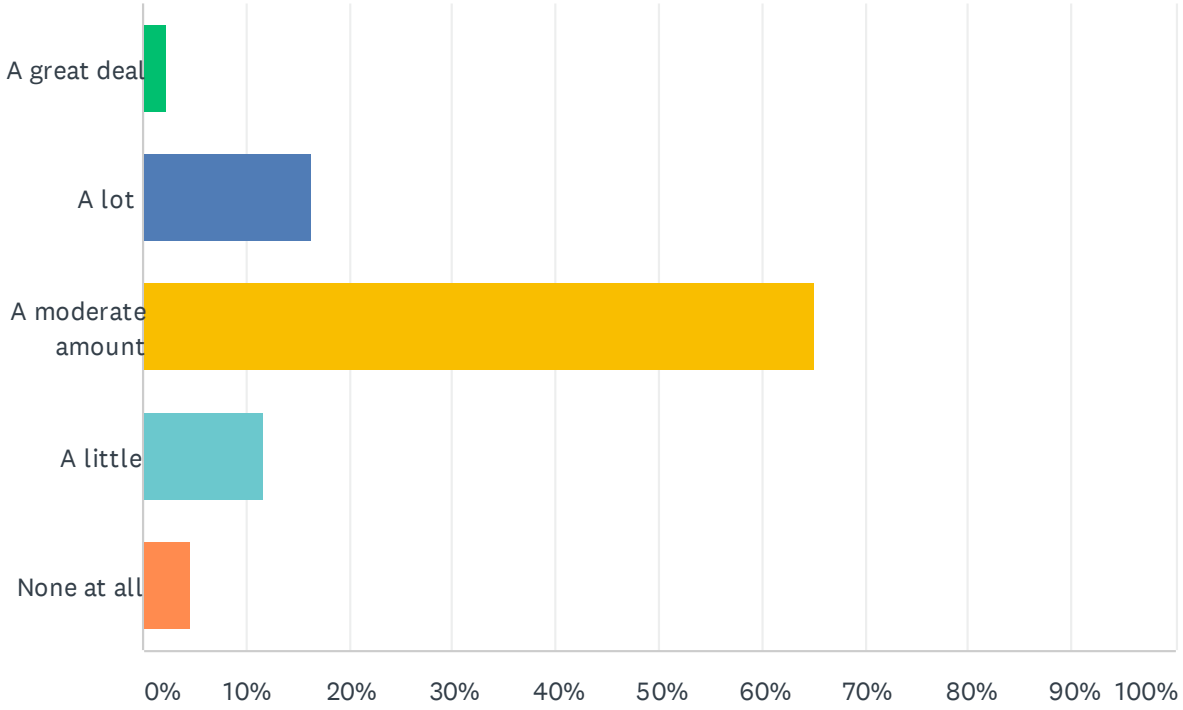
Q4 Which of the following resources have you used for your software and computing work or research? Select all that apply.

Answered: 44 Skipped: 0



Q8 Are the documentation tools you use adequate?

Answered: 43 Skipped: 1



#	PLEASE COMMENT
1	For Python, I can find easy to use resources and/or ask questions online and get ready answers. For ROOT, it feels like the documentation is outdated and esoteric, and trying to get answers online is merely an exercise in frustration.
2	I don't understand the question
3	Not always, no. But this may be a symptom of the code base I use not being supported (just bad timing).
4	Documentation in the two collaborations that I have worked with are awful, outdated, and only experts know what still works and what changes have been done.