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, the Circular Electron Positron Collider (CEPC) in China, and the Compact Linear Collider (CLIC) at CERN, 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 commensurate 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 versa). 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 Science Drivers

The LHC program at CERN is a flagship project in particle physics both internationally and for the US community. It seeks to probe the characteristics of basic physics processes at the frontier of energy and our understanding of the structure of these processes. Next-generation energy-frontier research facilities, based on novel circular or linear electron collider designs, are the next step in this scientific endeavor.

Atomic, nuclear, and now particle physicists have long pursued the discovery of the basic constituents of ordinary matter. This work has resulted in a very successful theory to describe the interactions (forces) among them. All atoms, and the molecules from which they are built, can be described in terms of these constituents. The nuclei of atoms are bound together by strong nuclear interactions. Their decays result from strong and weak nuclear interactions. Electromagnetic forces bind atoms together, and bind atoms into molecules. The electromagnetic, weak nuclear, and strong nuclear forces are described in terms of quantum field theories. The electromagnetic and weak nuclear interactions are intimately related to each other, but with a fundamental difference: the particle responsible for the exchange of energy and momentum in electromagnetic interactions (the photon) is massless while the corresponding particles responsible for the exchange of energy and momentum in weak interactions (the W and Z bosons) are about 100 times more massive than the proton. A critical element of this "Standard Model" (SM) is the prediction (made more than 50 years ago) that a qualitatively new type of particle, called the Higgs boson, would give mass to the W and Z bosons. Its discovery at the LHC by the ATLAS and CMS Collaborations in 2012 [1, 2] confirmed experimentally the last critical element of the SM.

The SM describes essentially all known physics very well, but its mathematical structure and some important empirical evidence tell us that it is incomplete. These observations motivate a large number of SM extensions, generally using the formalism of quantum field theory, to describe new physics beyond the SM (BSM) physics. For example, "ordinary" matter as described by the SM accounts for only 5% of the mass-energy budget of the universe, while dark matter, which interacts with ordinary matter gravitationally, accounts for 27%. While we know something about dark matter at macroscopic scales, we know nothing about its microscopic, quantum nature, *except* that its particles are not found in the SM and they lack electromagnetic and nuclear interactions. BSM physics also addresses a key feature of the observed universe: the apparent dominance of matter over anti-matter. The fundamental processes of leptogenesis and baryongenesis (how electrons and protons, and their heavier cousins, were created in the early universe) are not explained by the SM, nor is the required level of CP violation (the asymmetry between matter and anti-matter under charge and parity conjugation) present. Constraints on BSM physics come from "conventional" HEP experiments plus others searching for dark matter particles either directly or indirectly.

The last decadal planning process by the U.S. particle physics community produced a strategic plan [3] which was submitted to the NSF and the DOE in May 2014. The report identifies "five compelling lines of inquiry that show great promise for discovery over the next 10 to 20 years". These are the primary science drivers for our research:

- *Use the Higgs boson as a new tool for discovery*
- *Pursue the physics associated with neutrino mass*
- *Identify the new physics of dark matter*
- *Understand cosmic acceleration: dark matter and inflation*
- *Explore the unknown: new particles, interactions, and physical principles.*

The energy-frontier program provides essential scientific tools to address the first, third, and fifth of these. This program could provide physicists with access to orders of magnitude more data acquired compared to the Higgs discovery over the next two decades. As Figure 3 illustrates,

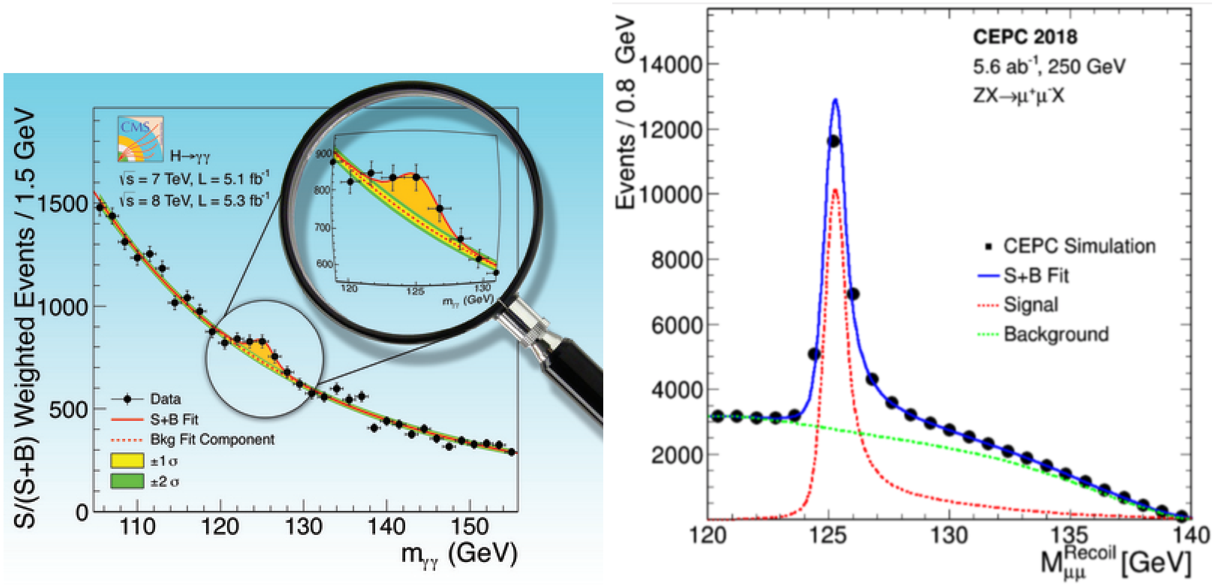

Figure 1: Searches for new physics at the LHC often involve very small signals with very large backgrounds. Next-generation facilities will provide massive data samples with much improved signal-to-noise ratios. This will allow for precise studies needed to uncover underlying physical phenomena.

researchers will evolve from needing a needle-in-the-haystack approach to having overwhelming data sets. This amazing scale of data sample is not limited to Higgs physics; instead it would be a major opportunity for a huge range of high-energy physics researchers. The potential prize given by next-generation colliders is one or even two orders of magnitude in improved precision across electroweak physics observables.

Energy-frontier experiments publish hundreds of scientific papers per year by collecting enormous data sets using a complex apparatus whose pieces are each a major scientific undertaking to build and operate. Data science researchers can impact research practices worldwide. We seek to connect research communities across the energy-frontier communities to enable this scientific undertaking, and at the same time provide a new opportunity for students and young researchers to learn about this science, its community and contribute to its evolution.

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 new 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 five 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 2 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

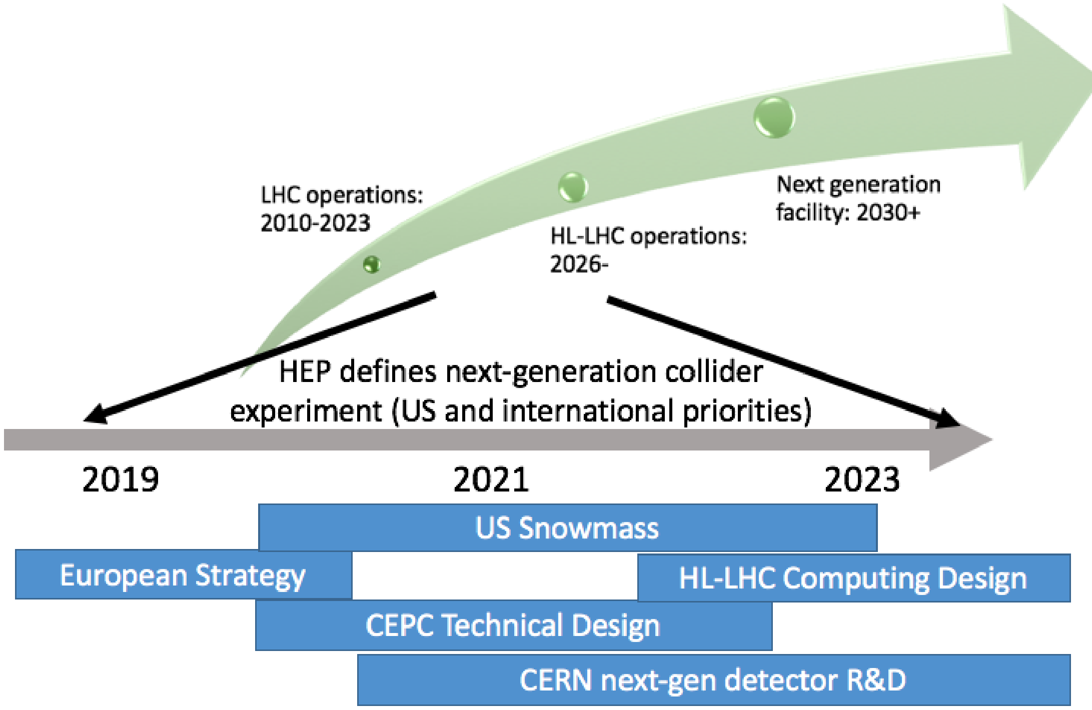

Figure 2: Near-term timelines for research and development towards potential next-generation HEP collider experimental facilities. HEP researchers are working to define requirements and designs for next-generation facilities through an international community process during the next three to five years.

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.

3 Computational and Data Challenges

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 biggest) of data-science tools in basic science research.

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

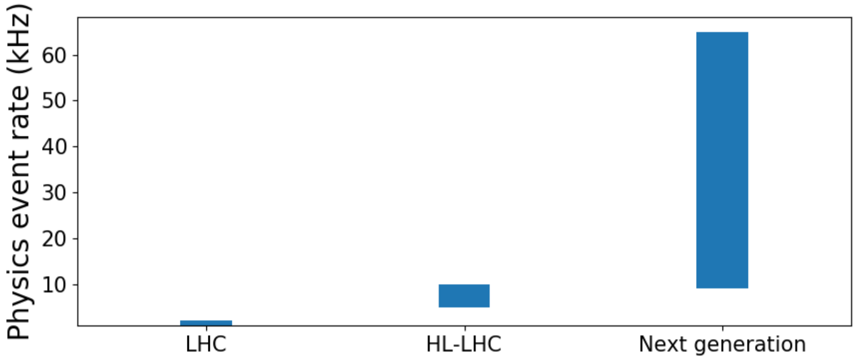

Figure 3: Event rates for LHC and future collider experiments. The blue bars illustrate the range of physics event rates, which will be better determined as facility parameters (and particle cross sections) are better measured. These rates are after the vast majority of particle collisions are rejected by the real-time data processing system, and thus represent the rates of data seen by data analysts.

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.

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, there 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.

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.Community tools for geometry description include DD4HEP [**?**] Geant4 [**?**] has become the community standard for modeling the interactions of particles with detector materials, including relevant physics-process descriptions spanning many orders of magnitude of energy needed to accuracy predict patterns observed in experimental apparatus. While these standardized tools exist, few researchers are proficient in their use, as they are typically not essential for researchers doing data analysis.
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 hermiticity, and mechanical 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. No single approach or implementation has become the standard in this area. This creates a difficulty for research groups looking to be "users" rather than "developers" of these tools.

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. Some components of algorithms are inherently detector design specific, however projects such as ACTS [?] aim to provide an infrastructure (toolbox) for tracking developments.

4 Networks in High Energy Physics

Collaboration networks work differently for high-energy physics experimentalists than in many scientific research fields. Here we describe typical interactions, and missed opportunities, in both running experiments as well as experiments in their research and development phase. The environment is quite different in the two cases.

Modern HEP experiments require contributions from hundreds or thousands of researchers to be successful. Groups are formulated around common ideas to initially propose and eventually build the experimental apparatus, collect data and eventually analyze and publish it. Beyond the "idea" stage, this process requires a establishing a formal scientific collaboration, which today is international in nature, to enable researchers to work together to create, and mutually benefit from, a common infrastructure represented by the experimental facility, computing resources and software. Large experiments require decades from start to finish. Different from many fields, the entirety of data gathered from long data-taking runs are used by many analysis groups.

4.1 Networks in and across operating experiments

By the operating phase of an experiment, researchers contribute to a common software base. For example, the LHC experiments each have codebases with several millions of lines of source code and hundreds of active contributors at any given time. A common code base is required because the scale of the HEP computing challenge requires experiments to have a common data reduction phase, done together with corresponding detector Monte Carlo simulation samples, before individual research groups start to analyze data. This data reduction step is far beyond the resource capacity of any single analysis group.

This process facilities the natural growth of regional networks as well as international networks of researchers with a specific interest. For example, collaborative groups will form around each of the data reduction challenges in the experiment, having the goal to establish the best algorithm possible. These networks mean that when a researcher wants to evaluate the feasibility of a new idea, be it a new way of operating a detector, a new algorithmic approach, or something else, that researcher has a substantial infrastructure and knowledge base to draw from. The established network enables this new question to be answered quickly without having to "reinvent the wheel".

In this way, researchers within a single HEP experiment have long been well connected, forming an international networks. As the scale of experiments has grown, networks have also grown across experiments. In particular for the LHC experiments, cross-experiment networks have grown in the software-development areas including detector simulation algorithms, event reconstruction techniques, and analysis frameworks. Similar to experimental networks, these have grown out of the realization of common needs, and eventually the establishment of common projects.

The LHC community of experts has thus established a fruitful network, thanks in part to the creation of the HEP software foundation [?]. The HSF has helped researchers establish working groups, and holds an annual workshop for face-to-face discussions. It is community driven, largely by researchers from the LHC community.

4.2 Networks for proposed experiments

The collaborative network is quite different HEP project still in its proposal stage. From early in the process, collaborations of researchers focused on research and development of detector components must work together to formulate a new concept. These groups tend to be strongly regionally focused as detector components are typically built by countries or small groups of countries. An important piece of a detector proposal is the optimization of detector geometries and segmentation, including the interplay between all the detector components that make up a full experimental apparatus.

It is this last consideration that often leads to an involved sensitivity study. It typically requires doing a full detector simulation in order to mock up experimental measurements that are needed to differentiate one set of design parameters from another, or to establish that a detector design meets the physics benchmarks desired. It is these small groups of detector advocates that typically carry out these analyses as part of establishing the most appropriate detector design for the proposed experimental facility.

Networks of researchers responsible for software infrastructure develop much more slowly. Distinct from an operating experiment, there is typically no, or a very minimal network in place to facilitate these needed detector simulation, event reconstruction and analysis studies. For example, a group proposing an alternative detector design may be asked to make their design more realistic in order to demonstrate that it is in fact better than an existing baseline design. Contrary to the situation in an operating experiment, these researchers are faced with doing it outside of any organized software framework. Instead tools are gathered together via regional networks. This naturally leads to duplicated work, incomplete results, and longer timescales for completion (or just giving up on potentially game-changing ideas.

This is not to say that the international community does not have sufficient knowledge, infrastructure or effort. Rather, that in the research and development phase, the cross-regional, cross-detector, or cross-facility network of networks is lacking. This is natural, as major funding for researchers to build up experimental infrastructures comes only after detector concepts are agreed and funded. Instead operating experiments, i.e., the LHC experiments, have well established infrastructure and knowledge to build upon.

It is this *network of networks* that we seek to build for future collider detector concepts at the HEP energy frontier.

4.3 Community benefits of building a software network-of-networks

The vibrant *network of networks* connecting researchers working on physics application software components in across current-generation and next-generation experiments in HEP would be an enabling capability for the HEP collider research community. The most apparent scientific benefit will be a dramatic improvement in the way these new experiments are designed and benchmarked. Establishing accurate estimates for the physics reach of proposed experiments pays dividends for a decades long experimental program as it helps to ensure the right facilities are constructed and the right experiments are run.

In addition, this network of networks would enable researchers on current facilities to be involved in design decisions taken as next-generation facilities are realized. Their practical experience can ensure that decisions taken result in sustainable and performant systems. This is especially important at completely new collaborations, as greenfield opportunities are a rare opportunity to dramatically improve the experience of HEP researchers with their experimental data.

More immediate benefits come to researchers building up the facility designs, and specifically detector designs, for next generation facilities. These researchers will be able to get detector studies done and decisions taken more quickly. Their results will be more reproducible, and their tools and techniques will be more easily shared with colleagues across the community.

A network-of-networks in software will also build expertise in the US. This expertise will directly open areas of collaboration once projects get funded and a US program gets established. The US has a leading role in the software and computing infrastructure of the LHC experiments. Building on this expertise as next-generation facilities are funded makes sense both from the scientific capability point of view, but also as a way to further built the US cyberinfrastructure that enables large scale data analysis.

5 The FRESCO network of networks

The FRESCO project will build a *network of networks* designed to bring together researchers across the HEP collider community towards a common goal of enabling design choices of future electron colliders at the energy frontier. Our approach has three main thrusts: **community-driven training** on software tools that are critical for detector design; **workshops**, supported ion part by FRESCO and its international partners, to discuss and build up a software infrastructure to enable researchers to quickly prototype and study new detector designs; and **researcher exchange** supporting young researchers to work together on a short-term, but intensive, collaborative project with specific goals.

FRESCO will establish and focus on specific goals to help the future-collider community discuss its approach from facility concepts towards a community-driven decision making process. This section will explain the methods, timelines and goals for each of these project thrusts.

5.1 Community training

As in other sciences, the set of software that must be mastered by HEP researchers continues to grow and become more complex. Researchers investigating next-generation facilities must be proficient with the lower level tools typically not used by researchers whose focus is purely on data analysis. The most apparent examples are detection response simulation packages (e.g., Geant4 [**?**]) and detector geometry definition packages (e.g., dd4hep [**?**]). However there are many other tools, from physics event generators, to event-processing frameworks whose functionality must often be augmented by researchers exploring a new scientific realm (even if just at a simulation level).

These tools tend to be domain-specific tools developed and supported by HEP researchers. As such there is considered field expertise to build upon, and instead knowledge transfer is the critical piece. FRESCO aims not to invent new training materials or curricula, but rather to connect together existing training programs and opportunities, while augmenting courses to ensure that the needs of researchers pursuing Snowmass can be met.

Specifically, FRESCO plans to

- Form working group of current and next-generation collider researchers to identify opportunities and areas of particular need.
- Seek to co-locate important training programs with other community events, such as the workshops described below, to reduce travel burdens and to increase student engagement at existing community meetings
- Prioritize instructor and student recruitment to ensure programs pull from an appropriately diverse applicant pool.

FRESCO targets one or two topical training events each year reaching 30 to 50 students. FRESCO will use participant funds to facilitate the travel of instructors and 15 to 30 students annually, with partners supporting the travel of other researchers. FRESCO will build upon infrastructure in existing training-in-HEP projects, such as FIRST-HEP [4], to gain student and instructor feedback before and after each training event. It is important to understand how best to meet the needs of students by identifying strengths and weaknesses in training materials and program design.

5.2 Workshops

FRESCO and its partners will co-organize approximate two or three community workshops annually. Workshops will be each organized around a single specific topic, and will have a specific goal or set of goals to be achieved. The idea is to engage an community that is interested in the topic and likely to engage in a series of events on each topic or theme, with FRESCO workshops touching on each theme about once per year. Focusing on a single theme is meant to enable workshop participants to be more fully engaged and to come away from the few day event having established new collaborations and outcomes to work towards. FRESCO aims to avoid the sort of workshop that is more a "good time" for participants to then return to the work they were doing already before.

FRESCO will organize exit surveys to judge effectiveness of the organization and agenda towards achieving both their technical goals and progress towards building a HEP network of networks. The survey will also help understand attendee engagement and goals for subsequent workshops. We will use the survey results and their trends to track community interest and engagement, and to improve the workshop effectiveness over the course of the project.

As with the training events, FRESCO will use participant funds to facilitate the travel of of 30 to 50 researchers per year. FRESCO will focus on early-career researchers, which are both the most engaged in current research and development goals, and the most likely to lack travel funding for such workshops.

5.3 Researcher exchange

In addition FRESCO will facilitate student, postdoc and young faculty researchers to more closely collaborate with international experts. We will establish an exchange program to fund US scientists travel for extended co-location with international experts (typically for 2-4 weeks at a time). This will enable periods of close collaboration sufficient to complete the project itself and to enable further research building upon the newly gained knowledge base.

FRESCO expects to fund approximately six researchers per year for periods averaging two weeks in duration. Actual project duration will be defined in order to ensure project goals can be accomplished, as to ensure a positive outcome from each investment.

FRESCO will employ a transparent process for soliciting community ideas and identifying projects of interest. Workshop events will be used to target areas of high interest or of special need. Student projects would be solicited twice per year and reviewed via a selection process. The PIs will assemble a small teams of community experts for each proposal round to identify the most promising projects to select. Primary criteria will be established before project review. Criteria will include considerations including quality and feasibility of the proposed work; diversity of the selected projects both in terms of topic area and researcher; and importance of a selected outcome to the community. Proponents (or their groups) would be excluded from being involved in any funded project during a given round of project selections.

Project applications will be tracked in order to follow community interest and to identify areas where advertisement and outreach can be used to identify opportunities across the community. Selected researchers will be required to complete a short project summary outcome

5.4 Leveraging existing community events

Training events and workshops would also be organized to take advantage of other opportunities to leverage an existing event (i.e., conferences or community meetings) that would also benefit from a broader attendance across the researcher network of networks engaged by FRESCO. Example events with which FRESCO workshops would be co-located include

- Annual Hong Kong Institute of Advanced Study program on High-Energy physics (Hong Kong, January): This program already consists of two weeks of topical workshops followed by a conference, all focused on the physics of future circular electron colliders. The international community is primarily engaged through HEP theory researchers, while the experimental facility community from Asia is well represented. FRESCO and its partners will organize a topical workshop focused on tools and techniques for detector design. The IAS program will support travel costs for international researchers. FRESCOs role will be to co-organize the workshop (two or three days) and to recruit experts from the LHC community to attend and contribute their expertise.
- Annual HEP software foundation meeting (location varies, typically US or Europe): The HSF meeting brings together software experts from the international community, largely from the LHC experiments. FRESCO and its partners will organize a co-located workshop with the goal of not only working towards a common software basis for the future research collider community, but engaging them with the broader HSF program. This should facilitate a closer engagement of the future collider community with new tools and techniques as well as modern software practices.
- Annual workshop/hackathon in US. Bringing together groups of researchers for a short period of time to work towards goals centered on a common infrastructure has proven to be very succcessful method of collaborative development and learning [**?**]. FRESCO would organize one annual workshop in the form of a hackathon, most likely either at Princeton University or the University of Washington.
- Annual CEPC and/or FCC collaboration weeks (Europe): Both the CEPC and FCC communities have annual (or more frequent) collaboration meetings. FRESCO will engage with the meeting organizers to host a tools and techniques workshop. The most natural meeting would be the CEPC meeting in Europe, as it facilitates attendance by the dominantly Asian CEPC community as well as the dominantly European FCC community. While nominally organized as collaboration meetings, these meetings are normally open to the HEP community (as there is not yet results from physics data)

As with the proposed community workshops, FRESCO will build on existing researcher-exchange opportunities. We anticipate to both better leverage FRESCO funding and to extend existing opportunities to more US researchers. Example programs include:

- Summer student programs such as the CERN summer program, are well established in HEP as an essential mechanism for introducing students to new facilities. Student projects can often be extended beyond the, for example to make up a students research thesis, at minimal cost. Thus FRESCO can build upon successful summer projects and give students a new opportunity to continue their research.
- Student and researcher exchange programs such as the DOE-INFN program [**?**] are existing opportunities for US researchers to work abroad on future collider facility research. FRESCO will work to identify projects and researchers well suited to take advantage of these programs.
- Short-term projects specifically focused on detector development or software development were funded by IHEP-Beijing to support the recently completed CEPC Design Report. The primary goal has been to bring international university researchers to work with IHEP re-

searchers on specific research topics. These opportunities continue as CEPC enters its next phase of design studies.

5.5 FRESCO evolution after the Catalytic period

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 to 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.

We believe that a vibrant network of networks would continue to succeed Beyond the funded phase of FRESCO. HEP has the natural seeds in the form of large collaborations to continue established workshops and events. The HSF is also a critical, established, piece that will enable a community built around software infrastructure to build on the work of FRESCO. We view the HSF as a necessary component beyond the experimental collaborations to foster the sort of community built by FRESCO.

6 Proponents and Collaborators

FRESCO will accomplish its goals by connecting the international detector design and optimization community and the international HEP software development community. As a Catalytic project, the initial phases of FRESCO will be to foster this collaboration and build a network of these existing networks. Nevertheless, the PIs have established a number of collaborations across the US and international community to bootstrap the FRESCO project.

Specific collaborations are described in more detail in the Princeton University Facilities, Equipment and Other Resources document. Here we summarize the types of collaborations established. Geographically, collaborators span the most active regions of HEP experimental groups, concentrated in US, Europe and Asia.

Collaborators with FRESCO include **??? write more here on each bullet - probably more on the organizations**

- University groups in the US
- International university and laboratory groups

- Projects responsible for training in HEP
- Projects developing event reconstruction code in HEP

FRESCO would co-fund the participant travel, workshop expenses and researcher exchange travel allowances together with its partners.

7 FRESCO team and Coordination plan

The PIs have primary responsibility of coordinating the FRESCO network of networks. As discussed in the previous section, they have established a number of US and international collaborators who will help with the initial coordination efforts. The coordination of FRESCO will evolve dynamically as the network of networks itself evolves. This is important in order to evolve FRESCO in directions best suited for the Snowmass exercise and the broader needs of the next-generation facility community.

PI Hsu is... Atlas and please mention future collider R&D (as Lange lacks that)???

PI Lange is an expert in event reconstruction and simulation algorithms, currently working as a member of the IRIS-HEP software institute and as a member of the CMS collaboration. He leads the IRIS-HEP area focusing on developing algorithms for event reconstruction applications for HL-LHC experiments.

Both Hsu and Lange primarily work with FRESCO stakeholders to establish and grow the community network of networks. They will be the primary contacts for the scheduling and logistics of workshops and training events with the help of a steering group to be formed with international representation across the network of networks. Membership of the steering group will be seeded from established FRESCO collaborations and will evolve with time as the network of networks evolves. This group will be essential for ensuring that FRESCO can effectively reach the community, and its needs, with a vision beyond that of the two PIs alone. The steering board will help identify workshop opportunites and schedules. We expect to meet approximately monthly as needed to facilitate FRESCO activities.

Separately, the PIs will organize the researcher exchange program. They will leverage Collaboration and institutional connections, in addition to the FRESCO network of networks, to advertize opportunties (planned on a twice per year basis). The PIs will organize a review process with the help of FRESCO members to ensure a fair and transparent review of applications. Finally they will be respsonible for ensuring the success of funded projects and the community adoption of results.

A postdoctorial researcher (to be identified) will help both with workshop and training event organization and execution (including exit surveys), and to seed research and development projects of particular need to the FRESCO network of networks. **can you add a sentense here with a specific thought???**

8 Involving early career researchers and encouraging diversity

Facilitating the success of students and postdocs is a critical component of the FRESCO network of networks. Young researchers are the biggest component of the workforce tasked with developing analyses of future facilities in HEP. FRESCO aims to enable their work through increased training opportunities as well as opportunities for direct interactions with experts. This is especially important as young researchers are almost always involved only at a part time level. One downside to the long timelines of HEP collider experiments is that young researchers must work on multiple experiments in order to experience a full project lifecycle.

As such, the participant funds in FRESCO will primarily target students, postdocs, and junior faculty researchers. These researchers have both the most to gain from, and the most to contribute back to this opportunity.

HEP in general, and more so the software development areas in HEP, are struggling to reach a diverse workforce. Discussions on techniques and ideas for improving are a frequent component to international conferences as well as collaboration meetings of experiments. FRESCO will strive to reach a diverse network by having a transparent process for distributing participant funds and through outreach designed to ensure that the full community is aware of, and feels engaged with, this opportunity. FRESCO will work to ensure that all researchers are welcome, and that all participants are heard and feel confident that they can freely express their opinions and be treated with respect regardless of their background.

9 Intellectual Merit of the Proposed Work

The proposed FRESCO network-of-networks enable HEP researchers to efficiently and effectively use software tools to develop concepts for next-generation collider facilities. It will connect groups across an international community of software experts and researchers. FRESCO will enable the design studies required to build an international consensus on the formulation of facilities that would provide HEP researchers with data sets with unprecedented discovery reach, as well as a solid basis of experiments for decades to come. For US researchers, a successful FRESCO program will be, in the short term, an enabling component of the Snowmass exercise. In the long term, FRESCO would help US researchers take leadership roles in determining and developing the cyberinfrastructure for HEP in for the next decade and beyond.

The proposed FRESCO program would bring together the tremendous synergy between advancement of early-career researcher engagement in science and technology pursuits, with the opportunities to developing novel cyberinfrastructure for use by domain experts. Via IRIS-HEP and other initiatives, a major new investment in software and computing has started in the US with the hope to enable the promised transformative science program of the HL-LHC. FRESCO will enable the broader HEP community to take advantage of that research and development as it matures into enabling capabilities for large-scale data processing and analysis.

10 Broader Impacts of the Proposed Work

Developing a research opportunity for students and junior researchers to work with international experts on next-generation experimental facilities will help generate broader interest in science and technology fields across the US. Our program will draw from researchers across the US and aims to engage a diverse group of researchers.

The training and cyberskills gained during FRESCO training events and workshops will directly contribute to the broader STEM workforce, and not be limited to the HEP community itself. HEP researchers are active in data science research with other scientific areas as well as with industrial partners. FRESCO would be an important new opportunity to build skills of US researchers pursuing data science careers and other research areas will impact fields far beyond the LHC or HEP communities.

FRESCO includes an outreach program for high-school teachers via the QuarkNet program [5]. QuarkNet is a long running educational program designed to introduce high-school students and teachers to experimental particle physics. FRESCO will host QuarkNet staff and teachers participating in QuarkNet to attend our data science and software workshop and brainstorming ideas of Future Collider outreach activities.

11 Results from Prior NSF Support

PI Lange has not previously been supported as PI or co-PI by the NSF. Since 2016 he has however received salary support from PHY-1624356 ("U.S. CMS Operations at the LHC", 1/01/2017-12/31/2021, $54,749,942, PI Marlow), as well as previous award PHY-1120138 (1/01/2012-6/30/2017, $43,794,718, PI Marlow), OAC-1450377 ("Collaborative Research: SI2-SSI: Data-Intensive Analysis for High Energy Physics (DIANA/HEP)", 05/01/2015-04/30/2019, $1,145,564, PI Elmer). He is senior personnel on OAC-1836650 ("S2I2: Institute for Research and Innovation in Software for High Energy Physics (IRIS-HEP)", 09/01/2018-08/31/2023, $25,000,000, PI Elmer). We report here some results from this salary support.

Intellectual merit: Advancing scientific methods for organizing and operating large distributed software and computing infrastructures in support of international scientific experiments. Specifically, Lange has enabled the physics activities of CMS at the LHC. He was recently the CMS offline software and computing co-coordinator (2014-2016) for the operations and development of the simulation, reconstruction and analysis software, and the entire distributed computing system used by CMS. Lange has led the CMS software release effort (2016-2017) [6–8], and has worked to modernize the CMS analysis software environment,including promoting the use of Python tools developed across data science [9]. Lange is an active contributor to the HEP Software Foundation, and worked as co-editor of the software trigger and event reconstruction chapter of the Community White Paper (CWP) [10,11]. The CWP resulted in a research and development roadmap for HEP, derived from a community process, towards the HL-LHC and other future experiments.

Broader Impacts: Training the national and international community of physics and computer science students using HEP data. Lange has mentored students as part of the CERN summer student program and the Google Summer of Code program for numerous years. Recent, he has collaborated with CMS groups in Republic of Georgia (ongoing) and Thailand (2016-2017) mentoring their students on CMS projects [12].

PI Hsu....

Intellectual merit:

Broader Impacts:

12 Proposal notes

unique opportunities and/or novel connections that will be catalyzed and what will be achieved through the activities that could not be achieved with single group or individual support. In addition, proposals at the Full-Scale Implementation level must include a mapping of the field(s) on which the network of networks is focused and explain how the proposed network of networks relates to existing efforts and research frameworks

overarching shared vision Proposals at the Catalytic level must describe the intended process to identify and establish common goals

For Catalytic Projects, if the linkages and/or goals catalyzed are expected to become a long-term network of networks, discuss how the network of networks will evolve its efforts beyond the award period.

International Collaborations and Contributions: explain that HEP is fundamentally international

Coordination Plan: Proposals at both the Catalytic and Full-Scale Implementation levels should present plans for coordination and communication across the proposed network of networks, including a description of the procedures, practices, and other protocols that will foster meaningful collaborations and enable the network of networks goals to be met.

Describe the processes to be used to prioritize activities consistent with the goals of the network of networks to allocate funds and resources, and to coordinate IT or other essential research infrastructure. Include plans to maintain cohesion and to address potential areas of conflict and evolving leadership needs across the linked networks. Define the specific roles and responsibilities of the collaborating PIs, co-PIs, other Senior Personnel,

Student and Early-Career Development Plan: Describe the proposed activities and opportunities within the network of networks to build or enhance professional skills and global research perspectives of students, particularly graduate students, postdoctoral scholars, and early-career researchers, including plans for international exchange programs or other substantive education and training

Broadening Participation Plan: Describe the approach the proposed network of networks will take to increase diversity, broaden participation, and encourage the involvement of underrepresented groups within its activities, including engaging participants at a diverse range of institutions.

Evaluation: Proposals should describe plans for the assessment of the proposed network of networks, including self-evaluation of progress toward the goals of the linked networks.

Proposers should budget funds for up to four network representatives to attend a PI meeting during year 1 of the project.

Personnel and Partner Organizations (3-page limit): This information provides NSF and reviewers with a comprehensive list of the personnel and organizations involved in key components of the network of networks. Provide current, accurate information for all personnel and organizations with roles in the leadership, coordination, research, education, dissemination, and assessment activities. The list should include all PIs, co-PIs, Senior Personnel, paid/unpaid Consultants or Collaborators, Subawardees, Postdoctoral scholars, and project-level advisory committee members. This list should be numbered and include (in this order) Full name, Organization(s), and Role in the project, with each item separated by a semi-colon. Each person listed should start a new numbered line. For example:

Mary Smith; XYZ University; PI John Jones; University of PQR; Senior Personnel Jane Brown; XYZ University; Postdoctoral Scholar; project coordination Bob Adams; ABC Community College; Paid Consultant; professional development training Susan White; DEF Corporation; Unpaid Collaborator Tim Green; ZZZ University; Subawardee; assessment

data management plan

Each letter of collaboration must only include the following statement: If the proposal submitted by Dr. [insert the full name of the Principal Investigator] entitled [insert the proposal title] is selected for funding by NSF, it is my intent to collaborate and/or commit resources as detailed in the Project Description or the Facilities, Equipment or Other Resources Section of the proposal.

References

[1] G. Aad *et al.* [ATLAS Collaboration]. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, 716:1, 2012.

[2] S. Chatrchyan *et al.* [CMS Collaboration]. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett. B*, 716:30–61, 2012.

[3] Particle Physics Project Prioritization Panel. Building for Discovery: Strategic Plan for U.S. Particle Physics in the Global Context. `http://science.energy.gov/~/media/hep/hepap/pdf/May%202014/FINAL_DRAFT2_P5Report_WEB_052114.pdf`.

[4] FIRST-HEP project (homepage). `http://first-hep.org`.

[5] QuarkNet website. `https://quarknet.org`.

[6] D. J. Lange. Software integration and development tools in the CMS experiment. International Conference on Computing in High Energy and Nuclear Physics (`https://indico.cern.ch/event/35523/contributions/840037/`), 2009.

[7] D. J. Lange, M. Hildreth, V. N. Ivantchenko and I. Osborne. Upgrades for the cms simulation. J. Phys.: Conf. Ser. 608 012056 (`http://iopscience.iop.org/article/10.1088/1742-6596/608/1/012056/pdf`) 2015.

[8] D. J. Lange. Simulation and reconstruction upgrades for the cms experiment. International Conference on Computing in High Energy and Nuclear Physics (`https://indico.cern.ch/event/304944/contributions/1672685/`), 2015.

[9] D. J. Lange. Building a scalable python distribution for hep data analysis. International Workshop on Advanced Computing Analysis Techniques in Physics Research (`https://indico.cern.ch/event/567550/contributions/2638685/`), 2017.

[10] Johannes Albrecht et al. HEP Community White Paper on Software trigger and event reconstruction: Executive Summary. 2018. `https://arxiv.org/abs/1802.08640`.

[11] Antonio Augusto Alves et al. A Roadmap for HEP Software and Computing R&D for the 2020s. 2017. [arXiv 1712.06982] `https://arxiv.org/abs/1712.06982`.

[12] J. Pivarski, D. Lange, and T. Jatuphattharachat. Toward real-time data query systems in HEP. arXiv:1711.01229 (`https://arxiv.org/abs/1711.01229`), 2017.