AccelNet: Future Research Software Collaboration for High Energy Physics

David Lange, Shih-Chieh Hsu

1 Introduction

The quest to understand the fundamental building blocks of nature and their interactions is one of the oldest and most ambitious of human scientific endeavors. Unanswered questions that are within reach for scientists to answer include the origin of dark matter, the hierarchy of known particle masses, and the properties of the recently discovered Higgs Boson. These are some of the science drivers identified by high-energy physics researchers as being the highest priority to address by next-generation experimental facilities.

Scientific facilities under construction like the high-luminosity Large Hadron Collider (HL-LHC) at CERN promise a major step towards answering these and other important questions. Beyond the HL-LHC, proposals for the next generation of facilities, including the Future Circular Collider (FCC) at CERN, the International Linear Collider (ILC) in Japan, and the Circular Electron Positron Collider (CEPC) in China, promise to provide scientists a very detailed understanding of the Higgs and any new discoveries made at experiments already underway.

Looking forward, the international research community will work together on the detailed analyses needed to reach consensus opinions that define these next-generation facilities. The APS Division of Particles and Fields (DPF) is organizing a community planning, "Snowmass", exercise to use as input in the funding agency prioritization process. This will take place in the early 2020s, together with similar international processes. These studies require many of the same novel software tools as the HL-LHC community is developing.

To fully realize the discovery potential of these experiments a comensurate software investment is required. Collider experiments in HEP provide researchers with massive data pipeline that is a challenge for real-time data collection, for reconstructing the underlying physical interactions from sensor data, and for analysts understanding the ensemble of results. HEP cyberinfrastructure researchers are enabling scientists to "harness the data revolution" at the Exabyte scale. While a full scale system is relevant only as experiments are realized, tools with the capability and flexibility to make detector design choices quickly are essential. The community is pursuing R&D required,including developing novel approaches to methods for detector simulation, data reconstruction algorithms, data analysis techniques, and processing frameworks. This investment builds on prior conceptualization work that actively build collaborations around an R&D plan in the area of application software for HEP.

We propose the Future Research Software Collaboration (FRESCO) to enable budding international research and development connections to become the strong collaborations needed for today's R&D to also support the Snowmass process towards defining the next-generation experiments. This will establish a direct link from researchers focused on current-generation experiments to the needs and goals of future facilities (and, importantly, visa vera). HEP has two kinds of well established networks where researchers form close collaborations: international experimental collaborations; and country-specific or regionally based networks. FRESCO will build a network of these networks linking software domain experts and research communities that require their expertise. The overarching goal is to extend current software community-building efforts to include also researchers working towards the next-generation of experiments in high-energy physics.

2 Motivation

Proposals for next-generation collider facilities promise to provide scientists a very detailed understanding of the Higgs and any new discoveries at HL-LHC. Goals include operating with sufficient intensities to produce huge data samples, with hundreds of thousands of Higgs particles and hundreds of billions of Z bosons per year. Achieving this means designing systems to orchestrate the analysis of 100 times more data than at today's experiments.

This scientific opportunity would be unprecedented in HEP. To fully realize the discovery potential of these experiments, a major computational challenge must be addressed. For example, the HL-LHC will provide close to an exabyte of data to analysts each year even after selecting only the most interesting 0.01% of all detected events. This represents one of the largest scale data science challenges in scientific research today and in the future. High-energy physics researchers are both understanding how to apply data-science community developed tools at this scale, and developing new tools that solve these unique problems. Researchers in HEP are very much"harnessing the data revolution" at an international scale. HEP, and specifically energy-frontier collider experiments, are among the largest drivers (if not the bigger) of data-science tools in basic science research.

Construction and operations of next-generation facilities are still in the future. For example, proponents of both CEPC and FCC have recently published a Conceptual Design Report describing the compelling scientific-discovery opportunities, as well as the detector and engineering challenges of building such a facility. The next step towards a potential new facility is R&D needed to understand the detector design parameters required to capture the full physics potential of a potential that facility. Facilities could start operations on a 10 to 20 year timescale, support researchers over a 15 year operational period, while also being the infrastructure backbone for another 50 years of cutting-edge experiments. While this may seem to be the distant future, it means that the next 5 years are critical for defining the detector design and configurations needed for researchers to later realize the science potential with the first round of experiments at these facilities. Figure 1 illustrates some of the near-term R&D goals will impact the long term.

On this same timescale, the APS Division of Particles and Fields (DPF) is organizing a community planning, "Snowmass", process to use as input in the funding agency prioritization process. While planning is still to be finalized, this exercise is expected to take place during 2021 and 2022 and would engage the entire high-energy physics community. Similar processes will take place internationally, including that of the European Strategy Group. The international research community will work together to make the detailed analyses needed to reach consensus opinions that define these next-generation facilities. Given the massive scale of any new facility in HEP, one or at most two of the concepts for future colliders would be pursued by the community over the next decade. A cross-facility and cross-regional network, a network of networks, is essential for achieving the best overall outcome.

A robust set of software tools and a capable, trained, researcher community are fundamental, for this prioritization to be able to take decisions based on the best available information from the proponents of each facility concept. Mature simulation, reconstruction and analysis pipelines will be required in order to carry out studies to support design choices that must be driven by achieving the best physics outcome possible. Analyses must be at a fidelity sufficient to reliably differentiate between detector design choices and to validate the expected scientific sensitivity. Before any of this can happen, tools for detector simulation and event reconstruction must be in place.

Whereas the HL-LHC program has built up a team with more than a decade of experience, next-generation experimental groups rely on small development teams to provide these results. As critical funding decisions for the project have yet to be made, funding for dedicated technical personnel is not yet in place. For this reason and others, a large field-wide benefit could be derived if different communities recognize the need to move away from local solutions to a more sustainable common toolkit that the broader international network of researchers can take advantage of.

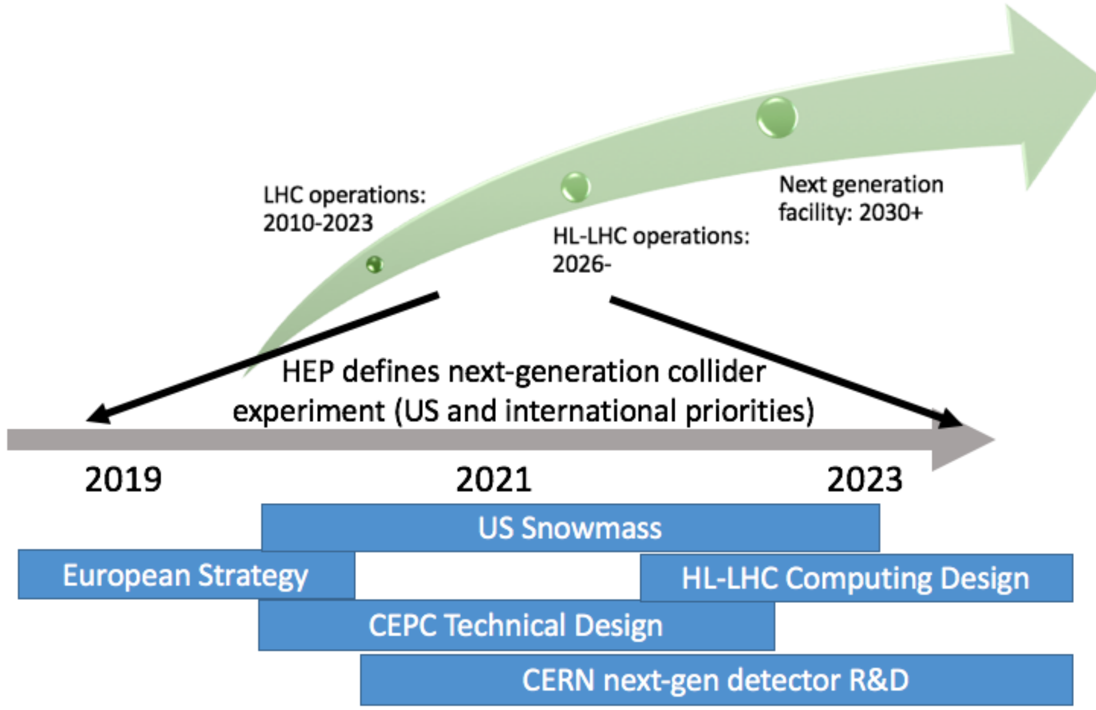

Figure 1: Neartime timelines for research and development towards potential next-generation HEP collider experimental facilities.

We view the need for short-term software to support detector design choices for next-generation facilities together with a unique design opportunity where the HL-LHC software and computing community can bring its expertise and ensure that high-energy physics benefits long term through forward-looking work. The near-term Snowmass milestone, and related international milestones, requires substantial community engagement to succeed. Contrary to the past, the appears to be engaged towards a common approach to evaluating related next-generation experiments, built up from common tools and simulations. This can be particularly beneficial given the commonality between detector concepts that are proposed for each accelerator concept.

3 Project Goals

We propose a Future Research Software Collaboration for High Energy Physics (FRESCO) to enable the budding international research and development connections to grow into the strong collaborations needed for today's R&D to be a success. FRESCO will connect networks of researchers, which are typically focused on a single experiment or region, and enable new collaborations.

The FRESCO program will consist of three areas:

1. Topical workshops: designed to enable discussions and working meetings towards a specific research area and milestone.
2. Internations researcher exchange program: designed to enable short-term, intensive, collaboration between researchers to finish a collaborative development project or milestone.
3. Training events: designed to enable researchers to get the knowledge they need to effectively collaborate internationally while reaching their detector design goals.

Each of these will be focused on enabling the work on early-career researchers (from students through junior faculty). Researchers in this area must be proficient in a complex suite of tools in order to successfully perform detector performance studies. These are even more broad, and typically less well documented, than those encountered by data analysts in current-generation experiments. Researchers must be properly equipped in order to efficiently and effectively use high-energy physics tools and the data-science tools that are essential to their studies.

We see three areas that span the HEP computing challenge where community collaboration is particularly critical:

1. Geometry and simulation: The base layer of software that supports easily adaptable detector designs appropriate for detector concept studies.
2. Calorimetry and full event reconstruction: The process of going from calorimetric hit information, to estimates of particle trajectories, and finally to fully reconstructed event (essentially a "picture" of each interaction) is an ultimate goal of advanced detector designs. Finite segmentation, lack of total hermicity, and echanical structures are examples that complicate the algorithm design for any proposed detector concept. At the same time, excellent energy and spatial resolution must be achieved.
3. Track and vertex reconstruction algorithms: The inner-most detector components are designed to provide accurate kinematic information for all charged particles produced in collisions. At these energies, detector designs are challenged to obtaining excellent momentum resolution and position resolution while maintaining low material profiles. Tracking and vertexing algorithms must reconstruct particle momentum, origin vertex and trajectory information from patterns left in the detector. Algorithms that can easily incorporate changes in detector design are especially valuable in evaluating which detector designs meet design criteria.

If successful, FRESCO will enable the US and international high-energy communities to make fully informed choices towards its next generation of accelerator facilities. This network of networks will build on a significant program in the US whose focus is HL-LHC. There, a vibrant and international research and development program is beginning to address the computational and data-science challenges that must be met in order to maximize the scientific potential of the HL-LHC data. Projects are beginning in the US, including the NSF-funded IRIS-HEP project, and internationally with the goal of defining the computing model for experiments in the coming three to five years. Reaching across experimental and facility boundaries, and connecting the US program with related international efforts are essential for the successful completion and deployment this R&D. Building upon these networks to form a broader collaboration will help ensure the scientific discovery potential of the HL-LHC program.

FRESCO will help form a basis for further discoveries in high-energy physics; facilitate the adoption, and effective usage, of state-of-the-art computational and data-science techniques and technologies by HEP researchers; and help the US research community lead international collaborations in HEP.

FRESCO is proposed to be a three year program enabling a network of networks to grow across the high-energy physics community. Its primary mission will be to connect HL-LHC researchers with the US and international communities evaluating the near-term challenges of next-generation collider facilities. Beyond the near-term goals of Snowmass or technical design specifications of facilities, FRESCO will facilitate US researchers gain critical expertise should it decide to pursue one of the international accelerator facilities as the centerpiece of its future energy-frontier accelerator program. This decision would be the outcome of an effective Snowmass process.