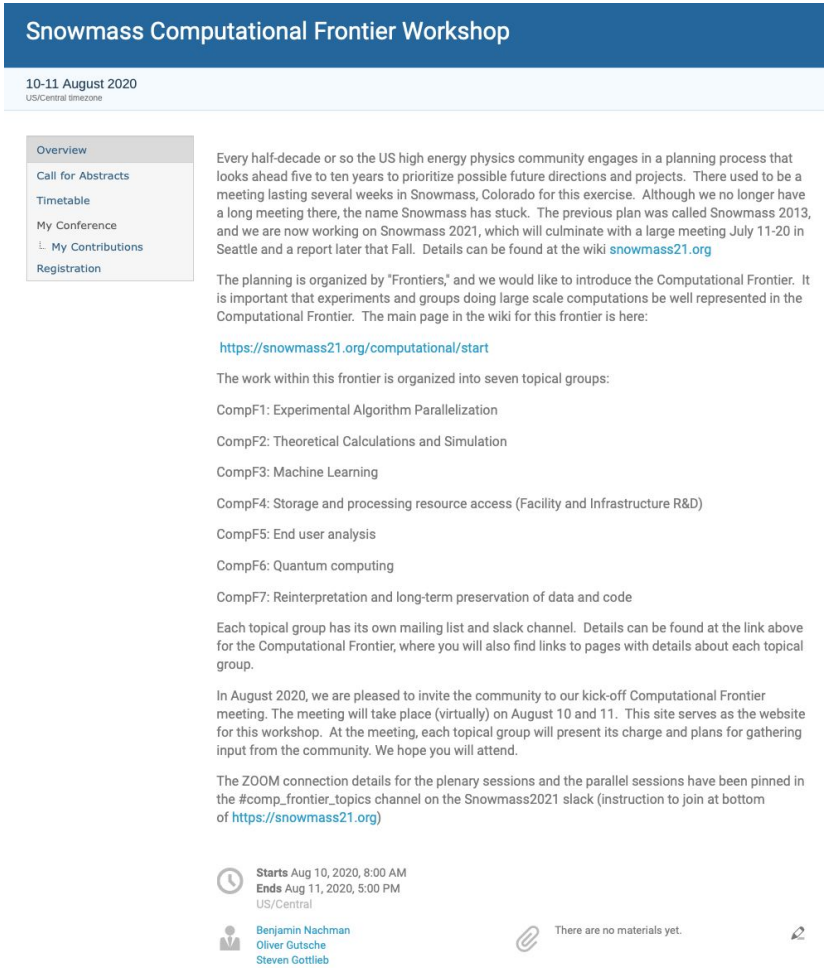
Snowmass 2021: Computational Frontier

Monthly meeting of the WLCG Grid Deployment Board
September 9, 2020

Steve Gottlieb (Indiana U), Ben Nachman (LBNL),
Oliver Gutsche (FNAL)

Disclaimer

- All slides copied from presentations
  - Either marked specifically
  - Or taken from the Snowmass Computational Frontier Workshop (10-11 August 2020)
    - https://indico.fnal.gov/event/43829/
- All credits go to the authors
- All mistakes are mine!



Snowmass Computational Frontier Workshop

10-11 August 2020
US/Central timezone

Overview
Call for Abstracts
Timetable
My Conference
My Contributions
Registration

Every half-decade or so the US high energy physics community engages in a planning process that looks ahead five to ten years to prioritize possible future directions and projects. There used to be a meeting lasting several weeks in Snowmass, Colorado for this exercise. Although we no longer have a long meeting there, the name Snowmass has stuck. The previous plan was called Snowmass 2013, and we are now working on Snowmass 2021, which will culminate with a large meeting July 11-20 in Seattle and a report later that Fall. Details can be found at the wiki snowmass21.org

The planning is organized by "Frontiers," and we would like to introduce the Computational Frontier. It is important that experiments and groups doing large scale computations be well represented in the Computational Frontier. The main page in the wiki for this frontier is here:

https://snowmass21.org/computational/start

The work within this frontier is organized into seven topical groups:

CompF1: Experimental Algorithm Parallelization

CompF2: Theoretical Calculations and Simulation

CompF3: Machine Learning

CompF4: Storage and processing resource access (Facility and Infrastructure R&D)

CompF5: End user analysis

CompF6: Quantum computing

CompF7: Reinterpretation and long-term preservation of data and code

Each topical group has its own mailing list and slack channel. Details can be found at the link above for the Computational Frontier, where you will also find links to pages with details about each topical group.

In August 2020, we are pleased to invite the community to our kick-off Computational Frontier meeting. The meeting will take place (virtually) on August 10 and 11. This site serves as the website for this workshop. At the meeting, each topical group will present its charge and plans for gathering input from the community. We hope you will attend.

The ZOOM connection details for the plenary sessions and the parallel sessions have been pinned in the #comp_frontier_topics channel on the Snowmass2021 slack (instruction to join at bottom of https://snowmass21.org)

Starts Aug 10, 2020, 8:00 AM
Ends Aug 11, 2020, 5:00 PM
US/Central

Benjamin Nachman
Oliver Gutsche
Steven Gottlieb

There are no materials yet.

Snowmass 2021

U.S. Strategic Planning Process for Particle Physics

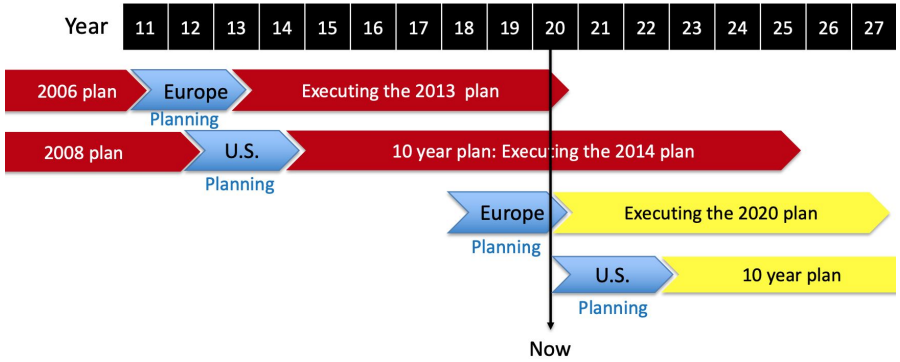
~year-long process
Snowmass Community-Wide "Science" Study
Organized by Division of Particles and Fields (DPF) of APS

Input to P5

~year-long process
P5 (Particle Physics Project Prioritization Panel)
formulate a 10-year plan (20 year vision) within funding constraints
Subpanel of HEPAP, High Energy Physics Advisory Panel for DOE/NSF funding agencies

Particle Physics is Global

- Europe and U.S.
  - Frequency: 7 years (Europe), ~8 years (U.S.)
  - Process: ~2 years in total (~1 year on science + ~1 year priorities)
    - ~1 year on science: Snowmass (U.S.)
    - ~1 year priorities: P5 (U.S.)
- Japan, Canada, China, …



From Young-Kee Kim @ Snowmass Town Hall Meeting @ APS Virtual April Meeting: https://indico.fnal.gov/event/23601/

Learn more about the history and spirit of Snowmass in "How to Snowmass" written by Chris Quigg): https://indico.fnal.gov/event/45207/attachments/133652/164937/How_to_Snowmass-final-links.pdf

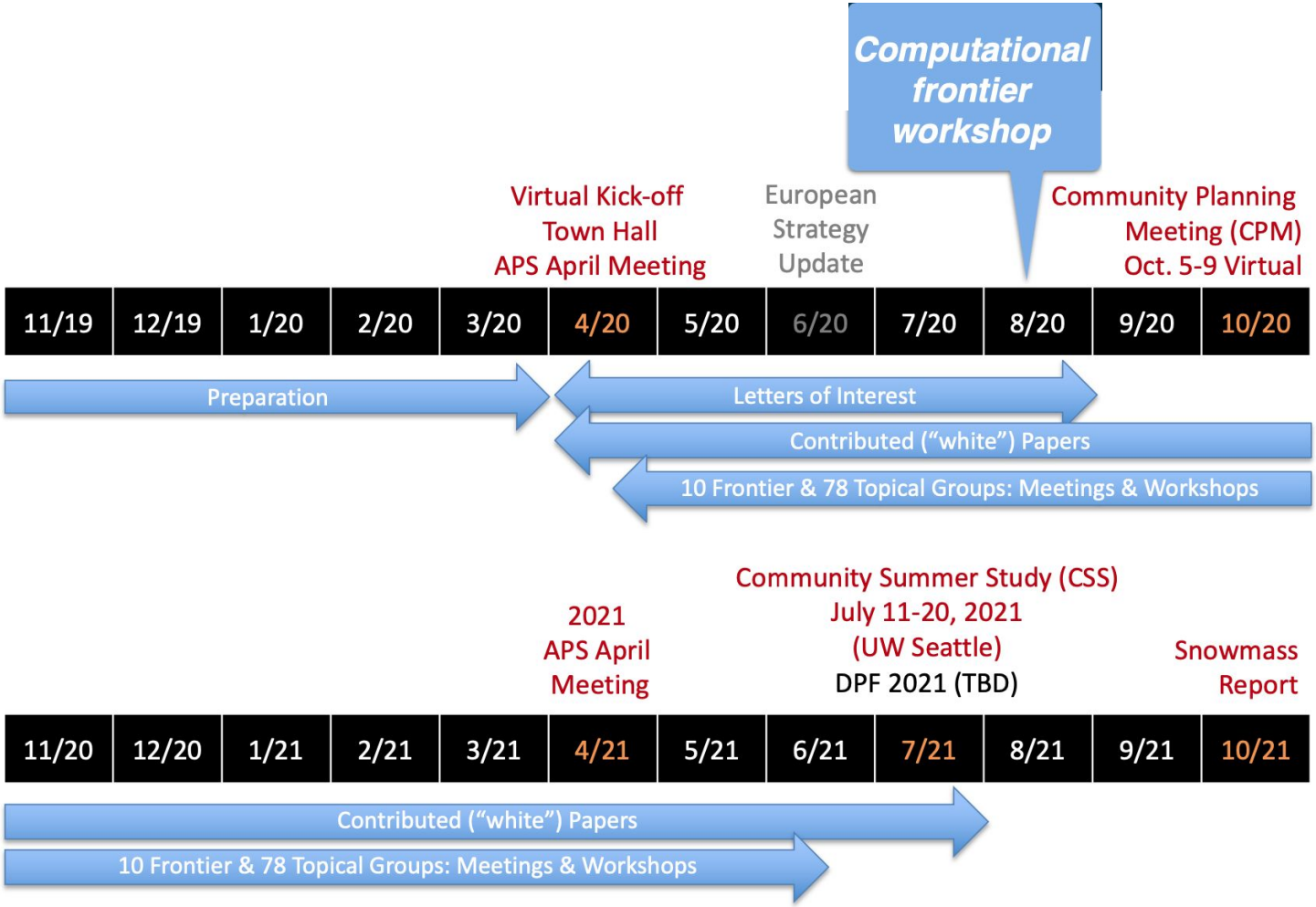
Snowmass = Particle Physics Community Planning Exercise

- Goals
  - To define the most important questions for the field of particle physics
  - To identify promising opportunities to address them

- Do & Do-Not
  - Do: Address the questions the particle physics community wants to answer over the next two decades and how we plan to answer them
  - Do-Not: Prioritize activities (this is the goal of the P5)

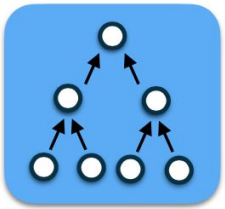
- The Snowmass process could include
  - Develop a framework of scientific questions that can form the basis of a future program
  - Survey experiments, facilities, and capabilities that would address these questions

from Young-Kee Kim, DPF chair



from Young-Kee Kim, DPF chair

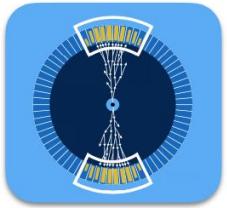
Computational Frontier: Topical Working Groups



CompF01

Experimental Algorithm Parallelization

Guiseppi Cerati (FNAL), Katrin Heitmann (ANL), Walter Hopkins (ANL)



CompF02

Theory Calculations & Simulation

Peter Boyle (BNL), Daniel Elvira (FNAL), Ji Qiang (LBNL)

CompF03

Machine Learning

Phiala Shanahan (MIT), Kazu Terao (SLAC), Daniel Whiteson (Irvine)

CompF04

Storage and Processing Resource Access

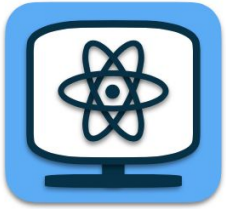
(Facility and Infrastructure R&D)

Wahid Bhimji (NERSC), Rob Gardner (U. Chicago), Frank Würthwein (UCSD)

CompF05

End User Analysis

Gavin Davis (U. Mississippi), Peter Onyisi (U. Texas at Austin), Amy Roberts (UC Denver)



CompF06

Quantum Computing

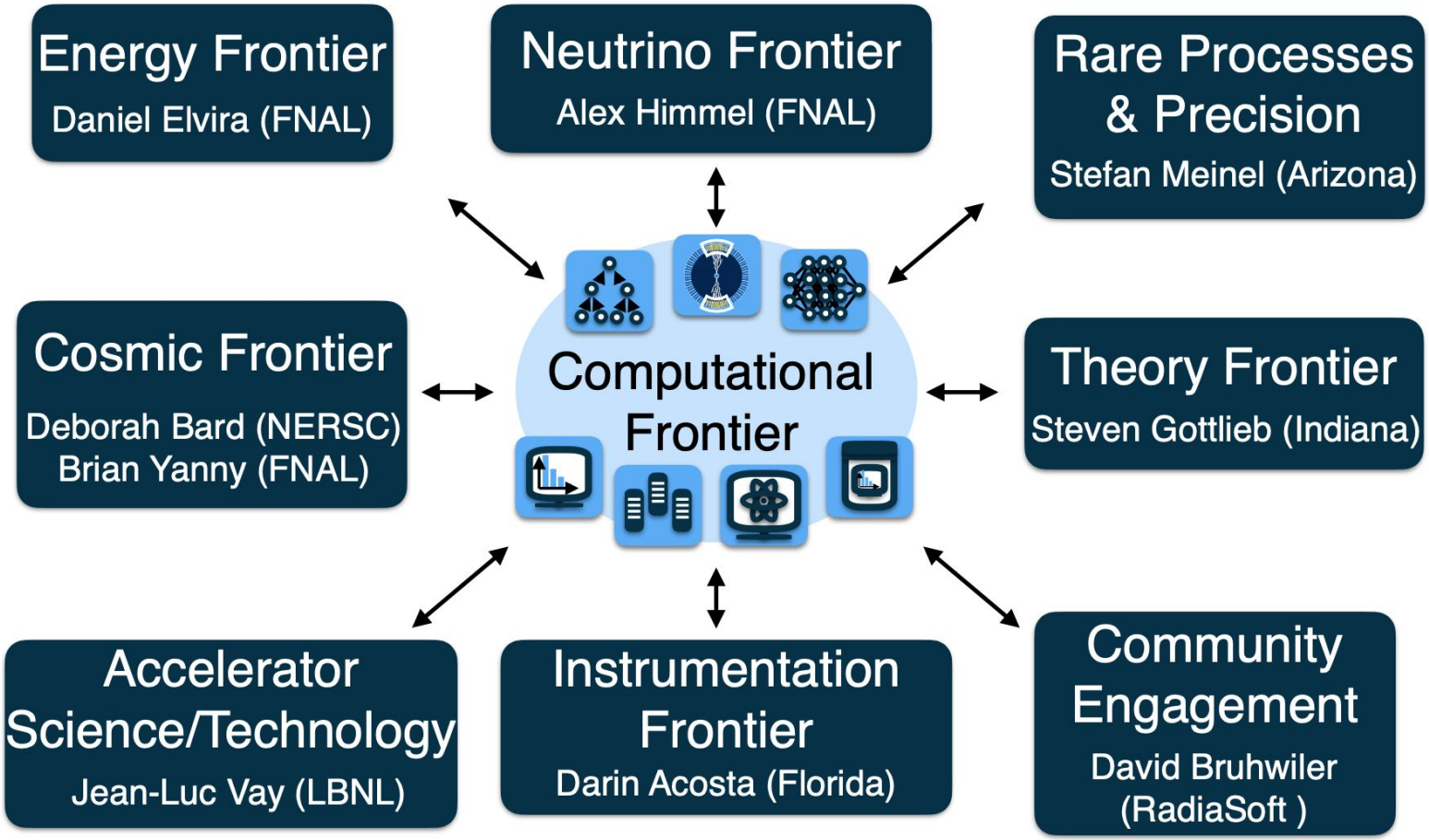
Travis Humble (ORNL), Gabriel Perdue (FNAL), Martin Savage (U. Washington)

CompF07

Reinterpretation & Long-term Preservation of Data and Code

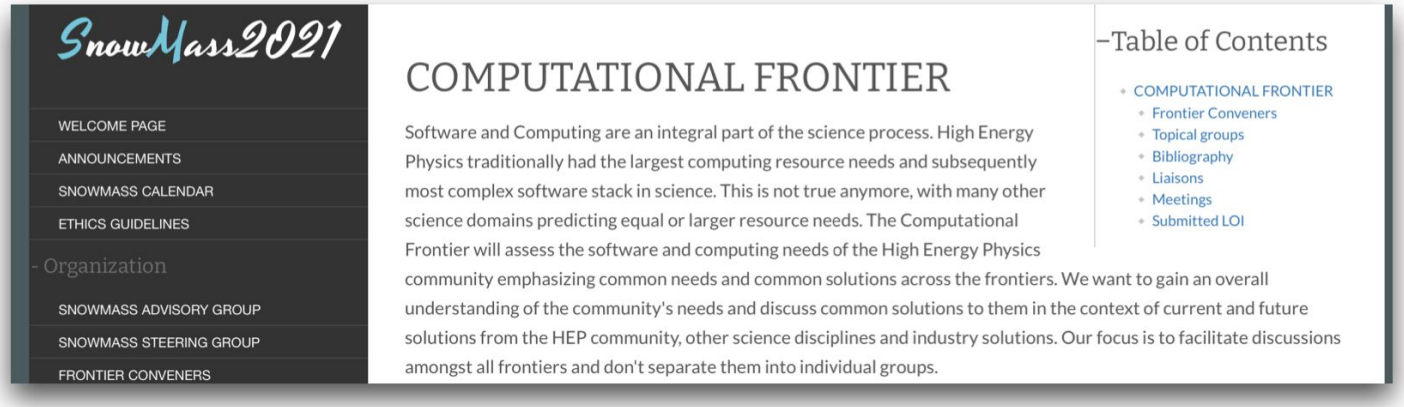
Kyle Cranmer (NYU), Mike Hildreth (Notre Dame), Matias Carrasco Kind (Illinois/NCSA)

Computational Frontier: Liaisons



Snowmass 2021: communication

https://snowmass21.org/computational/start



Join our Slack channels!

comp_frontier_topics
compf01-expalgos
compf02-theorycalcsim
compf03-ml
compf04-storeandprocess
compf05-useranalysis
compf06-quantum
compf07-preservation

Join our topical group meetings!

Join our email lists!

Topical groups

Name	Email List	Slack Channel
CompF1: Experimental Algorithm Parallelization	snowmass-compf01-expalgos[at]fnal.gov	#compf01-expalgos
CompF2: Theoretical Calculations and Simulation	snowmass-compf02-theorycalcsim[at]fnal.gov	#compf02-theorycalcsim
CompF3: Machine Learning	snowmass-compf03-ml[at]fnal.gov	#compf03-ml
CompF4: Storage and processing resource access (Facility and Infrastructure R&D)	snowmass-compf04-storeandprocess[at]fnal.gov	#compf04-storeandprocess
CompF5: End user analysis	snowmass-compf05-useranalysis[at]fnal.gov	#compf05-useranalysis
CompF6: Quantum computing	snowmass-compf06-quantum[at]fnal.gov	#compf06-quantum
CompF7: Reinterpretation and long-term preservation of data and code	snowmass-compf07-preservation[at]fnal.gov	#compf07-preservation

- Instructions to join a mailing list
- Instructions to join the Snowmass2021 Slack (at the end of the page)

Computational Frontier: Scope & Outcome

Our main time horizon should be ~10 years (HL-LHC, DUNE, LSST, etc.), but it is also useful to think about the next-to-next experiments and what R&D/funding opportunities we may need to be ready for the computing of the future.

Outcome:

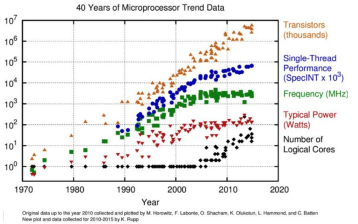
- Every topical group writes a document about their findings and points out opportunities and challenges
- The Computational Frontier writes one document consolidating all topical working groups
- Snowmass 2021 writes one document consolidating all frontiers

CompF1: "Experimental Algorithm Parallelization"

Computing landscape

The computing landscape has been transforming in the last few years: end of Dennard scaling, emerging of GPUs, building of exascale machines.

This means that adiabatic improvements from past solutions may not work or may be suboptimal. This is an opportunity to re-think how we process our data, and define new solutions for a higher science throughput.

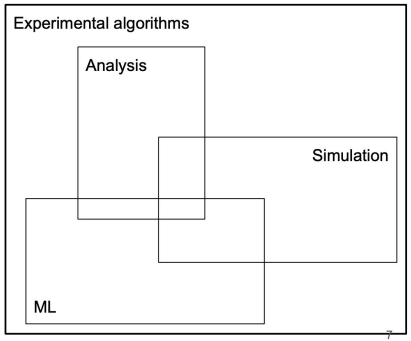


Relationships with other working groups

The definition of “experimental algorithms” is broad, covering the topics of other WGs.

We’ll focus on the area not covered by others. It means central (i.e. not analysis specific), non-ML algorithms whose inputs are experimental data (both offline and software trigger). This may have different meaning for different physics frontiers!

Frameworks are not specifically covered in other groups, and we’d be happy to discuss implications of parallel execution for frameworks in our WG.



Functional areas of our working group

- Parallelization of detector reconstruction algorithms, physics object reconstruction/calibration algorithms
- Utilization of CPU, accelerator hardware and what comes next in 5-10 years
- Developing better algorithms in addition to parallelization
- Portability solutions that support the same algorithm implementation on multiple hardware architectures

CompF1: workshop parallel sessions

CompF1: Experimental Algorithm Parallelization

Conveners: Giuseppe Cerati (Fermilab), Katrin Heitmann (Argonne National Laboratory), Walter Hopkins (Argonne National Laboratory)

Google doc for live ...

2:00 PM	**Cosmic Frontier Report** **Speaker**: Brian Yanny (FNAL) Comp-Cosmic liais...	15m	
2:25 PM	**Intensity Frontier Report** **Speaker**: Alex Himmel (Fermilab) 2020-08-10-dune-re...	15m	
2:50 PM	**Energy Frontier Report** **Speaker**: Vyacheslav Krutelyov (Univ. of California San Diego (US)) snowmassRECO_0...	15m	

CompF1: Experimental Algorithm Parallelization

Conveners: Giuseppe Cerati (Fermilab), Katrin Heitmann (Argonne National Laboratory), Walter Hopkins (Argonne National Laboratory)

Google doc for live ...

12:00 PM	**HEP-CCE Report** **Speaker**: Meifeng Lin (BNL) hep_cce_SnowMas...	15m	
12:25 PM	**IRIS-HEP Report** **Speaker**: Heather Gray (UC Berkeley/LBNL) Snowmass_IRISHE...	15m	
12:50 PM	**Challenges and opportunities**	40m	

CompF1: Workshop parallel session summary

Challenges discussion / All

- Definition of metrics for experimental algorithms and how to weigh them
  - Ease of implementation and portability
  - Performance
  - Longevity (design/optimize for machine available now or try to develop long lasting implementation)
- Traditional grid resources vs HPC centers
  - Main limitation of the grid: cannot point code to machine with a particular set of features
  - Using a variety of centers makes verification challenging given the diverse set of resources
- Training
  - Have to improve teaching of software development to be able to face future challenges
  - Most students focus on high-level programming, while efficient algorithm development often needs deeper knowledge of programming paradigms

Cross-frontier themes

- Many different needs, from different experiments or different algorithms
  - difficult to have a one-fits-all solution, even within a frontier
  - possible exceptions: accelerated FFTs in Cosmic and Neutrino frontier, real time processing (trigger/broker applications)
- Transitioning from HTC to HPC (or using both)
  - evolution of the programming model
- (Optimal) use of heterogeneous resources
  - how to keep the GPUs busy?
- Interplay between ML and traditional reco algorithms
  - switch to ML approaches vs rewriting algorithms
  - avoid separate workflows, ensure feedback between the two

CompF2: Theoretical Calculations and Simulation

- Six subtopics:
  1. Event Generators
  2. Accelerator Modelling
  3. Detector Modelling
  4. Theoretical calculations (Perturbative)
  5. Theoretical calcualtions (Lattice Gauge Theory)
  6. Cosmic simulations
- 98 subscribers to mailing list
  - Live google doc minute taking
- Biweekly teleconference meetings on Fridays 3pm EST, O(50) participants

- Background of preexisting community papers (HSF, USQCD) in areas
- Varying degrees of activity depending on community
  - High degree of self organisation in accelerator/beam modelling
    https://snowmass-compf2-accbeammodel.github.io
  - Event generators, Detector modelling appear to be self organising
  - Community LoI (Lattice) + multiple collaboration submissions
  - One CF topical phone call with perturbative community
    - Plan of record is a multiple Frontier community letter
  - Cosmic Frontier only really represented by HACC structure formation to date
    - Have attempted to reach further but not no wider engagement

CompF2: workshop parallel sessions

CompF2: Theoretical Calculations and Simulation

Conveners: Daniel Elvira (Fermilab), Peter Boyle (Brookhaven National Laboratory), ji qiang (LBNL)

2:00 PM **Event Generators** 15m

Status of LOI(s)

Speakers: Hugh Gallagher (Tufts University), Stephen Mrenna (Fermilab), Steven Dytman (Univ. of Pittsburgh), Steven Gardiner (Fermilab)

compf2-neutrino-a... Event Generator LO...

2:15 PM **Accelerator Modeling** 15m

Status of LoI(s)

Speaker: Jean-Luc Vay (Lawrence Berkeley National Laboratory)

Accel_Comp_LOIs....

2:30 PM **Detectors/Beam Modeling** 15m

Status of LoI(s)

Speakers: Krzysztof Genser (Fermilab), Vincent Pascuzzi (LBL)

SMCF02_DetSim_2...

2:45 PM **Theoretical Calculations (Perturbative)** 15m

Status of LoI(s)

Speaker: Prof. Andreas von Manteuffel (University of Mainz)

CF02-WorkshopAu...

3:00 PM **Cosmic Simulations** 15m

Status of LoI(s)

Speaker: Salman Habib (Argonne National Laboratory)

cosmo_sims_snow...

3:15 PM **Theoretical Calculations (Lattice Gauge Theory)** 15m

Status of LoI(s)

Speaker: Andreas Kronfeld (Fermilab)

ConpF2-20200810....

CompF2: Theoretical Calculations and Simulation

Conveners: Daniel Elvira (Fermilab), Peter Boyle (Brookhaven National Laboratory), ji qiang (LBNL)

12:00 PM **Theoretical Calculations and Simulation Panel Discussion (CF2)** 1h 30m

Panel Members

Event Generators - Hugh Gallagher, Stephen Mrenna
Accelerator Modelling - Eric Stern
Detector Modelling - Vincent Pascuzzi, Krzysztof Genser
Theory (Lattice) - Andreas Kronfeld
Theory (Perturbative) - Andreas von Manteuffel
Cosmic Simulations - Salman Habib

Moderator Questions

Classical computing:

Q1. For each field, please explain what science do you want to do, and do you expect to be able to achieve your goals with projected computing resources?

Q2. For each field, please estimate the fraction of the computing cycles that use algorithms that are highly parallel, and viable to port to parallel architectures (vector/accelerator etc..).

Q3. Do you currently use HTC, or HPC, or a mix of computing resources? (How) do you expect this to change in future?

Q4. How much human effort is required to support software development or adaptation for new machines?
(e.g.
How big is US effort? Is it part of an international effort?
How does this compare and fit in? Size of code?
Number of FTE years to port to acceleration? Languages considerations such as OpenMP offload, SYCL, or CUDA etc..?
Any difficulties? Plan for long-term code user support?)

Q5. Do you need DOE computing lab expert support for (software) R&D; funding such as ECP or SciDAC.
(e.g. How much? Do you have collaboration with applied math people? Is there any need for advanced numerical methods?)

Q6: What will your requirements for data storage be?
(e.g. Volume? Bandwidth? Distribution? Integrity guarantees? Life cycle? Data sharing?)

Machine Learning

Q1 For each field, do you expect to use machine learning in your main algorithms 10 years from now? What application benefits do you expect from MI in your area?

Q2 Please describe the degree to which you expect to use machine learning in 10 years. What level of certainty do you have?

Q3 Are you able to use commercial ML packages, like TensorFlow, Baidu, Theano, Torch, or do you need custom software? Do you need a programme of education in ML methods?

Quantum Computing

Q1 For each area, do you expect to engage with the development of quantum computing as scientific activity?

Q2 For each area, do you expect quantum computers to help solve your computational problems in the next 10 years? 20 years? Are quantum algorithms understood?

Q3 Is there activity or engagement with quantum algorithm programming?

CompF2: Workshop parallel session summary

Classical Computing

Event generation: Physics goals are extending parton showers, higher orders. Software challenges to achieve goals, entire rewrites, technically hard. Experiments need to trust to use.

Neutrinos: interactions with nuclear physics, lots of knowledge & intellectual input built in myriad bits of code

Work needed is several orders of magnitude more to control systematic errors

Pythia: software/physics improvements can be made, adaptation to HPCs would need significant development, verification/validation effort. Some elements are parallelizable others are not.

Accelerator modelling: Portion of community using HPC, with SciDAC and ECP support. Cross over to other energy accelerators and nuclear.

Detector modelling: Prior project parallelising (part of) Geant4 (GeantV) sets scale of effort. Celeritas is the current R&D effort to adapt G4 to HPC/GPU (ORNL, FNAL, ANL). Real thing may take in excess of 10 FTE if on time for HL-LHC

Fundamental role for nuclear physicists (Geant4 models). Need to estimate systematic errors.

Continuous development and support of Geant4 toolkit demands more effort.

Theory (Lattice): Able to use HPC, with SciDAC, ECP funding which need to continue/replace/evolve, along with local cluster investment. Funded algorithms, software development programs. ASCR serves purposes well.

Theory (Perturbative) Bespoke configurations of high memory long job run computing nodes need, not typically easy on HPC sites. Mathematics & algorithms even more important than computing power. Symbolic portions challenging, phase space integrals more HPC friendly. 100TB data not uncommon.

Cosmic simulation: Strong record of efficient HPC exploitation well served by ASCR, with SciDAC and ECP funding.

Data volume a significant issue. ~500k LOC applications.

Machine learning

Detector modelling:

Use ML in ATLAS fast calorimeter ~90% of runtime. Initially PCA, more moving to GANS. Very HPC friendly. Need education programme to disseminate across all ATLAS sites as so important. Education in summer schools. Some reinforcement learning.

Calorimeter simulation time LHCb 50% : move to ML.

Cosmic simulation:

Use in preconditioning, for example where algorithms self heal; learn a good solution guess.

Education is important, and direct interactions. Encourages physicists and computer scientists to be broadly interested and educated. We should be interested in each others work.

Theory (Lattice):

Attempts to use in MCMC where algorithm self heals; learn a good proposal. Breadth and interdisciplinary engagement. ECP has 3 applied math members, JLAB & William & Mary working together particularly well. ML LOI will be written.

Quantum Computing

Many interested in the science and quantum mechanics, but practical usage looks far off.
Cosmic: nature of bulk cosmic data inappropriate.
Event generation: Neutrinos (no), colliders some color reconnection possibilities and possible advances in parton showers.
Accelerator modelling: some proposals (Ji Qiang).
Lattice: intellectual interest in addressing sign problem, neutron star phases of matter, etc..Escape Euclidean space limitations. Not expected for workhorse calculations.

Career paths

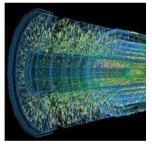
Ubiquitous problem in all sub-topics. Physicists develop the software with expert knowledge in both physics, algorithms and computing. Retention can be difficult, as often on soft money.

CompF3: Machine Learning

Particle physics-specific ML

Particle physics often has unique stats challenges
- Symmetries, boundaries, limits
- Data on manifolds or subsurfaces
- Sensitivity to uncertainties
- Heterogenous data structures

Simulation and ML



Simulation very expensive
- Fast simulation with ML
- Limitations and possibilities

Interpretability/validation

What has the machine learned?
- Reverse engineering ML strategy
- Exactness proofs
- Uncertainty measures
- Data reconstruction

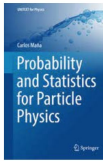
Community Tools

Standard tools and packages
- Software
- Data structures
- Adapting industry tools

Resource needs

What computing resources are needed?
- Real-time ML
- Clusters with GPUs, FPGA, etc
- Cloud processing
- Operations intelligence

Education



What do physicists need to know?
- Physics-ML specific courses?
- Outreach to community (ie CS/ML)
- Curation of open data
- Ethics and safety of AI

CompF3: workshop parallel sessions

CompF3: Machine Learning

Conveners: Daniel Whiteson (UC Irvine), Kazuhiro Terao (SLAC National Accelerator Laboratory), Kazuhiro Terao, Phiala Shanahan (Massachusetts Institute of Technology)

2:00 PM **Introduction** 10m
Speakers: Daniel Whiteson (UC Irvine), Kazuhiro Terao (SLAC National Accelerator Laboratory), Phiala Shanahan (Massachusetts Institute of Technology)

2:15 PM **Particle-Physics-Specific ML** 20m
Speaker: Patrick Komiske (MIT)
PTK_Snowmass_M...

2:40 PM **Interface between Simulation and ML** 20m
Speaker: Anja Butter
Snowmass_AButter...

3:05 PM **Interpretability and Validation of ML models** 20m
Speaker: Andrew Larkoski (MIT)
CompF3_Larkoski_...

CompF3: Machine Learning

Conveners: Daniel Whiteson (UC Irvine), Kazuhiro Terao (SLAC National Accelerator Laboratory), Kazuhiro Terao, Phiala Shanahan (Massachusetts Institute of Technology)

12:00 PM **Community tools and standards for ML work** 20m
Speaker: Daniel Guest
ML pipelines for H...

12:30 PM **Resource needs and management** 20m
Speakers: Javier Duarte (University of California San Diego), Javier Duarte
Snowmass_Resour...

1:00 PM **Education and engagement of ML skills** 20m
Speaker: Savannah Thais (Yale University)
snowmass_compu... snowmass_compu...

CompF3: Workshop parallel session summary

- Physics-specific Machine Learning (talk link)
  - Reproducibility and public dataset (CompF07) for a coherent development for shareable, reusable tools/algorithms
  - Data models/structures specific to physics research (CompF04)
- Interface with simulation (talk link)
  - Nice review of how ML used to boost existing event generation workflow, and generative models to the extent of unfolding with LHC example.
  - More ways of simulation and ML interface? Involve physics frontier liaisons to get more inputs
- ML resources and management (talk link)
  - Online (fast-ML) and offline (distributed ML etc.), commercial cloud, HPC, grid resources discussed: how do they scale in future?
  - Strong correlation to CompF01 (algo. parallelism) and CompF04 (storage/processing resource access)
- Education and engagement of ML skills (talk link)
  - Education/Career development, outreach/community building, public/benchmark dataset (CompF7), ethics and safety of AI
  - Large cross-cut with most of CommF working groups in all aspects above
- Interpretability and validation (talk link)
  - Use domain knowledge to maximize machine performance. Great example: more survey of instances from physics frontier liaisons?
  - Look outside HEP for “learning physics from machines” (CommF groups)
- ML community tools (talk link)
  - Many ML packages, options to distribute software environment, and ways to construct training and inference pipeline: deal with caos!
  - Accept evolving ecosystem, avoid mono-culture, and find common standards across experiments: survey explicitly collaborations through physics frontier liaisons?

CompF4: Storage and processing resource access (Facility and Infrastructure R&D)

CompF4: Functional areas

- Provide access to data for large scale central workflows
- Provide access to data for end user analysis

- Hierarchical storage
  - Access to long-term high-latency storage (tape ...)
  - Access to low latency storage (disk, ssd, nvme ...)

- Access to
  - CPU resources ➜ GRID, HPC, Cloud
  - Accelerator resources ➜ GRID, HPC, Cloud
  - Specialized AI hardware

- Interconnecting everything through Network

CompF4: Mandate

- What are the workflows related to storing and accessing data of the stakeholders?
- What different storage solutions/technologies are used, will be used?
- What is the technology evolution of storage solutions/technologies? What R&D is needed?
- What are the storage and access needs of the stakeholders?
- What are the storage resource needs of the stakeholders?
- What is the role of the network in these workflows?
- How are the solutions used by the community embedded/derived from solutions from industry/other science domains
- Recruit community members to represent the physicist/analysis perspective.

Future CompF4 activity

- During fall 2020 collect:
  - Further background material / reports
  - Wide community exemplars of resource and access needs
  - Questionnaire for resource needs and research topics

- ~Oct-Nov: Focussed workshop to further gather and synthesise these requirements

CompF4: workshop parallel sessions

CompF4: Storage and processing resource access (Facility and Infrastructure R&D)

Conveners: Frank Wuerthwein (UCSD), Robert Gardner (University of Chicago), Wahid Bhimji (NERSC, Berkeley Lab)

Monday Video | Monday ZOOM chat | Tuesday Video | Tuesday ZOOM chat

2:00 PM — **Discussion of group mandate** — 30m
Speakers: Frank Wuerthwein (UCSD), Robert Gardner (University of Chicago), Wahid Bhimji (NERSC, Berkeley Lab)
Community Comm... | snowmassAugust1...

2:30 PM — **Edge services** — 20m
Speaker: Joe Breen
Edge service Slides

2:50 PM — **Storage Evolution** — 20m
Speaker: Shigeki Misawa (BNL)
FrontierStoragePre...

3:10 PM — **Analysis Facilities** — 20m
Speaker: Brian Bockelman
Snowmass-AF-1.pdf

CompF4: Storage and processing resource access (Facility and Infrastructure R&D)

Conveners: Frank Wuerthwein (UCSD), Robert Gardner (University of Chicago), Wahid Bhimji (NERSC, Berkeley Lab)

Monday Video | Monday ZOOM chat | Tuesday Video | Tuesday ZOOM chat

12:00 PM — **Upcoming computing facilities for science** — 20m
Speaker: Nick Wright (NERSC, LBNL)
NWright_Snowmas...

12:30 PM — **AI hardware and facilities** — 20m
Speaker: Wahid Bhimji (NERSC, Berkeley Lab)
AI hardware slides

12:50 PM — **Future landscape of Science Networking** — 20m
Speaker: Eli Dart
20200811-dart-net...

CompF4: Workshop parallel session summary

The Talks

- Edge Services … Joe Breen (U.Utah)
- Storage Evolution … Shikegi Misawa (BNL)
- Analysis Facilities … Brian Bockelman (Morgridge/UW-Madison)
- Upcoming Computing Facilities … Nick Wright (NERSC)
- AI Hardware and Facilities … Wahid Bhimji (NERSC)
- Future Landscape of Science Networking … Eli Dart (ESnet)

3

What's next?

- We are considering a second workshop dedicated to CompF4 this "fall".
  - maybe with invites for specific topics to other working groups.
    - If conveners from other groups are interested, please get in touch!
- We encourage people to send us LOI's if you want to influence both the existence and agenda of such a possible 2nd workshop.

CompF5: End user analysis

Mandate

Consider:

- Types of resources needed for analysis facilities
- Use of accelerators
- Analysis libraries
- Data storage formats & dataset bookkeeping
- Programming languages
- Software for collaboration: version control, messaging
- "Real-time" analysis
- Long-term reproducibility and preservation

While also considering:

- Sustainability, both technical and human
  - Documentation
  - Training
  - Long-term software support & development
  - Integration with the broader ecosystem
  - Broad applicability to the field
  - Hardware facility evolution
- Interaction with other scientific fields & industry
  - Role of proprietary technology
  - Potential contributions to computing outside HEP
- User experience
  - Ease of use & setup
  - Scalability of technologies
  - Required training and broad applicability of training to other domains

Working Group Goals

- Produce a document that identifies impediments to end user analysis and potential ways to address such issues
  - With broad scope, e.g. fragmentation of knowledge across platforms, or lack of documentation as an equity concern
  - Informed by feedback from the broad user community
- Bear in mind that the detailed landscape will certainly change in the next 5-10 years, but hopefully requirements change more slowly
- Highlight potentially transformative avenues for R&D efforts
  - Including identifying gaps not covered by current work
- LOIs/white papers are encouraged to take a "big picture" stance
  - What core issues are being addressed?
  - It's OK to have a LOI that just identifies problems without having specific solutions in mind

End user analysis survey

Still collecting responses! Fill it out at https://forms.gle/rzvtNEGxhoXYAfKj7.

We'd love to see more

- Dark matter community voices
- Nuclear physics community voices
- Theory voices
- Experimental and Test Facility voices
- Early-career voices
- YOUR VOICE, if you haven't already filled out the survey!

Very early results will be shown at the start of the parallel session

CompF5: workshop parallel sessions

CompF5: End user analysis

If you're looking for connection information, register to get an email with connection information.

If you run into any issues, contact us through our slack channel #compf05-useranalysis or mailing list.

Email listserv@fnal.gov, no subject, message body: subscribe SNOWMASS-COMPF05-USERANALYSIS Firstname Lastname

Conveners: Amy Roberts (CU Denver), Gavin Davies (University of Mississippi), Peter Onyisi (University of Texas at Austin)

Monday Video | Monday ZOOM chat | Tuesday Video | Tuesday ZOOM chat

2:00 PM **Introduction** — 15m — CompF5 Zoom
Speakers: Amy Roberts (CU Denver), Gavin Davies (University of Mississippi), Peter Onyisi (University of Texas at Austin)
compf05-useranaly... | GDoc for questions...

2:15 PM **Equity issues & LOIs** — 10m
Speaker: Mu-Chun Chen (University of California, Irvine)
CEF_Townhall_CF3... | GDoc for questions...

2:25 PM **View from the HEP Software Foundation** — 10m
Speaker: Khoo
DAWG_ Scope & Ac... | GDoc for questions... | Google slides

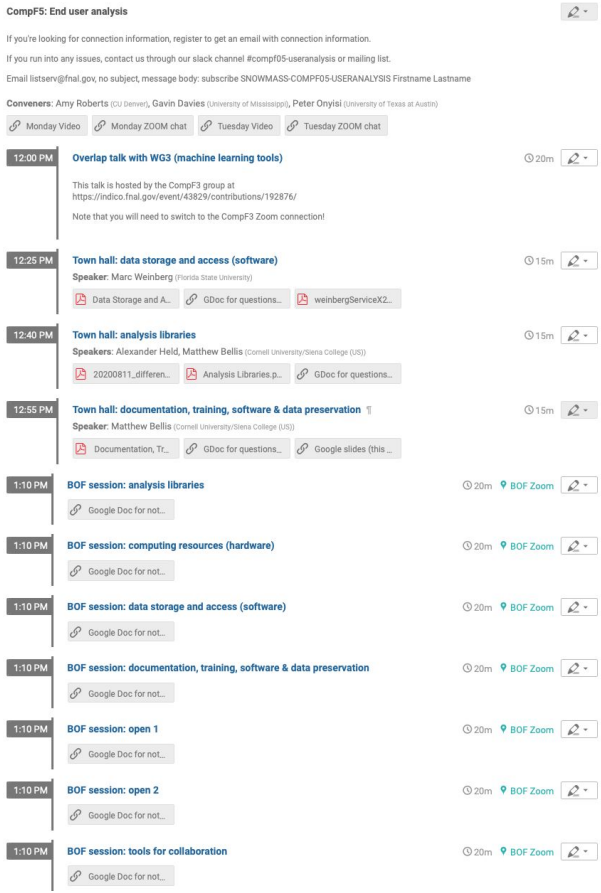
2:35 PM **Town hall: tools for collaboration** — 15m
compf05-useranaly... | GDoc for questions...

2:50 PM **Town hall: computing resources (hardware)** — 15m
compf05-useranaly... | GDoc for questions...

3:10 PM **Overlap talk with WG4 (analysis facilities)** — 20m

This talk is listed with the CompF4 group at https://indico.fnal.gov/event/43829/contributions/193065/

Note that you will have to switch to a different Zoom connection!



CompF5: Workshop parallel session summary

Resources & Data Access

- Need to connect large datasets to user code. What are the best models for that?
  - Many options - university/lab clusters, Grid submissions, central experiment servers...
  - Not obvious there is a one-size-fits-all solution across domains. Focusing on requirements for each domain is critical to answering this question.
- “Analysis facility” intriguing but needs definition
  - How interactive is “interactive”? Are people willing e.g. to reserve timeslots to get guaranteed resources?
  - Are there benefits to providing a higher-level interface? (Analysis-as-a-service, not generic batch queues with user-installed software?)
- Is there a need to figure out how to adapt more analysis code to GPU, if the CPU/GPU balance tilts?

Analysis & Inclusion

- Almost all HEP students & postdocs will have to deal with analysis software (both general and experiment-dependent). This has a major impact on their productivity and experience
- When software requires restricted knowledge to use, those who are not plugged in suffer
- Documentation & providing working examples of preserved complete analysis workflows may help to provide guidance to young scientists. This may be an area where containerization can help.

Data Formats, Libraries & Languages

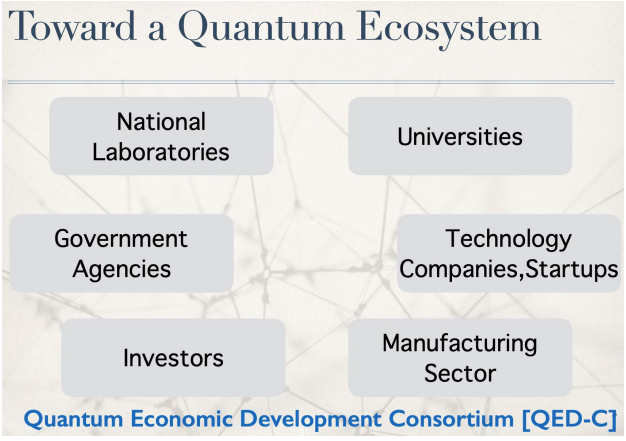
- Sometimes need to get data to/from other ecosystems, which don’t speak ROOT (ML being a famous example)
  - Mixed opinions about what else one might use
- We depend on many open-source libraries and projects
  - How does the community support work on this critical infrastructure?
- We also generate many small-scale projects
  - How do we ensure long-term support for software maintenance?
  - What is the appropriate balance of “spontaneous” vs. “centrally-planned” software in HEP?
- Analysis seems to be a mixed Python/C++ world (and this is where many people get trained). What are the opportunities and costs for other languages?

CompF6: Quantum computing

Why Quantum Computing / Quantum Information Science

- Quantum information science (QIS) is a major area of research emphasis in the DOE and for the nation at large.
- HEP has historically had a number of important roles in this field and we continue to play an important role in theoretical developments.
- We are becoming more important on the hardware side and have important contributions to make to QIS.
- Furthermore, there are a number of deeply interesting science questions quantum technologies enable us to ask in HEP.
  - In short - we can play an important role in this endeavor and our science will benefit.

Toward a Quantum Ecosystem



Areas for Snowmass

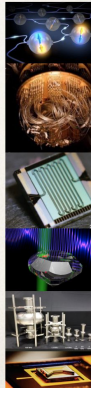
Thank you for the presentations and contributions

- This is a new area for Snowmass, HEP long-range planning
- Looking to identify and assess all "quantum" areas impacting and advancing HEP during the next decade
  - Quantum networks
  - Data analysis in HEP with quantum computers
  - NISQ-era quantum devices for HEP
  - Simulation of quantum field theories
  - Quantum simulation and hardware co-design
  - Tensor networks
  - Quantum information, error correction and holography
  - QFTs on AdS
  - Search strategies for new particles using SRF cavities
  - Quantum computing for event generators
  - Quantum algorithms for quantum sensing
  - Algorithm development for beyond NISQ-era devices

Quantum Comp. and Tech. for HEP

Summary

- QIS : an emerging and disruptive impact on HEP
- HEP anticipated to impact and be impacted
- Simulation, Communication, Sensing, Data
- Close collaboration with theory, HPC and experiment
- Close collaboration between Labs, Universities, Tech companies
- Close collaboration with other domain sciences, QIS
- New to Snowmass, next decade expected to be transformative



CompF6: workshop parallel sessions

CompF6: Quantum computing

Conveners: Gabriel Perdue (Fermilab), Martin Savage (INT), Travis Humble (Oak Ridge National Laboratory)

Live Google Docum... Video

2:00 PM **Introductory Remarks** 10m

Introductory Remarks

Speakers: Gabriel Perdue (Fermilab), Martin Savage (Institute For Nuclear Theory), Travis Humble (Oak Ridge National Laboratory)

2:10 PM **Tensor Networks** 12m

Speaker: Yannick Meurice (U. of Iowa)

meuriceTNQCHEP....

2:22 PM **Simulation of Quantum Field Theories** 12m

Snowmass_2020_...

2:34 PM **Quantum information, error correction and holography** 12m

2:46 PM **QFTs on AdS** 12m

Speaker: Daniel Harlow (MIT)

snowmasscomp.pdf

2:58 PM **Data Analysis in HEP with Quantum Computers** 12m

DataAnalysisinHEP...

3:10 PM **Quantum Computing Co-design for HEP** 12m

Speaker: Travis Humble (Oak Ridge National Laboratory)

Quantum Computi...

CompF6: Quantum computing

Conveners: Gabriel Perdue (Fermilab), Martin Savage (INT), Travis Humble (Oak Ridge National Laboratory)

Live Google Docum... Video

12:00 PM **Quantum simulation and hardware co-design** 12m

Speaker: Raphael Pooser (Oak Ridge National Laboratory)

Pooser_Quantum_...

12:12 PM **Quantum computing for event generators** 12m

Speaker: Christian Bauer (LBNL)

EventGeneration.pdf

12:24 PM **Quantum algorithms for quantum sensing** 12m

Comments from Andrew Sornborger

Speaker: Andrew Sornborger

12:36 PM **NISQ-era Quantum Devices for HEP** 12m

Speaker: Norbert Linke (Joint Quantum Institute, University of Maryland)

NML_snowmass_o...

12:48 PM **Algorithm Development for beyond NISQ-era devices** 12m

Speaker: Nathan Wiebe

1:00 PM **Quantum Networks for HEP** 12m

Nicholas A. Peters, Michael McGuigan, Panagiotis Spentzouris

Speaker: Nick Peters

quanutm networks ...

1:12 PM **Issues in HEP relevant to QML, decoherence, and quantum foundations** 12m

Andreas Albrecht, Andrew Sornborger, Patrick Coles

Speaker: Andreas Albrecht

Albrecht Snowmas...

1:24 PM **Search Strategies for new particles with SRF cavities** 12m

Search Strategies for new particles with SRF cavities

Speaker: Alexander Romanenko (Fermilab)

CompF6: Workshop parallel session summary

Overlaps

- Strong overlaps with the Theory Frontier and Instrumentation (quantum sensing)
  - May be worth pursuing joint workshops.
  - We are already in conversation with the Theory Frontier about joint and cross-listed whitepapers - need to begin this conversation with the Instrumentation Frontier.
- Within the Computing Frontier, potentially interesting overlaps with
  - CF2 - theory and simulations (how can quantum computing improve “traditional” HEP simulation, how can we cover the “gap” between detector simulation and pure QFT simulation, etc.?), SciDAC-esque collaborative teams?
  - CF4 - interesting facilities questions: how to spend the marginal “access” dollar in HEP? Do we prefer commercial cloud access to quantum resources? Should we buy in bulk and offer access through something like INCITE? Should we co-design and build HEP-centric hardware? (Or rather - given that we WILL build hardware, how do we invest in it for “production”? Do we?) Should we “commission” hardware from industry?

CompF7: Reinterpretation and long-term preservation of data and code

Functional (Focus) Areas:

- Public data
  - (comes in many forms … HepData, public likelihoods, CERN OpenData, data for education/outreach)
  - Tools for generating annotated public data and software
  - Tools for sharing data and software
- Not-yet-public Data
  - Tools for generating annotated “private” data and software
- Tools for combining results across experiments and frontiers
- Tools for archiving and re-running analyses (RECAST/REANA, …)
  - Internal-to-experiment and external “public” preservation

❖ Obvious overlap with all physics groups, as well as other computational areas
❖ Will try to join/convene as many joint sessions as possible moving forward

Group Mandate, Activities, Questions:

- Define the stakeholders and consumers of the data and software
  - What are the needs/requirements of the stakeholders?
    - (probably most difficult question to answer)
- What resources are needed?
  - e.g. long-term storage with external access, infrastructure for preserving executable code, etc.
  - metadata infrastructure
- What technologies are available or will be available, what is the technology evolution of these tools?
  - To be discussed in common with CompF5: End User Analysis:
    - version control
    - Containers/VMs
    - proprietary software/licenses

Group Mandate, Activities, Questions (cont.):

- How are/will the stakeholders use these technologies?
- What are the workflows that are used to combine results across experiments and frontiers?
- What tools are used/needed by the stakeholders to combine results across experiments and frontiers?
- What will the technological evolution of these tools look like?
- How are other science domains handling this topic?
- What are other science domains using, what is industry using?

Overall Goals:

- Raise awareness/visibility of preservation issues across frontiers
- Communicate current efforts/technologies to other groups/frontiers
- Mediate incorporation of these concepts and objectives into *all* reports and guidelines (where appropriate)
- Production of general guidelines (aspirations?) for preservation of scientific results

CompF7: workshop parallel sessions

CompF7: Reinterpretation and long-term preservation of data and code

Conveners: Kyle Cranmer (NYU), Matias Carrasco Kind (NCSA/University of Illinois), Mike Hildreth (Notre Dame University)

Discussion Notes | zoom_chat_Monda... | zoom_chat_Tuesda... | zoom_video_Mond... | zoom_video_Tuesd...

2:00 PM	**Introduction/Overview** **Speakers**: Kyle Cranmer (NYU), Matias Carrasco Kind (NCSA/University of Illinois), Mike Hildreth (Notre Dame University)	10m
2:10 PM	**Open is not Enough / Analysis Preservation and Reinterpretation (RECAST)** **Speaker**: Lukas Heinrich (New York University) Snowmass2020.pdf	20m
2:35 PM	**REANA, Data and Analysis Preservation at CERN** **Speaker**: Tibor Simko (CERN) preserve-to-reuse-s...	15m
2:55 PM	**Publishing likelihoods** **Speaker**: Matthew Feickert (Southern Methodist University) Feickert_2020-08-1... | GitHub source repo | talk: Likelihood Pub...	10m
3:10 PM	**Evolution of Preservation: Cloud Services and Analysis Facilities**	20m

CompF7: Reinterpretation and long-term preservation of data and code

Conveners: Kyle Cranmer (NYU), Matias Carrasco Kind (NCSA/University of Illinois), Mike Hildreth (Notre Dame University)

Discussion Notes | zoom_chat_Monda... | zoom_chat_Tuesda... | zoom_video_Mond... | zoom_video_Tuesd...

12:00 PM	**Analysis Preservation Bootcamp** **Speaker**: Leonora Vesterbacka (NYU) APWorkshop_Sno...	15m
12:20 PM	**Experiences with Open Data** **Speaker**: Cari Cesarotti (Harvard) openData_snowma...	10m
12:30 PM	**Experiences with Open Data** **Speaker**: Anthony Badea (Harvard) The World of Open ...	10m
12:40 PM	**Town Hall discusion on Astrophysics, Dark matter, Cosmology, etc.**	15m
12:55 PM	**Joint Town Hall with CF5: Data Analysis**	15m
1:10 PM	**Organization moving forward**	20m

Thank You

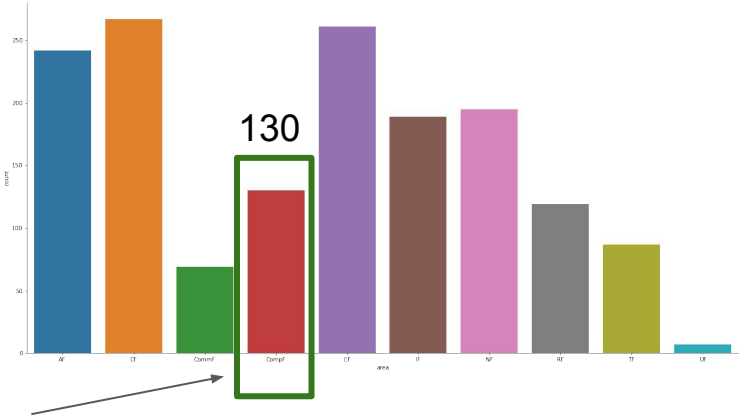
- ✦ We would like to express our thanks to all those who made this a successful workshop:
- ✦ Our topical group conveners
- ✦ Our liaisons to the other frontiers and Snowmass Early Career
- ✦ Especially, all the community members who joined this workshop. We had over 300 registrants!

- ✦ We Identified a number of issues that might benefit from their own workshops
  - Performance Portability Challenge
  - Analysis Facilities
  - Education and Engagement; Career Development
  - Quantum Computing
  - Computing Models and Opportunities for Technological Progress
  - Fostering International Exchange/Coordination in Software & Computing

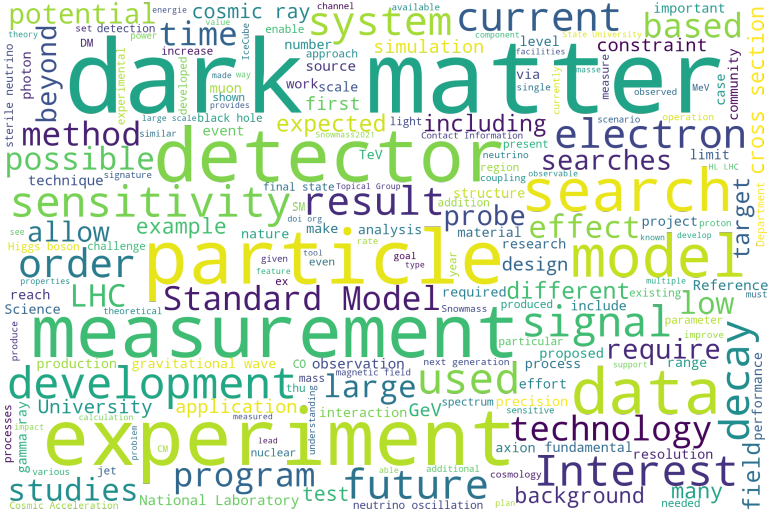
- ✦ Possible workshops (to be arranged):
  - Performance portability
  - Programming/scripting languages
  - Future hardware landscape
  - Algorithm-hardware co-design
  - Your great ideas…

LoIs - Over 1800 LoIs submitted by August 31

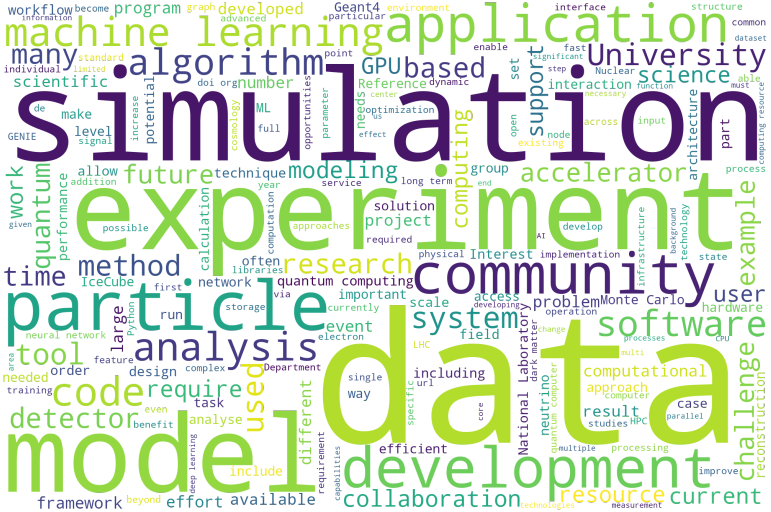
Letters of Interest (LoIs) are informal documents intended to be useful in the first stages of the Snowmass study. They will help Snowmass conveners to prepare the Snowmass Community Planning Meeting that will take place on October 5-8, 2020 virtually. LoIs could include opinions, interests and proposals that could further be studied. They should contain a maximum of 2 pages of text, plus relevant bibliography. There is no further requirement on the format, but **the primary contact information (name and email address)** should be included in the LoI. Please make these as simple and easy to read as possible. Authors of the letters are welcome to make a full writeup for their work as a contributed paper and submit it to the Snowmass proceedings. However, a contributed paper is not required.



All frontiers



Computational Frontier



Credits: Gordon Watts: https://gordonwatts.github.io/snowmass-loi-words/

Snowmass Community Planning Meeting - Virtual

5-8 October 2020
Virtual
US/Central timezone

Contact

kiburg@fnal.gov

boj@fnal.gov

Over the next year, the U.S. particle physics community will be engaged in Snowmass 2021, an in-depth process to define the most important questions for our field and to identify promising opportunities to address these questions in a global context.

The primary goal of the Community Planning Meeting is **to develop plans and steps to take ("Snowmass Planning") between October 2020 and the Snowmass Community meeting in July 2021, leading to a final report in October 2021.**

Other goals include:

- Inspire the community about the field, and encourage them to engage broadly in the Snowmass process
- Inform the community about plans from other regions and from related fields and planned Snowmass activities
- Listen to the community
- Provide space for members across the field to talk to each other and to discuss, promote, and develop new ideas
- Establish cross working-group connections and identify gaps

We very much look forward to the community's strong participation in the Community Planning Meeting. Any information concerning Snowmass is available at snowmass21.org.

Snowmass is an APS process. All participants of the CPM must agree to follow the APS Code of Conduct.

Please note that all times in the timetable are displayed in **US Central Time (UTC-6)**.

Starts Oct 5, 2020, 8:00 AM
Ends Oct 8, 2020, 5:00 PM
US/Central

Virtual

https://indico.fnal.gov/event/44870/