Evolving Research Software towards Next-Generation High-Energy Physics Experiments

Peter Elmer

Princeton University

13 May 2025 - IIT Bombay

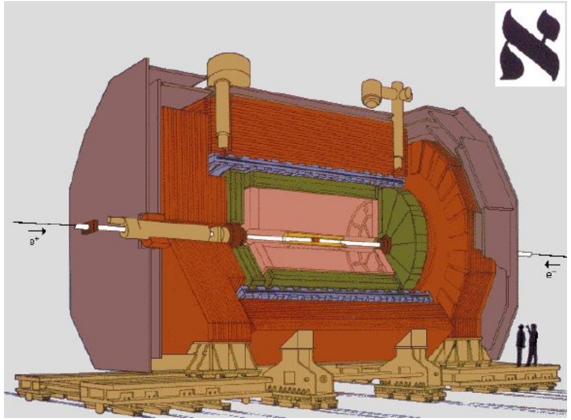
iris hep
Institute for Research & Innovation in Software for High Energy Physics

Introductions…. Who am I?

I am an experimental particle physicist (Princeton Physics since 2001, but based in Geneva, Switzerland) focused on computational and data science problems in my field, along with the software/computing systems to solve them.

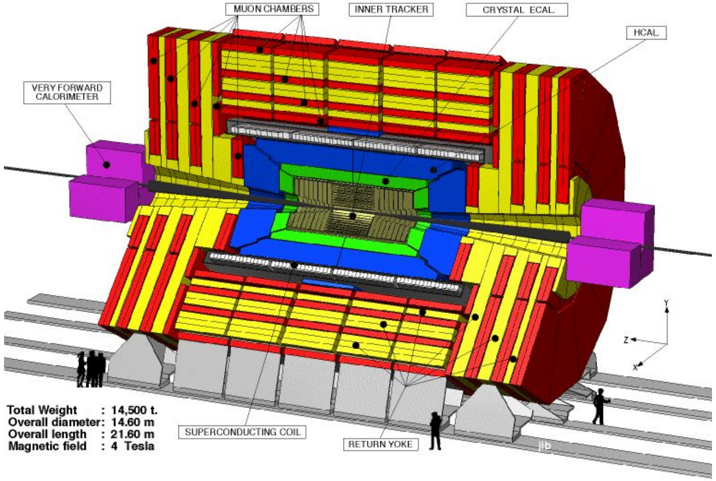
Researchers in experimental particle physics tend to introduce themselves to each other with reference to the series of experiments with which they have collaborated. So here is my own version of that:

Aleph@CERN



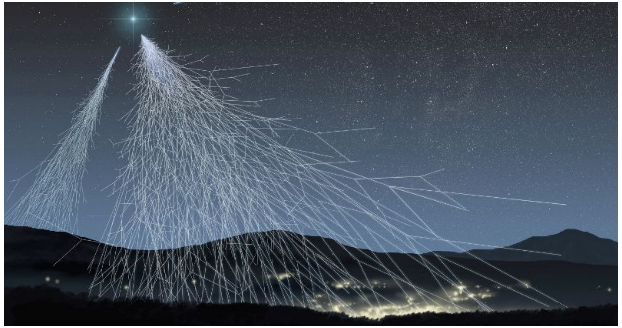
BaBar@SLAC

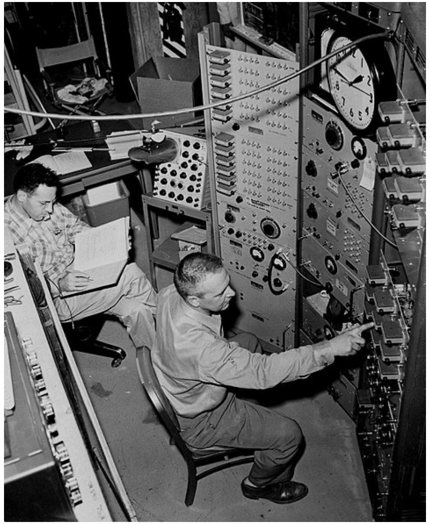
CMS@CERN

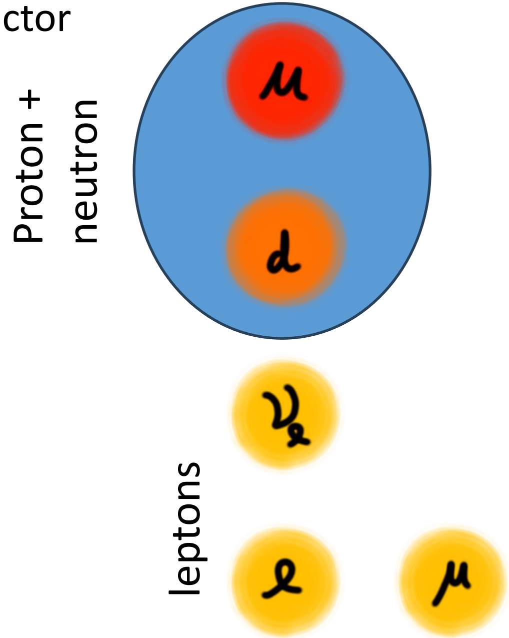


Fundamental Particles of Matter - “Particle Zoo”

1937: Discovery of the muon (Anderson and Neddermeyer)
    a copy of the electron but with 200 times the mass ($m_{\mu} = 200 \times m_e$)
    -and- the positron (Anderson) in cosmic rays
1947: Charged pion discovery and Kaon discovery, 1949: neutral pion discovery
1955: Discovery of the antiproton, at the Bevatron
1956: Discovery of the neutrino (Cowan, Reines, et al), using nuclear reactor
1960s: Quark Model and Deep Inelastic Scattering Experiments (SLAC)

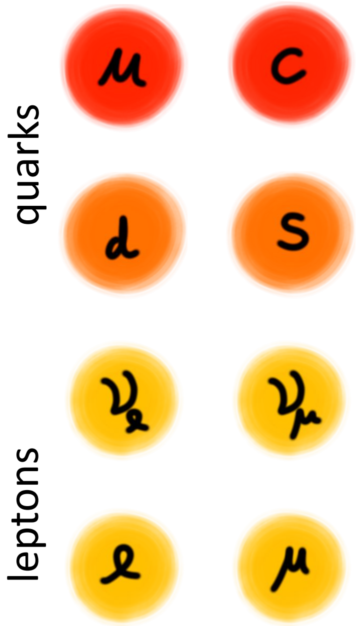






A first surprise - Muons “Who ordered that?”

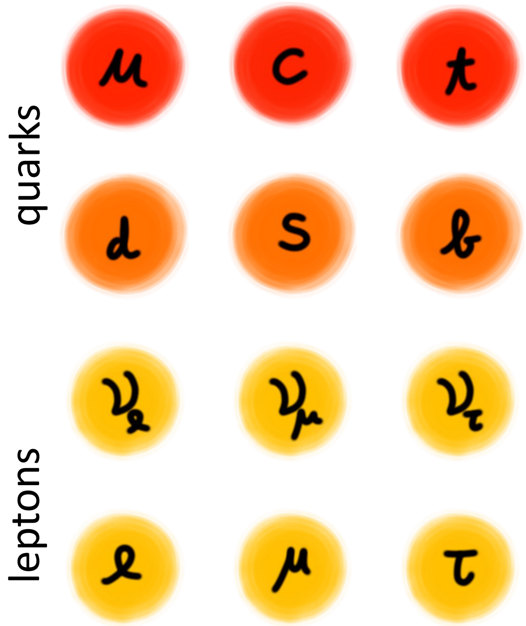
November "Charm" Revolution: Two families of fermions?



+ anti-particles
(i.e. copies of opposite charge)

Fast forward….

Three complete families of fermions

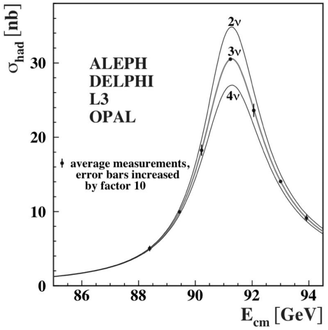


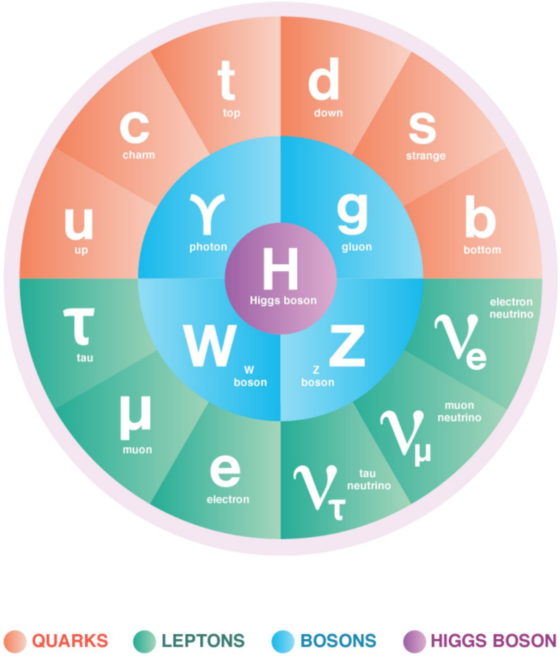
	Electrical Charge	Spin
	+2/3	1/2
	-1/3	1/2
	0	1/2
	-1	1/2

Intrinsic Angular Momentum

+ anti-particles

(i.e. copies of opposite charge)



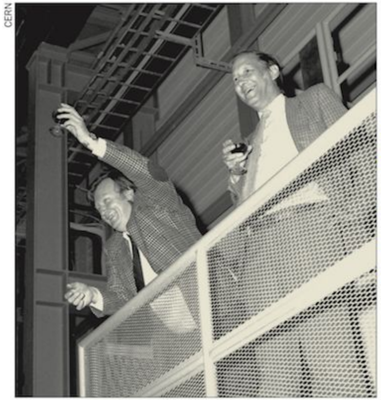


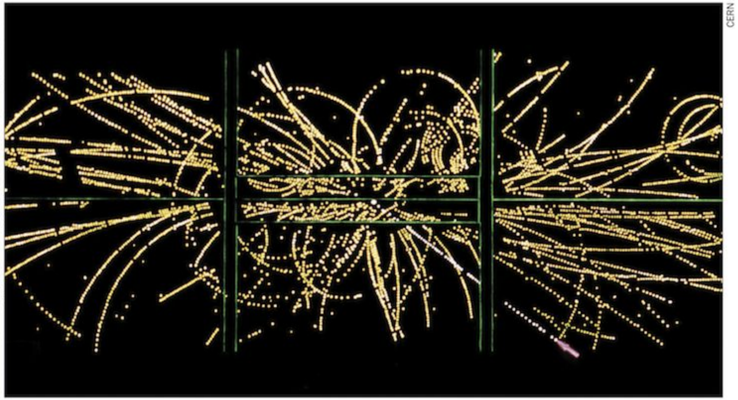
Higgs Particle Discovery Announcement July 4th, 2012

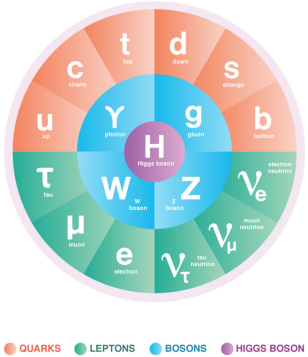
Something that looks like the Standard Model Higgs boson was discovered in 2012, Mass ~ 125 GeV

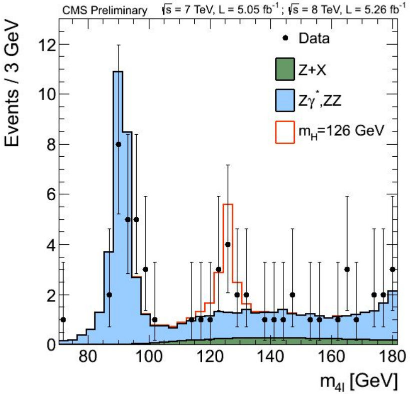
Nobel prize in physics in 2013

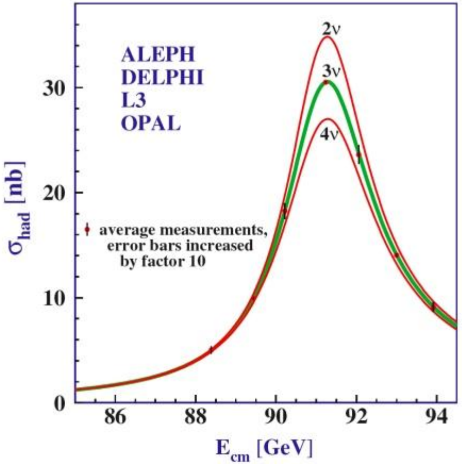
Experimental Development of the Standard Model

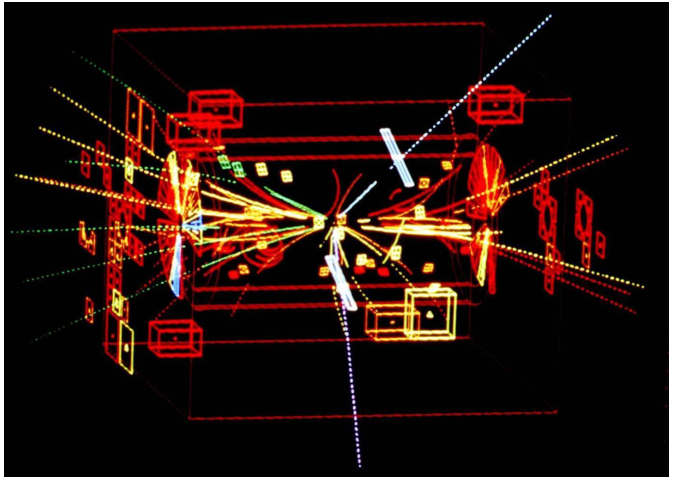


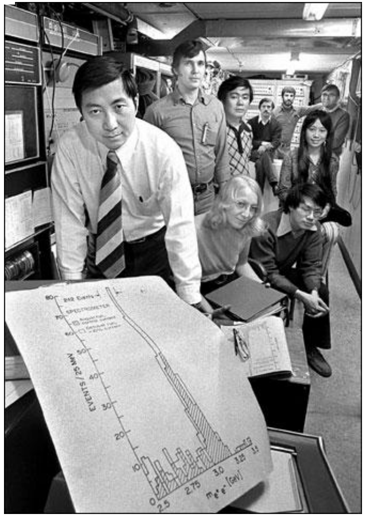












How did we get here?
(As experimentalists)

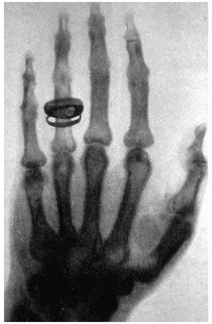
Tools ⟵⟶ Discoveries

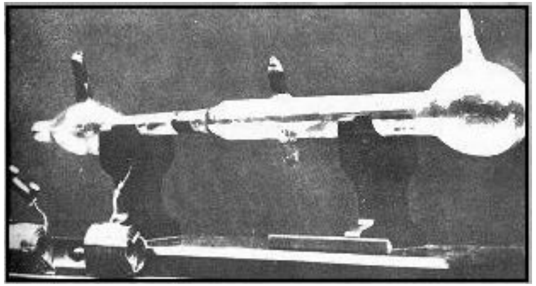
1895 - Roentgen discovers x-rays using cathode ray tubes

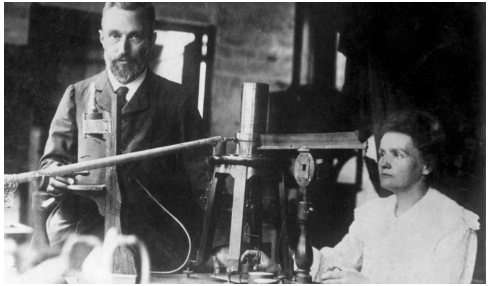
1896 - Bequerel accidentally discovers radioactivity in uranium, trying to explain x-rays

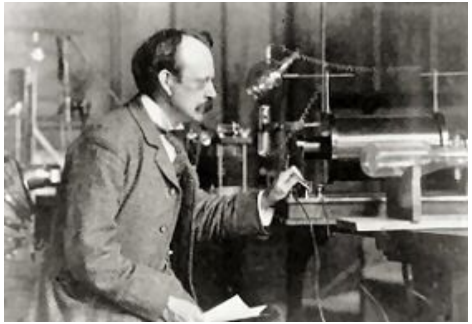
1898+ - Marie and Pierre Curie discover and explore radioactivity in other elements (polonium, thorium, radium)

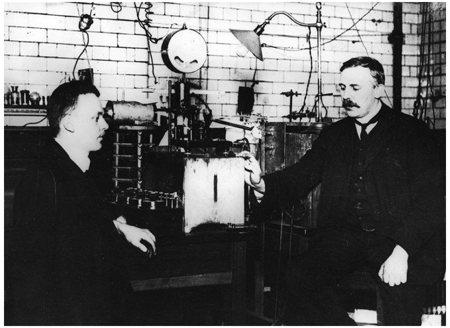
1911 - Rutherford, Geiger, Marsden use radioactivity to explore atomic structure and discover the nucleus









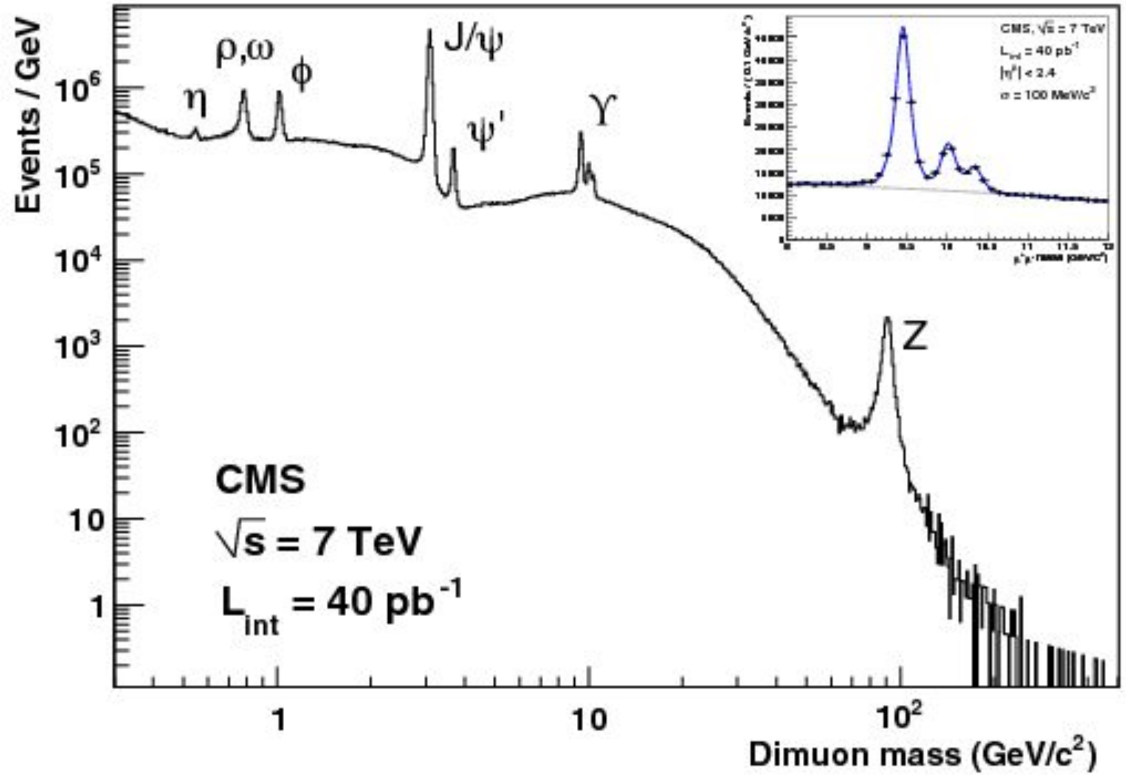


<u>Evolution in particle physics</u>

<u>Discovery of a particle</u> leads to Nobel prize and full exploration of the properties of the particle.

On the subsequent experiments the <u>particle is used for calibration or as a tool itself (e.g. in a beam)</u>.

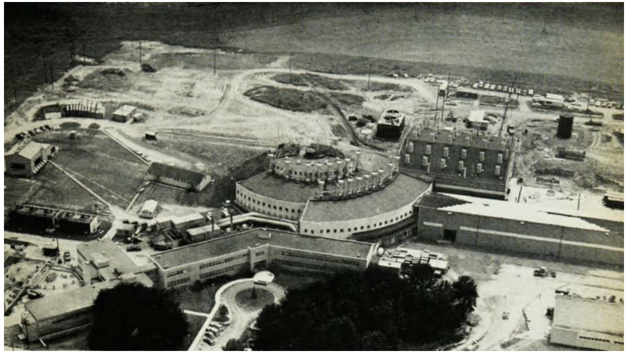
On later experiments the <u>particle becomes a background</u>.



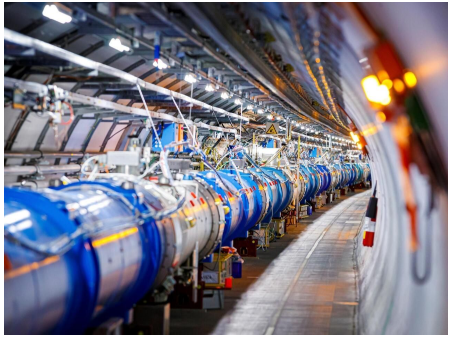
Big Science and Accelerators - larger and larger facilities

E.O. Lawrence and the cyclotron
at U.C. Berkeley

The birth of the "National Accelerator
Laboratory" in the US (now Fermilab)



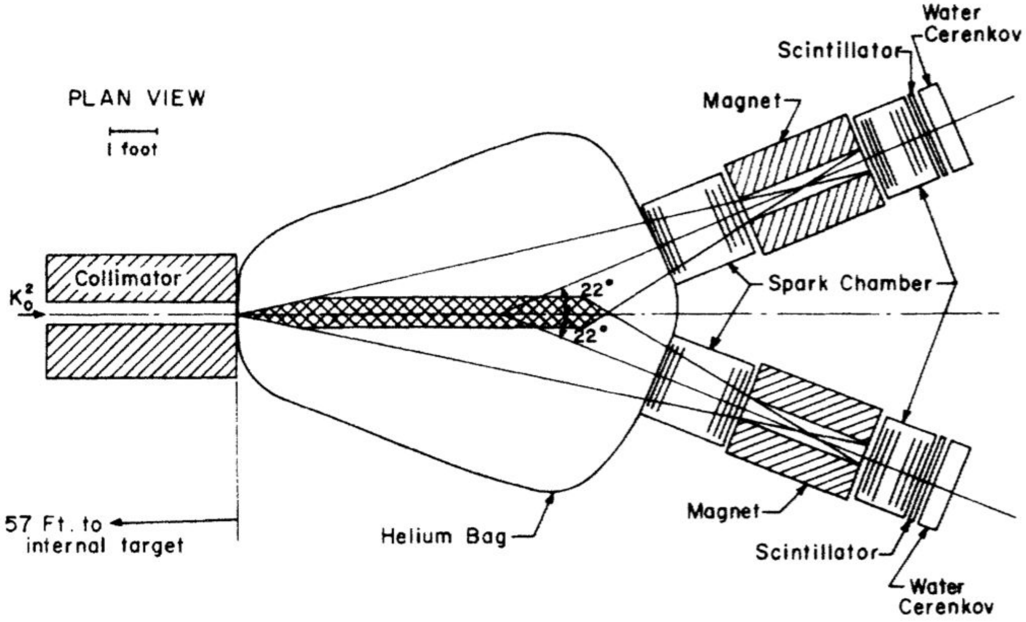
The Princeton-Pennsylvania Accelerator,
Milton White 1964 Physics Today 17(8): 27



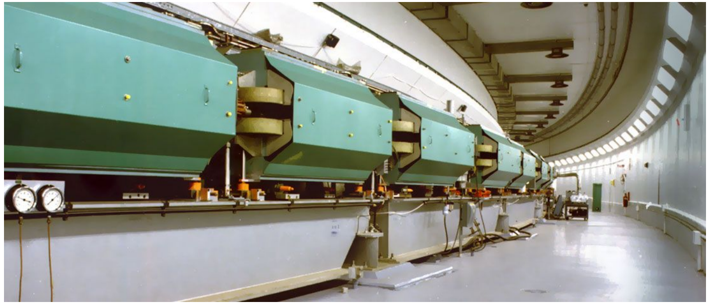
The Large Hadron Collider and
CERN first as a European laboratory
and then as a "world laboratory"

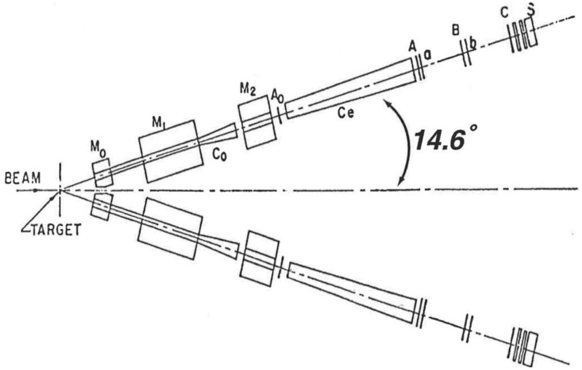
EVIDENCE FOR THE 2π DECAY OF THE K_2^0 MESON*†

J. H. Christenson, J. W. Cronin,‡ V. L. Fitch,‡ and R. Turlay§
Princeton University, Princeton, New Jersey
(Received 10 July 1964)

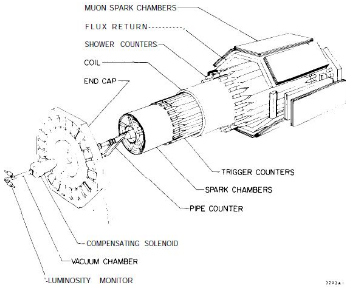


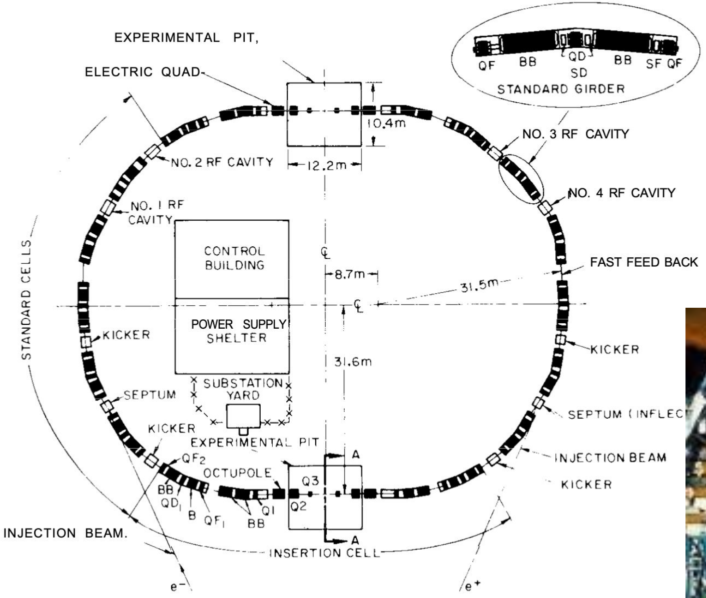
AGS at BNL delivering protons on a fixed target





Spear e+e- collider at SLAC and Mark 1 detector

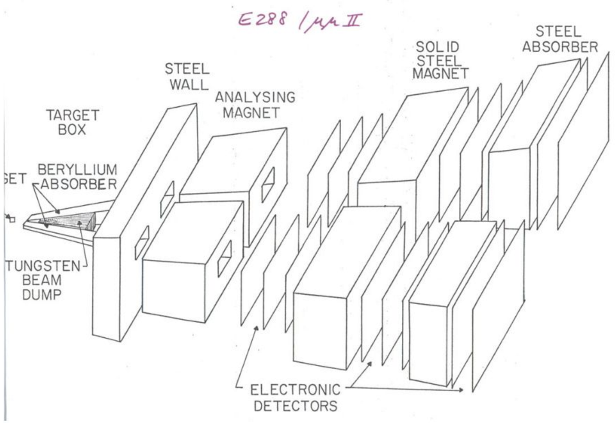




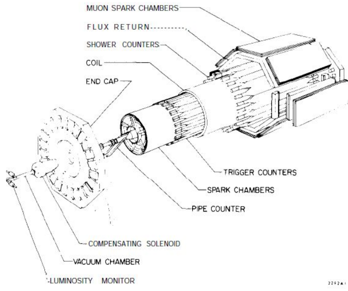
1. Schematic of the SPEAR storage ring.

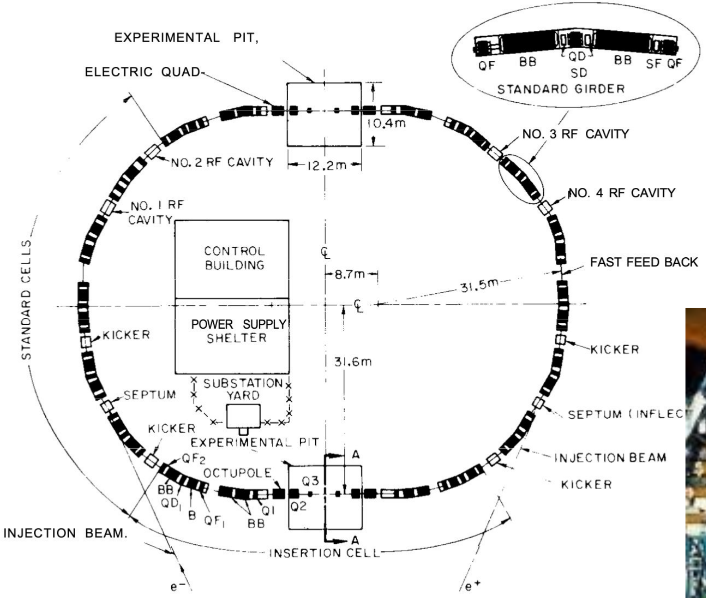
J and Psi discoveries in 1974

Protons at FNAL on a Uranium fixed target



Spear e+e- collider at SLAC and Mark 1 detector

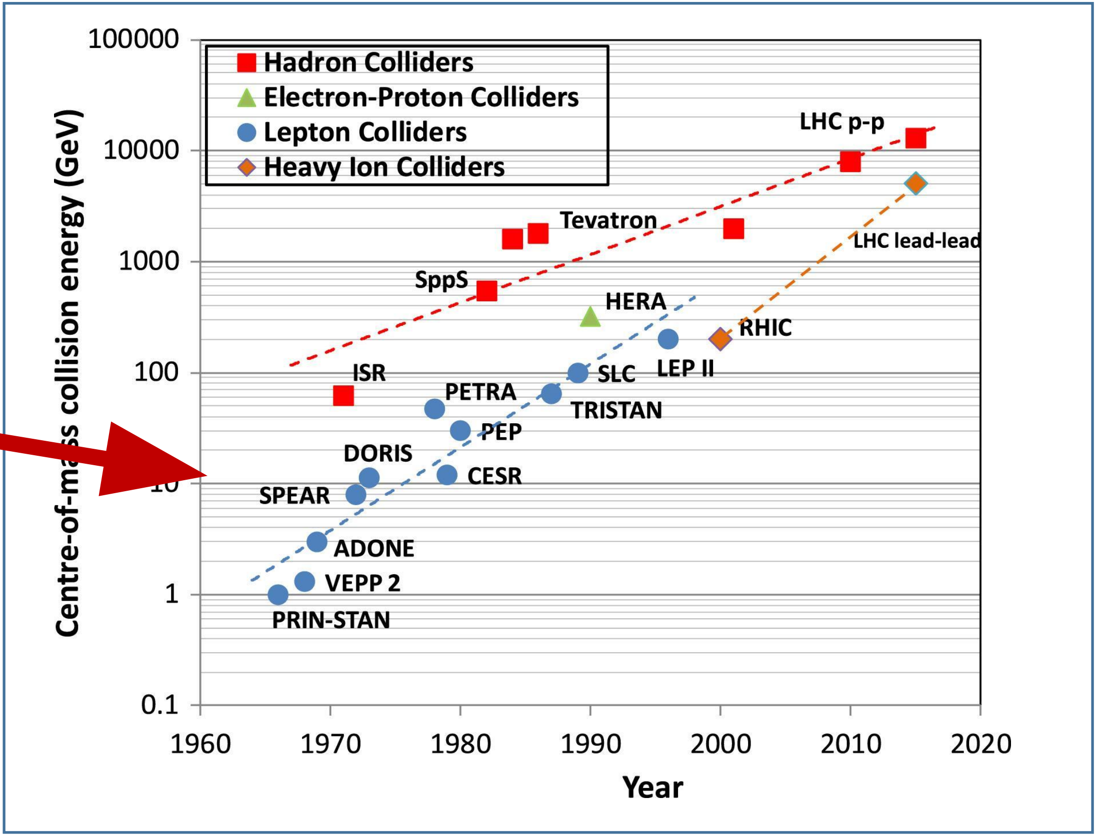




1. Schematic of the SPEAR storage ring.

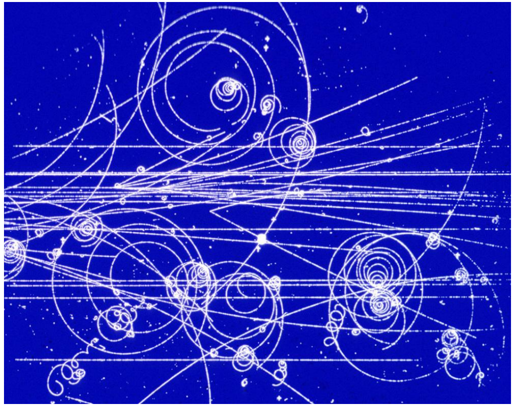
Bottom and Tau discoveries in 1977 and 1975

High Energy Physics is a facilities driven science



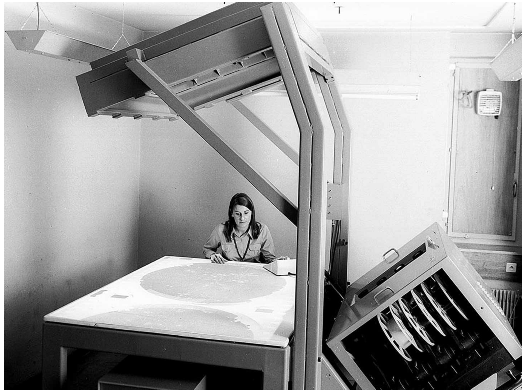
Last month:
https://indico.slac.stanford.edu/event/9040/

Instrumentation - Detectors (+ Electronics)

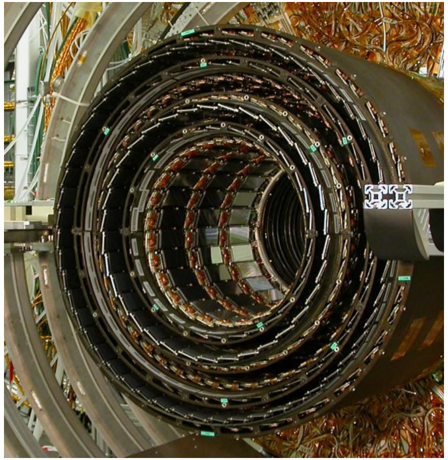


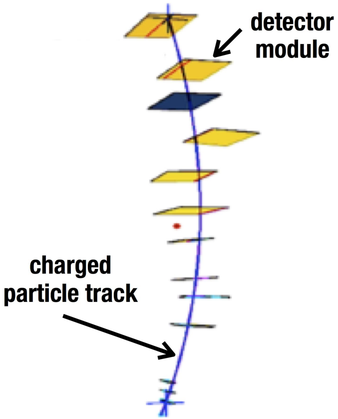
Bubble chamber photography

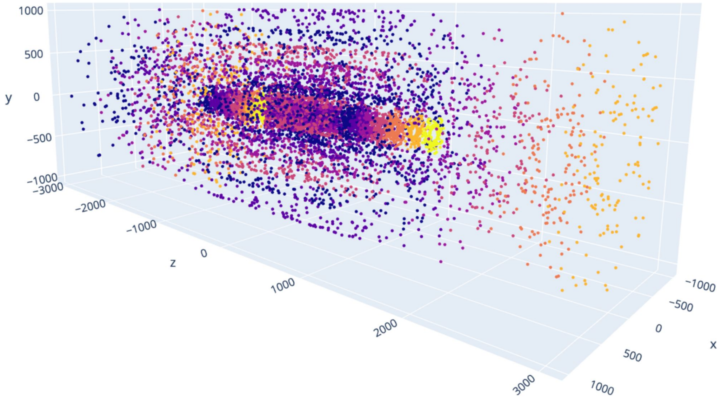
By-hand "scanning" of the photos



Modern detectors with (digital) electronic readout



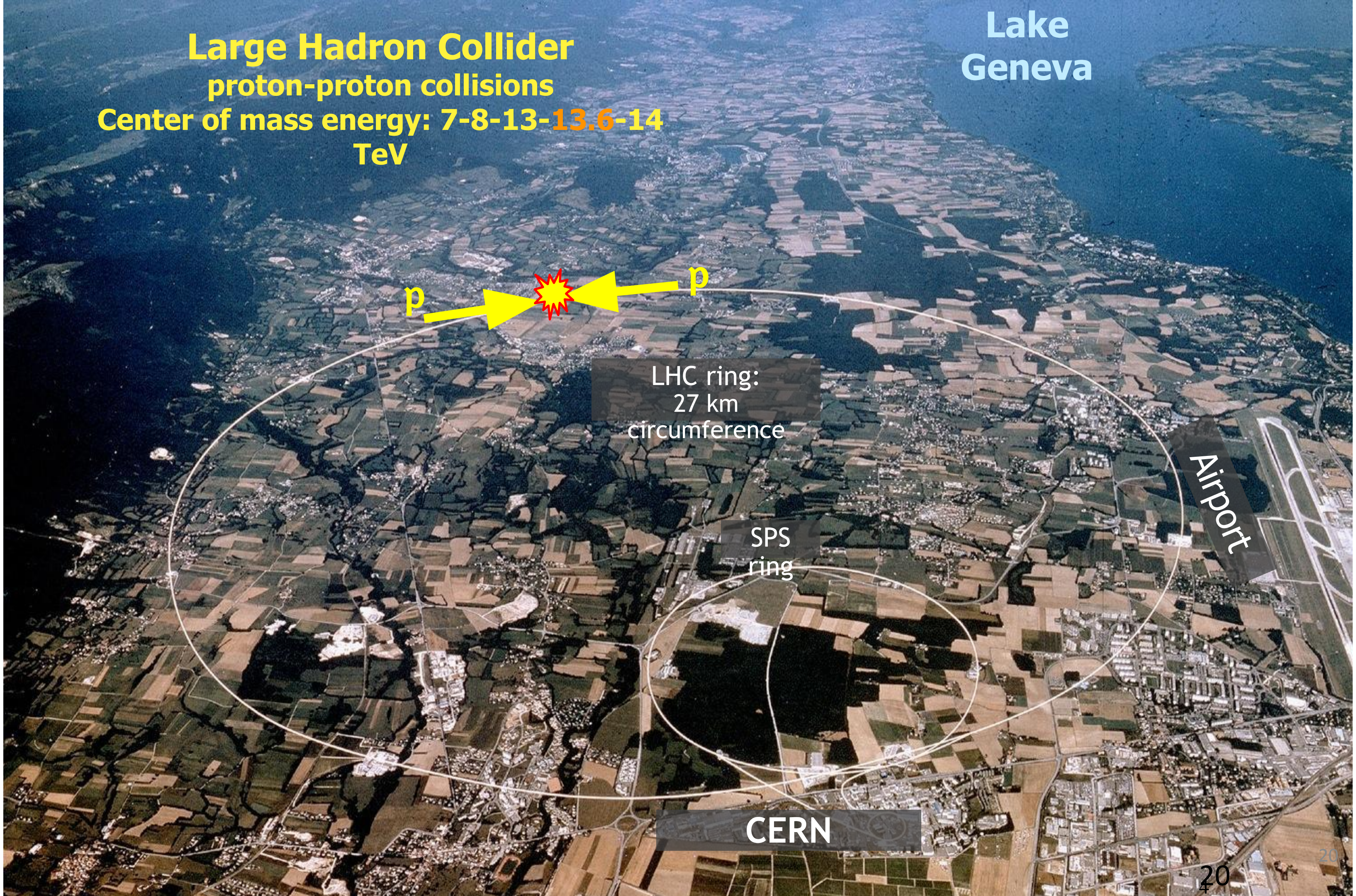




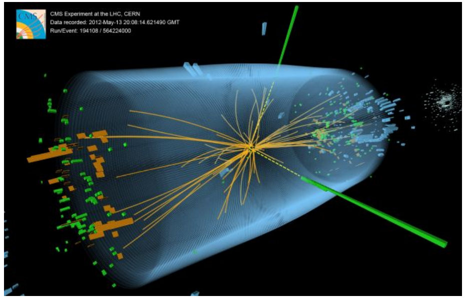
"New directions in science are launched by new tools much more often than by new concepts. The effect of a concept-driven revolution is to explain old things in new ways. The effect of a tool-driven revolution is to discover new things that have to be explained." - *Freeman Dyson*

Physics Nobel Prizes → Tools open doors to new research avenues

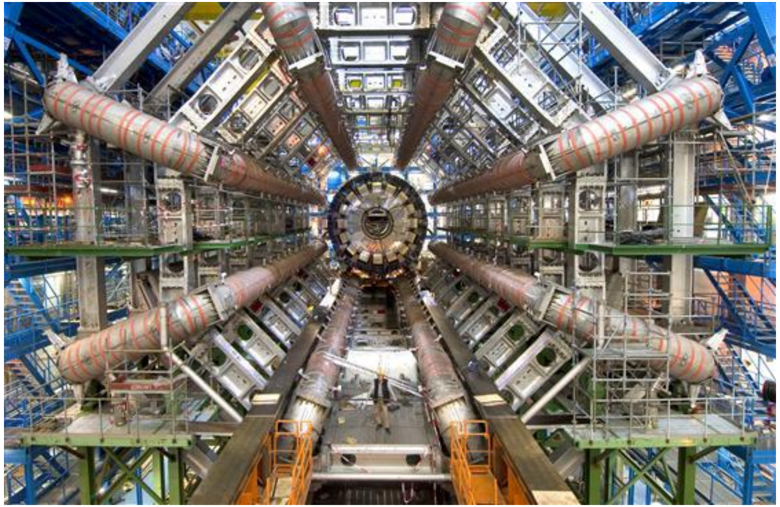
- **2024** (Hopfield, Hinton) - foundational discoveries and inventions that enable machine learning with artificial neural networks
- **2023** (Agostini, Krausz, L'Huillier) - development of experimental methods that generate attosecond pulses of light for the study of electron dynamics in matter
- **2018** (Ashkin) - invention of optical tweezers and their application to biological systems
- **2018** (Mourou, Strickland) - invention of a method of generating high-intensity ultra-short optical pulses
- **2014** (Isamu, Hiroshi, Nakamura) - invention of efficient blue light-emitting diodes, which has enabled bright and energy-saving white light sources
- **2012** (Haroche, Winelar) - development of methods that enable measuring and manipulation of individual quantum systems
- **2009** (Boyle, Smith) - invention of the CCD sensor, an imaging semiconductor circuit
- **2009** (Kao) - achievements concerning the transmission of light in fibres for optical communication
- **2005** (Glauber) - contributions to the field of optics
- **2005** (Hall, Hänsch) - contributions to the development of laser spectroscopy
- **2000** (Kilby) - development of the integrated circuit (microchip)
- **2000** (Alferov, Kroemer) - development of fast semiconductors for use in microelectronics
- **1997** (Chu, Cohen-Tannoudji, Phillips) - process of trapping atoms with laser cooling
- **1992** (Charpak) - invention of a detector that traces subatomic particles
- **1989** (Dehmelt, Paul) - development of methods to isolate atoms and subatomic particles for study
- **1989** (Ramsey) - development of the atomic clock

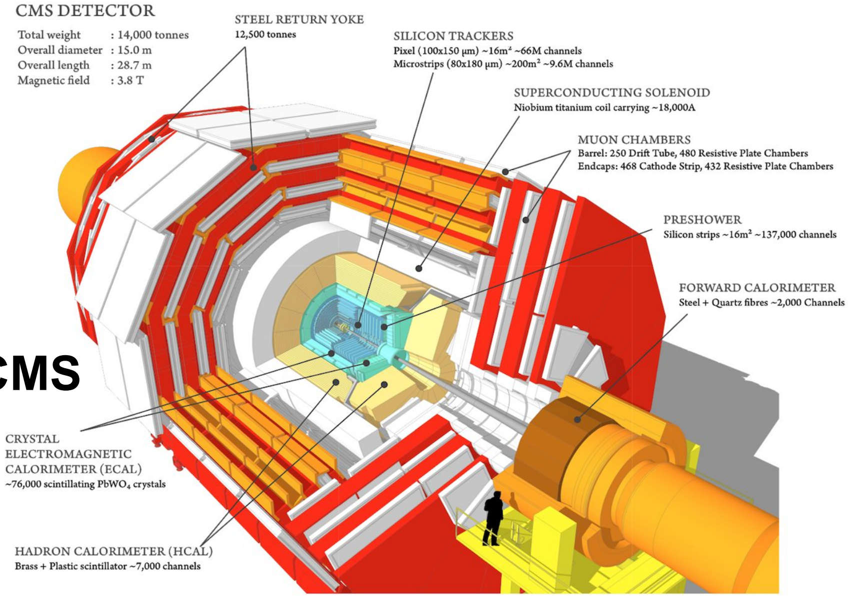

Large Hadron Collider
proton-proton collisions
Center of mass energy: 7-8-13-13.6-14 TeV
Lake Geneva
p
p
LHC ring:
27 km circumference
Airport
SPS ring
CERN

Large Hadron Collider Experiments Are Massive Data Generators



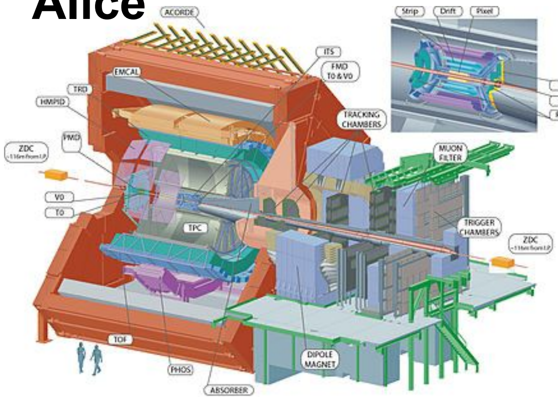
Atlas

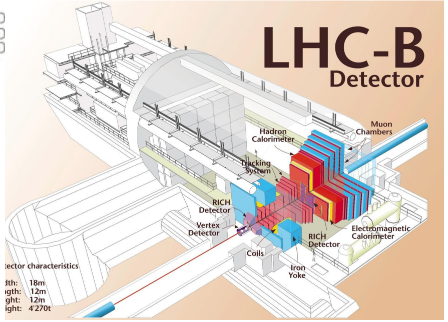




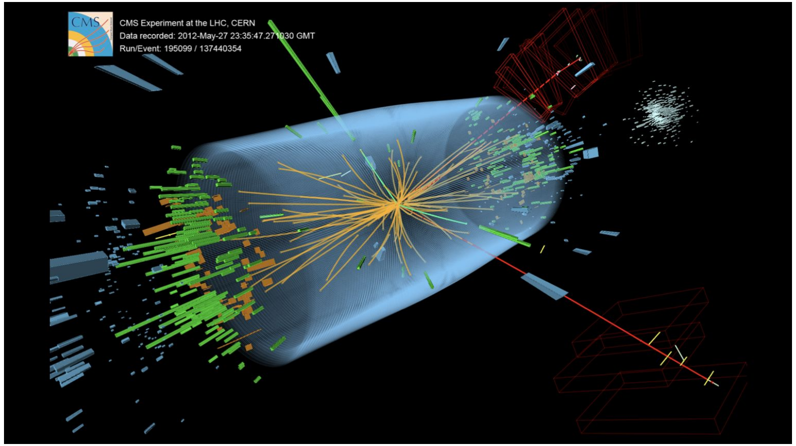
CMS

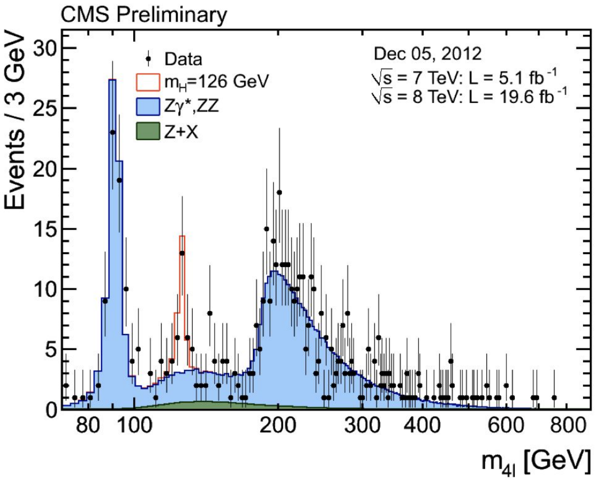
Alice



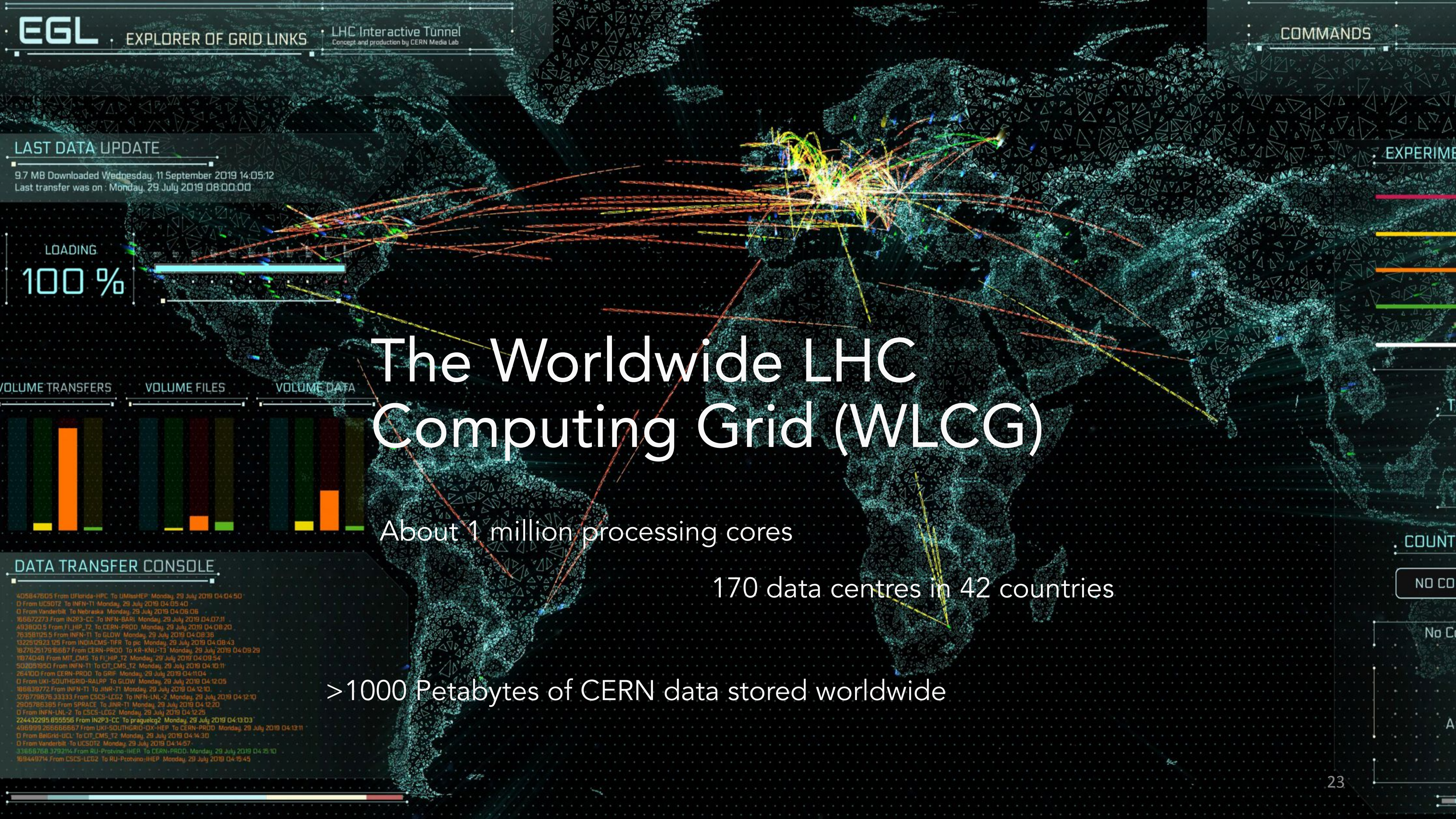


Just as with facilities, HEP scientists rely on large computing infrastructures to do their science





Large scale software
and computing
infrastructures



The Worldwide LHC Computing Grid (WLCG)

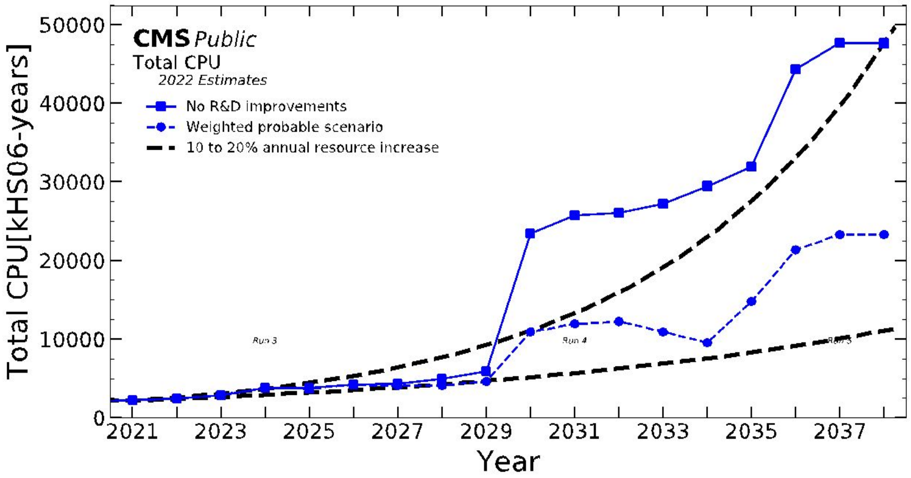
About 1 million processing cores

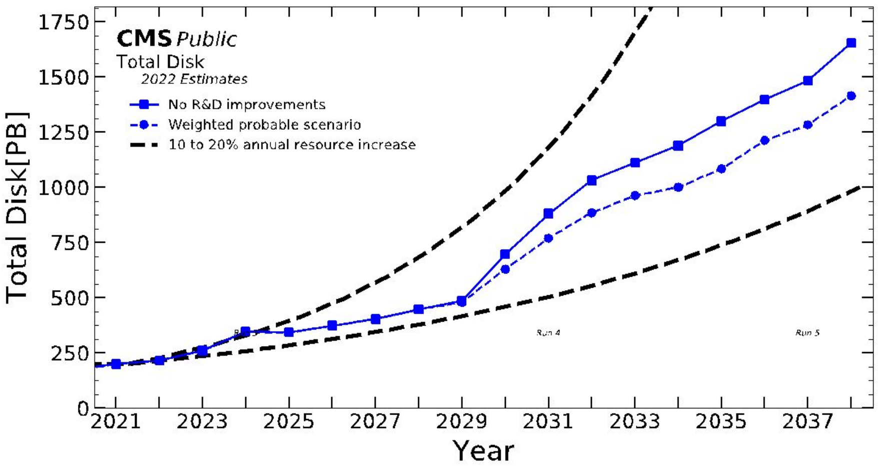
170 data centres in 42 countries

>1000 Petabytes of CERN data stored worldwide

“Cyberinfrastructure”

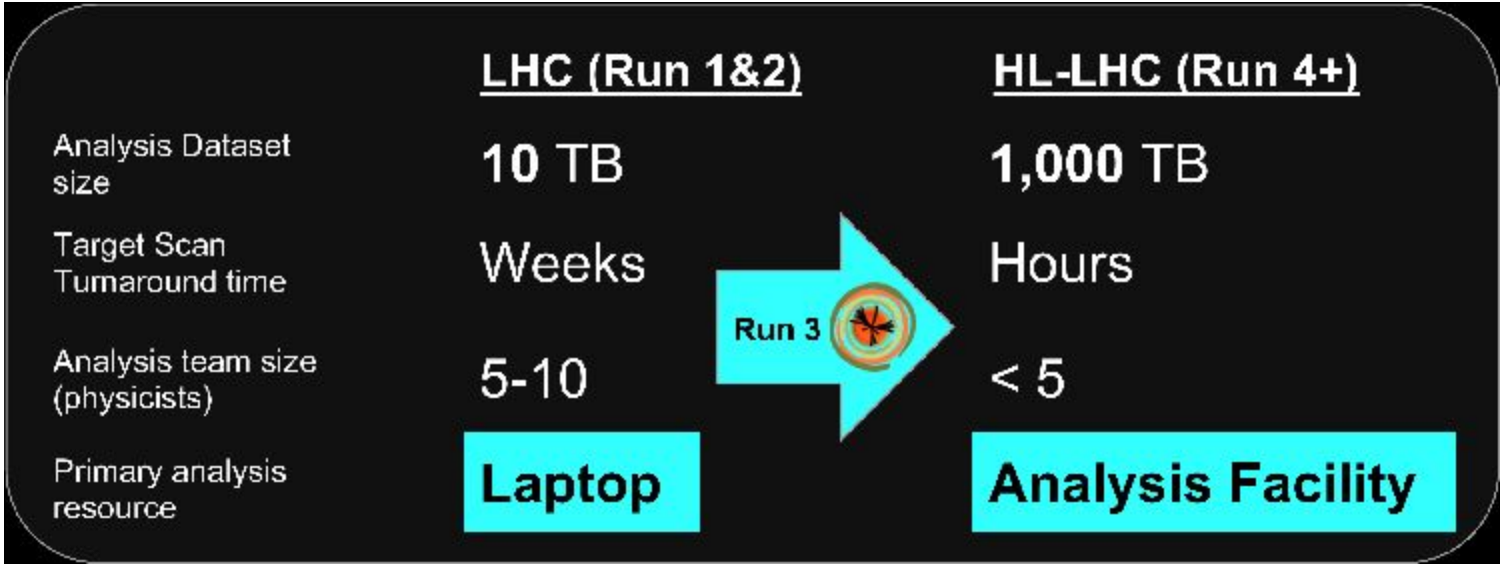
Future experiments pose even larger computing challenges





- A naive extrapolation from today's computing model and techniques, even after assuming Moore's Law increases in capabilities, is insufficient to meet the expected resource needs for HL-LHC
  - Technology evolution for processors and storage is an additional challenge
  - New ideas and methods are needed, and software is the key ingredient

Human time is critical: Optimizing analysis is about more than just about pure resources



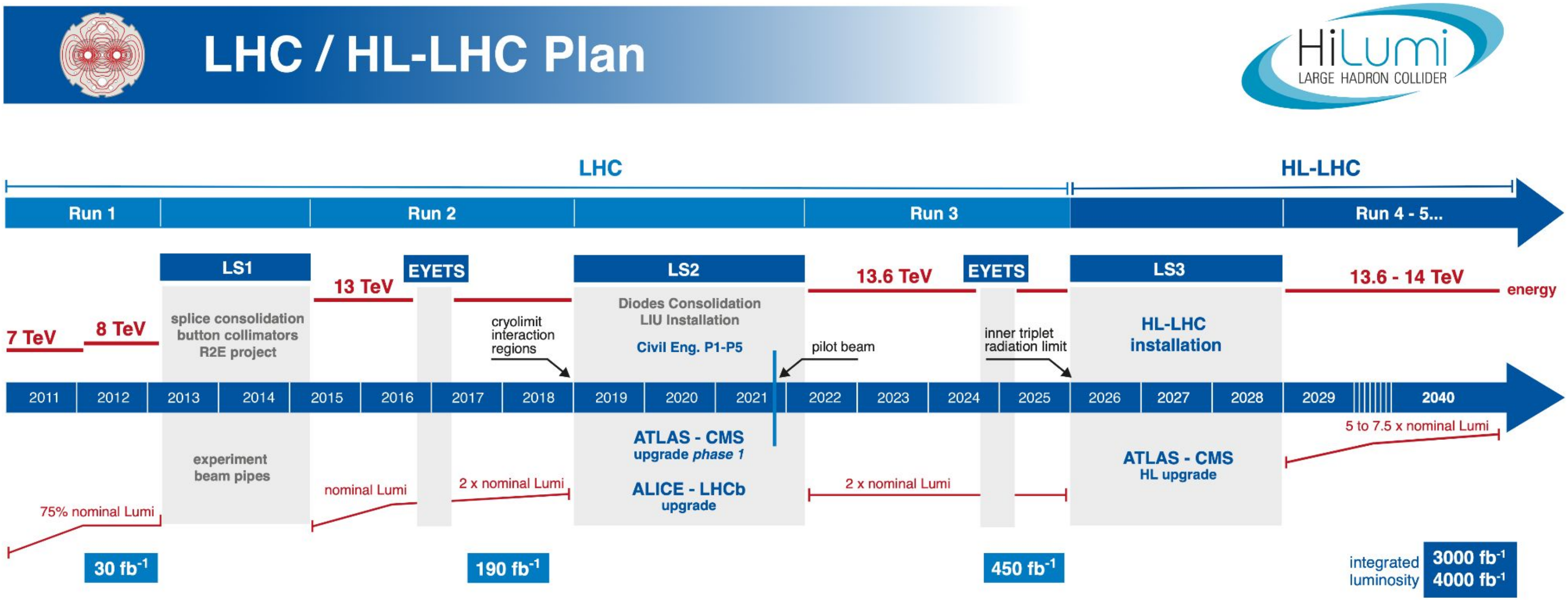
LHC analysis:

- Search & Precision Physics
- Simple ML techniques (BDT)
- Reproducibility in its infancy

HL-LHC analysis:

- Very High Precision Physics
- Modern ML (Deep Learning)
- Reproducible and Open Data

Experimental timescales span decades



Experiment designs start far before data taking. CMS was formed in 1992 (more than 30 years ago!), expects to run through 2040 and do data analysis for years after that

Experimental Observation of a Heavy Particle J†

J. J. Aubert, U. Becker, P. J. Biggs, J. Burger, M. Chen, G. Everhart, P. Goldhagen, J. Leong, T. McCorriston, T. G. Rhoades, M. Rohde, Samuel C. C. Ting, and Sau Lan Wu
Laboratory for Nuclear Science and Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

and

Y. Y. Lee
Brookhaven National Laboratory, Upton, New York 11973
(Received 12 November 1974)

We report the observation of a heavy particle J, with mass $m = 3.1$ GeV and width approximately zero. The observation was made from the reaction $p + \mathrm{Be} \rightarrow e^+ + e^- + x$ by measuring the e^+e^- mass spectrum with a precise pair spectrometer at the Brookhaven National Laboratory's 30-GeV alternating-gradient synchrotron.

Discovery of a Narrow Resonance in e^+e^- Annihilation*

J.-E. Augustin,† A. M. Boyarski, M. Breidenbach, F. Bulos, J. T. Dakin, G. J. Feldman, G. E. Fischer, D. Fryberger, G. Hanson, B. Jean-Marie,† R. R. Larsen, V. Lüth, H. L. Lynch, D. Lyon, C. C. Morehouse, J. M. Paterson, M. L. Perl, B. Richter, P. Rapidis, R. F. Schwitters, W. M. Tanenbaum, and F. Vannucci‡

Stanford Linear Accelerator Center, Stanford University, Stanford, California 94305

and

G. S. Abrams, D. Briggs, W. Chinowsky, C. E. Friedberg, G. Goldhaber, R. J. Hollebeek, J. A. Kadyk, B. Lulu, F. Pierre,§ G. H. Trilling, J. S. Whitaker, J. Wiss, and J. E. Zipse

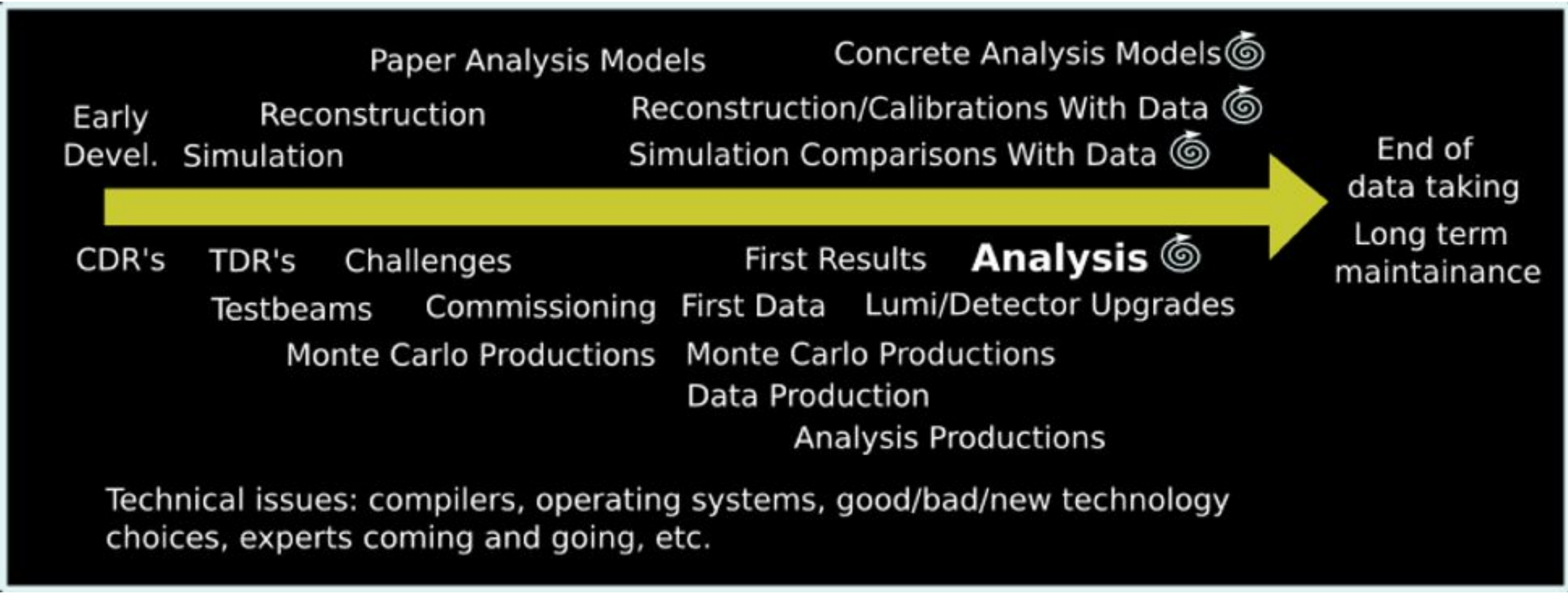
Lawrence Berkeley Laboratory and Department of Physics, University of California, Berkeley, California 94720

(Received 13 November 1974)

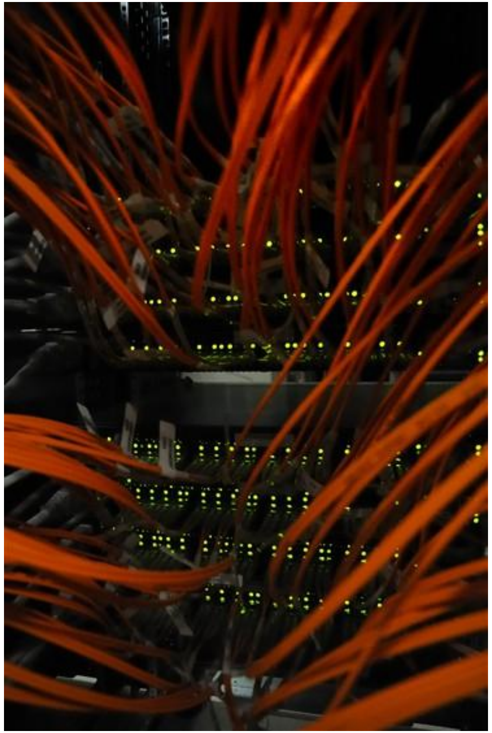
We have observed a very sharp peak in the cross section for $e^+e^- \rightarrow$ hadrons, e^+e^-, and possibly $\mu^+\mu^-$ at a center-of-mass energy of 3.105 ± 0.003 GeV. The upper limit to the full width at half-maximum is 1.3 MeV.

CMS has ~4300 Scientists, Engineers and technicians
(including 800 PhD students) from 41 Countries and 179 institutes

HEP software lifecycle



Cyberinfrastructure?

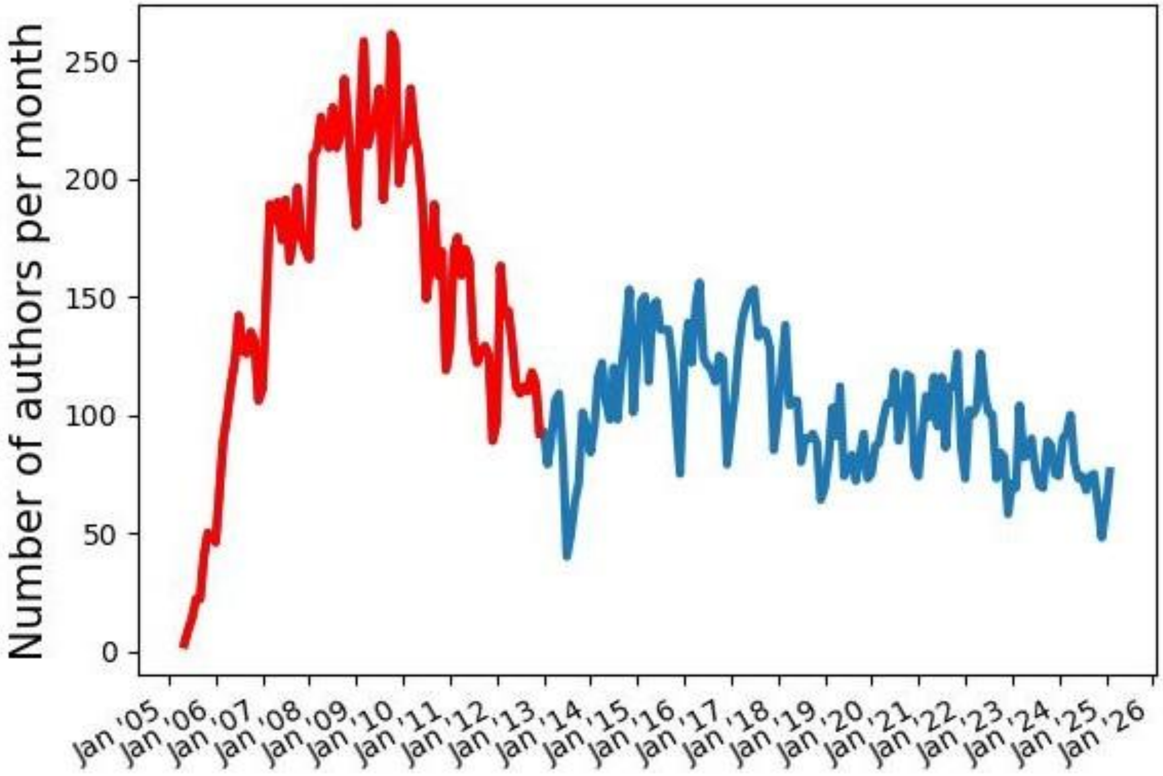


Conceptual motivations behind the HEP Software Foundation

Computer hardware is a consumable. Software is the actual "cyberinfrastructure" in the long run.

More importantly software is also an *intellectual product* of our research, not just a tool.

Large scale collaborative software development in HEP



Many developers, typically a handful of true experts

Millions of lines of code for CMS - And this excludes most data analysis code, event generators, detector simulation codes, and others…

http://iris-hep.org

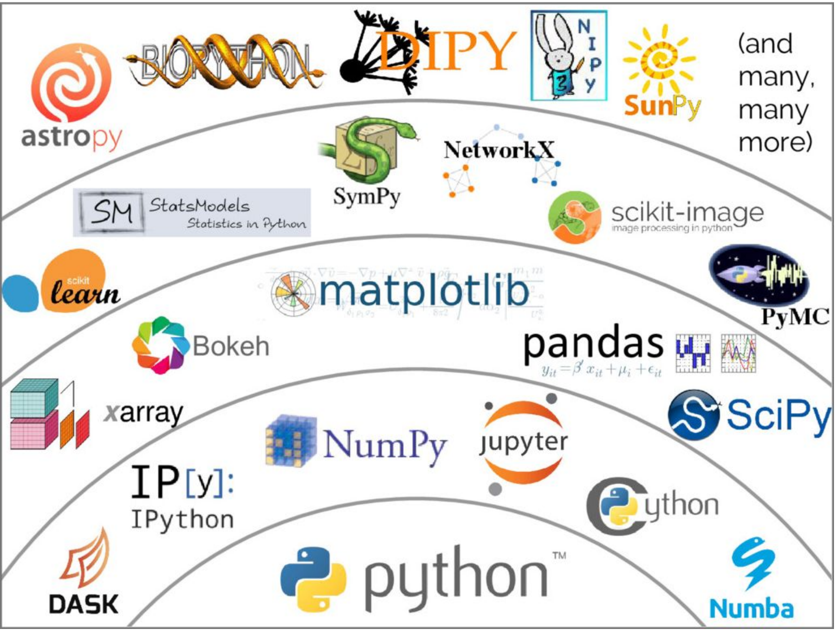
Conceived as a **"software upgrade"** project and guided initially by the "Community White Paper" roadmap developed in 2016-2017: it involves 21 universities, spanning ATLAS, CMS and LHCb.

IRIS-HEP is supported by the U.S. National Science Foundation through the **Office of Advanced CyberInfrastructure** in the Directorate for Computer and Information Science and Engineering and the **Division of Physics** in the Directorate for Mathematical and Physical Sciences.

10-year project: Originally funded in 2018 as OAC-1836650 and renewed in 2023 through 2028 as PHY-2323298.

Leveraging data science for HEP analysis

Scientific Python / PyData vision/ecosystem



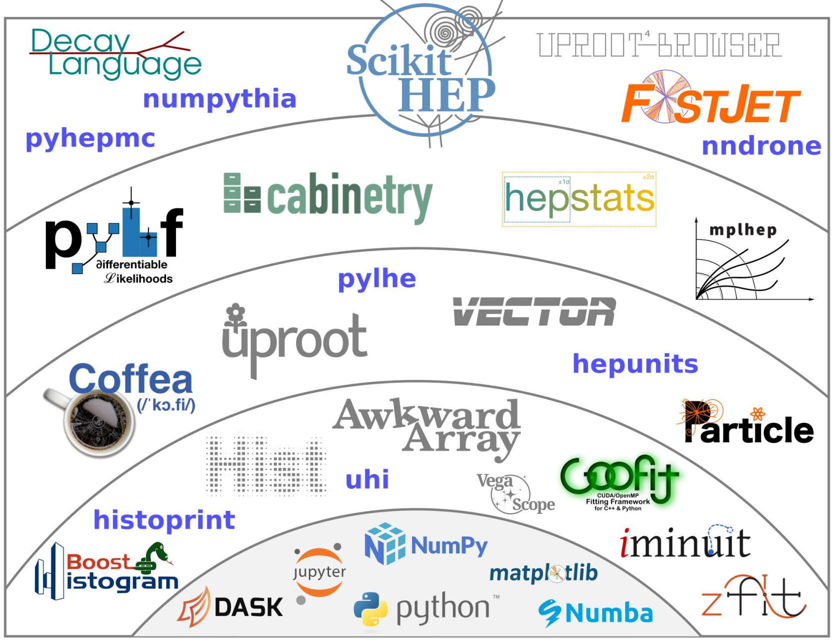
Application Specific

Domain Specific

Technique specific

Foundational

Developing HEP data analysis ecosystem



PyHEP development

HSF | Activities | Meetings | Communication | Projects & Support | About

PyHEP - Python in HEP

The PyHEP working group brings together a community of developers and users of Python in Particle Physics, with the aim of improving the sharing of knowledge and expertise. It embraces the broad community, from HEP to the Astroparticle and Intensity Frontier communities.

Conveners

- Eduardo Rodrigues (LHCb, University of Liverpool)
- Jim Pivarski (CMS and IRIS-HEP, Princeton)
- Matthew Feickert (ATLAS and IRIS-HEP, University of Wisconsin-Madison)
- Nikolai Hartmann (Belle II, LMU Munich)

All coordinators can be reached at hsf-pyhep-organisation@googlegroups.com.

Getting Involved

Everyone is welcome to join the community and participate, contribute, to the organised meetings and by means of the following communication channels:

- Gitter channel PyHEP for any informal exchanges.
- GitHub repository of resources, e.g., Python libraries of interest to Particle Physics.
- PyHEP Workshop Twitter handle: @PyHEPConf

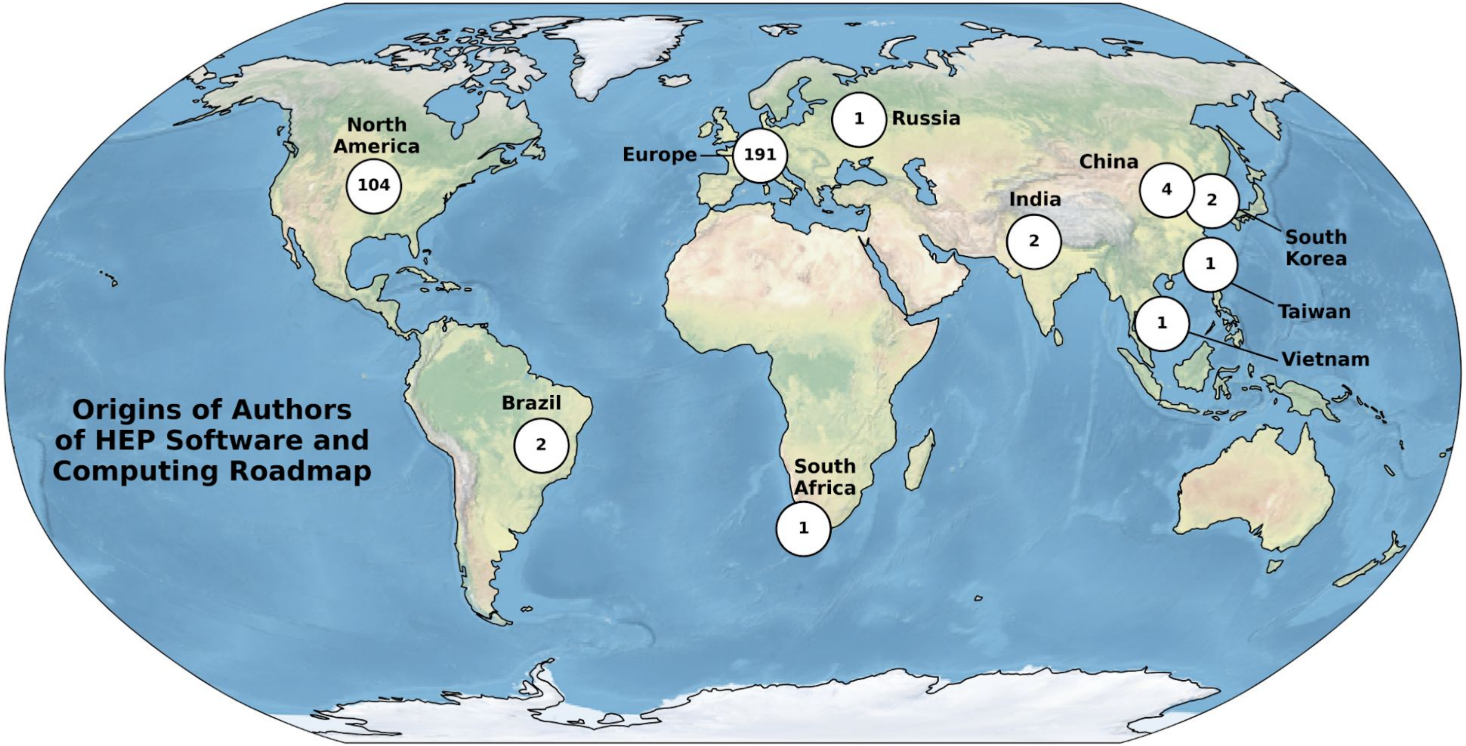
Extra Gitter channels have been created by and for the benefit of the community:

- PyHEP-newcomers for newcomers support (very low entry threshold).

The HSF-India Project

http://research-software-collaborations.org/

However, nearly all authors of the HSF Community Roadmap were from institutions in Europe and the US



Team – We wrote a proposal to the US National Science Foundation aiming to broaden this community to include Asia – Starting with India




Princeton University: Peter Elmer, David Lange (PI)
University of Massachusetts, Amherst: Rafael Coelho Lopes de Sa, Verena Martinez Outschoorn

Facilitating international research software: The “HSF-India” project

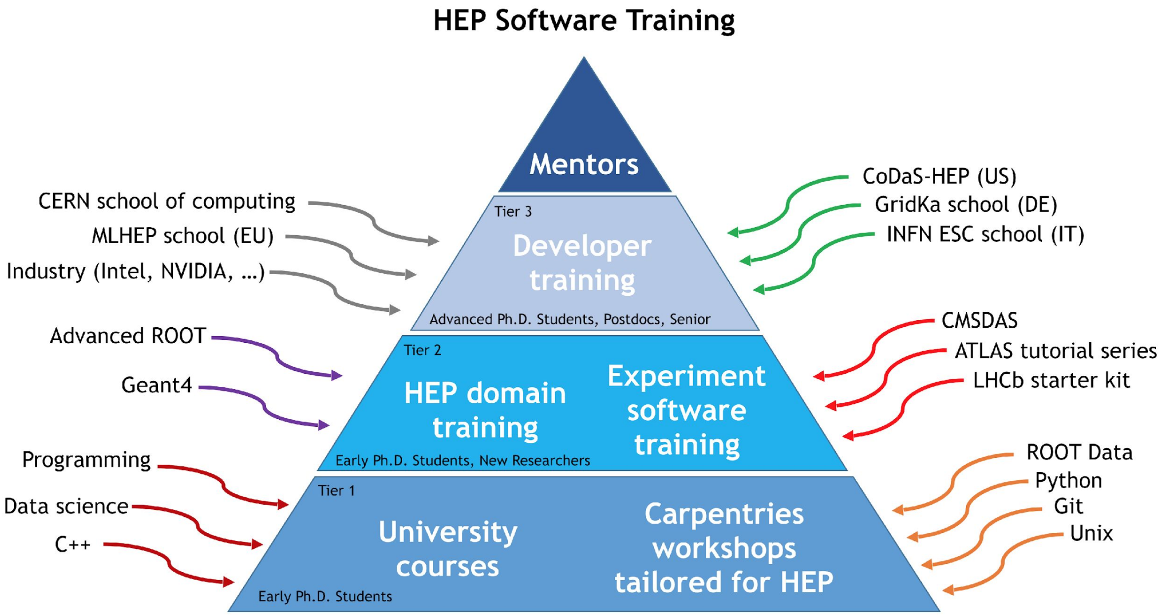
- Given the growing complexity of our scientific data and collaborations, building and fostering collaborations are increasingly important to raise the collective productivity of our research community.
- HSF-India project aims to build international research software collaborations between US, European, and India based researchers to reach the science goals of experimental particle, nuclear and astroparticle research.

Intended as a long-term investment in international team science with a broad research scope

Rather than directly fund a specific research activities, much of our funding is to facilitate research collaborations

- Training in research software skills
- Bidirectional research exchanges
- Student programs

Bootstrap collaboration through software training



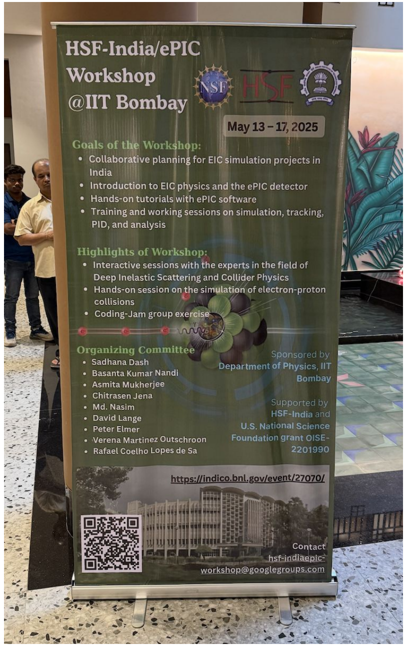
- A vision for training in HEP: researchers progress (vertically) from basic skills training, through user training in existing software to training in skills needed to develop new research software.

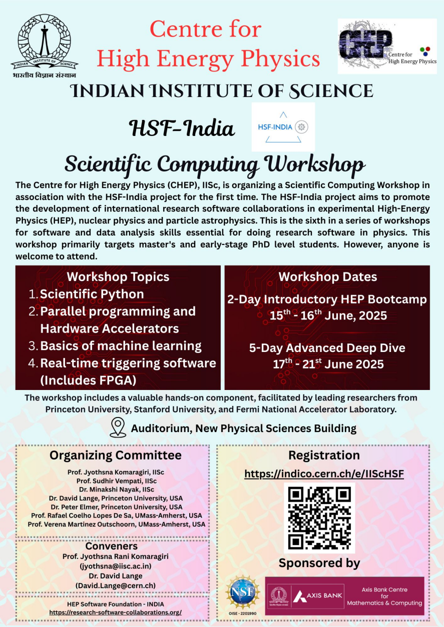
We have run software workshops in Mumbai (TIFR), Bhubaneswar (NISER), Delhi (University of Delhi), Kolkata (VECC) and U.Hyderabad

- Regionally organized, primarily targeting MS/PhD level students.
- Mix of lectures and hands-on exercises
- Mix of local and US instructors
- Jupyter notebook based materials derived from/patterned after HSF training courses

We want to organize these events regionally to make it easier for interested students to attend

Recent and upcoming events





HSF-India Software Workshop
U.Hyderabad - 13-17 Jan 2025
https://indico.cern.ch/event/1394564/

EIC Detector and Physics Simulation Workshop and Tutorial at IIT Bombay
13 -17 May 2025

HSF-India workshop in IISC Bangalore
16-22 June 2025
https://indico.cern.ch/event/1519117/

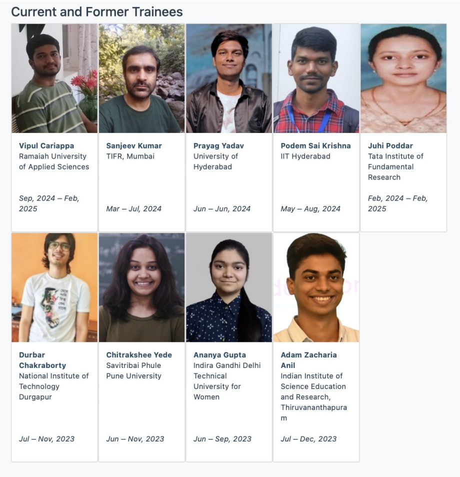
Also: CMSDAS IIT Hyderabad - 23-27 June 2025 (perhaps a collaboration with HSF-India)

We are exploring additional possibilities for “hackathon” events to further build skills, as well as Physics/CS collaborations.

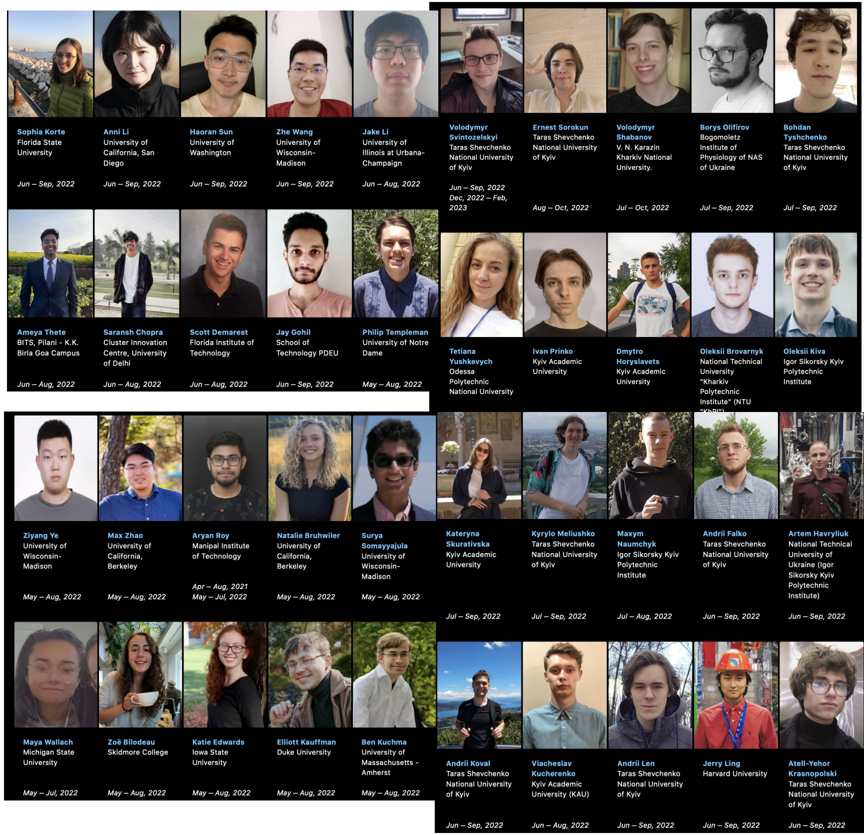
3-6 month project Fellows Program

- **Project focused** aiming to bring students into contact with “mentors” to work on a specific, pre-defined project, allowing them to grow their software skills and experience working in large projects
- These short term projects that build **longer term collaborations** in research software and foster **scientific career progression**
- Our program is open for applications for either full time (eg, during semester breaks) or part time expressions of interest

[HSF-India Research Software Trainees](#)



Also IRIS-HEP Fellows: https://iris-hep.org/fellows.html



Bidirectional Research Exchange Program

We also have funding for “research exchanges” that support travel costs for 1-3 months. These are meant for very senior PhD students and more senior researchers that have already

Who can we support?

- Researchers affiliated to a US university/lab exchange based in India
- Researchers affiliated to an university/lab in India exchange based in US or CERN to work with a US affiliated group

We are looking for either project/host ideas or those interested in doing an exchange. If you have ideas for projects that interest you, we can help identify matches with US researchers

Conclusion and Opportunities

- Experimental science is enabled by tools innovation, which become facilities enabling further science. Our science has been driven by accelerator and detector innovation.

- As facilities get larger, scientific collaborations are created to build and exploit those facilities. These large international collaborations of scientists allow experimental endeavors to exploit regional strengths.

- Just as with other aspects of our science, software teams that are inherently international are most likely to develop performant, highly usable, and sustainable **research software ecosystems**, which are a new kind of “facility” and an *intellectual product* of our research.

- HSF-India is a project which aims to further catalyze global collaboration in research software in Physics.
    - http://research-software-collaborations.org

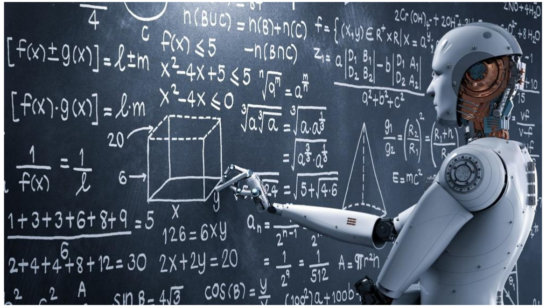
And the Future for HEP?

WLCG
Worldwide LHC Computing Grid

The **Worldwide LHC Computing Grid (WLCG)** project is a global collaboration of around 170 computing centres in more than 40 countries, linking up national and international grid infrastructures. The mission of the WLCG project is to provide global computing resources ... **[2000's and 2010's era]**

HSF
HEP Software Foundation

The **HEP Software Foundation** facilitates cooperation and common efforts in High Energy Physics software and computing internationally. **[2010's and 2020's era]**



What collaborative research efforts will the rest of the 2020's and the 2030's produce for the future "facilities" for the HEP community?