Computing within the 2021 Snowmass process

V. Daniel Elvira[1,*] on behalf of the Snowmass Computational Frontier and topical group conveners, and of all participants.

[1]Fermi National Accelerator Laboratory, Batavia, P.O. Box 500, Batavia, Illinois, 60510, USA.

Abstract. We present different elements of the 2021–2022 Computational Frontier (CompF) effort within the Snowmass process, including organization, events, and highlights of the final report [1]. With focus in the next 10–15 years, we analyze the impact of newly established and emerging technologies, identify challenges within theoretical, experimental and observational research programs, and present the report recommendations. We emphasize the main recommendation, which is the formation of a Coordinating Panel for Software and Computing (CPSC).

1 Introduction

This article presents the highlights of the Computational Frontier (CompF) activities and final report [1] in the context of the 2021–2022 High Energy Physics (HEP) Snowmass process in the United States. In the interest of space, we will refer to the CompF Report most of the times, rather than reproduce its long list of references here. Section 2 briefly describes the CompF organization and programmed events, including those taking place during the Community Summer Study Workshop (CSS) held in Seattle in July 2022 [2]. The impact of newly established and emerging technologies such as heterogeneous computing, Artificial Intelligence (AI) and Quantum Computing (QC) is discussed in Sect. 3. Section 5 reproduces the recommendations listed in Ref. [1], the main one being the establishment of an American Physical Society–Division of Particles and Fields (APS–DPF) sponsored Coordinating Panel on Software and Computing (CPSC). Next, in Sect. 6, the CompF report is placed in the global context by drawing comparisons with the recommendations of the European Strategy for Particle Physics [3]. To summarize, Sect. 7 addresses the post-Snowmass era, including the process of creation of the CPSC, and the interaction with the Particle Physics Projects Prioritization Panel (P5) [4].

2 The process

Information about organization details, events, and documentation, including Letters of Intent (LOIs) and White Papers (WPs) is available in the Snowmass twiki page [5].

The CompF activities were split in seven groups categorized either by S&C topic or technology. The latter category recognizes the fact that some computing technologies have become or have the potential to become transformational, improving significantly the precision

*e-mail: μdaniel@fnal.gov

of our theoretical calculations and experimental measurements while simultaneously boosting computing performance. This motivated the creation of working groups with focus on Machine Learning (ML) and Quantum Computing (QC). The groups targeting S&C topics relevant to HEP were: Experimental Algorithm Parallelization, Theory Calculations and Simulation, Storage and Processing Resource Access, End-user Analysis, and Reinterpretation and Long-term Preservation of Data and Code.

Computing is an enabler of the HEP physics programs and the HEP field consistently makes contributions to computer science, shaping technologies and methods while developing HEP-specific solutions in partnership with computing researchers. Communication across frontiers was considered a high priority, thus liaisons with all the physics and other enabling frontiers were appointed, as well as with the Community Engagement Frontier. The former included the Theory, Energy, Rare Processes and Precision, Neutrino, Cosmic, Undergroud Facilities, Instrumentation, and Accelerator Science and Technology Frontiers. A series of events were held to organize the process and discuss progress. Biweekly conveners meetings started in April 2020, a CompF Workshop was held in August 2020, and many topical-group-specific meetings took place. A workshop dedicated to the S&C needs of small experiments was also held. CompF received 136 LOI and 71 WP submissions. The main Snowmass event was the July 2022 Community Summer Study (CSS) in Seattle [2], where CompF organized individual sessions for each topical and technology group, sessions for small and for big experiments, a joint meeting with industry partners, and either organized or co-organized plenary sessions on AI, the future of computing in HEP, quantum science and technology.

The final report was written between the months of April and November 2022 [1].

3 Newly established and emerging technologies

This section delves into the evolution of computer hardware and its impact on HEP computing infrastructure and software. It also discusses the irruption of artificial intelligence and the revolutionary potential of quantum computing.

3.1 Computer hardware evolution

Dennard Scaling is a law in computing that states that the power used by a silicon device per unit of volume is independent of the number of transistors within the volume. Moore's law states that, historically, transistor density in a silicon device has doubled every approximately two years. Simultaneously, clock speed in CPUs increased approximately 1000 times between 1970 and 2000. All this combined has allowed computer speed to double every two years while the cost of hardware has halved. Sometime in the early 2000's, these laws started to break down due to the effect of leakage current, and the size of the devices approaching the atomic scale. Faster clocks demand significantly more power and generate more heat, therefore bringing budgetary and environmental concerns. The result is a paradigm shift in computing architectures, an evolution towards heterogeneous systems with multi-core machines, using co-processors (e.g., GPUs) and more complex memory configurations. Making GPU resources available for prototyping, finding code portability solutions across platforms, as well as negotiating with HPC centers a resource allocation process amenable to HEP schedules, are essential steps for a smooth and successful technology transition.

A consequence of computer heterogeneity is a departure from the stability of the past, when the same old code would run cheaper and faster in future machines without too much adaptation or re-engineering. Research investment is fundamental in most projects to bridge the gap between resource needs and availability, and development effort is critical to deliver upgraded production-quality software on time. These tasks are heavy in person-power with rare and expensive talents, making recruitment, training, and retention an organizational and financial challenge.

3.2 Artificial intelligence

Artificial intelligence in HEP has gone from an emerging to a mainstream technology. It is bringing revolutionary changes in the way we operate facilities, perform theoretical calculations, and analyse experimental data. Although multivariate analysis has been commonplace in HEP since the 1990's, modern Machine Learning (ML) techniques were not even mentioned in the Snowmass 2013 Report [6]. The charge to the ML CompF group was to identify the many ML tools and applications to HEP in both online and offline settings, understand what software development and specialized hardware would be needed, and evaluate how different methods would be applied in large and small scales.

Even if industry is driving ML research these days, the level of precision and the need for uncertainty quantification, validation of simulated data, and the interpretation of experimental results in the context of theory predictions, require HEP-dedicated solutions. HEP direct contributions to the most popular ML platforms would therefore be beneficial, as well as partnerships with industry to develop HEP custom solutions based on their advanced hardware accelerators.

An incomplete list of examples of ML applications to HEP includes generative models to accelerate various steps in the simulation chain (physics generators, detector simulation, lattice gauge theory), unsupervised classification for anomaly detection in searches for new physics, gradient descent in instrumentation to optimize detector designs, ultra low-latency inference for control in particle accelerators.

3.3 Quantum computing

Quantum computing (QC) is an emerging technology with the potential to produce a paradigm shift, a revolution in every field of research, including particle physics. Quantum computers (QCs) use interference and entanglement in calculations where the states are represented with qubits, bits in a two-dimensional Hilbert space. A fundamental obstacle to make QCs viable at a large scale is the environmental noise that affects the quantum states, which results in quantum decoherence. Presently, quantum computing is in the near-term noisy intermediate-scale quantum era (NISQ). There is a rapid development of software for quantum computers with interested parties also working on benchmark examples that would eventually serve to tailor quantum computers for HEP applications. The field should prepare for the post-noisy era in the timescale of the next Snowmass, given the possibility that fault tolerance is achieved in about a decade.

4 A brave new world

Computing in HEP is a global endeavor accelerating discovery across different domains, as well as across nations. As a consequence, effort must be coordinated with partners. The complexity of HEP experimental and observational instruments and the size of data volumes demand more computing power, data storage, and expensive algorithms in a world of flat budgets. In this context, S&C in HEP has evolved into an integral part of the scientific apparatus. It is not a service but an essential component of a detector or a telescope. The complexity, the physics content, and impact of computing in HEP is commensurable with that of the rest of the experimental instruments, thus we need to exploit the synergies between detector and software design. While it is customary to use simulation to design and optimize detectors for best physics performance, it is important to invest more effort to simultaneously optimize detectors for computing performance.

As noted before, hardware evolution calls for a redesign of the computing model for HEP. Experimental software frameworks need to use heterogeneous resources, locally and remotely, including government and commercial facilities; the data management model needs to be revisited to handle data access, transfer, and processing across a diverse set of computing systems; almost every piece of software, including common tools for event generation, detector simulation, and reconstruction needs to be adapted or re-engineered; suitable portability tools need to be identified and utilized to support different computing platforms without code duplication; infrastructure for AI/ML needs to be developed.

5 Findings and recommendations

The goal of the Snowmass community planning exercise was to provide a 20-year global vision and a 10-year execution plan for HEP. Consequently, the full report discusses mostly facilities to be built after 2035. However, near-term programs are not a given, thus the CompF report focused mostly on the upcoming 10–15 years to make the case for the funding to bridge the needs versus resources gap of near-future programs. Additionally, recommendations address the near-term computing needs of feasibility studies, as well as of accelerator and detector R&D for future facilities. The actual computing needs of post-2035 physics programs were not addressed, since rapidly evolving technologies make projections inaccurate.

Before listing the CompF findings and recommendations, it is important to highlight two distinctive features of computing. One is that computing technology and software paradigms change on a much faster scale than the interval between Snowmass community planning exercises. The second is that S&C efforts are neither funded nor managed as projects, unlike facilities and experimental devices. These features motivate in a natural way the main CompF recommendation for a Coordinating Panel for Software and Computing (CPSC). In addition to the main ask, CompF has identified critical challenges that are limiting the physics output of the US HEP community, deserve immediate action, and lead to four additional recommendations (1–4).

5.1 A Coordinating Panel for Software and Computing (CPSC)

Main recommendation: CompF recommends the creation of a standing Coordinating Panel for Software and Computing (CPSC) under the auspices of DPF, mirroring the panel for advanced detectors (CPAD) established in 2012. The charge and scope of this panel will be discussed within the community in subsequent months. The panel would promote, coordinate, and assist the HEP community on Software and Computing, working with scientific collaborations, grassroots organizations, institutes and centers, community leaders, and funding agencies on the evolving HEP S&C needs of experimental, observational, and theoretical aspects of the HEP programs. The scope should include research, development, maintenance, and user support. There is also a recommendation for the CPSC to setup a study group on Diversity, Equity, and Inclusion (DEI) in HEP computing. The CPSC will be neither a center nor an institute and will not manage budgets.

5.2 Software improvement, maintenance, and user support

Finding #1: Long-term development, maintenance, and user support of essential software packages cutting across project or discipline boundaries are largely unsupported.

- A new structure is needed to fund modernization, maintenance, and user support of existing tools. Grants typically only fund ground-breaking R&D or development of new software.
- Examples: Event generators and simulation tools such as Geant4[7], which do not belong to a particular facility, experiment, or survey; S&C tools associated to one or more experiments; data and software preservation for utilization after an experiment has ended.

Recomentation #1: The US HEP community should take a leading role in long-term development, maintenance, and user support of essential software packages with targeted investment.

5.3 Research and development

Finding #2: R&D for software and computing cutting across project or discipline boundaries receive insufficient support.

- Computational HEP is a vehicle for cross-cutting R&D. Supporting research in this area at a variety of scales would be broadly impactful.
- Examples: S&C for theoretical calculations and physics generators; cosmological, accelerator, and detector modeling; machine learning methodology and hardware ecosystems; and algorithms and packages across experiment boundaries.

Recommendation #2: Through existing, reshaped and expanded programs, R&D efforts cutting across project or discipline boundaries should be supported from proof of concept, to prototype, and to production.

5.4 Heterogeneous computing

Finding #3: Scarcity of personnel and expertise jeopardizes full and optimal use of heterogeneous and high-performance computing (HPC) resources.

- Most HEP software runs on a single computing platform, making it difficult to use the diversity of hardware accelerators and computing resources, such as cloud, HPC, etc.
- To satisfy the needs of inherently serial algorithms that are still transitioning towards computing accelerators or are not cost-effective to port, an appropriate level of traditional CPU-based hardware should coexist with the modern systems.

Recommendation #3: Support computing professionals, researchers, and physicists to conduct code re-engineering and adaptation to enable use of heterogeneous resources effectively.

5.5 Training and career paths

Finding #4: Investment in training and career paths within HEP for S&C researchers is insufficient.

- Sustainable efforts in HEP computation require continual recruitment and training in the context of an environment that is diverse, inclusive, supportive, and welcoming.
- Faculty and staff positions for physicists with expertise in S&C for HEP are scarce and person-power shortfall endemic.

Recommendation #4: Make a strong investment in career development for HEP S&C researchers to ensure future success.

6 The global context

The US is a leading member of the HEP international community and contributes efforts which are well integrated into the global program. Recently, Europe carried out a process to define a strategy for particle physics. Unsurprisingly, the recommendations in the area of S&C for HEP in the two regions resonate with each other. Computing challenges are immediate and need to be addressed through a vigorous R&D program. Interestingly, the Europeans also emphasize the need to take a more holistic approach to detector design, including the impact on computing resources. We list here a few quotes from the computing section of the European Strategy Physics Briefing Book [3]. The document points out that "computing challenges are immediate and need to be addressed through a vigorous R&D program", and "there is a clear need to strengthen existing R&D collaborative structures, and create new ones, to address future experimental challenges of the field post HL-LHC". It recognizes that "it is becoming increasingly vital to take a more holistic approach to detector design which includes impact on computing resources". It also identifies training and career development challenges as posing a growing risk to the field of particle physics, and calls attention to "a limited amount of success in attracting, developing and retaining instrumentation and computing experts".

Since the 2018–2020 European Particle Physics Strategy Update, initiatives were launched to address the computing challenges. One of them is the European Open Science Cloud (EOSC) environment [8] for hosting and processing research data to support European Union (EU) science. This is a federation of existing infrastructures, national data centers and research infrastructures that allows seamless access to data and interoperable services, enabling researchers and citizens to use the data and publish the results of their analysis. A second initiative is the European Science Cluster of Astronomy and Particle Physics Research (ESCAPE) [9], a program to develop common infrastructure solutions for both particle physics and astronomy facilities, push state of the art data management and computing R&D, as well as to address challenges for data preservation, sustainability, and open access to data.

7 Final thoughts

The Snowmass CompF exercise was carried out in an open and transparent manner; it was a successful participatory experience. A primary aspiration of the conveners is that the report represents the thinking of the community faithfully. It will be a primary tool to communicate what computing can do to accelerate particle physics discovery, and what the computing in HEP community needs to succeed in pursuing that goal. We hope the report serves as valuable input to the Particle Physics Projects Prioritization Panel (P5) [4], as well as to funding agencies.

Computing plays a fundamental role in the execution of current and future HEP theoretical, experimental, and observational programs. Computing should not be understood as a service any longer, but accepted as an element of the scientific apparatus in HEP experiments and surveys. Modern computing architectures, as well as newly established and emerging technologies are changing the face of particle physics, but they are also bringing significant transitional challenges. It is therefore essential to provide sufficient and timely support for software, from the R&D and prototyping stages, to production-quality delivery and deployment. Recruitment and retention of individuals with the rare and expensive talents needed to deliver computing for HEP must be improved by promoting an inclusive, diverse, and welcoming environment. Training opportunities and viable career paths must be created for the young people who choose to develop computing expertise within HEP. We need to strike a balance between R&D and improvement, as well as maintenance, of existing tools.

The Snowmass participants recommended the creation of a Coordinating Panel for Software and Computing (CPSC). The CPSC will work within the mandate of the DPF and DPF by-laws. It will be a standing committee of DPF and will report to its Executive Committee (EC). Its main goal will be to facilitate communication among S&C stakeholders, to help identify issues and problems, and to coordinate responses among subsections of the HEP computing ecosystem. At the time of submission of this paper, the DPF leadership was close to announce the plan to appoint a Formation Task Force (FTF) and start the member nomination process. The FTF will function as an *ad oc* subcommittee of the DPF EC. The charge to the FTF is to write and submit a report to the EC that will serve as a formal mandate and organizational plan for the CPSC.

8 Acknowledgments

V. Daniel Elvira is supported by Fermilab, operated by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy. This paper is for the most part a summary of the Snowmass Computational Frontier Report [1], authored by: V. Daniel Elvira, Steven Gottlieb, Oliver Gutsche, Benjamin Nachman (frontier conveners), S. Bailey, W. Bhimji, P. Boyle, G. Cerati, M. Carrasco Kind, K. Cranmer, G. Davies, R. Gardner, K. Heitmann, M. Hildreth, W. Hopkins, T. Humble, M. Lin, P. Onyisi, J. Qiang, K. Pedro, G. Perdue, A. Roberts, M. Savage, P. Shanahan, K. Terao, D. Whiteson, F. Wuerthwein (topical group conveners).

References

[1] V. D. Elvira, S. Gottlieb, O. Gutsche, B. Nachman, *et al.*, "The Future of High Energy Physics Software and Computing", Computatinal Frontier Report contribution to Snowmass 2021, https://arxiv.org/abs/2210.05822 [hep-ex], https://doi.org/10.48550/arXiv.2210.05822 (accessed 28/12/2023).

[2] Snowmass Community Summer Study Workshop, July 17-26, 2022, University of Washington, Seattle, USA. http://seattlesnowmass2021.net (accessed 28/12/2023).

[3] R. K. Ellis *et al*, "Physics Briefing Book: Input for the European Strategy for Particle Physics Update 2020", https://arxiv.org/abs/1910.11775; CERN-ESU-004 (accessed 28/12/2023).

[4] Particle Physics Projects Prioritization Panel (P5) report in preparation, https://www.usparticlephysics.org/p5 (accessed 28/12/2023).

[5] Computational Frontier, Snowmass Process 2021, https://snowmass21.org/computational/start (accessed 28/12/2023).

[6] L. A. T. Bauerdick, S. Gottlieb, *et al*, "Planning the Future of U.S. Particle Physics (Snowmass 2013): Chapter 9: Computing", https://arxiv.org/abs/1401.6117 [hep-ex], https://doi.org/10.48550/arXiv.1401.6117 (accessed 28/12/2023).

[7] S. Agostinelli, *et al.*, "Geant4–a simulation toolkit", NIM **A 506** (2003), 250. https://doi.org/10.1016/S0168-9002(03)01368-8 (accessed 28/12/2023).

[8] European Open Science Cloud, https://eosc-portal.eu (accessed 28/12/2023).

[9] European Science Cluster of Astronomy and Particle Physics ESFRI, https://projectescape.eu (accessed 28/12/2023).