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

Sudhir Malik (UPRM)

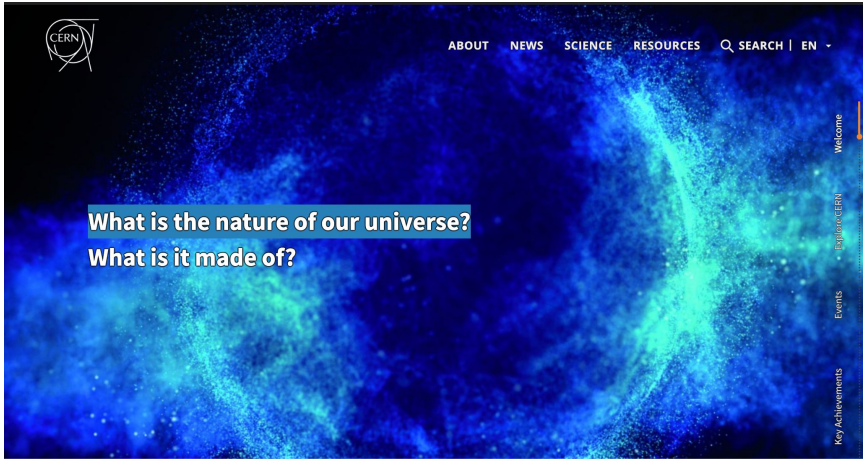
OAC-1836650
Proposal 2323298

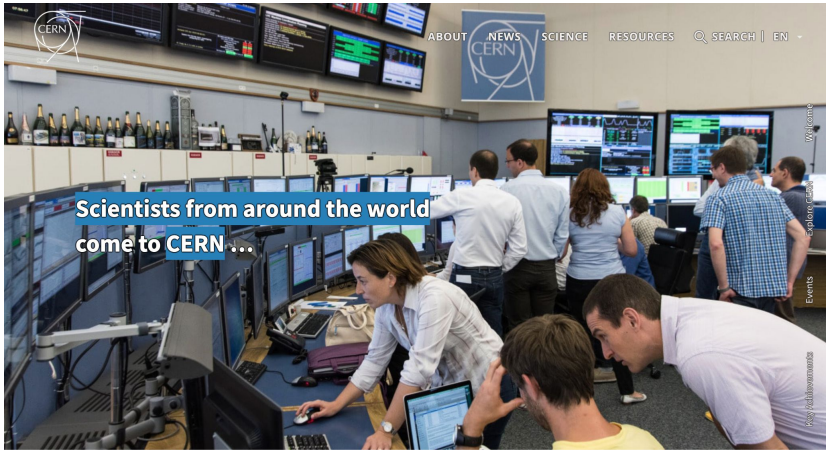
http://iris-hep.org

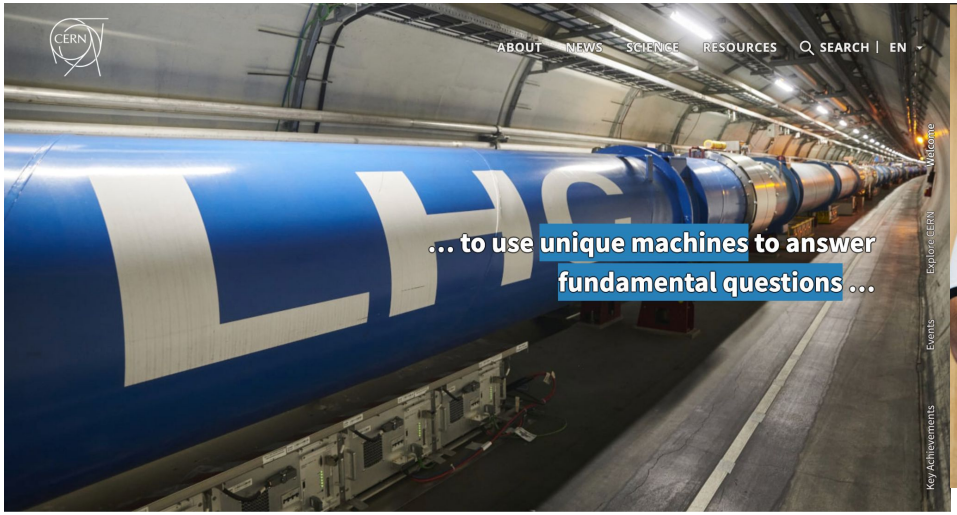
Context

Location


Ireland
Amsterdam
Netherlands
London
Brussels
Belgium
Germany
Frankfurt
Berlin
Sign in
Prague
Czech Republic
Luxembourg
Paris
Munich
Vienna
Slovakia
Austria
Budapest
Hungary
France
Switzerland
Slovenia
Milan
Venice
Zagreb
Croatia
Bay of Biscay
Bosnia and Herzegovina
Florence
Monaco
Sarajevo
Montenegro
Italy
Andorra
Rome
Podgorica
Barcelona
Tirana
Albania
Madrid
Tyrrhenian Sea
Spain
Valencia
Seville
Granada
Algiers
Tunis
Málaga
Gibraltar
Malta

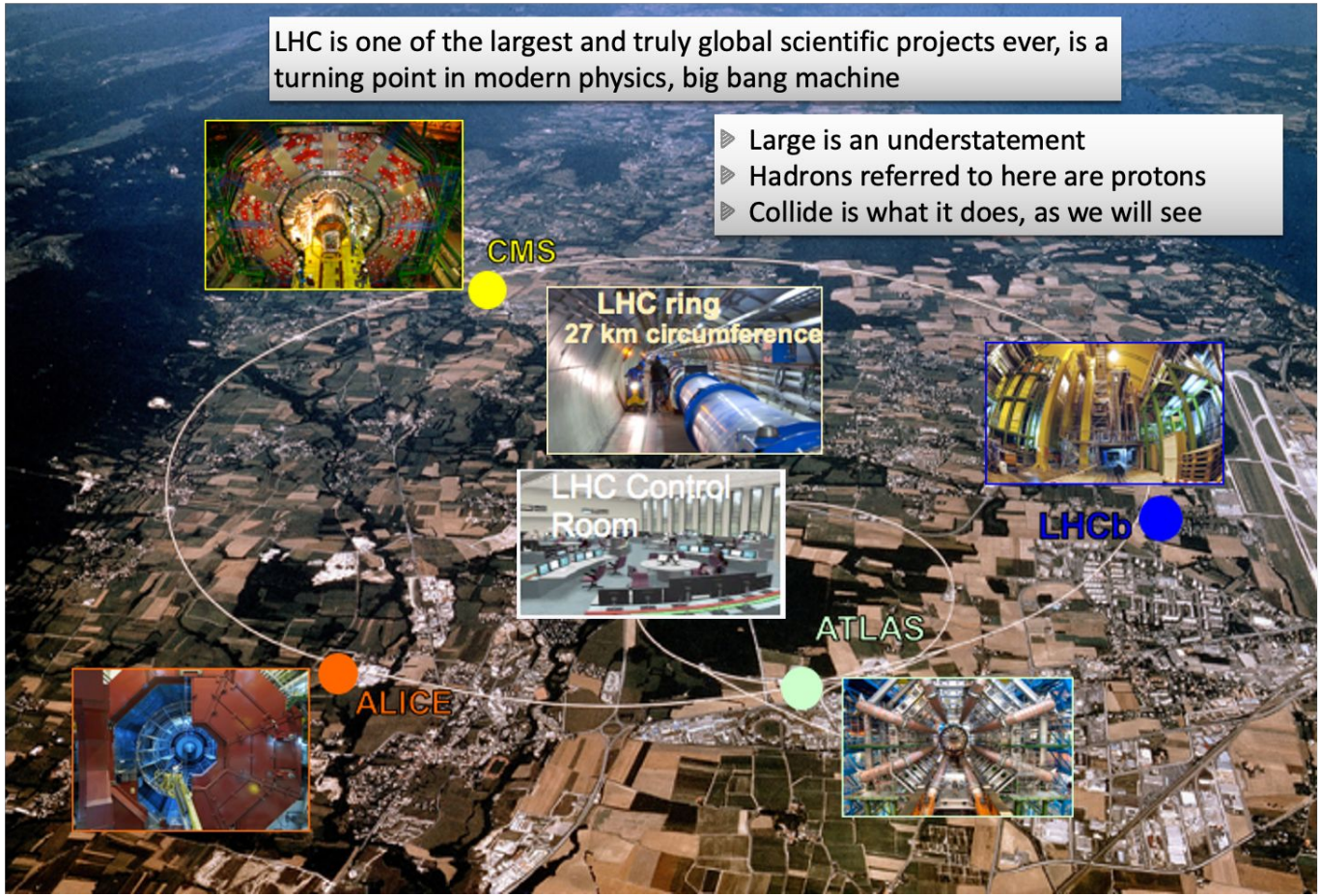

ABOUT NEWS SCIENCE RESOURCES SEARCH | EN
What is the nature of our universe?
What is it made of?
Welcome
Explore CERN
Events
Key Achievements


ABOUT NEWS SCIENCE RESOURCES SEARCH | EN
Scientists from around the world come to CERN …
Welcome
Explore CERN
Events
Key Achievements


ABOUT NEWS SCIENCE RESOURCES SEARCH | EN
… to use unique machines to answer fundamental questions …
Welcome
Explore CERN
Events
Key Achievements


CERN KNOWLEDGE TRANSFER
Maximising the global positive impact of CERN on society
ABOUT NEWS SCIENCE RESOURCES SEARCH | EN
… pushing the limits of technology for the benefit of society.
Welcome
Explore CERN
Events
Key Achievements

The tools to achieve the mission to find new physics



The Giant CMS Detector

Using the largest and most complex detectors ever b
With incredible precision

Like a 14,000 tonne 3D camera with ~75megapixels taking 40 million photos per second!
And it is 100m beneath us!

Crowning Achievement

The Nobel Prize in Physics 2013
François Englert, Peter Higgs

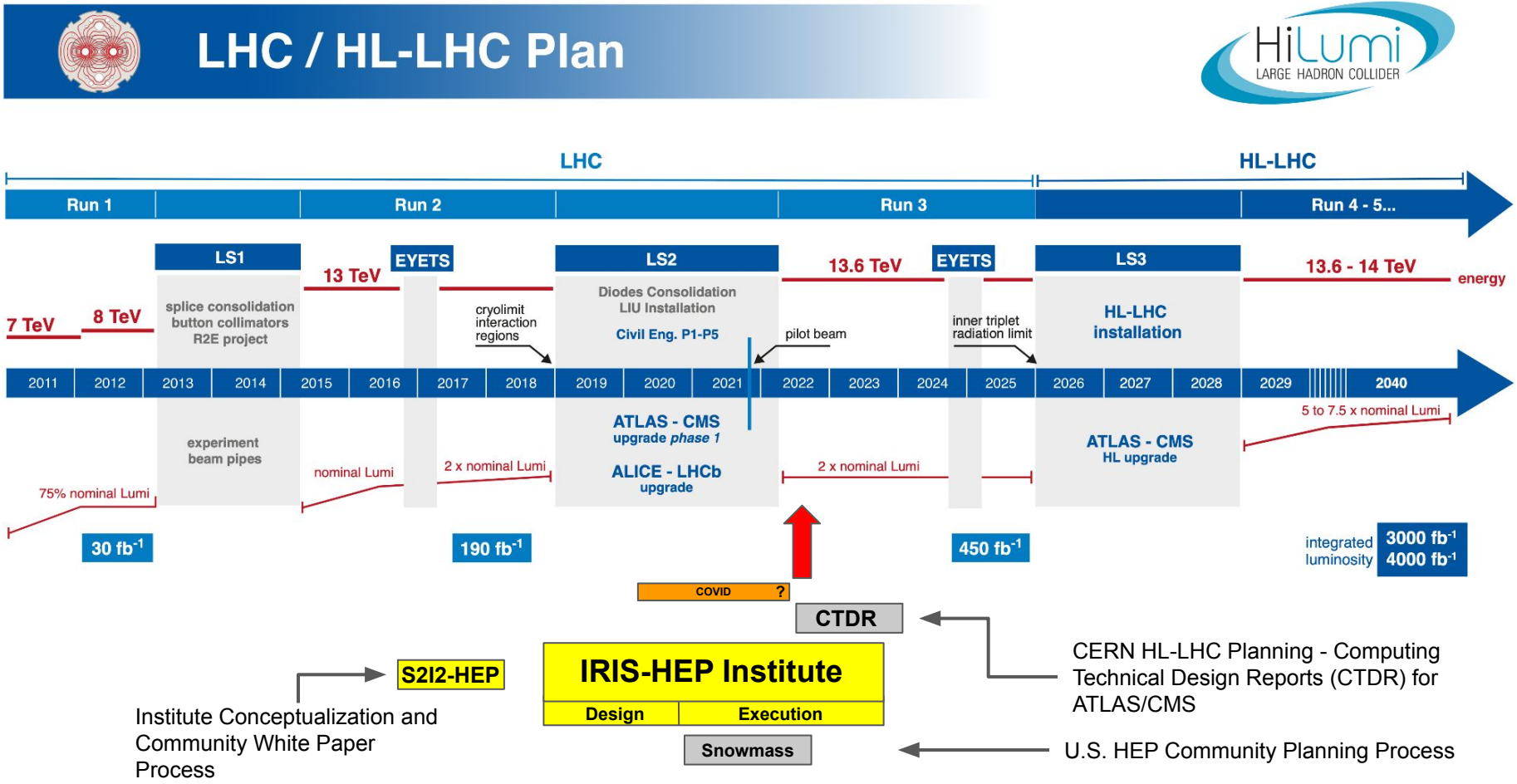
The Nobel Prize in Physics 2013

Photo: A. Mahmoud
François Englert
Prize share: 1/2

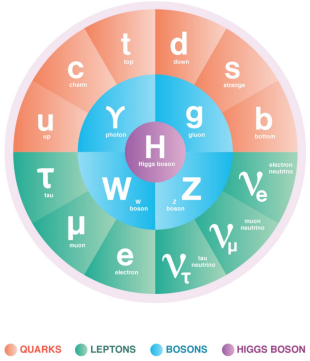
Photo: A. Mahmoud
Peter W. Higgs
Prize share: 1/2

The Nobel Prize in Physics 2013 was awarded jointly to François Englert and Peter W. Higgs *"for the theoretical discovery of a mechanism that contributes to our understanding of the origin of mass of subatomic particles, and which recently was confirmed through the discovery of the predicted fundamental particle, by the ATLAS and CMS experiments at CERN's Large Hadron Collider"*

HL-LHC Future

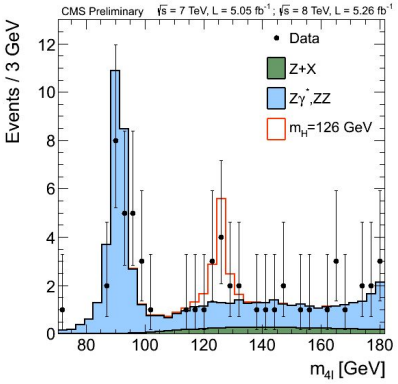


Science Driver: Discoveries beyond the Standard Model of Particle Physics

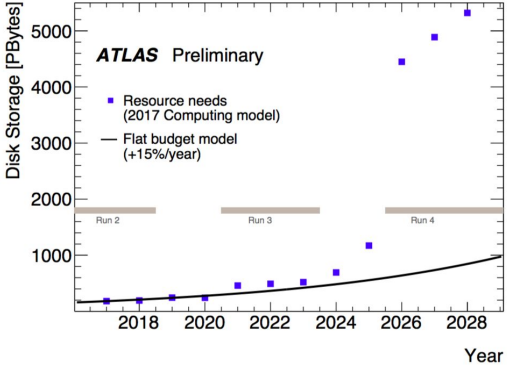


From "Building for Discovery - Strategic Plan for U.S. Particle Physics in the Global Context" - Report of the Particle Physics Project Prioritization Panel (P5):

1) Use the Higgs boson as a new tool for discovery
2) Pursue the physics associated with neutrino mass
3) Identify the new physics of dark matter
4) Understand cosmic acceleration: dark energy and inflation
5) Explore the unknown: new particles, interactions, and physical principles

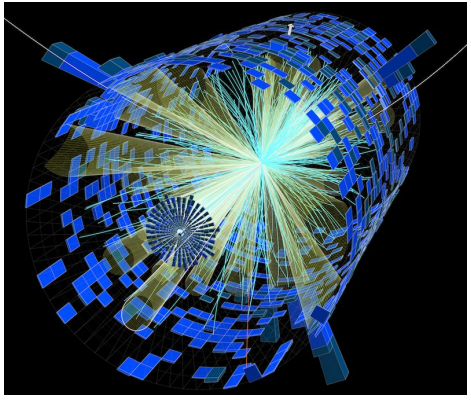


Computational and Data Science Challenges of the High Luminosity Large Hadron Collider (HL-LHC) and other HEP experiments in the 2020s



The HL-LHC will produce exabytes of science data per year, with increased complexity: an average of 200 overlapping proton-proton collisions per event.

During the HL-LHC era, the ATLAS and CMS experiments will record ~10 times as much data from ~100 times as many collisions as were used to discover the Higgs boson (and at twice the energy).

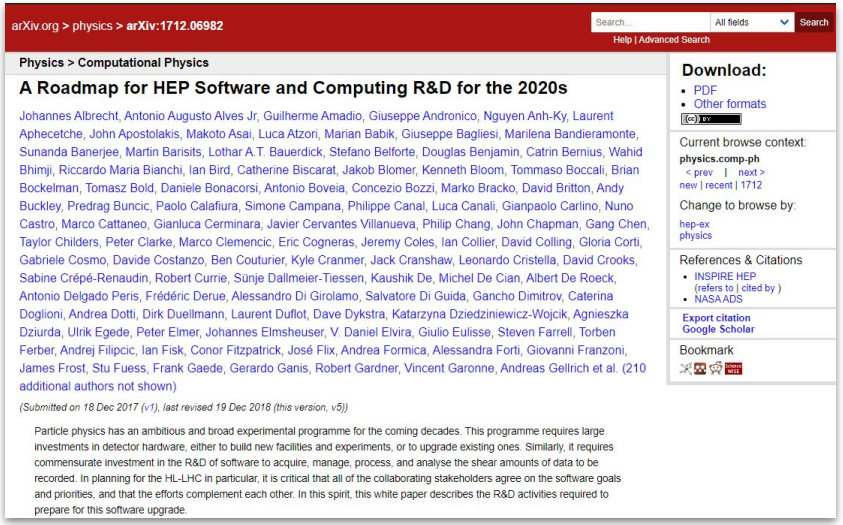


Birth of IRIS-HEP

Community White Paper

January 2017
UCSD

June 2017
Annecy



arXiv.org > physics > arXiv:1712.06982

Search... All fields Search

Help | Advanced Search

Physics > Computational Physics

A Roadmap for HEP Software and Computing R&D for the 2020s

Johannes Albrecht, Antonio Augusto Alves Jr, Guilherme Amadio, Giuseppe Andronico, Nguyen Anh-Ky, Laurent Aphecetche, John Apostolakis, Makoto Asai, Luca Atzori, Marian Babik, Giuseppe Bagliesi, Marilena Bandieramonte, Sunanda Banerjee, Martin Barisits, Lothar A.T. Bauerdick, Stefano Belforte, Douglas Benjamin, Catrin Bernius, Wahid Bhimji, Riccardo Maria Bianchi, Ian Bird, Catherine Biscarat, Jakob Blomer, Kenneth Bloom, Tommaso Boccali, Brian Bockelman, Tomasz Bold, Daniele Bonacorsi, Antonio Boveia, Concezio Bozzi, Marko Bracko, David Britton, Andy Buckley, Predrag Buncic, Paolo Calafiura, Simone Campana, Philippe Canal, Luca Canali, Gianpaolo Carlino, Nuno Castro, Marco Cattaneo, Gianluca Cerminara, Javier Cervantes Villanueva, Philip Chang, John Chapman, Gang Chen, Taylor Childers, Peter Clarke, Marco Clemencic, Eric Cogneras, Jeremy Coles, Ian Collier, David Colling, Gloria Corti, Gabriele Cosmo, Davide Costanzo, Ben Couturier, Kyle Cranmer, Jack Cranshaw, Leonardo Cristella, David Crooks, Sabine Crépé-Renaudin, Robert Currie, Sünje Dallmeier-Tiessen, Kaushik De, Michel De Cian, Albert De Roeck, Antonio Delgado Peris, Frédéric Derue, Alessandro Di Girolamo, Salvatore Di Guida, Gancho Dimitrov, Caterina Doglioni, Andrea Dotti, Dirk Duellmann, Laurent Duflot, Dave Dykstra, Katarzyna Dziedziniewicz-Wojcik, Agnieszka Dziurda, Ulrik Egede, Peter Elmer, Johannes Elmsheuser, V. Daniel Elvira, Giulio Eulisse, Steven Farrell, Torben Ferber, Andrej Filipcic, Ian Fisk, Conor Fitzpatrick, José Flix, Andrea Formica, Alessandra Forti, Giovanni Franzoni, James Frost, Stu Fuess, Frank Gaede, Gerardo Ganis, Robert Gardner, Vincent Garonne, Andreas Gellrich et al. (210 additional authors not shown)

(Submitted on 18 Dec 2017 (v1), last revised 19 Dec 2018 (this version, v5))

Particle physics has an ambitious and broad experimental programme for the coming decades. This programme requires large investments in detector hardware, either to build new facilities and experiments, or to upgrade existing ones. Similarly, it requires commensurate investment in the R&D of software to acquire, manage, process, and analyse the shear amounts of data to be recorded. In planning for the HL-LHC in particular, it is critical that all of the collaborating stakeholders agree on the software goals and priorities, and that the efforts complement each other. In this spirit, this white paper describes the R&D activities required to prepare for this software upgrade.

Download:
- PDF
- Other formats

Current browse context: physics.comp-ph
< prev | next >
new | recent | 1712

Change to browse by: hep-ex physics

References & Citations
- INSPIRE HEP (refers to | cited by)
- NASA ADS

Export citation
Google Scholar

Bookmark

Many workshops, involving a diverse group
- International participants
- Computing Management from the Experiments and Labs
- Individuals interested in the problems
- Members of other compute intensive scientific endeavors
- Members of Industry
- http://s2i2-hep.org/
- https://hepsoftwarefoundation.org/

Individual Papers on the arXiv:

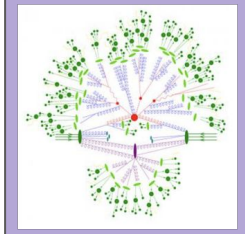
Careers & Training, Conditions Data, DOMA, Data Analysis & Interpretation, Data and Software Preservation, Detector Simulation, Event/Data Processing Frameworks, Facilities and Distributed Computing, Machine Learning, Physics Generators, Security, Software Development, Deployment, Validation, Software Trigger and Event Reconstruction, Visualization

Community White Paper & the Strategic Plan

arXiv 1712.06982 arXiv 1712.06592

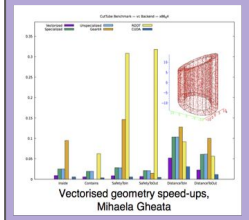
IRIS-HEP

HEP software ecosystem and priorities



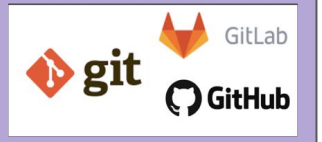
Physics Event Generators

Data, Software,
Analysis Preservation

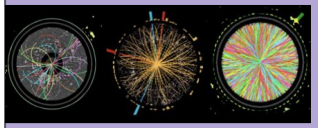


Detector Simulation

Security

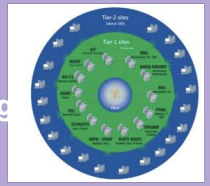


Software Development



Trigger,
Event Reconstruction

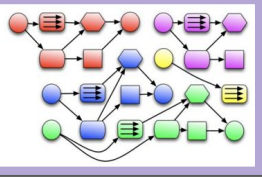
Data Analysis
Interpretation



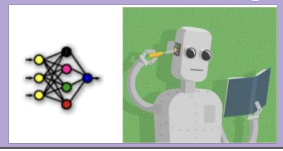
Facilities,
Distributed Computing

Visualization

Data Processing
Frameworks



Machine Learning

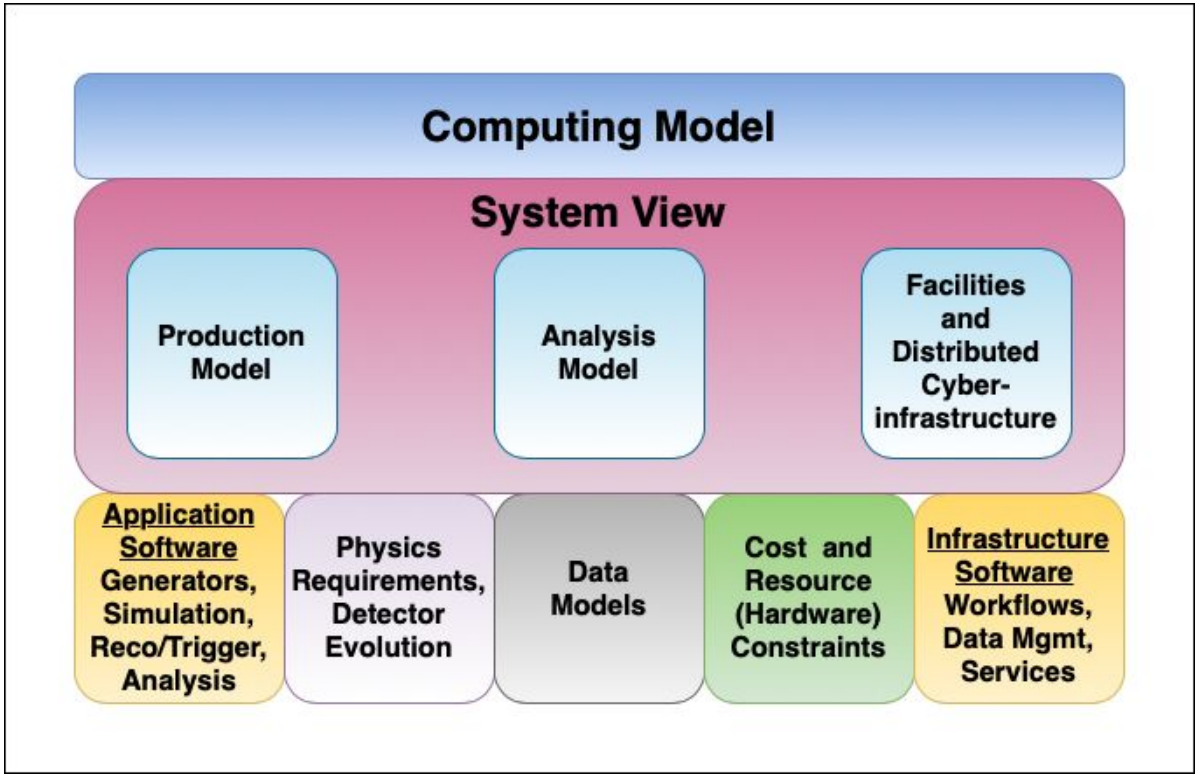


Data Management
Organisation Access

x

Why an Institute?

Software, System View & Computing Models



We aim to deliver more than software.
Big (Team) Science projects need “Computing Models”.

https://iris-hep.org/

Education & Training

Challenges and **Opportunities**

Not all HEP students can attend university-offered software courses

No standard curricula for HEP students exists

Experiments need **Cyberinfrastructure professionals** and lifelong learners

Democratize science by making software prerequisites accessible to everyone

We need a unified, scalable, and sustainable software training framework powered by the entire community

IRIS-HEP is leading training efforts and powered the **HSF Training WG**

HSF HEP Software Foundation

Challenges and Opportunities

We need a unified, scalable, and sustainable software training framework

Unified

- Material and events should be **centrally listed** and **discoverable**
- Concentrate efforts by prioritizing **cross-experiment** content
- A community must **guide**, **support**, and **coordinate**

Scalable

- The training initiatives need to reach **O(1000) students**/year
  ⇒ material must be teachable by **multiple instructors**
- **Self-study** must not be an afterthought

Sustainable

- Material must be **open source** and **maintained collaboratively**
- **Incentives & recognition** important motivators

A **unified** Training Center for HEP

Basics

The UNIX Shell
A guide through the basics of the file systems and the shell.
Start learning now!
Contribute!

Version controlling with git
Track code changes, undo mistakes, collaborate. This module is a must.
Start learning now!
Contribute!

Programming with python
Get started with an incredibly popular programming language.
Start learning now!
Contribute!

SSH
Introduction to the Secure Shell (SSH)
⚠ Status: Early development
Start learning now!
Contribute!

Machine learning
Get behind the buzzword and teach machines to work for you intelligently!
Start learning now!
Watch the videos!
Contribute!

Matplotlib for HEP
Make science prettier with beautiful plots!
✱ Status: Beta testing
Start learning now!
Contribute!

ROOT
The most famous data analysis framework used in HEP.
Start learning now!
Contribute!

Software Development and Deployment

Version controlling with git
Track code changes, undo mistakes, collaborate. This module is a must.
Start learning now!
Contribute!

Advanced git
Learn to work with branches and more with this interactive webpage.
Start learning now!
Contribute!

CI/CD (gitlab)
Continuous integration and deployment with gitlab.
Start learning now!
Watch the videos!
Contribute!

CI/CD (github)
Continuous integration and deployment with github actions.
Start learning now!
Watch the videos!
Contribute!

Docker
Introduction to the docker container image system.
Start learning now!
Watch the videos!

Singularity
Introduction to containerization with Singularity/Apptainer.
✱ Status: Beta testing
Start learning now!
Watch the videos!
Contribute!

(truncated)

- Training Center currently lists **21 training modules** (including material developed by IRIS-HEP/HSF, The Carpentries, and individual authors)
- Goal: Training Center as a **focal point for HEP Training activities**
- Central **list of training events**

Current and Upcoming Training Events

- **17 Jul - 21 Jul 2023** - CODAS-HEP 2023 - **Deadline:** 26 Apr 2023
- **7 Aug - 11 Aug 2023** - Open Science Grid User School 2023 - **Deadline:** 17 Apr 2023

Past Events

- **6 Mar - 10 Mar 2023** - 6th HEP C++ Course and Hands-on Training - The Essentials HSF
- **16 Jan - 20 Jan 2023** - Analysis Preservation Workshop HSF
- **11 Oct - 13 Oct 2022** - 5th HEP C++ Course and Hands-on Training - Advanced C++ HSF
- **3 Oct - 8 Oct 2022** - ESC22 EFFICIENT SCIENTIFIC COMPUTING
- **28 Sep - 30 Sep 2022** - HSF/IRIS-HEP Software Basics Training HSF
- **4 Sep - 17 Sep 2022** - CERN School of Computing 2022
- **13 Jul - 15 Jul 2022** - Software Carpentry (Virtual) HSF
- **19 Jun - 25 Jun 2022** - Thematic CERN School of Computing on *"Security of Research Computing Infrastructures"*
- **1 May - 7 May 2022** - Thematic CERN School of Computing on *"Scientific Software for Heterogeneous Architectures"*
- **21 Apr - 22 Apr 2022** - Matplotlib Training (Virtual) HSF
- **28 Mar - 30 Mar 2022** - Software Carpentry (Virtual) HSF
- **15 Mar - 17 Mar 2022** - 4th HEP C++ Course and Hands-on Training - The Essentials
- **13 Dec - 15 Dec 2021** - Software Carpentry (Virtual) HSF
- **10 Oct - 16 Oct 2021** - Thematic CERN School of Computing (tCSC autumn 2021)
- **25 Aug - 27 Aug 2021** - Software Carpentry (Virtual) HSF
- **17 Jul - 22 Jul 2021** - 2021 Intensity Frontier Computing Summer School
- **14 Jun - 18 Jun 2021** - Thematic CERN School of Computing (tCSC spring 2021 - online)

Scale and sustain

Everything open-source

Search or jump to... Explore

HEP Software Foundation Training Material

Training and educational material for the high energy physics community.

41 followers https://hepsoftwarefoundation.org/... @HsfTraining hsf-coordination@googlegroups.com

Overview Repositories 42 Projects 5 Packages Teams 3 People 28 Settings

README.md

Welcome to HSF Training 🤗

Weekly meetings

October 2022

24 Oct Training WG Planning Meeting

17 Oct Training WG Planning Meeting

10 Oct Training WG Planning Meeting

Community pages

Our community

Amber Roepe (she/her)

Slack

Monthly hackathons

HSF iris hep

Containerization Training Hackathon

The big goal!

we do a little hacking

Recognition

Authors

Thanks goes to these wonderful people (emoji key) who contributed

Meirin Oan Evans

amorenobr

Guillermo A. Fidalgo-Rodriguez

Marchela04

Most training modules are website built from easy-to-read source files (Markdown)

HSF Home Code of Conduct Setup Episodes Extras License Improve this page Search...

Machine Learning on GPU

Machine Learning on GPU - Setup

Watch on YouTube

< Complete video walkthroughs!

This tutorial explores Machine Learning using GPU-enabled PyTorch for applications in high energy physics.

It follows directly from the Introduction to Machine Learning lesson written by Meirin Evans.

Prerequisites

- A Kaggle account. Click here to create an account
- Basic Python knowledge, e.g. through the Software Carpentry Programming with Python lesson
- Basic ML knowledge, e.g. through the Introduction to Machine Learning lesson

^ Lessons build on each other

Introduction

For physicists working on analysis in data-intensive fields such as particle physics, it's quite common these days to start developing new machine learning applications. But many machine learning applications run more efficiently on GPU.

The aim of this lesson is to:

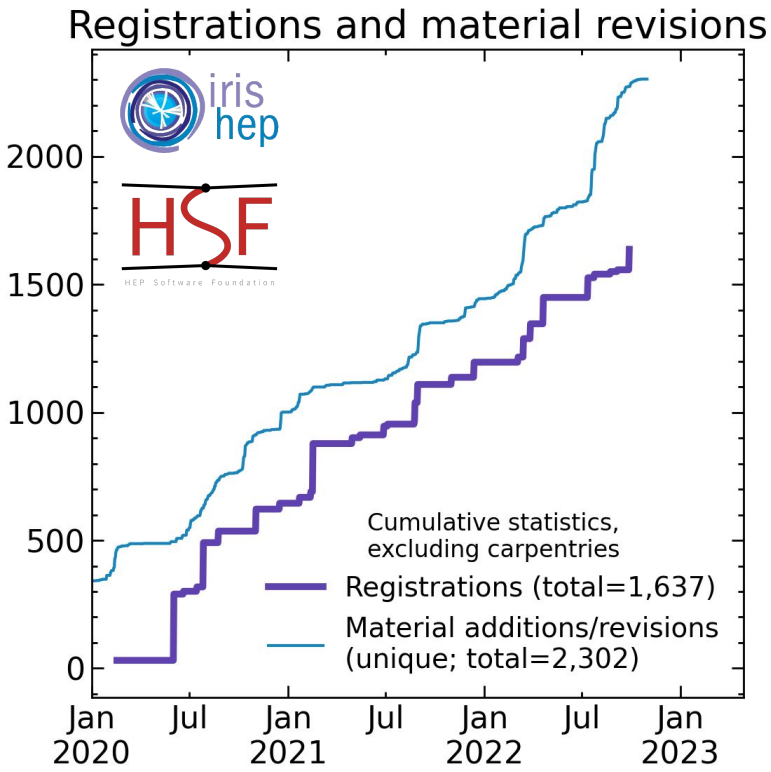
- demonstrate how to move an existing machine learning model onto a GPU
- discuss some of the common issues that come up when using machine learning applications on GPUs

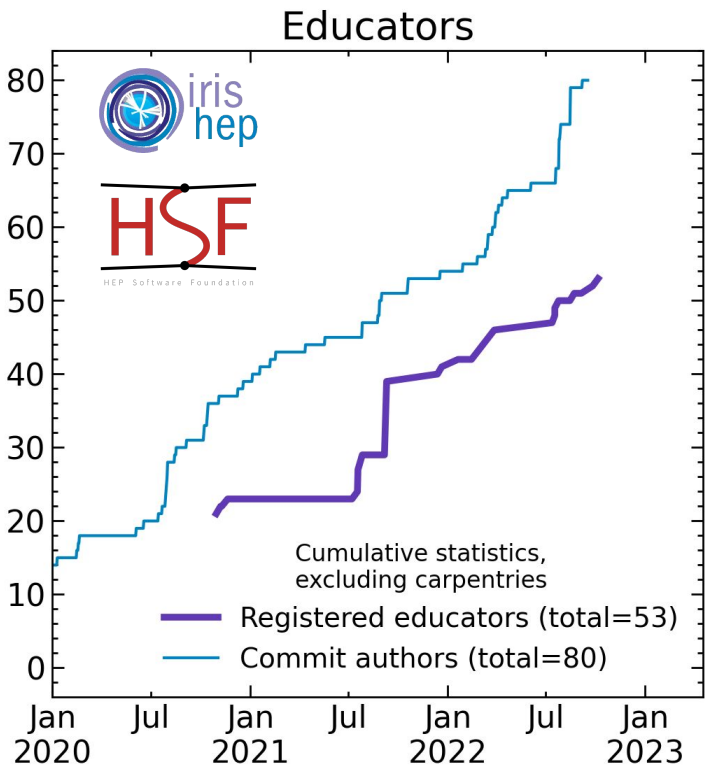
The skills we'll focus on:

1. Understanding a bit about GPUs
2. Using Python & PyTorch to discover what kind of GPU is available to you
3. Moving a machine learning model onto the GPU
4. Comparing the performance of the machine learning model between the CPU and the GPU

^ Enough verbosity for self-study

Scale and sustain





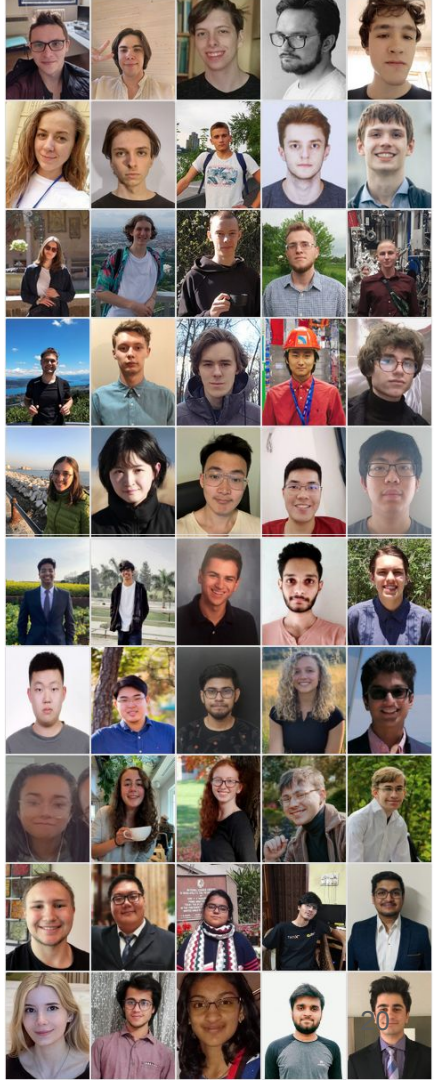
✅ Mentoring and developer training

Fellowship program: 3 months remote program for undergraduates

- More than **100 fellows** trained!
- Fellows work on software projects together with mentor
- IRIS-HEP serves as organizing entity; students supported by a variety of funding sources

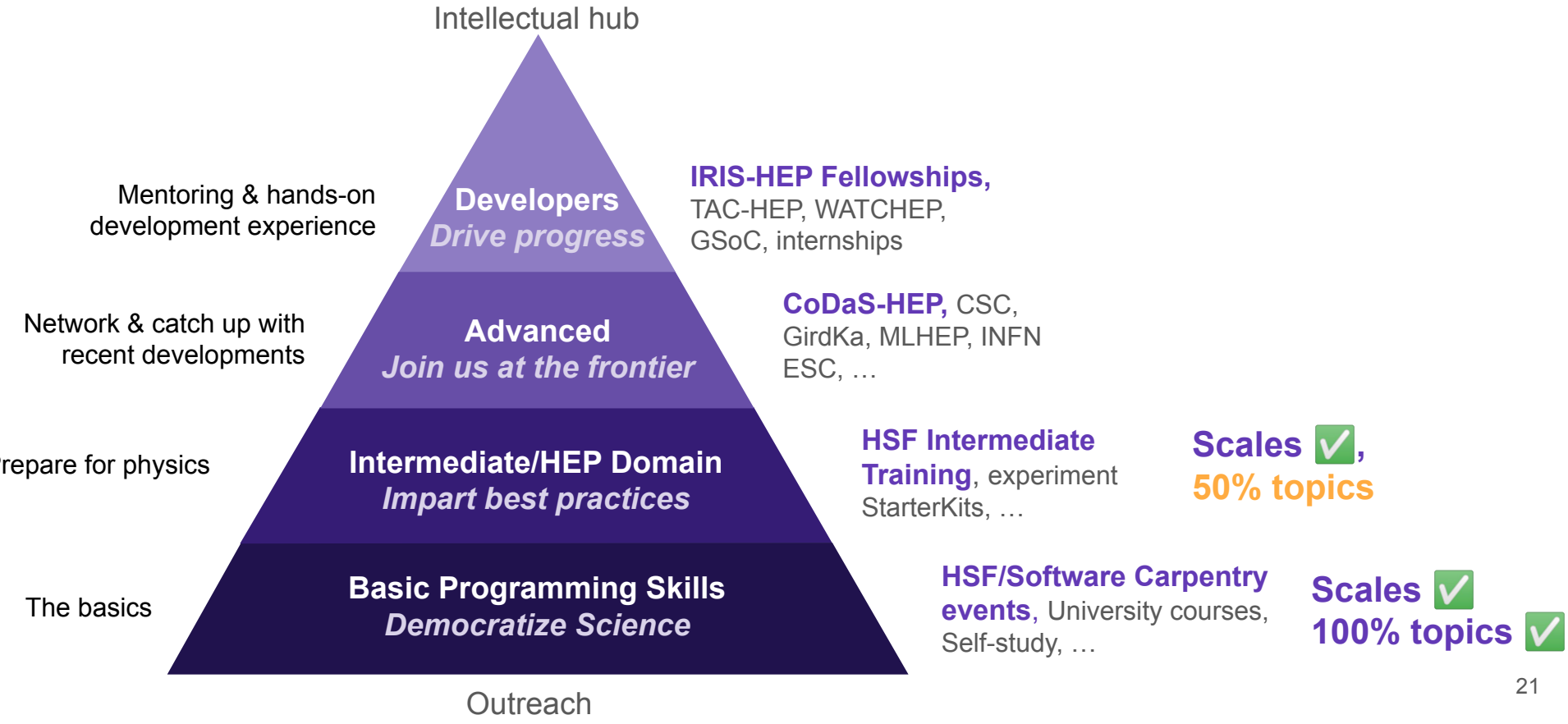
CoDaS-HEP: One week intensive in-person developer training

- Co-funded by CyberTraining grant
- Open to Ph.D. students and Postdocs
- 2017, 2018, 2019, 2022



Training Cyberinfrastructure professionals

IRIS-HEP is supporting the whole journey



Going **beyond**

✅ Basic-intermediate training

- ✅ Unified **community** of educators
- ✅ **100%** of basic prerequisites, **50%** of intermediate topics
- ✅ Demonstrated **sustainable scale out** using training modules (“Carpentry Style”)
- ✅ Successful **in-person training** and **hands-on experience** program

🚀 Scalable intermediate- advanced training

- **Plan:** Apply similar “Carpentry-style” paradigms; seed small working groups; strengthen collaborations with training groups beyond HEP
- **Challenging:** Fewer experts, fewer students, more mentoring required
- **Rewarding:** This trains the Cyberinfrastructure Professionals who drive software development!
- **Our strong suit:** Unify efforts, leverage existing community, use experience in creating sustainable & scalable material

The Training **Grand Challenge**

- Seed **topical WGs** for advanced material
- Rebuild **Training Center**
- New **liaisons** (HEP & beyond)

- **80%** of cross-experiment software topics for HEP Ph.D. students covered
- **50%** of HEP Ph.D. students received intermediate training

2018 - 2022 ✅

Democratized science, Scaled out basic/ intermediate training

2023

2024 - 2025

- **Teach first new** intermediate- advanced training modules

2026 - 2027

2027 - 2028

- **20%** of HEP Ph.D. students received advanced training
- **20%** of HEP Ph.D. students have supported one of the workshops

Outreach

Overview

Vision: - IRIS-HEP workforce development activities contribute to the preparation of a highly qualified STEM workforce

Activities: In the current phase of IRIS-HEP we built a set of core software training curriculum for teachers

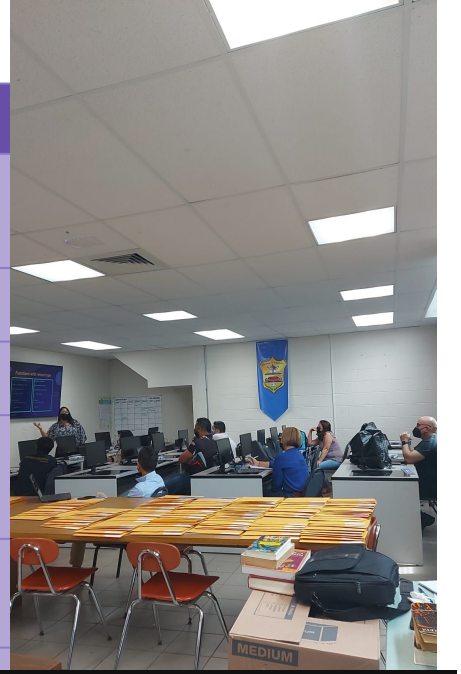
- 13 software training workshops (some overlapped with COVID era)

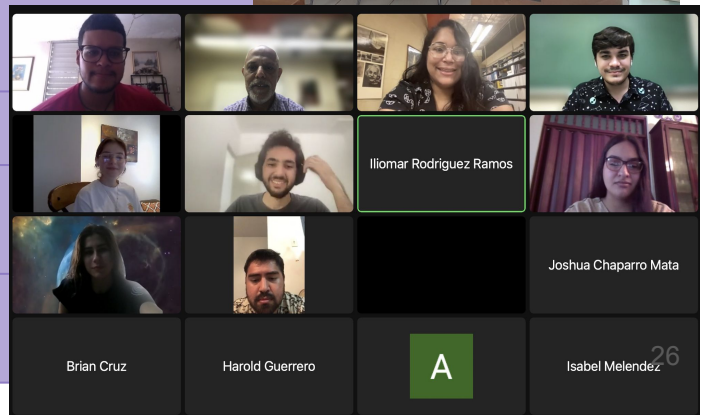
Impact

- Over 100 teachers software trained

Activities

Year	Month	Name	Participants/Tutors
2023	June/July	5 Coding Camps (Rice, Alabama, Washington, Fermilab, Puerto Rico)	
2022	July	Coding Camp at Fermilab (1 week) with quarknet	21/5
2022	Mar	Data Analysis for Lab Research (Virtual) (1 day)	20/3
2021	Aug	Arduino Micro Controller and C++ programming (STEM Teachers) (In person) (2 days)	9/3
	Feb	Machine Learning Basics for STEM teachers (Virtual) (2 days)	8/3
2020	July	Data Analysis for STEM teachers (Virtual) (2 days)	16/3
	June	Data Camp for STEM teachers (Virtual) (1 day)	11/3
2019	June	An introduction to programming for STEM teachers (2 days)	16/1
	April	Machine Learning Workshop/Hackathon (3 days)	25/1

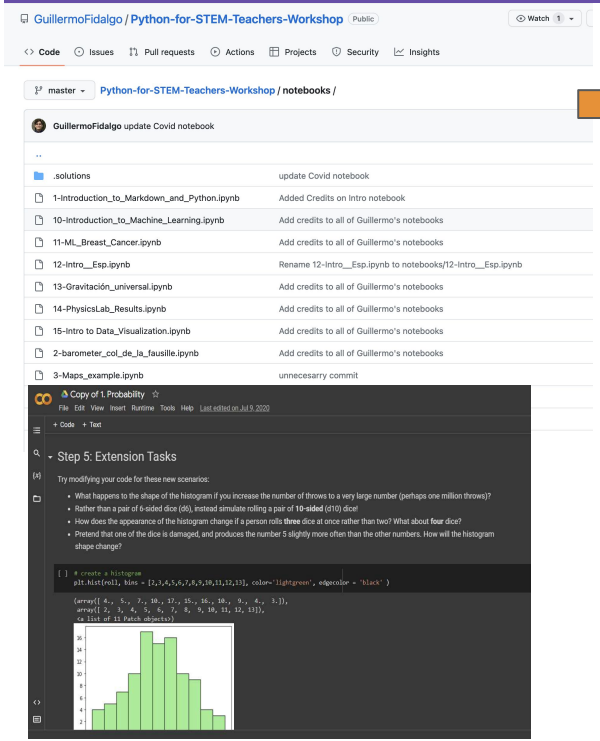




Software Training

Brings software experience to high school science teachers

Modules in GitHub



"student hat"
Engage, explore, cexplain

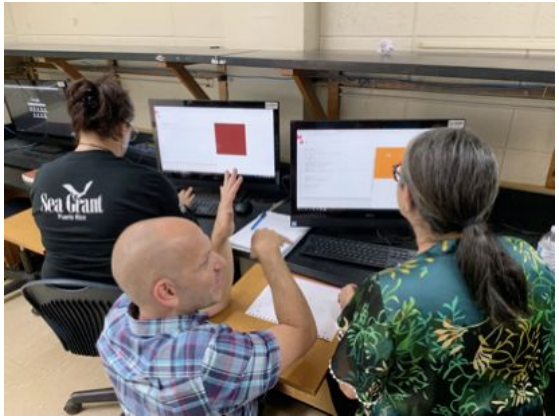
- Teachers work in groups
- Running Python code
- Using pre-Jupyter notebook
- Review basic coding
- Use CMS data

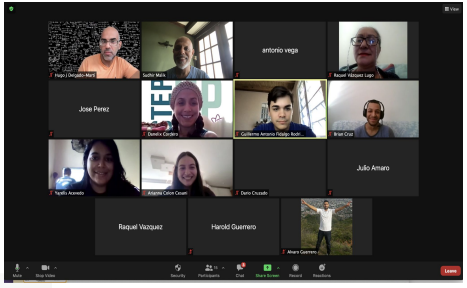
"teacher hat"
Elaborate and Evaluate

- Teachers develop implementation plans for their own classroom Writing Jupyter notebook
- Adapt and apply appropriately to their classroom

Workshop Example

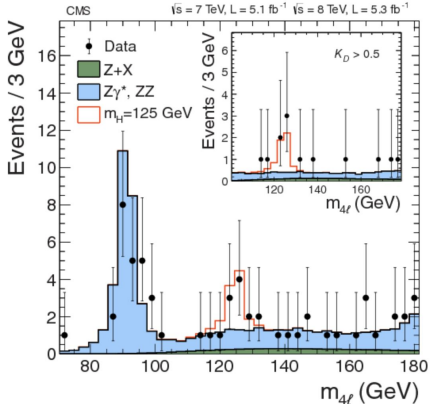
- Introduction to Programming
- Python, Jupyter notebooks, Colab
- Hands-on Data analysis
- HEP data preview with CMS Open Data Examples





Higgs hunting - an example of scientific research

```
plt.figure(figsize = (15,5))

# ttbar
ttbar_bar = plt.bar(center, ttbar, align = 'center', width = width, color = 'gray', linewidth = 0, edgecolor = 'b',
                    alpha = 0.5, label = r'$t\bar{t}$')

# DY
dy_bar = plt.bar(center, dy, align = 'center', width = width, color = 'g', linewidth = 0, edgecolor = 'black',
                 alpha = 0.5, bottom = ttbar, label = 'Z/$\gamma^{*}$ + X')

# ZZ
zz_bar = plt.bar(center, zz, align = 'center', width = width, color = 'b', linewidth = 0, edgecolor = 'black',
                 alpha = 0.5, bottom = ttbar+dy, label = r'ZZ $\rightarrow$ 4l')

# HZZ
hzz_bar = plt.bar(center, hzz, align = 'center', width = width, color = 'w', linewidth = 1, edgecolor = 'r',
                  bottom = ttbar+dy+zz, label = '$m_{H}$ = 125 GeV')

# Measured data
data_bar = plt.errorbar(center, hist, xerr = xerrs, yerr = yerrs, linestyle = 'None', color = 'black',
                        marker = 'o', label = 'Data')

plt.title('$ \sqrt{s} = 7$ TeV, L = 2.3 $fb^{-1}$; $\sqrt{s} = 8$ TeV, L = 11.6 $fb^{-1}$ \n', fontsize = 12)
plt.xlabel('$m_{4l}$ (GeV)', fontsize = 15)
plt.ylabel('Events / 3 GeV\n', fontsize = 15)
plt.ylim(0,25)
plt.xlim(rmin,rmax)
plt.legend(fontsize = 15)

plt.show()
```




```
[ ] # Let's set some values here in regards to the region we're looking at.

    rmin = 70
    rmax = 181
    nbins = 37

    M_hist = np.histogram(fourlep['M'], bins = nbins, range = (rmin,rmax))

    hist, bins = M_hist #hist=frecuency ; bins=Mass values
    width = 1.0*(bins[1] - bins[0])
    center = (bins[:-1] + bins[1:]) / 2
```

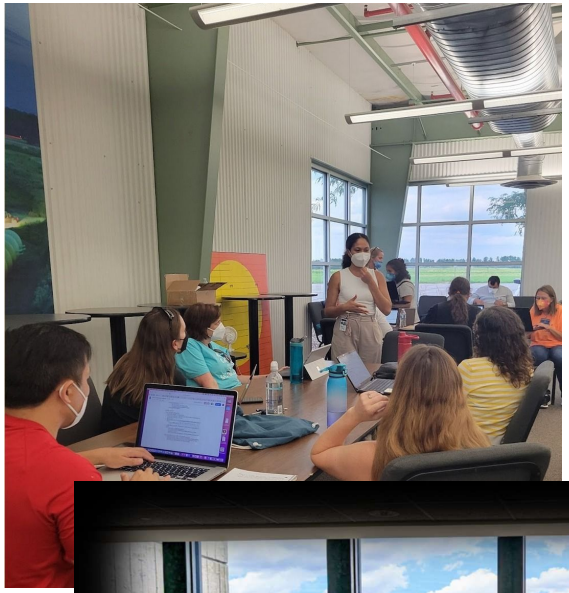
Let's look at some simulations from other processes there. Here are some Monte Carlo -simulated values for such events that have already been weighted by luminosity, cross-section and number of events. Basically we create a set of values that have some randomness in them, just like a real measurement would have, but which follows the distribution that has been observed in those processes.

```
dy = np.array([0,0,0,0,0,0.354797,0.177398,2.60481,0,0,0,0,0,0,0,0,0,0.177398,0.177398,0,0.177398,0,0,0,0,0,0,0,0,0,0,0.177398,0,0,0,0])
ttbar = np.array([0.00465086,0,0.00465086,0,0,0,0,0,0,0,0,0.00465086,0,0,0,0,0,0.00465086,0,0,0,0,0,0.00465086,0,0,0,0,0,0.00465086,0.00465086,0,0,0,0.0139526,0,0,0.00465086,0,0,
zz = np.array([0.181215,0.257161,0.44846,0.830071,1.80272,4.57354,13.9677,14.0178,4.10974,1.58934,0.989974,0.839775,0.887188,0.967021,1.07882,1.27942,1
hzz = np.array([0.00340992,0.00450225,0.00808944,0.0080008,0.00801578,0.0108945,0.00794274,0.00950757,0.0130648,0.0163568,0.0233832,0.0334813,0.0427229
```

Let's take a look at those numbers and how they contribute to what we'll measure in the accelerator.

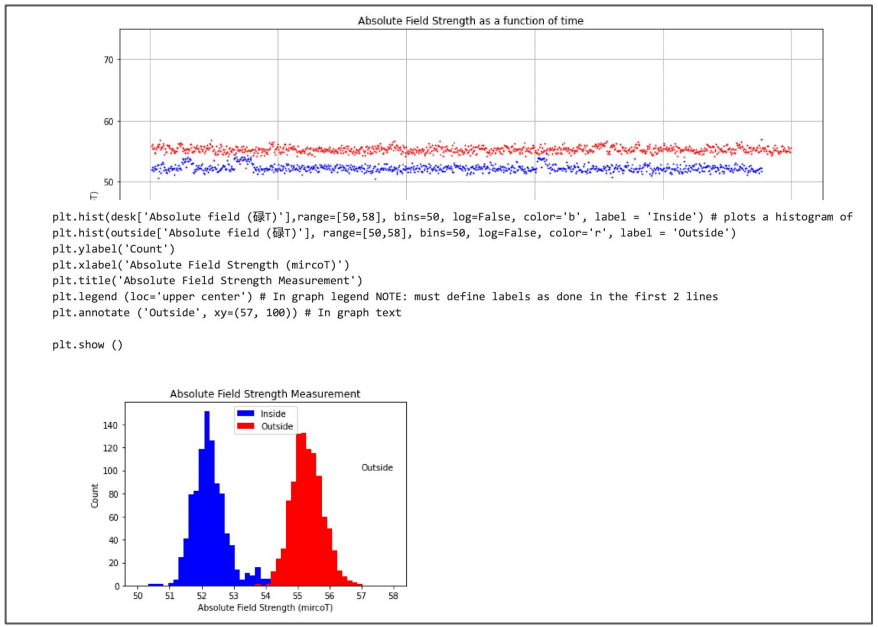
Scale up

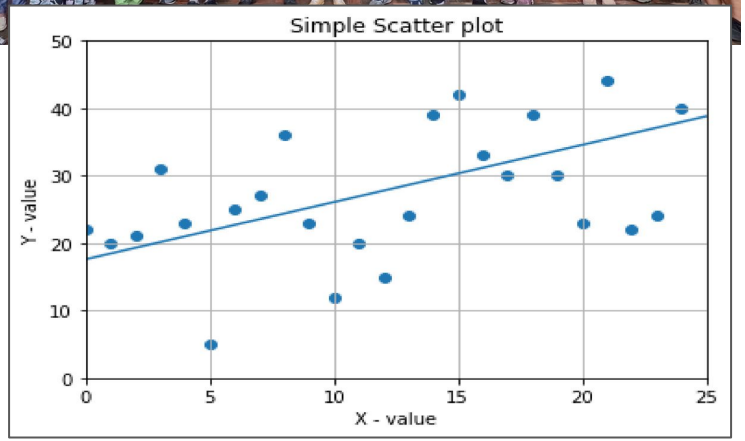
- Reach out to US wide community of high school teachers via Quarknet
- Provide software training for science teachers
- Coding Camps with Quarknet



Coding Camp

- Funding Support from IRIS-HEP
- Content creation and support
- Access to HEP software community





Building on Success

- Developed Software modules (teachers input involved)
- Software awareness and skill development among high school students via teachers
- Relation with community of teachers to expand and sustain our efforts
- Access to wider community of teachers to get software training
- Breaks barriers and enable diversity

```
import pandas as pd # Data manipulation
import numpy as np # Matrix operations
import statsmodels.formula.api as smf # Statistics
import matplotlib.pyplot as plt
import seaborn as sns
import matplotlib.style as style
from pandas.plotting import scatter_matrix
# Options for pandas
pd.options.display.max_columns = 50
pd.options.display.max_rows = 30
# Display all cell outputs
from IPython.core.interactiveshell import InteractiveShell
InteractiveShell.ast_node_interactivity = 'all'
from IPython import get_ipython
ipython = get_ipython()
# autoreload extension
if 'autoreload' not in ipython.extension_manager.loaded:
    %load_ext autoreload
%autoreload 2
```

```
/usr/local/lib/python3.6/dist-packages/statsmodels/tools/_testing.py:19: FutureWarning: pa
s in the public API at pandas.testing instead.
  import pandas.util.testing as tm
```

Data Wrangling

Data wrangling is the process where you manipulate the data in your favour, without dismissing any value a
In this step, you analyze your data to find and replace missing values acording to different techniques, clea

Data Exploratory Analysis

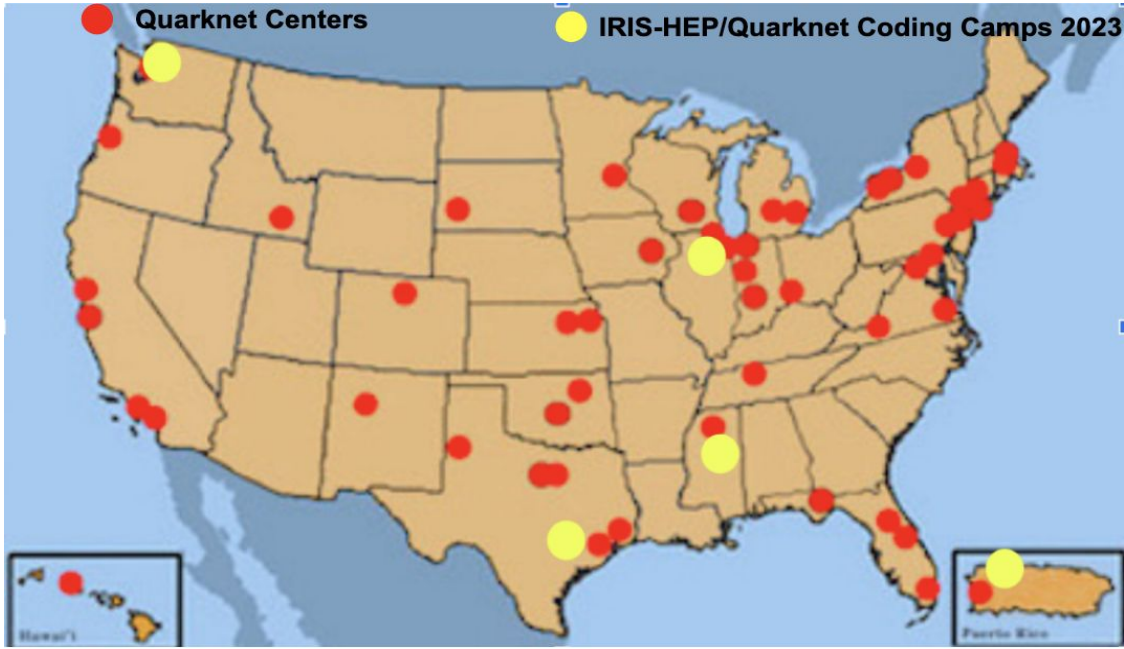
Milestones

End of 2 years

- Modules focussing High School Physics curriculum
- Develop workshops in Spanish language (enables diverse participation)
- 100% Coding Camps run by IRIS-HEP

By 5 years

- 100% Software Training across the nations 50 Quarknet Centers and beyond



Intellectual Hub

Leading the Field to Build Consensus

HEP Software Foundation

- Community White Paper Process in 2018-2019
  - Run by NSF funded planning grant that lead to IRIS-HEP
- Needed: Umbrella organization to “facilitate **cooperation and common efforts** in High Energy Physics software and computing internationally”

Analysis Ecosystem Workshops

- May 2017, May 2022
- First helped shape the Analysis Systems focus area of IRIS-HEP and HSF, and second report provides field’s blueprint for next ~5 years

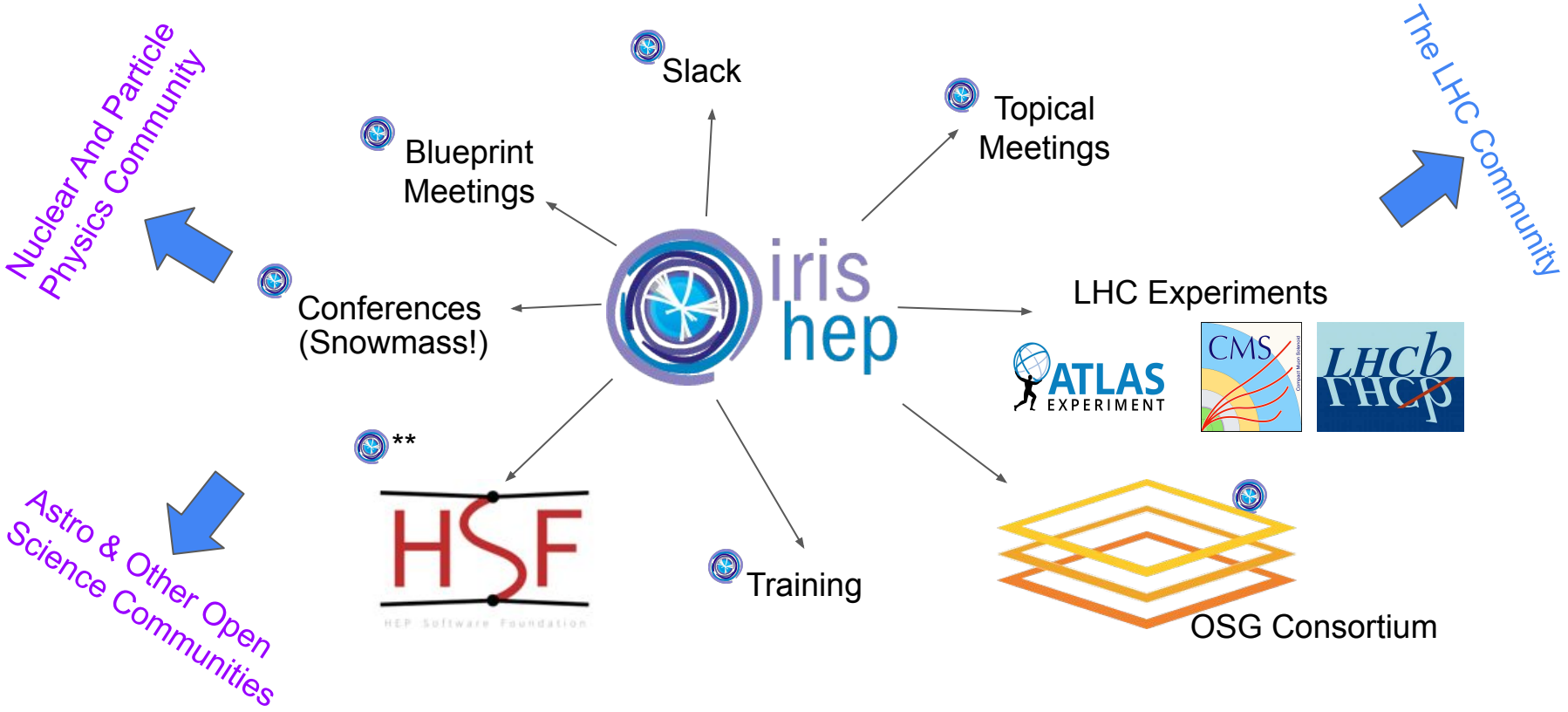
A Coordinated Ecosystem for the HL-LHC Computing R&D

- Oct 2019, Nov 2022
- Bring together institutes and grants working on computing R&D for the HL-LHC
- Groups from NSF & DOE (e.g. A3D3, IAIFI, HEP-CCE)
- Gaps, opportunities, collaborations possible

What Makes IRIS-HEP Successful?

1. The work we've done
2. Our Longevity and Scope
3. Our Connections with the HL-LHC Community

Intellectual Hub



Outreach & Impact on Other Communities

Projects that have **found uses** outside of HEP/LHC:

- Servicex (astro), iDDS (astro and other HEP) ACTS (nuclear physics), uproot/awkward (astro, theory, education/outreach), pyhf (other HEP)
- OSG Consortium (chemistry, bio, etc.)

We are always on the lookout for new collaborations:

- We attend conferences like SciPy that reach a broad scientific audience
- Well known in field that researchers from adjacent sciences contact us

Our contributions & Collaborations to OSS enable science and tools beyond HEP

pybind11 — H.Schreiner now a key contributor (182 merged PRs) and the release manager. We help build and manage a team including developers from Facebook and Google that are moving pybind11 forward.

matplotlib — MPL lead T.Caswell participates in IRIS-HEP Slack. Multiple IRIS-HEP team members are making PRs/issues.

DASK — Directly interacting with Anaconda to add the Awkward Array collection type, which is pushing changes in core Dask. (dask-awkward is the first high-level graph outside of core Dask.)

ANACONDA. — Several points of contact through Numba, Dask, and Pangeo, and Anaconda employees are active on the IRIS-HEP Slack. We also are heavily involved in conda-forge, including maintaining Numba's recipe.

kitware — We help maintain the CUDA support in CMake, have helped with OpenMP, and other performance related areas important to our field and many others. (9 merged PRs)

scikit-build — H. Schreiner joined scikit-build, created by kitware, and has been managing releases and fixing long standing issues. This includes CMake for Python (33 merged PRs) and related packages, used across the ecosystem.

pypa — H. Schreiner is a core developer on both build and cibuildwheel (116 merged PRs), key tools in the Python packaging ecosystem. Cibuildwheel has over 600 users, including NumPy, scikit-learn, and more. We have upstreamed parts of our developer guidelines to packaging.python.org documentation, and are helping with many of the pypa projects. We had a well attended talk at PyCon US 2022 over cibuildwheel, build, pybind11, and scikit-build.

APACHE ARROW — Awkward Array stresses cutting-edge features in Arrow and we have reported many issues.

PANGEO — Helping Pangeo projects adopt Awkward Array, including an analysis of ocean probes at FSU (OAC-2126413), refactoring spatialpandas, and selecting Argo data.

Numba — Awkward Array stresses cutting-edge features in Numba and Awkward Array's tests are now part of Numba's integration testing. We regularly attend Numba developer meetings.

PYODIDE — We have worked with the Pyodide developers (22 merged PRs) to improve their build system, adapt pypa/build, improve static checks for bugs, update packages, and added one our our packages to the default set of packages.

pypy — We work regularly with a lead developer of PyPy / NumPy to support PyPy as a first class target in cibuildwheel, build, setuptools, and pybind11.

International Projects Inspired

HSF — HEP Software Foundation

SWIFT-HEP

SIDIS Software Institute for Data-Intensive Sciences — Software Institute For Data Intensive Sciences

NSF — HSF-India (spin-off), FAIROS-HEP

Python Tool Ecosystem

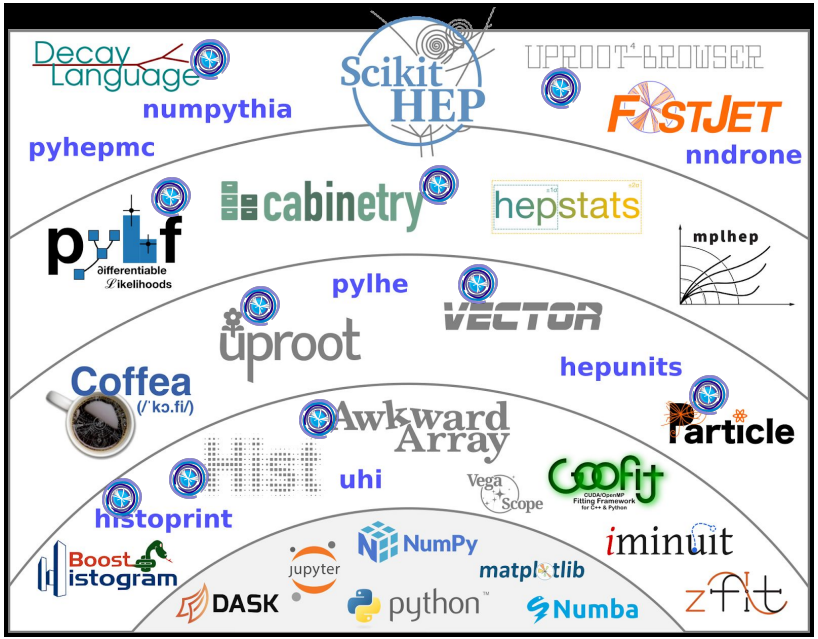
[The OSS Ecosystems for data]

Scikit HEP

A community-driven and community-oriented project providing Particle Physics at large with a Python package containing core and common tools.

Domain Specific

Foundational



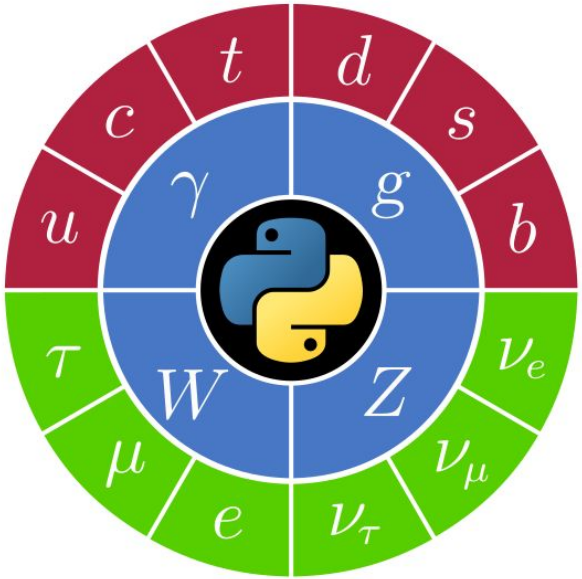
The PyHEP Conference Series

- Brings Developers & Users of the Python Eco-System together
- Future development directions
- Lead by IRIS-HEP members and others
- Sponsored by IRIS-HEP

Growth over the pandemic era

PyHEP.dev - Summer 2023 (link)

PyHEP - Fall 2023



2022 (link)
> 1000 participants

Intellectual Hub

A large group of people across the spectrum are working to build a **coherent ecosystem**

- Many communication tools are available (Slack, Mattermost, gitter, discourse, etc.)
- Our long-term role as intellectual hub allows it to fill this role.

(also email lists, etc.)

The IRIS-HEP Slack

Discussion channels on a broad range of topics inside and out of IRIS-HEP

- On packages we are authoring
- On broad experiment-related topics (e.g. systematic error handling)
- Gathering place for those working on related topics (e.g. coffee developers are in many channels)

We will continue to support grass-roots communications in the community.

Broadening Participation

High School STEM Students → Postdocs & Professors

Both target underrepresented minorities and HBCU's

U.S. CMS - PURSUE (Program for Undergraduate Research SUmmer Experience)

Tulika Bose[1], Sudhir Malik[2], Meenakshi Narain[3]
[1]Physics Department, University of Wisconsin Madison, Madison, WI 00682, USA
[2]Physics Department, University of Puerto Rico Mayaguez, Mayaguez, PR 00682, USA
[3]Department of Physics, Brown University, Providence, RI 02912, USA
E-mail: sudhir.malik@upr.edu

Direct Efforts **built on our well understood summer fellow mentoring program**.

- US CMS PURSUE
- Princeton's PACRI

Princeton's PACRI Program (HBCU's)

Next step: scaling up our pilot runs to a larger audience.

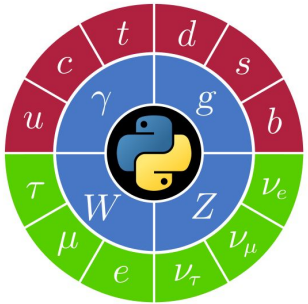
Tracking Progress

	Year 1	Year 2	Year 3	Year 4	Year 5
Training	First intermediate/ advanced Training Modules	**Scale** →		**80%** of cross-experiment software topics for Ph.D. S. covered **50%** of Ph.D. S. participated in intermediate/advanced training, **20%** have helped in event	
Outreach		Modules focussing High School Physics curriculum Develop workshops in Spanish language (enables diverse participation)		Coding Camps run at all Quarknet Centers	

Conclusions - Broader Impacts and Intellectual Hub

Large proposed program to reach many in the field, and students and STEM teachers:

- Reach 50% of all U.S. Ph.D.'s who have moved beyond beginner training modules
- Reach ~1000 students per year with our training
- Scale out our Coding Camp materials using QuarkNet to reach a larger audience of STEM teachers and students
- Build field-wide consensus around the path towards Analysis at the HL-LHC, a coordinated R&D program for computing at the HL-LHC in the USA, and Translational AI



HSF HEP Software Foundation

ATLAS EXPERIMENT

CMS

LHCb