Institute for Research and Innovation in Software for High Energy Physics (IRIS-HEP)

Undergraduate Fellows Webinar

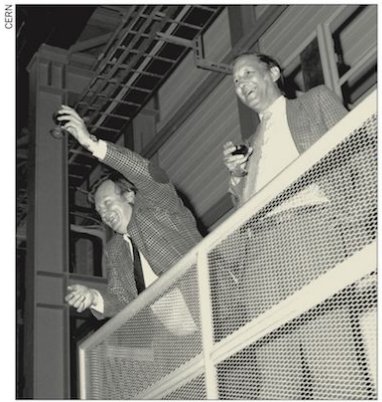
PI: Peter Elmer (Princeton), co-PIs: Brian Bockelman (Morgridge Institute), Gordon Watts (U.Washington) with UC-Berkeley, University of Chicago, University of Cincinnati, Cornell University, Indiana University, MIT, U.Notre Dame, U.Michigan-Ann Arbor, U.Nebraska-Lincoln, New York University, Stanford University, UC-Santa Cruz, UC-San Diego, U.Illinois at Urbana-Champaign, U.Puerto Rico-Mayaguez and U.Wisconsin-Madison

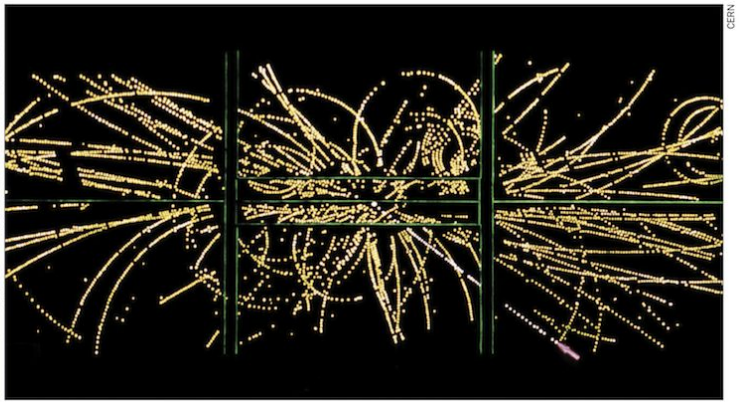
http://iris-hep.org

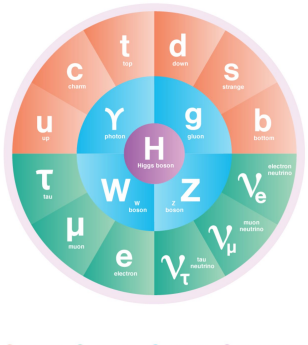
OAC-1836650

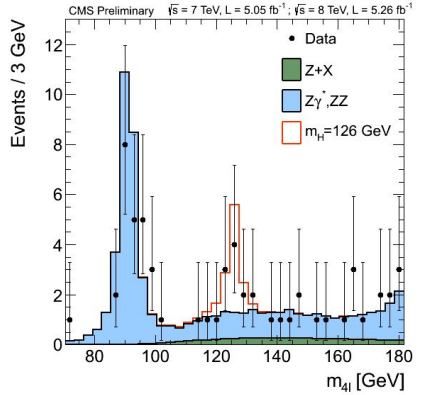
A bit of physics context for the Fellows program

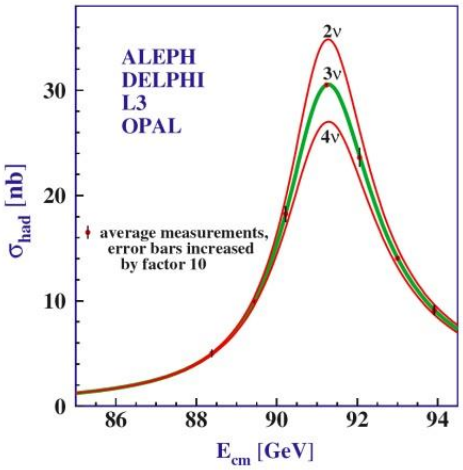
Experimental Development of the Standard Model of Particle Physics

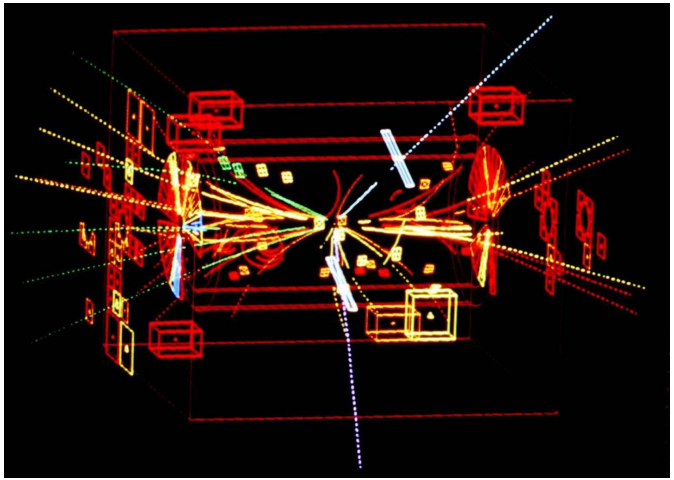


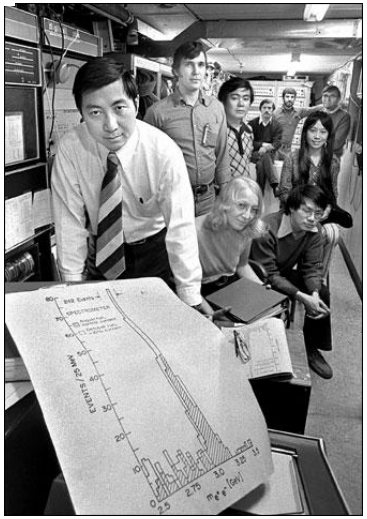




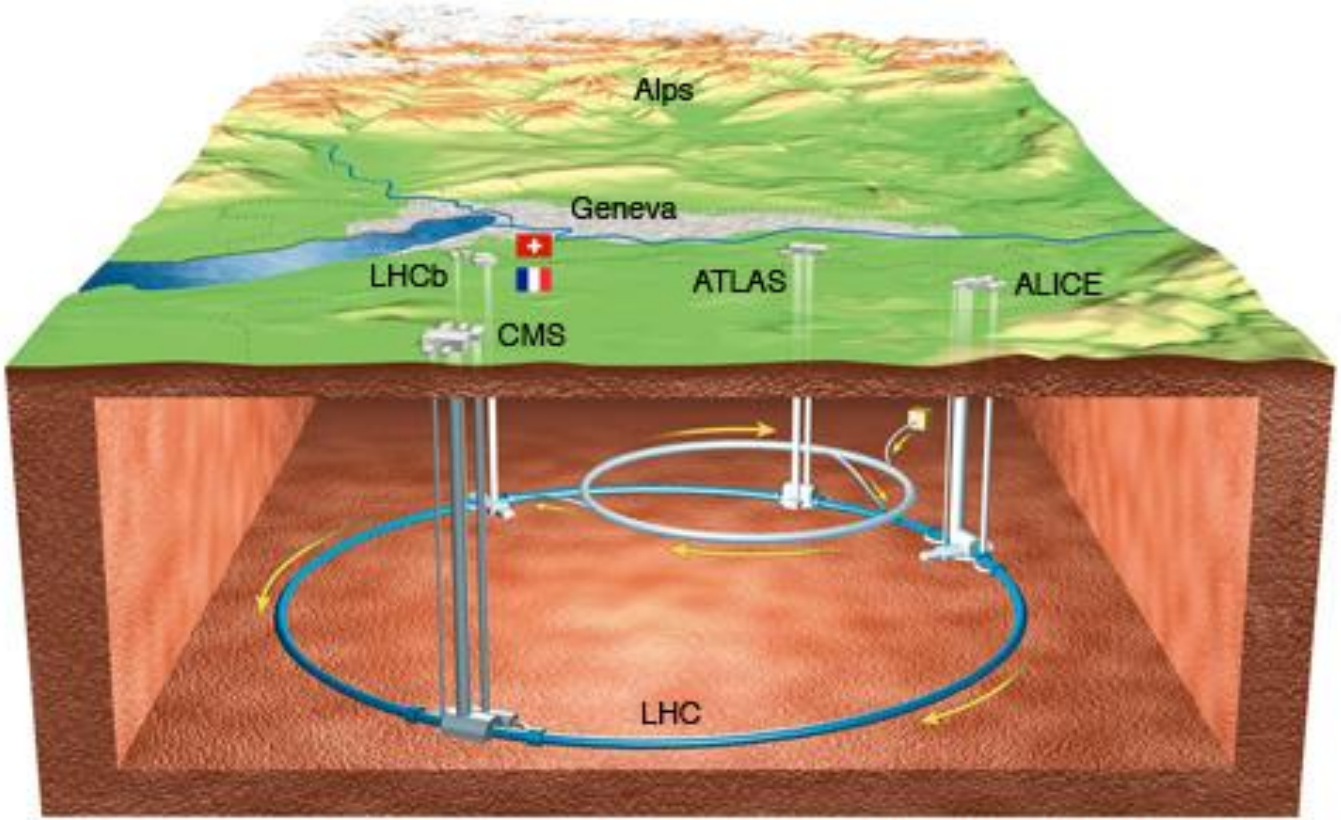


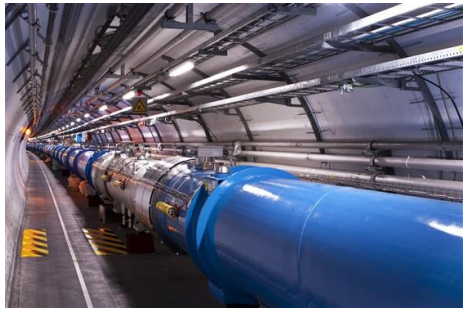






The Large Hadron Collider at CERN





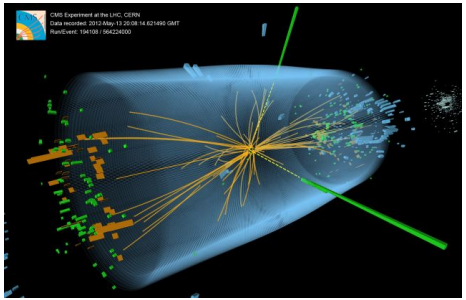
27km circumference tunnel, with larger caverns at points around the ring for detectors

100-150m underground

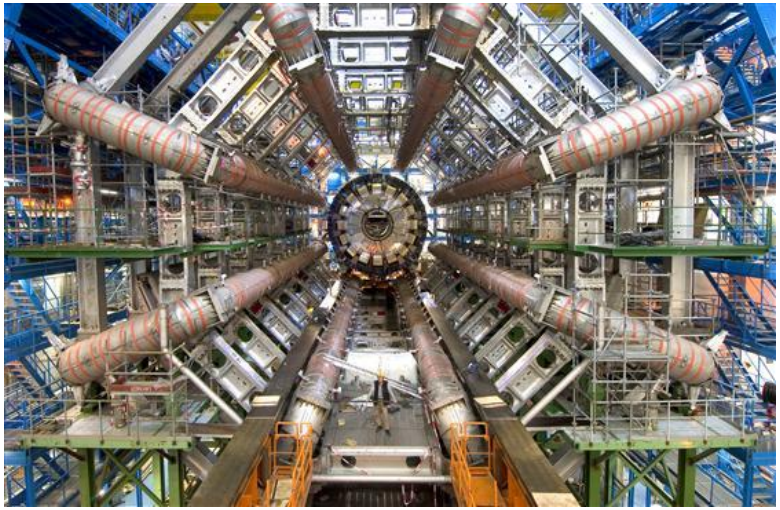
Built in the 1980s for a previous collider (LEP) and expanded with new caverns and access in the early 2000s

Highest energy collider (design 14 TeV, operating at 13TeV) currently available

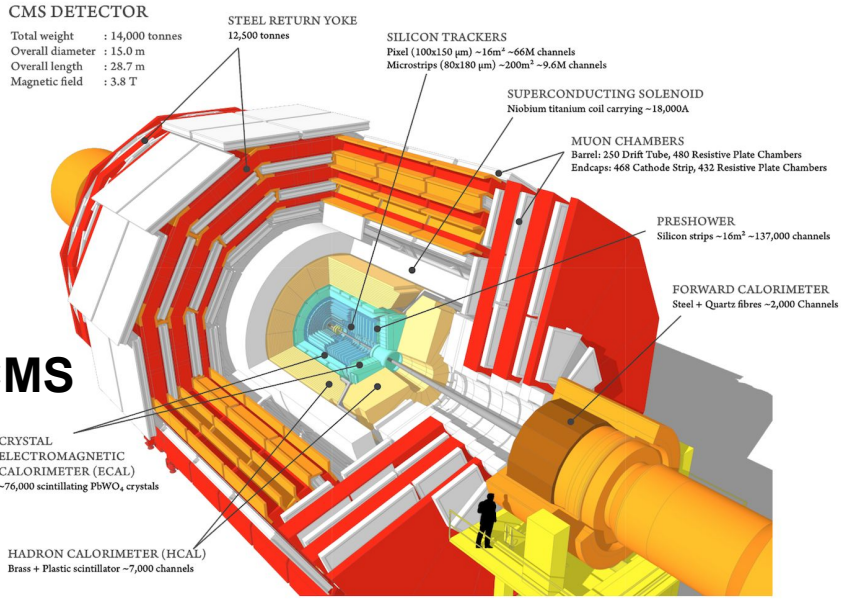
The Large Hadron Collider - Experiments



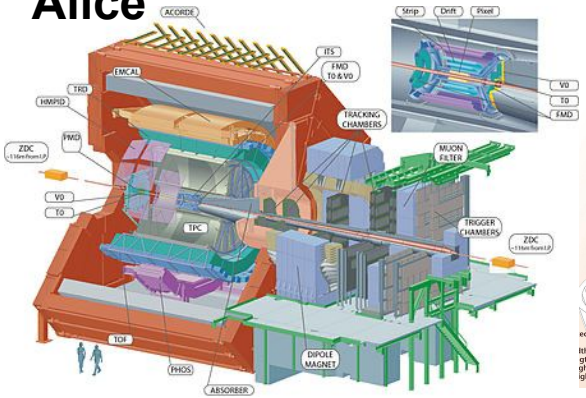
Atlas

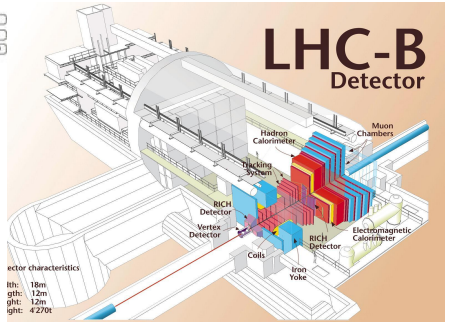


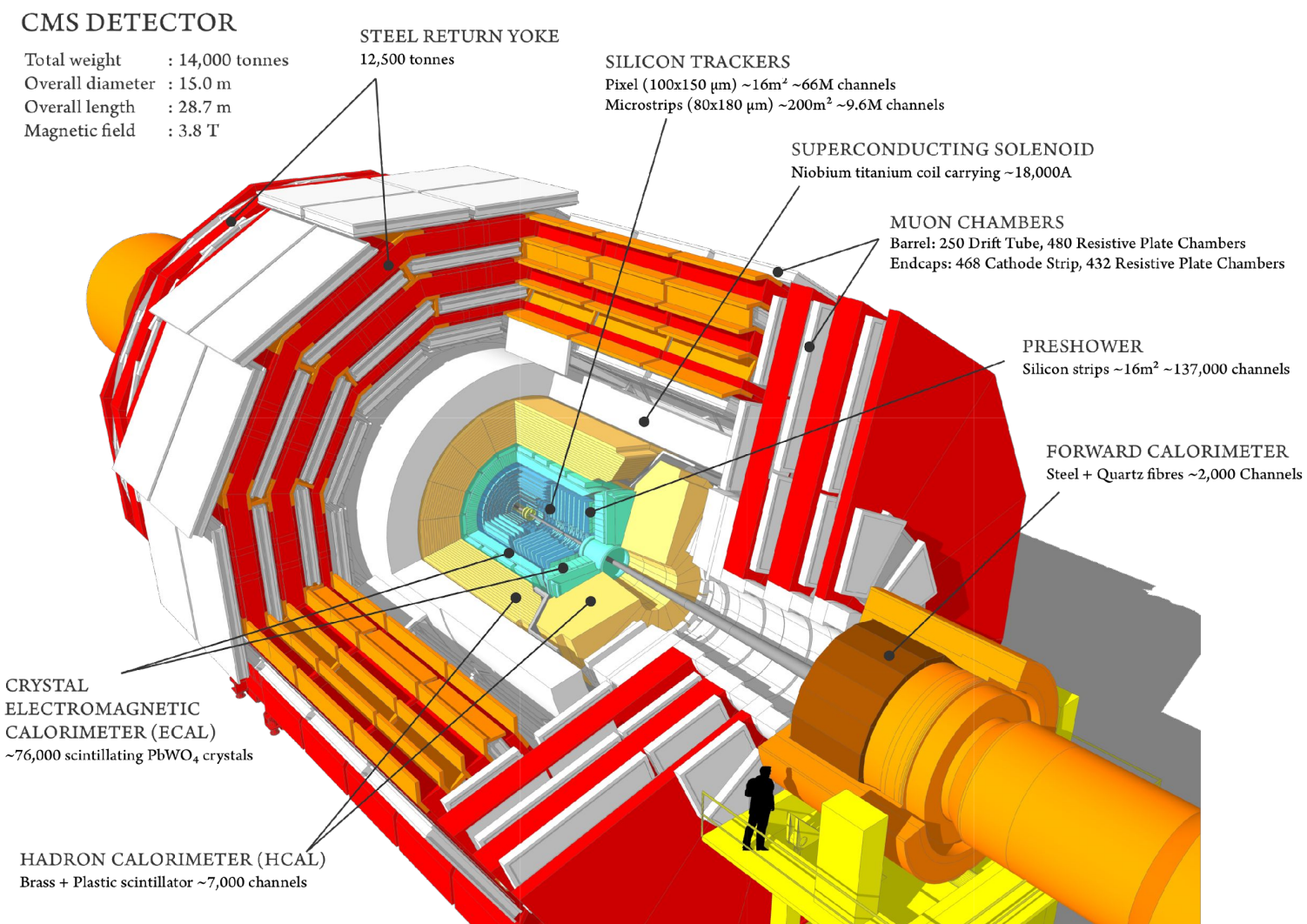
CMS



Alice





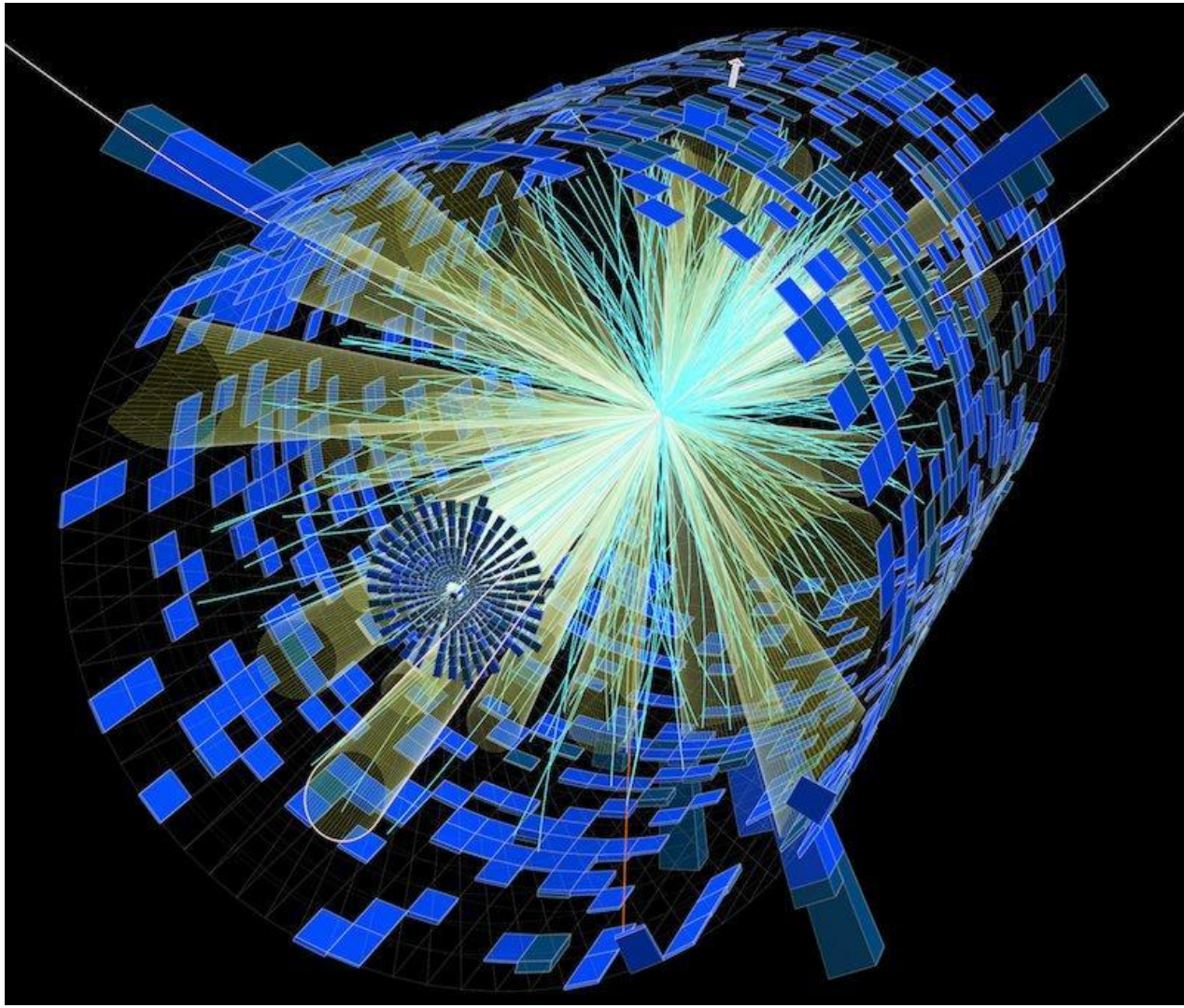

CMS DETECTOR
Total weight : 14,000 tonnes
Overall diameter : 15.0 m
Overall length : 28.7 m
Magnetic field : 3.8 T
STEEL RETURN YOKE
12,500 tonnes
SILICON TRACKERS
Pixel (100x150 μm) ~16m² ~66M channels
Microstrips (80x180 μm) ~200m² ~9.6M channels
SUPERCONDUCTING SOLENOID
Niobium titanium coil carrying ~18,000A
MUON CHAMBERS
Barrel: 250 Drift Tube, 480 Resistive Plate Chambers
Endcaps: 468 Cathode Strip, 432 Resistive Plate Chambers
PRESHOWER
Silicon strips ~16m² ~137,000 channels
FORWARD CALORIMETER
Steel + Quartz fibres ~2,000 Channels
CRYSTAL ELECTROMAGNETIC CALORIMETER (ECAL)
~76,000 scintillating $PbWO_4$ crystals
HADRON CALORIMETER (HCAL)
Brass + Plastic scintillator ~7,000 channels

The next big phase for the LHC is the “High Luminosity Large Hadron Collider” (HL-LHC), which will turn-on in 2029.

Much work is already underway to prepare upgrades to both the detectors and the accelerator complex.

Particle Physics experiments are massive data generators

The hardware upgrades are only part of the problem, significant **software upgrades** are needed to handle the much larger volume (and rate!) of data, which will also be more complex and be used to search for ever more subtle and rare things.

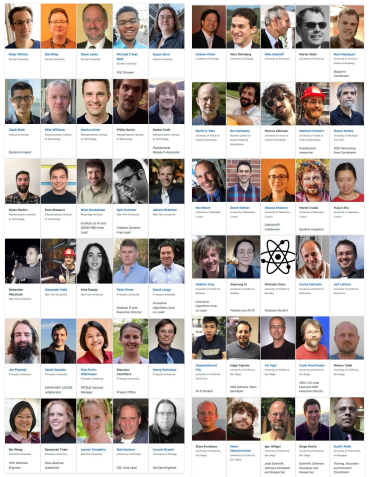


IRIS-HEP and the HL-LHC “Software Upgrade”

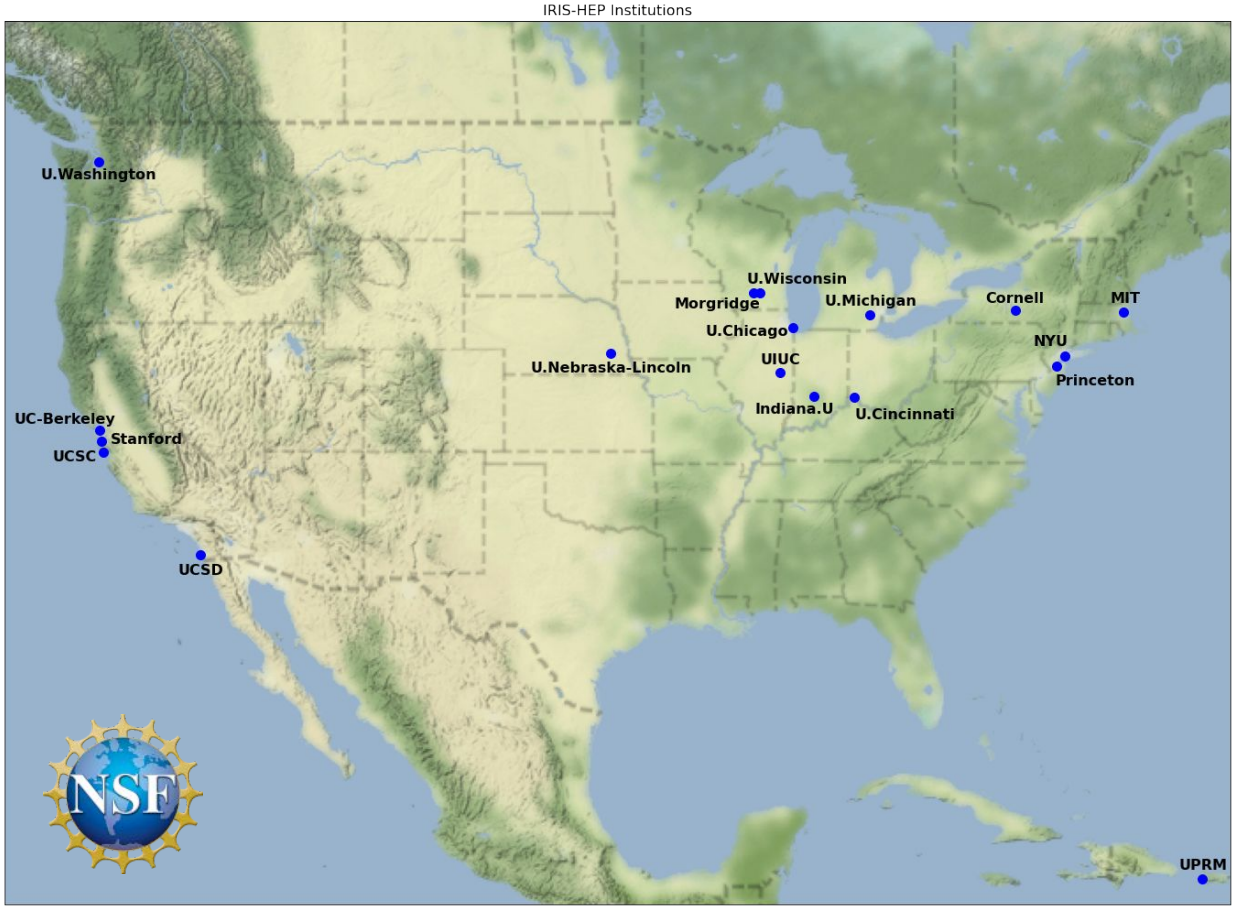
IRIS-HEP - A Virtual Software Institute

Funded by the National Science Foundation to work on “software upgrades” for the HL-LHC as well as training the next generation of research software practitioners

19 U.S. Universities, ~30 FTE’s distributed around the USA.

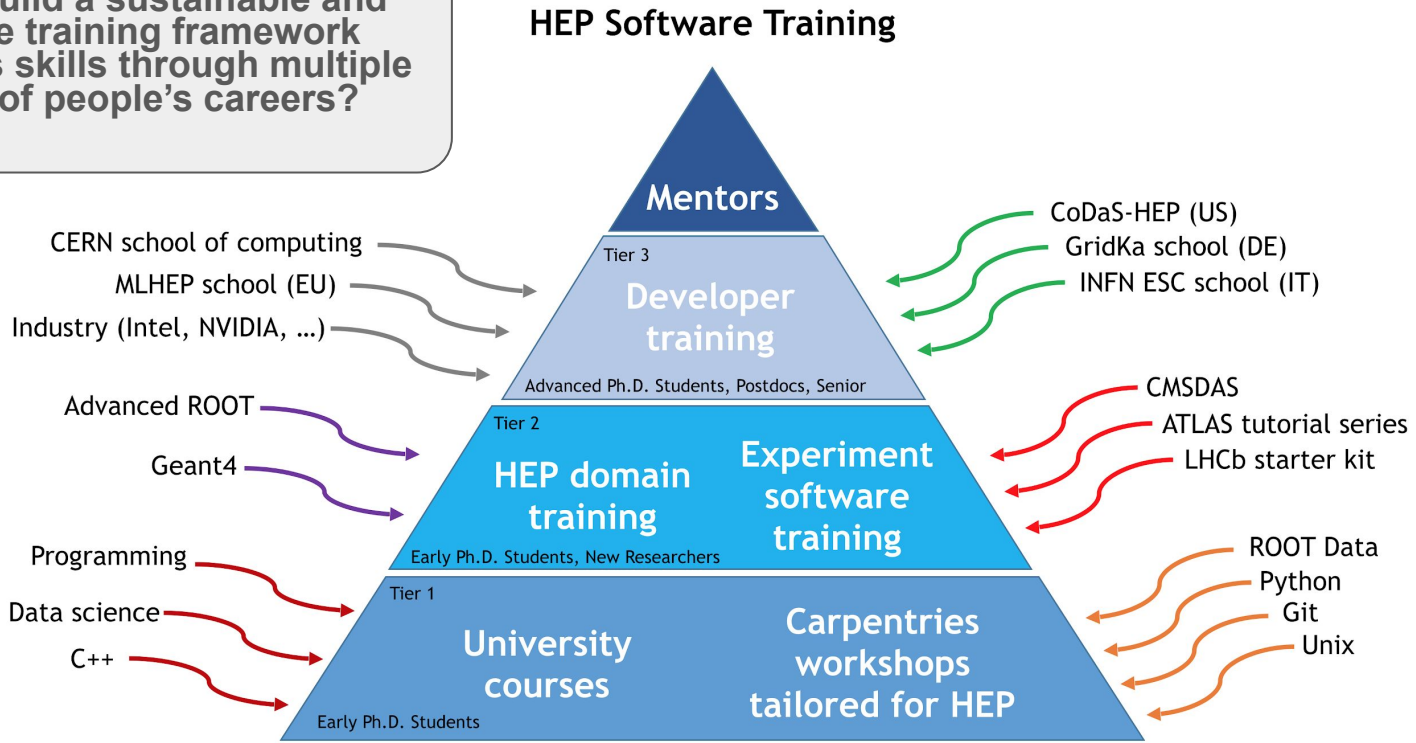


(many more but wouldn’t fit here!)



Training and Education - Sustainability/Scalability

How to build a sustainable and scalable training framework that grows skills through multiple stages of people's careers?



This is a general framework for training, but from the NSF we have funds from both IRIS-HEP (OAC-1836650) and a separate project FIRST-HEP (OAC-1829707, OAC-1829729, http://first-hep.org) which are working towards implementing this model.

Fellow Eligibility and Program Information

Eligibility:

- You must be enrolled in an undergraduate program at an accredited University or College and have completed at least 1 academic year by the start of the Fellowship.
- Our primary funding supports student Fellows enrolled in U.S. universities and colleges, and U.S. citizenship is not required.
- We also have some separate funding for a *small* number of students enrolled in international universities/colleges.

Program information:

- Duration of fellowship and total stipend: An IRIS-HEP Undergraduate fellowship will pay up to 12 FTE-weeks (full time equivalent) at $600/week (paid monthly) during Summer 2022. International Fellows will be paid a training stipend scaled for Purchasing Power Parity (PPP).
- We have some flexibility to accommodate different summer schedules and lengths. Fellowships should in general last 10-12 weeks (2.5 or 3 months).
- It is expected that the Fellow is available full time during the Fellowship period and will not have another significant activity (such as classes, another trainee position or a job) in the same time period.
- **Applications are now open - deadline 8 March**

An alternative for students not associated to US Institutions

https://hepsoftwarefoundation.org/activities/gsoc.html

Google Summer of Code 2022

Introduction

Google Summer of Code is a program that allows students to contribute to the development of open-source projects, mentored by participating organizations.

Particle physics is an exciting field where large collaborations of scientists collect and analyze petabytes of data from high-energy physics experiments, such as those at the Large Hadron Collider, hosted at the CERN laboratory in Geneva, Switzerland. Some of the questions that we collectively ask are:

- what are the fundamental blocks that make up our Universe?
- what is the nature of dark matter and dark energy?
- what is the nature of the asymmetry between matter and antimatter?
- what was early Universe like?

To answer these questions, particle physicists build software to simulate and analyze what happens in particle physics detectors.

The CERN software for experiments (CERN EP-SFT) group has participated in the GSoC since 2011. Since 2017 the program has expanded to involve the high-energy physics community under the umbrella of the HEP Software Foundation.

Information from last year's GSoC can be found here.

IRIS-HEP also partners with HSF/CERN through the GSoC program, including sharing of possible projects and training activities.

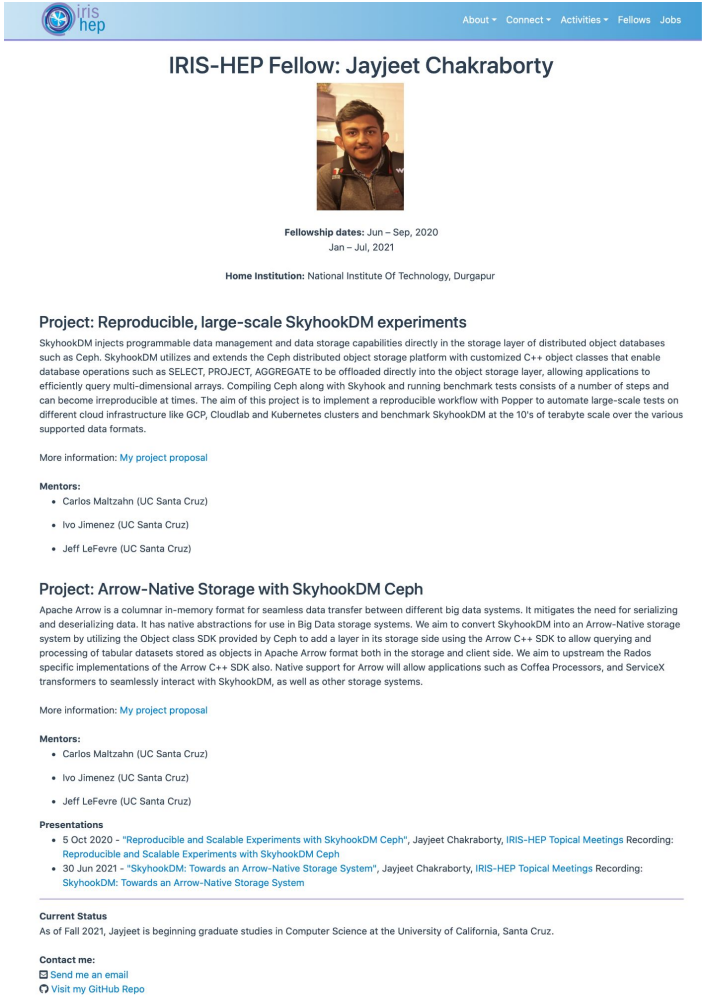
HSF/CERN GSoC also involves mentors from many time zones, but does have a slightly higher concentration at CERN (Geneva, Switzerland) and Europe.

Fellow activities and experience

Each Fellow works with their mentor remotely with daily communication via Slack, Github, etc. As needed zoom meetings will be organized, minimally once/week to plan and discuss. (An exception to the “remote” mentoring may happen if a Fellow is located near their mentor.)

Early in the summer we will organize some training sessions from the HSF curriculum for Fellows.

We will organize sessions at the beginning of the summer for the Fellow to present their project, the midpoint of the summer to give a status report and the end of the summer to make a final presentation. We may also have “social” events to allow Fellows and mentors to interact between projects.



iris hep

About Connect Activities Fellows Jobs

IRIS-HEP Fellow: Jayjeet Chakraborty

Fellowship dates: Jun – Sep, 2020
Jan – Jul, 2021

Home Institution: National Institute Of Technology, Durgapur

Project: Reproducible, large-scale SkyhookDM experiments

SkyhookDM injects programmable data management and data storage capabilities directly in the storage layer of distributed object databases such as Ceph. SkyhookDM utilizes and extends the Ceph distributed object storage platform with customized C++ object classes that enable database operations such as SELECT, PROJECT, AGGREGATE to be offloaded directly into the object storage layer, allowing applications to efficiently query multi-dimensional arrays. Compiling Ceph along with Skyhook and running benchmark tests consists of a number of steps and can become irreproducible at times. The aim of this project is to implement a reproducible workflow with Popper to automate large-scale tests on different cloud infrastructure like GCP, Cloudlab and Kubernetes clusters and benchmark SkyhookDM at the 10's of terabyte scale over the various supported data formats.

More information: My project proposal

Mentors:

- Carlos Maltzahn (UC Santa Cruz)
- Ivo Jimenez (UC Santa Cruz)
- Jeff LeFevre (UC Santa Cruz)

Project: Arrow-Native Storage with SkyhookDM Ceph

Apache Arrow is a columnar in-memory format for seamless data transfer between different big data systems. It mitigates the need for serializing and deserializing data. It has native abstractions for use in Big Data storage systems. We aim to convert SkyhookDM into an Arrow-Native storage system by utilizing the Object class SDK provided by Ceph to add a layer in its storage side using the Arrow C++ SDK to allow querying and processing of tabular datasets stored as objects in Apache Arrow format both in the storage and client side. We aim to upstream the Rados specific implementations of the Arrow C++ SDK also. Native support for Arrow will allow applications such as Coffea Processors, and ServiceX transformers to seamlessly interact with SkyhookDM, as well as other storage systems.

More information: My project proposal

Mentors:

- Carlos Maltzahn (UC Santa Cruz)
- Ivo Jimenez (UC Santa Cruz)
- Jeff LeFevre (UC Santa Cruz)

Presentations

- 5 Oct 2020 - "Reproducible and Scalable Experiments with SkyhookDM Ceph", Jayjeet Chakraborty, IRIS-HEP Topical Meetings Recording: Reproducible and Scalable Experiments with SkyhookDM Ceph
- 30 Jun 2021 - "SkyhookDM: Towards an Arrow-Native Storage System", Jayjeet Chakraborty, IRIS-HEP Topical Meetings Recording: SkyhookDM: Towards an Arrow-Native Storage System

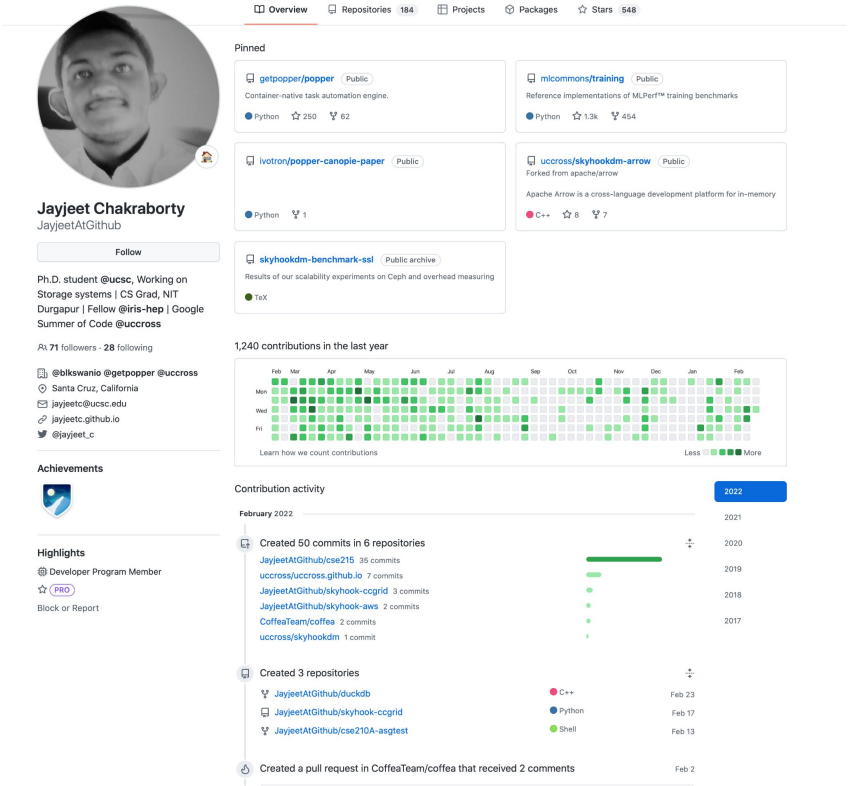
Current Status
As of Fall 2021, Jayjeet is beginning graduate studies in Computer Science at the University of California, Santa Cruz.

Contact me:
Send me an email
Visit my GitHub Repo

Software-centric experience

The value added by an IRIS-HEP Fellowship is:

- Improved software skills
- Experience working in collaborative teams
- Experience with cutting-edge research software
- Publicly visible coding experience (GitHub!) for your next step



Rough Timeline

- Wed 23 February - (This) Webinar
- Tue 8 March - deadline for applications
- By the end of March - first selection of applications to work with mentors on developing a proposal - this may involve a short interview and other follow-up
- mid-April - submission of proposals
- By the end of April - final selection of undergraduate Fellows for summer 2022
- May-Sep - Fellows work on projects (depending on specific academic year constraints)

IRIS-HEP Projects

Projects: Research software projects should in general be in the R&D areas listed above or (in some cases) be projects of more general interest to the HEP community. How do you find a project? There are several possibilities:

- Examples of currently open IRIS-HEP Fellow projects can be found on a separate webpage. These projects are of interest to IRIS-HEP researchers (who may then act as mentors). You can write to the mentor listed to discuss or reach out to fellows@iris-hep.org for more information. *This page may be updated from time to time, so check back and reload.*
- Additional example projects of interest in the wider HEP community can also be found on the HEP Software Foundation Google Summer of Code (HSF GSoC) webpage. While GSoC is a separate program, funded by Google, many of the projects could also be appropriate for IRIS-HEP Fellows if they are related to the R&D areas listed above.
- Recent Fellows and links to descriptions of their projects can be found below in the list of current and past Fellows. In some cases, there may be continuations of those projects.
- Some additional projects can be found on the open projects page of the Compiler Research project
- You can explore the links above to the IRIS-HEP R&D research activities. Most IRIS-HEP active projects list contacts with whom you can discuss to bring your own ideas.

Applying for the IRIS-HEP Fellow position

Prospective fellows can apply using this Application Google Form. You will need to provide:

- Your full name, email address, the name of your university or college and your current or planned major and/or area of study
- A resume/CV (in pdf format) with contact information
- An academic transcript - this can be unofficial, but should include course titles and overall GPA. At the time the Fellowship starts you must have completed at least 1 year of university/college.
- A short essay describing your interest in the internship program (maximum 1 page, pdf format). For example, you may wish to expand on 3 or 4 topics from the following list: your background, your skills, and strengths; what software, computing or scientific topics appeal to you; previous research experience, if any; what you may want to pursue as a future career; and what benefits you would like to gain from this program. If you already have a potential mentor and/or particular projects which interest you from the project lists above, you can also mention them here. It is however not required to have a mentor/project finalized to submit an application. Successful applicants will be connected to mentors to select and define their projects in a 2nd step following this application
- [Optional] The full name and email address of a reference. This is optional. Ideally it would be someone with whom you have interacted in a STEM context (e.g. a course or a previous research activity). You should contact the person in advance to confirm that they will write a letter for you and simply provide their name/email in the application form. After you submit the form, we will contact them to request the letter.