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

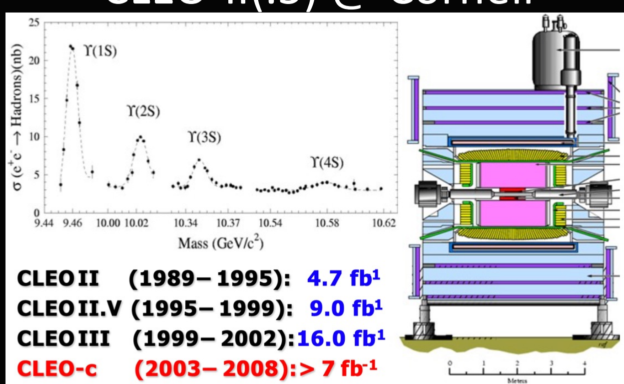
David Lange

Princeton University

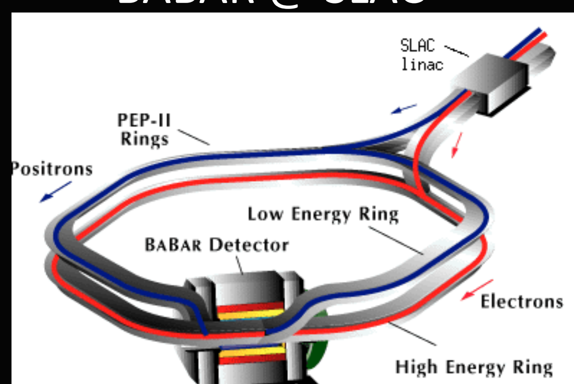
My career in high-energy physics

- 1999 PhD in experimental particle physics (University of California, Santa Barbara
- Postdoc and staff member at Lawrence Livermore Lab (1999-2015)
- Research staff at Princeton University (2016-present), currently residing at CERN (Geneva, Switzerland)

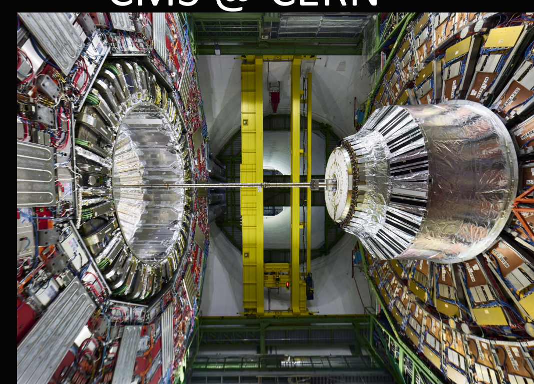
CLEO-II(.5) @ Cornell



BABAR @ SLAC

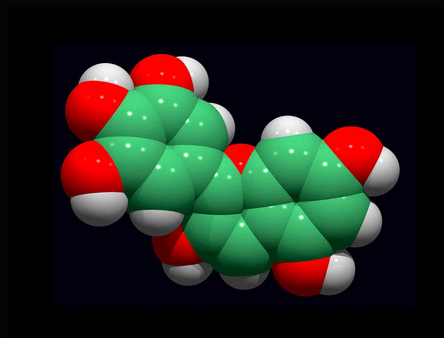


CMS @ CERN



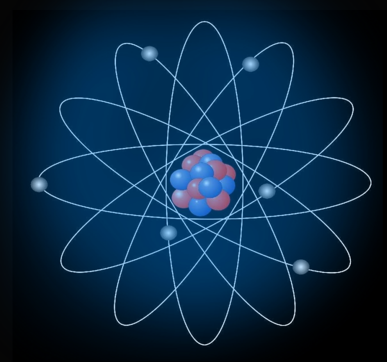
The Scales of Particle Physics

Molecule
10^{-9} m = 0.000 000 001 m



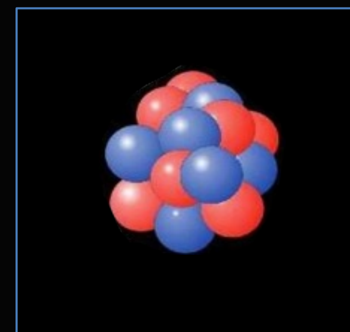
Delphinidin Molecule
(blue pigment of flowers and grapes)

Atoms
10^{-10} m = 0.000 000 000 1 m



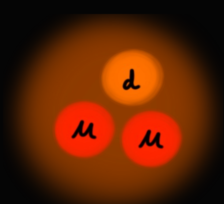
Composed of:
Nucleus and electrons

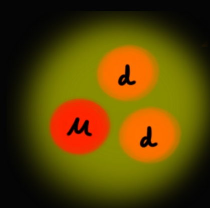
Nucleus
10^{-14} m = 0.000 000 000 000 01 m



Composed of:
Protons and neutrons

Protons and Neutrons
10^{-15} m = 0.000 000 000 000 001 m



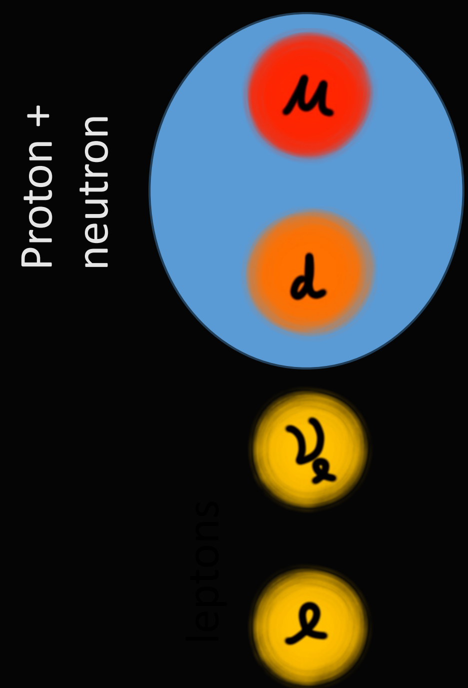


Composed of quarks

Quarks
$<10^{-18}$ m = 0.000 000 000 000 000 001 m

Quarks and electrons have
no dimensions
they look just like a point

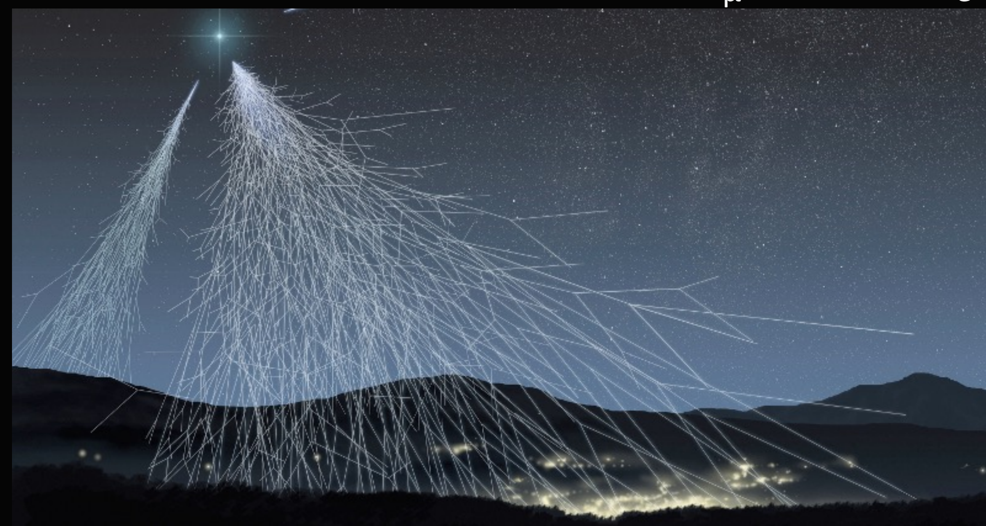
Fundamental Particles of Matter



Fundamental Particles of Matter

1937: Discovery of the muon (Anderson and Neddermeyer)

a copy of the electron but with 200 times the mass ($m_{\mu} = 200 \times m_e$)



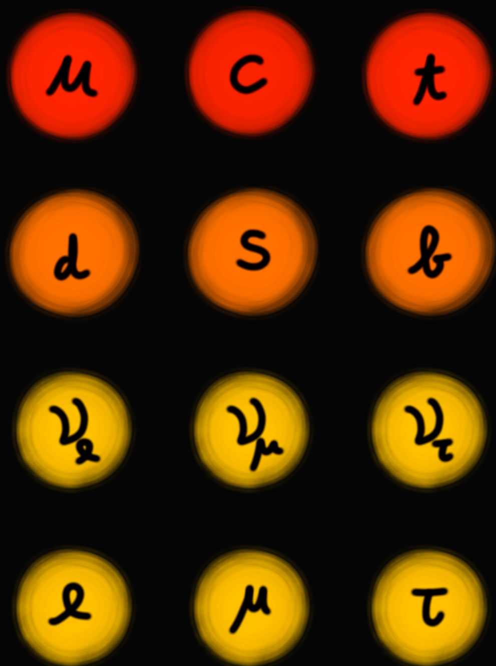
"A first surprise"

Fundamental Particles of Matter

Three complete families of fermions

quarks

leptons

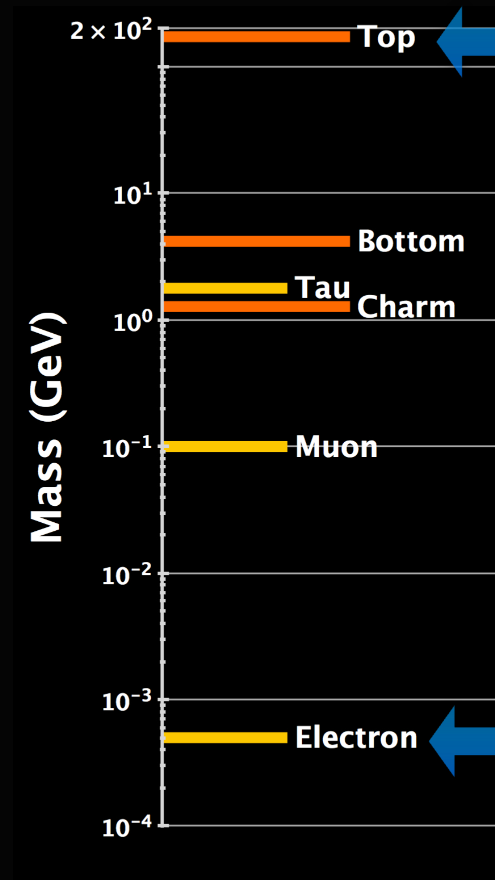


Electrical Charge	Spin
+2/3	1/2
-1/3	1/2
0	1/2
-1	1/2

Intrinsic Angular Momentum

+ anti-particles

(i.e. copies of opposite charge)



173 GeV

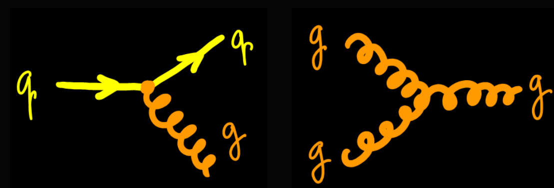
Equivalent to mass of a gold atom

79 protons & electrons
118 neutrons

0.5 MeV

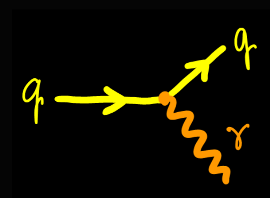
Interactions occur through exchange of bosons

0 1



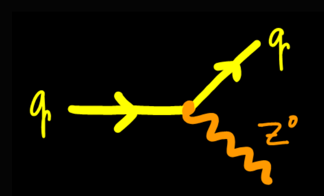
Strong force
(gluons)

0 1



Electromagnetic force
(photon)

0 1



Weak force
(W and Z bosons)

+1 or -1 1

Charge Spin

Beta decay:
$n \rightarrow p\, e^{-}\, \nu_e$

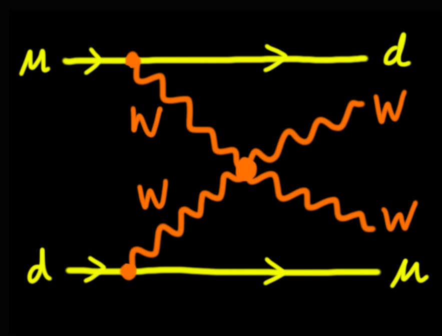
The weak nuclear force has a very small range ($10^{-18}\ m$) so its force carriers (W and Z boson) have to be massive

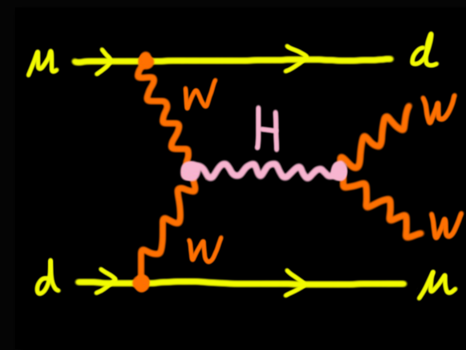
It is impossible to build a consistent theory for massive bosons like the W and Z without an additional particle.

The Higgs Boson

Solution proposed by several theorists in 1964

Higgs, Brout, Englert, Hagen, Guralnick and Kibble

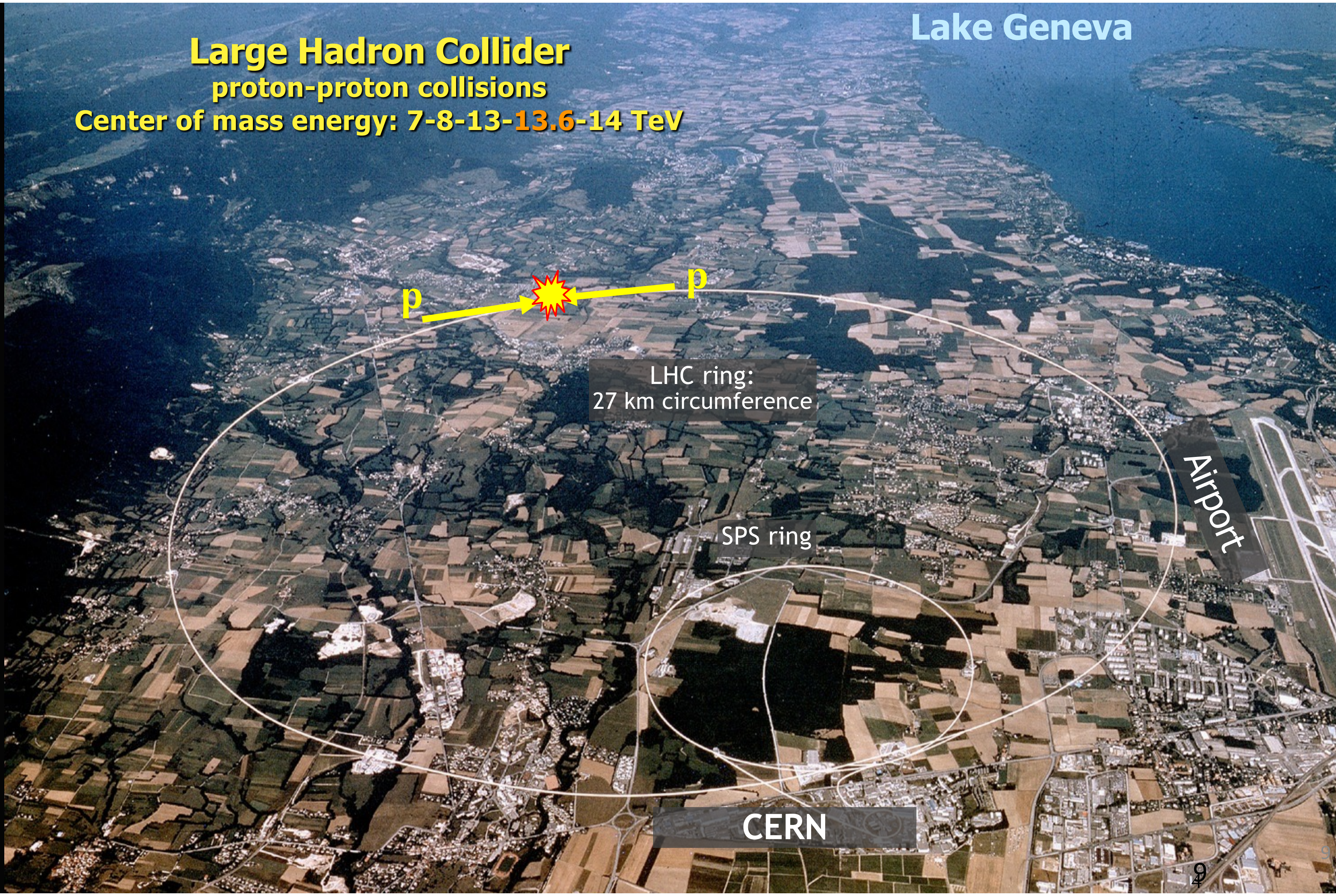


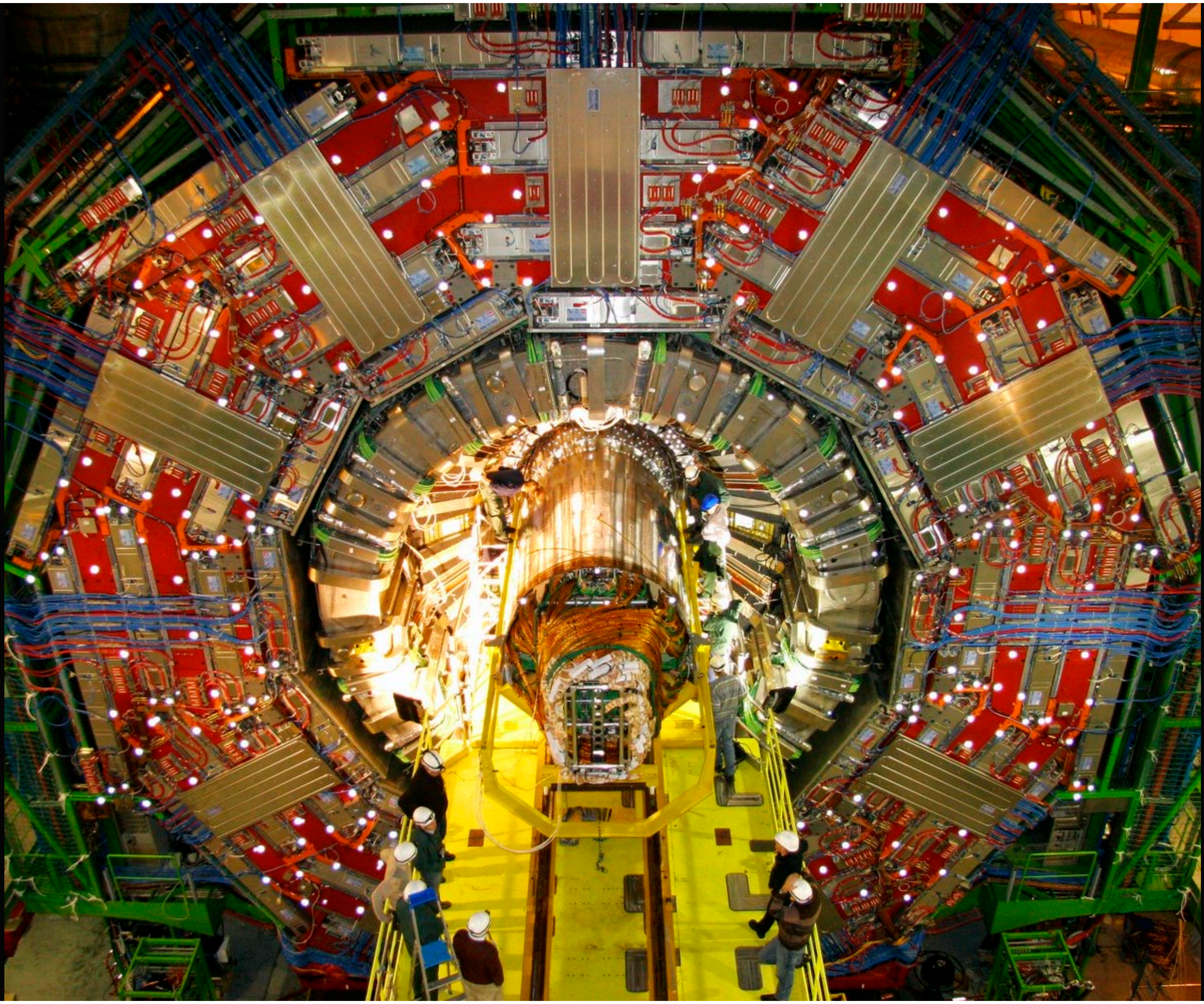


A new fundamental particle with spin 0
(the only one in the Standard Model)
could make the theory consistent again!

Electrical Charge	Spin
0	0

The LHC experimental facility was built to test this theory

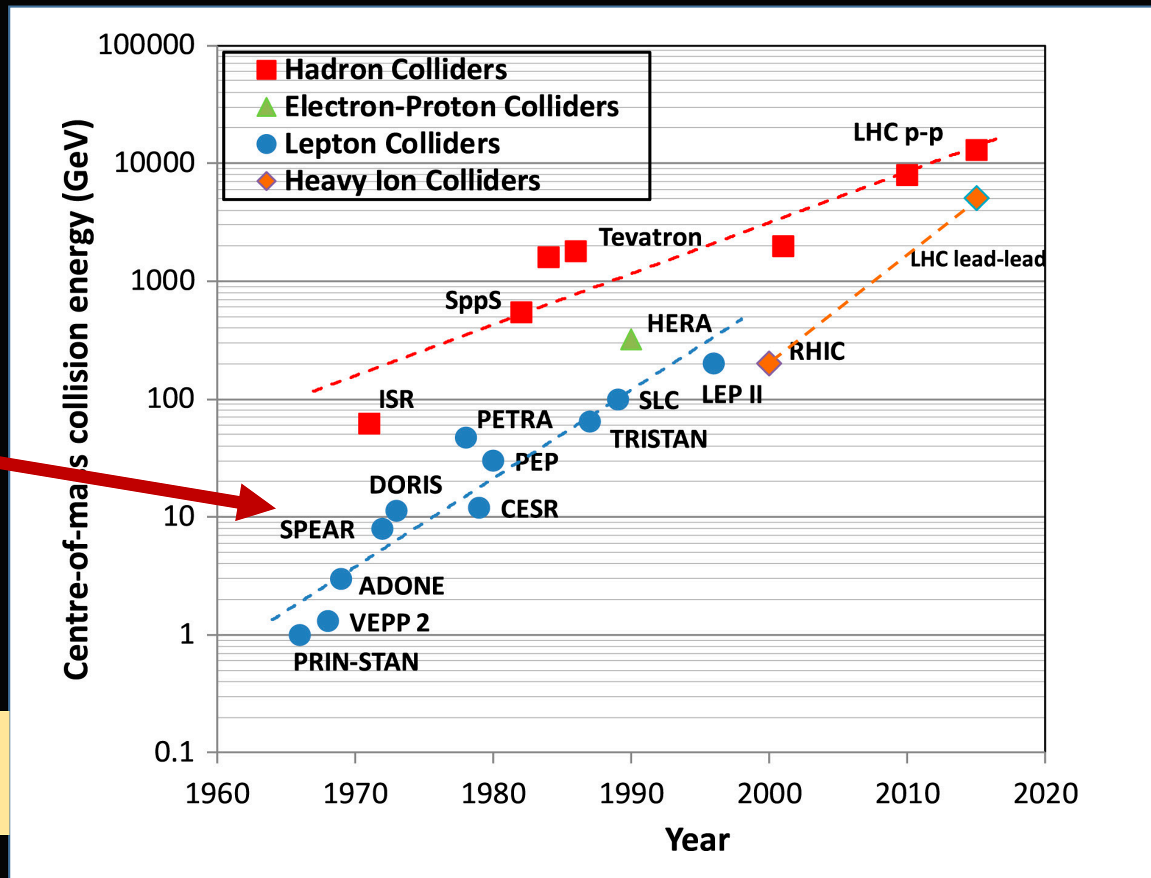

Large Hadron Collider
proton-proton collisions
Center of mass energy: 7-8-13-13.6-14 TeV
Lake Geneva
p
p
LHC ring:
27 km circumference
Airport
SPS ring
CERN



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

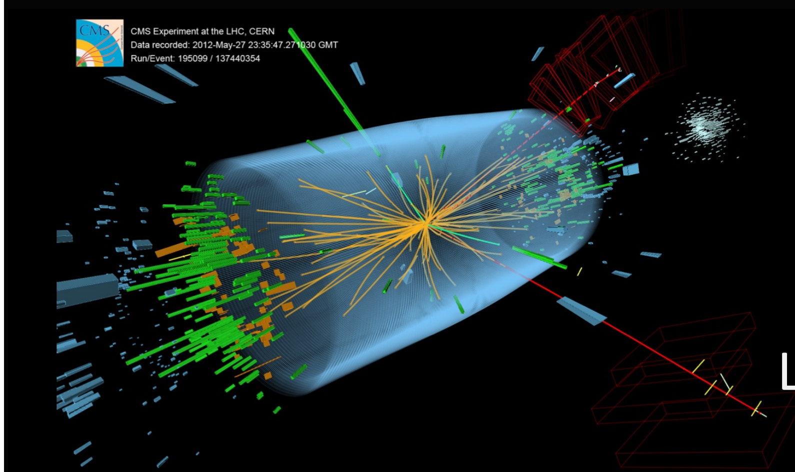

Higgs Particle Discovery Announcement July 4th, 2012
ICHEP, Melbourne
CERN, Geneva
Something that looks like the Standard Model Higgs boson was discovered in 2012, Mass ~ 125 GeV
Nobel prize in physics in 2013
RESERVED
12

High Energy Physics is a facilities driven science

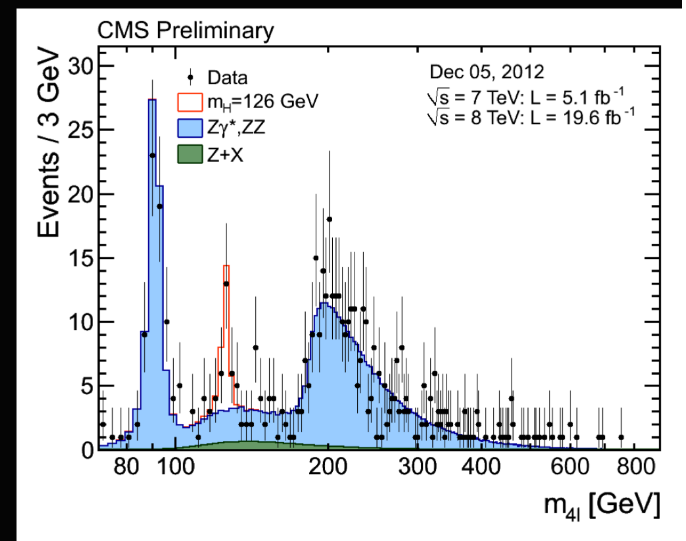


This week on Friday…
https://indico.slac.stanford.edu/event/9040/

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

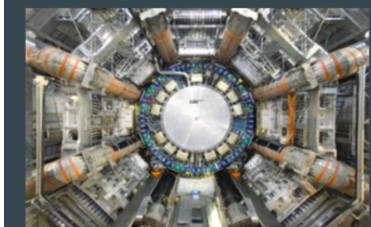


Large scale software and computing infrastructures



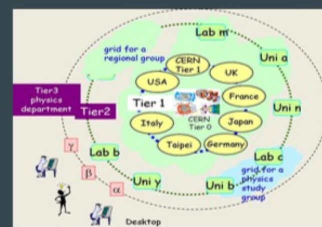
Our software takes data from collisions to physics results

Detector

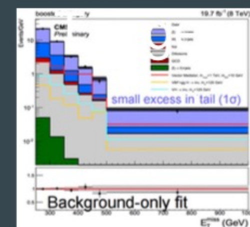


Trigger

Reconstruction



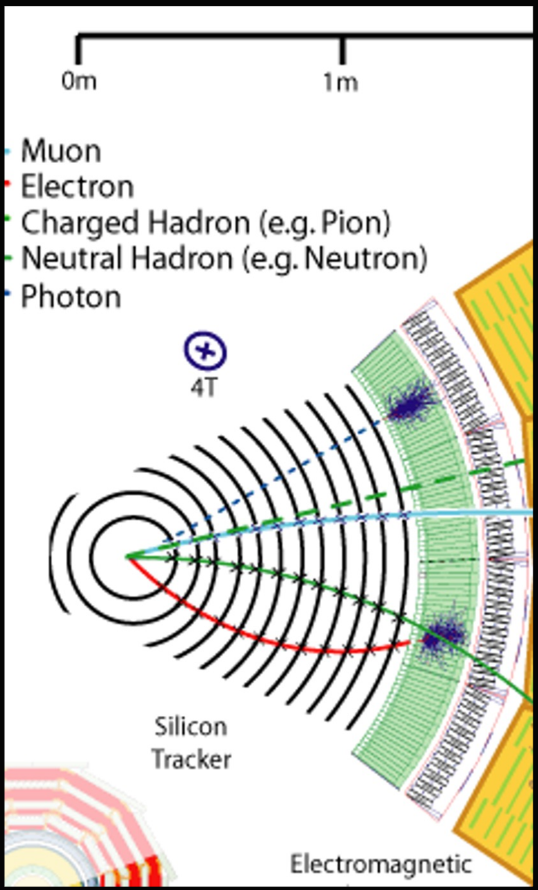
Analysis

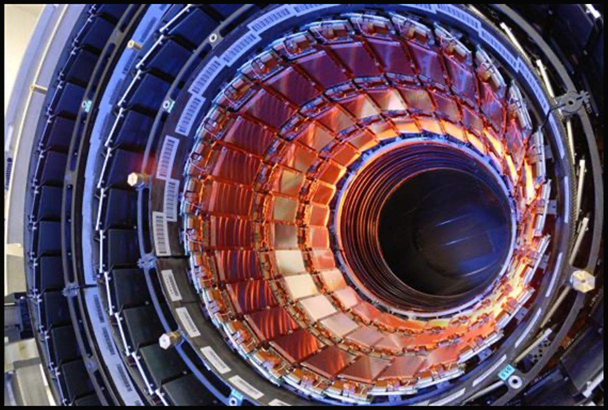


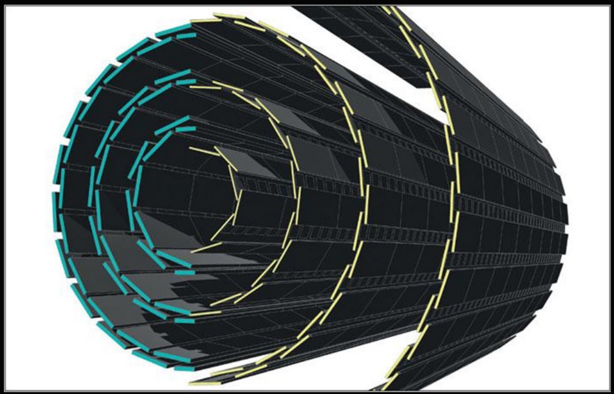
Paper

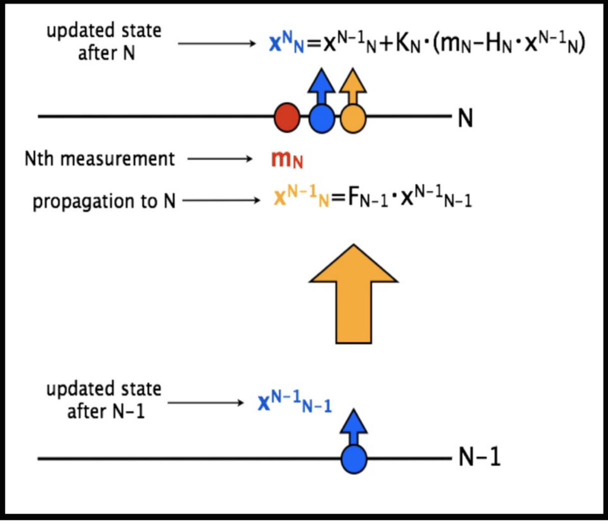
- Detector data collected at the rate of 60 TB/second must be reduced to ~5 GB/second in real time
- Pattern recognition algorithms find particle trajectories from the detector data ("reconstruction")
- In parallel, Monte Carlo methods are used to create simulations of how different physics processes appear in our detector
- Collisions are categorized by examining reconstructed particles for signs of specific physics processes (eg, Higgs decays) ("Analysis")

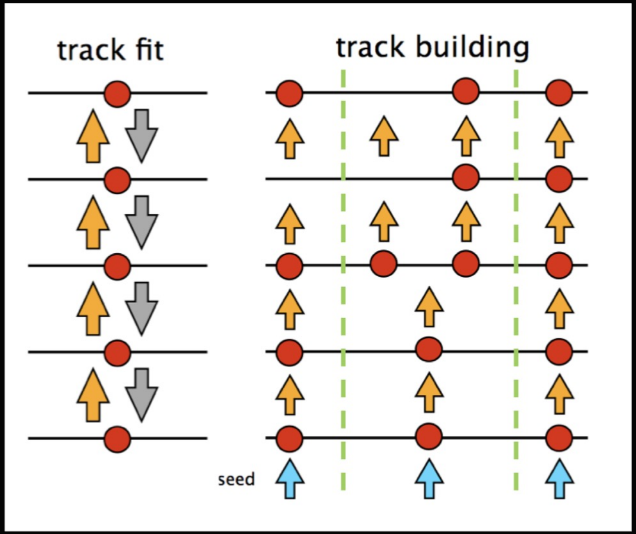
Reconstruction applications process RAW data into "physics objects" for analysis



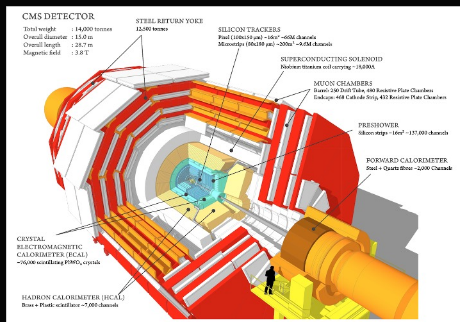




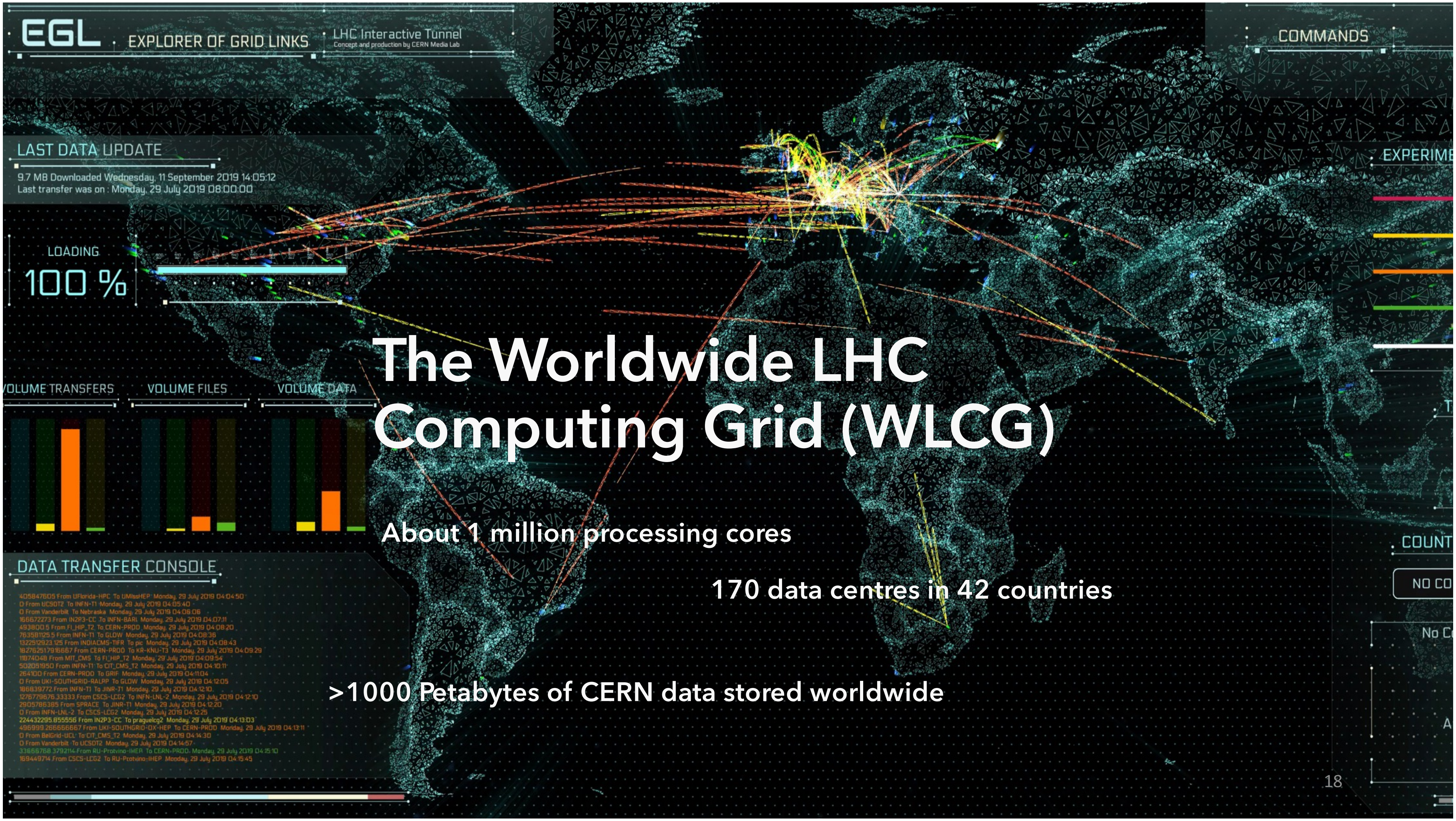




Tracking particles through CMS



	LHC events per year (/10^9)	Tracks per year (/10^{15})	Time for processing (Latency)
L1 Trigger (FPGA-based)	240000	1000	~10 us
"High Level" Trigger	6000	40	Seconds
Offline Processing	60	0.5	Days – can be redone if needed
Analysis	60	0.5	Years -- redone many times



The Worldwide LHC Computing Grid (WLCG)

About 1 million processing cores

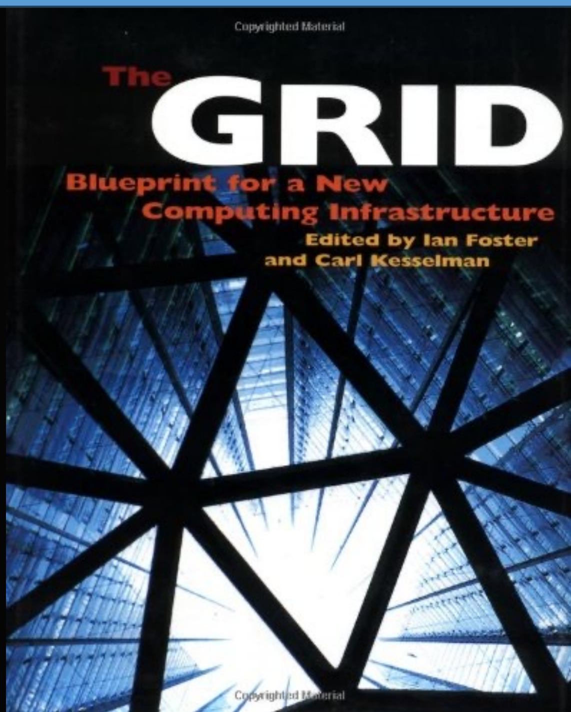
170 data centres in 42 countries

>1000 Petabytes of CERN data stored worldwide

What are the big challenges?

Scientific computing must make effective use of commodity computing which is driven by industry applications (today: AI/ML)

2005: "Data grid" vision

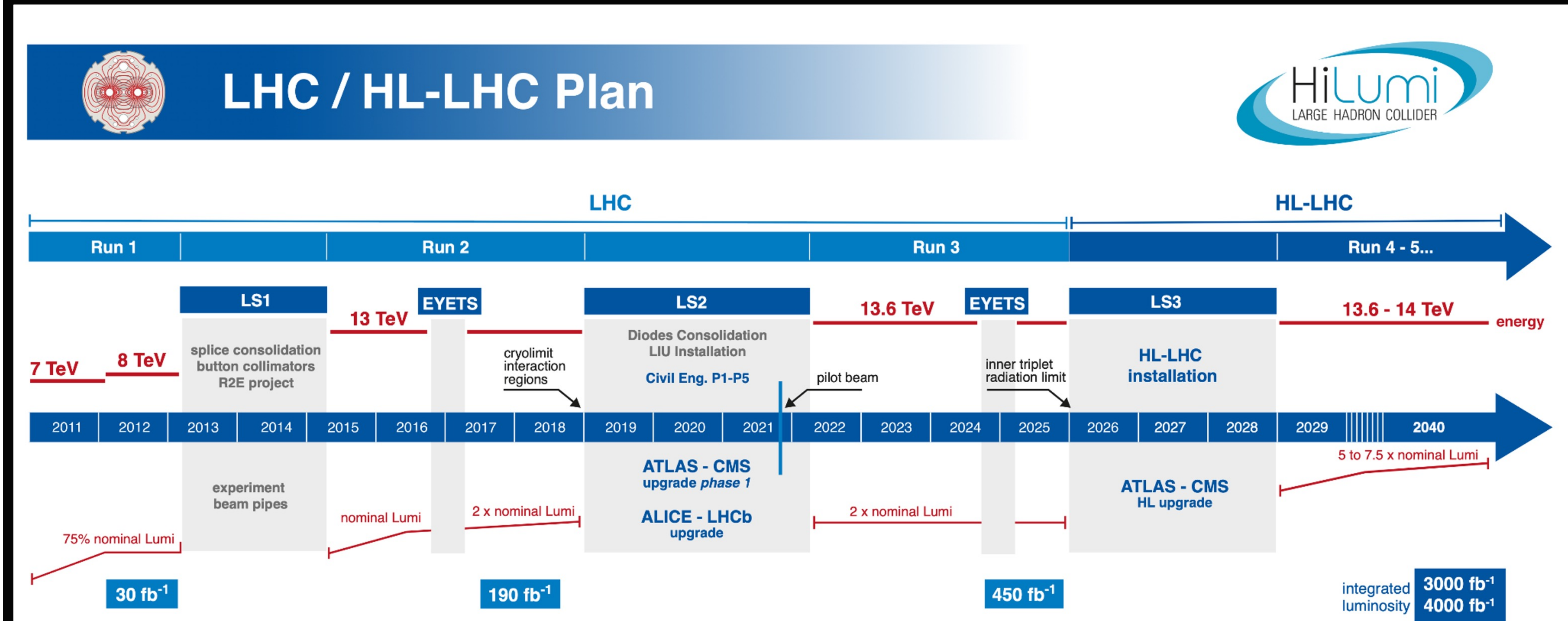


2025: Data centers dominate compute

aws

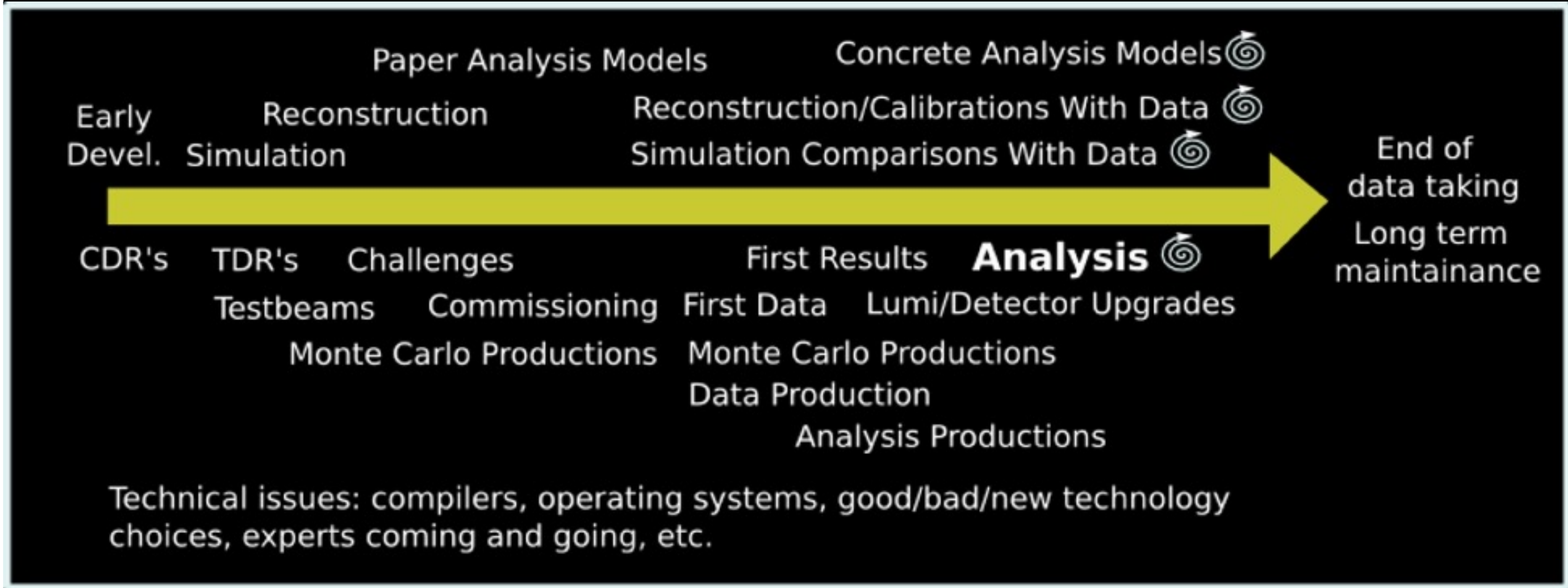
Google Cloud

Experimental timescales span decades

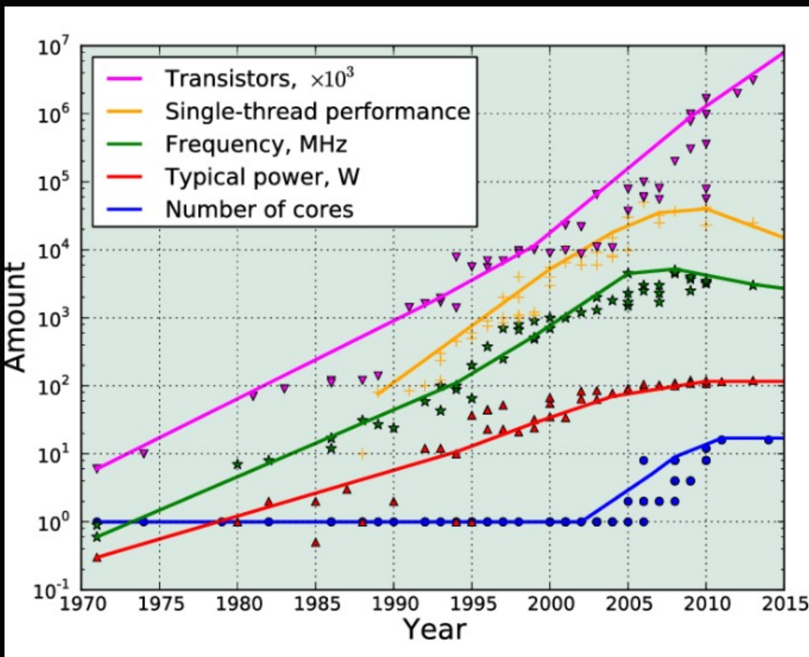


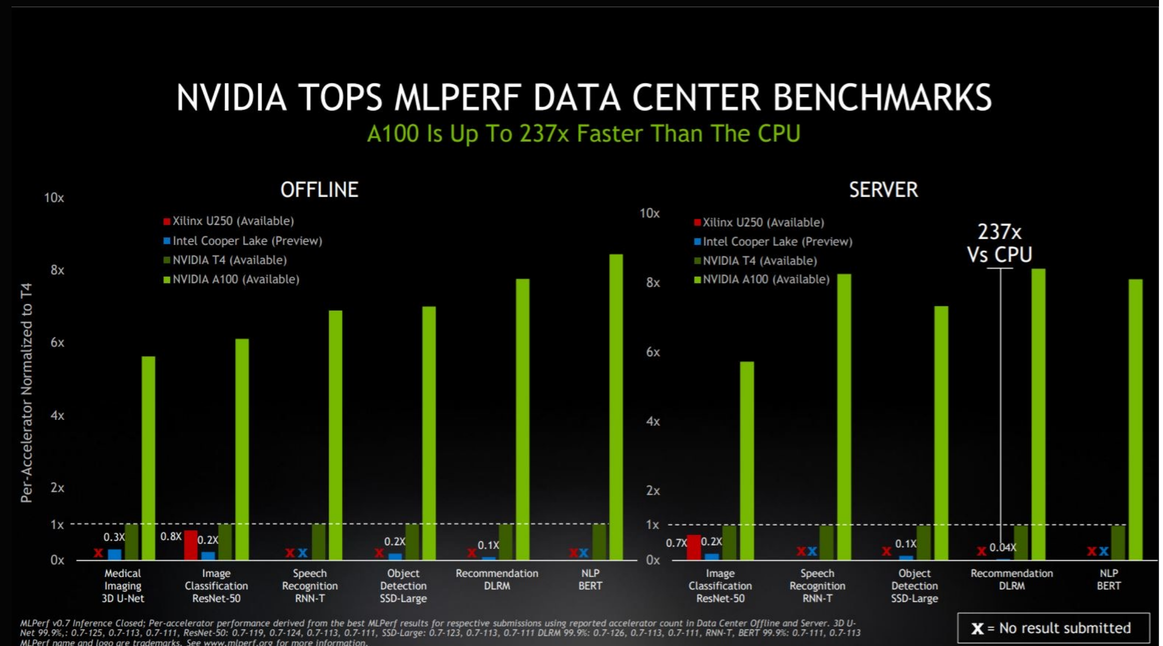
Experiment designs start far before data taking. CMS was formed in 1992 (more than 30 years ago!), expects to run through 2040 and do data analysis for years after that

HEP software lifecycle



Commodity resources evolve faster during experiment lifecycles

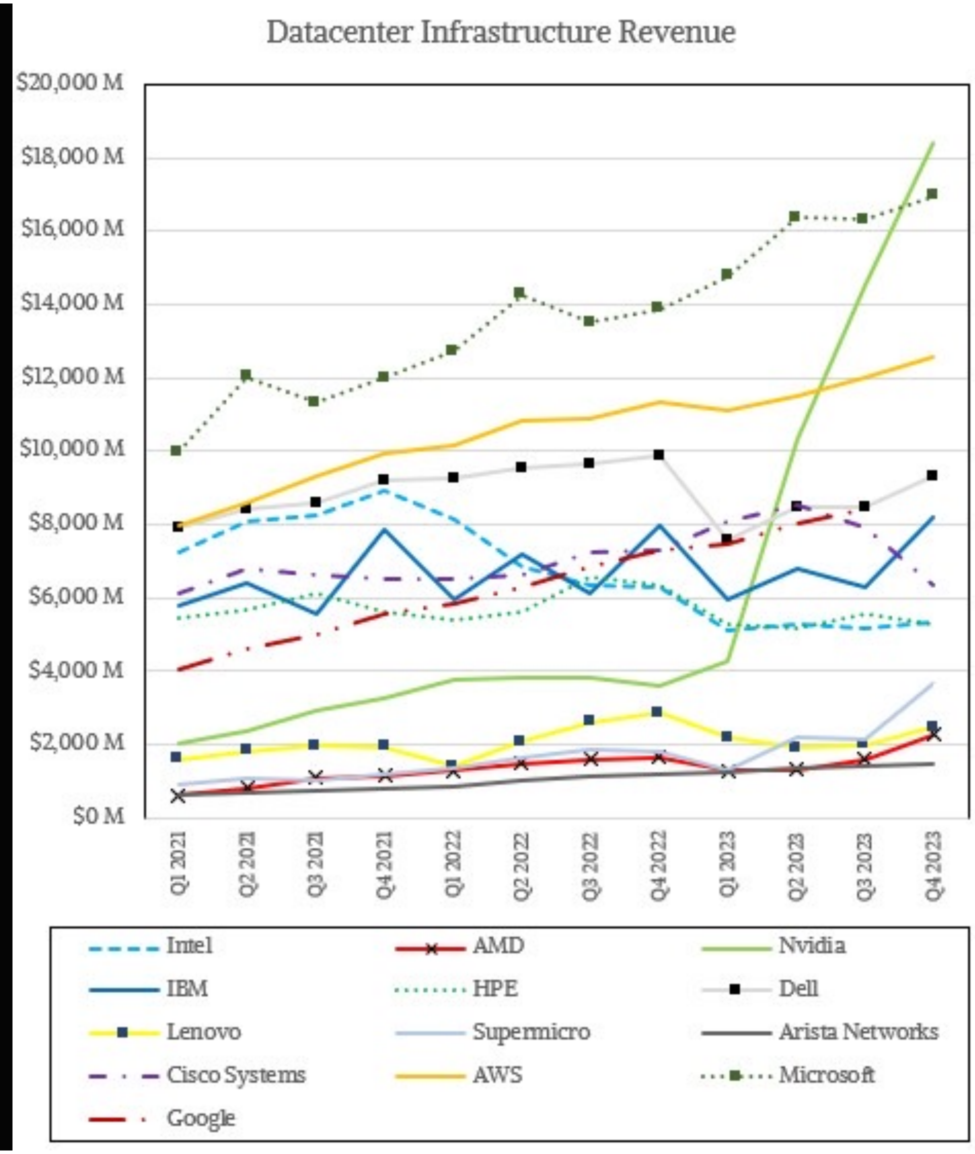




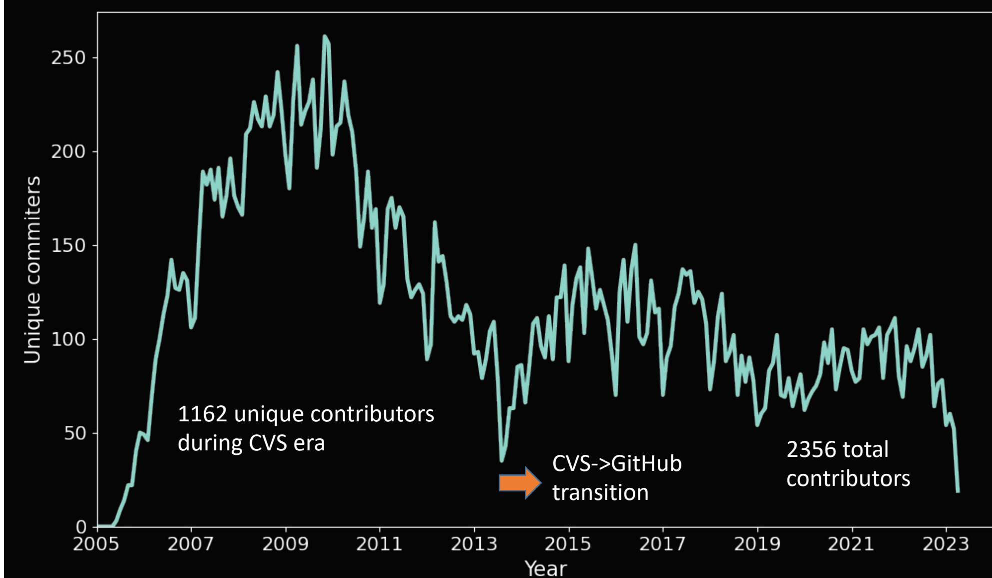
The end of Dennard Scaling: Parallelism has become the ways to faster performance in compute

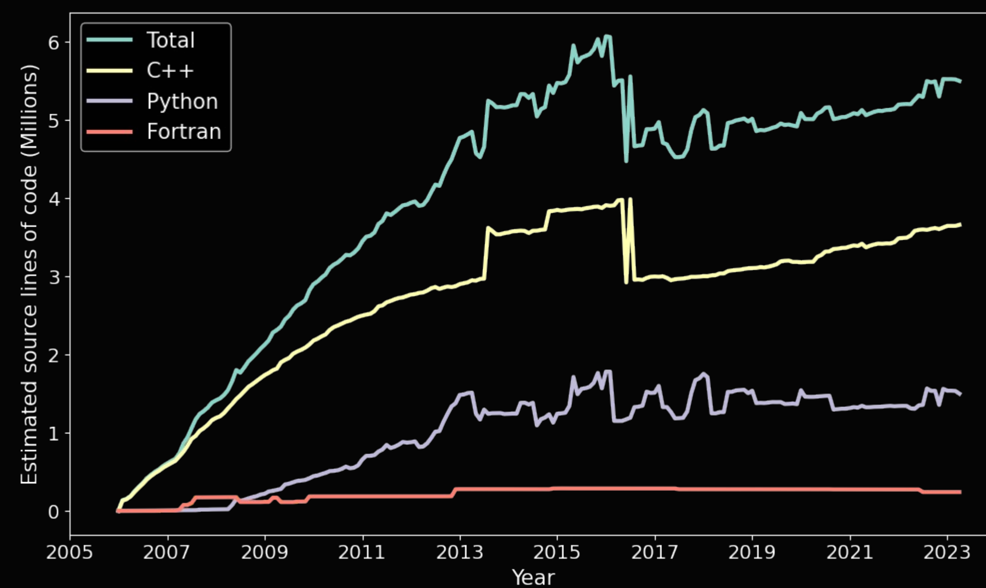
Parallelism on a chip: GPU performance currently triples every 18 months

Data center infrastructure spending reflects the rise in GPU performance



Large scale collaborative software development

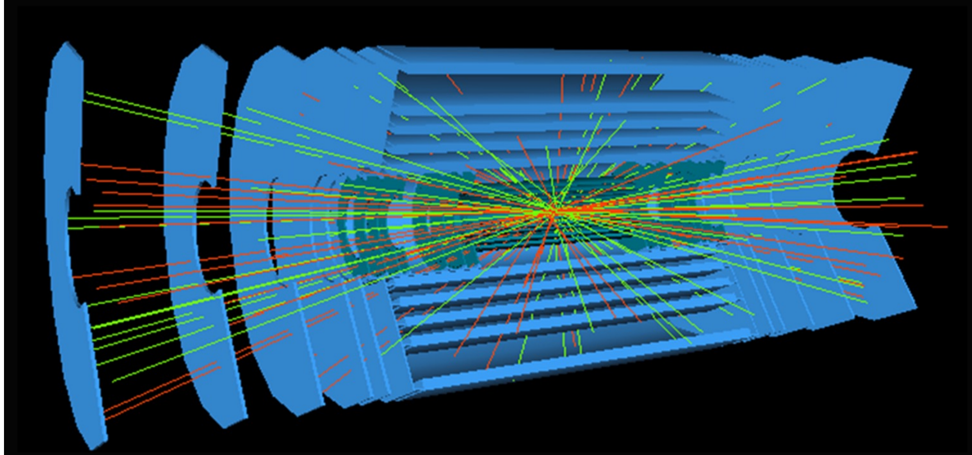


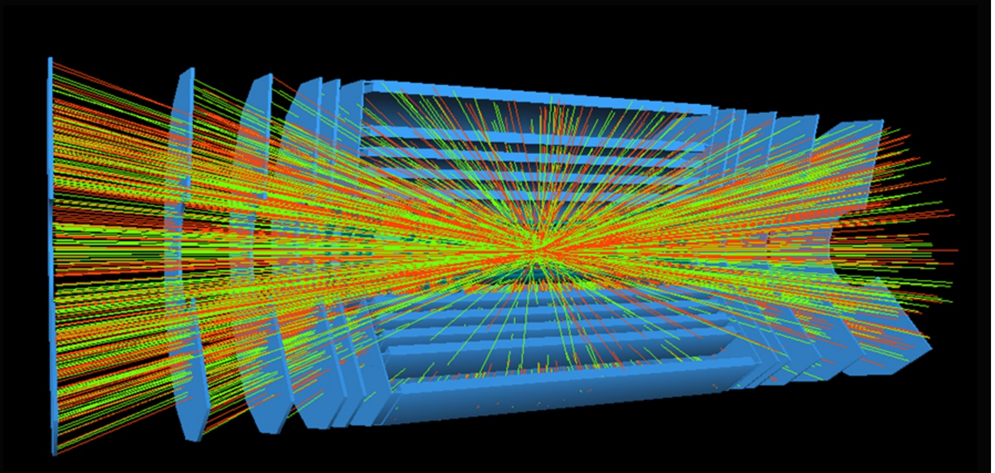


Many developers, typically a handful of true experts

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

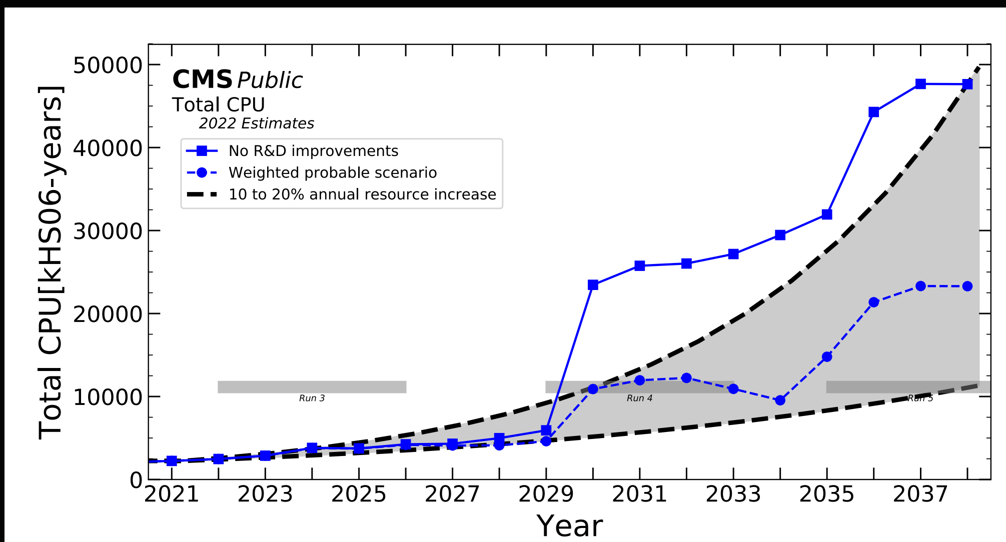
Challenge of next-generation, higher-luminosity or higher-intensity experiments

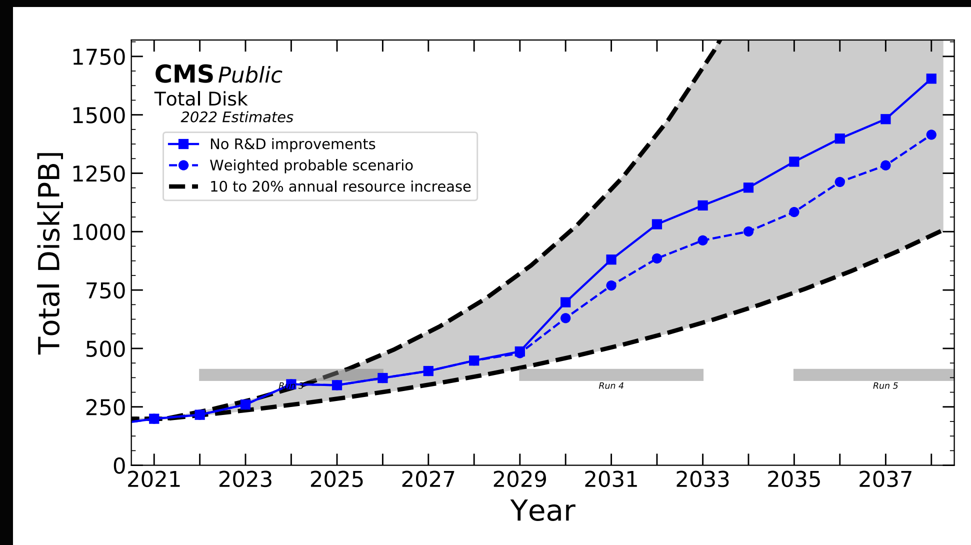




- HL-LHC expects to deliver 200 simultaneous interactions per bunch crossing. 5x more than today.
- More capable detectors are being built to facilitate finding ”needles” in this much bigger “haystack”
- Similarly, the analyst community must develop new and more capable approaches to prepare for much higher event rates, higher event complexity, and more detailed detector information.

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 (Run 1&2)	HL-LHC (Run 4+)
Analysis Dataset size	10 TB	1,000 TB
Target Scan Turnaround time	Weeks	Hours
Analysis team size (physicists)	5-10	< 5
Primary analysis resource	**Laptop**	**Analysis Facility**

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

Organizing the HEP community to address these challenges

The HSF was created as a means for organizing our community to address the software challenges of future projects such as the HL-LHC

- Catalyze new common projects
- Promote commonality and collaboration in new developments to make the most of limited resources
- Provide a framework for attracting effort and support to S&C projects
- Provide a structure to set priorities and goals for work in common projects

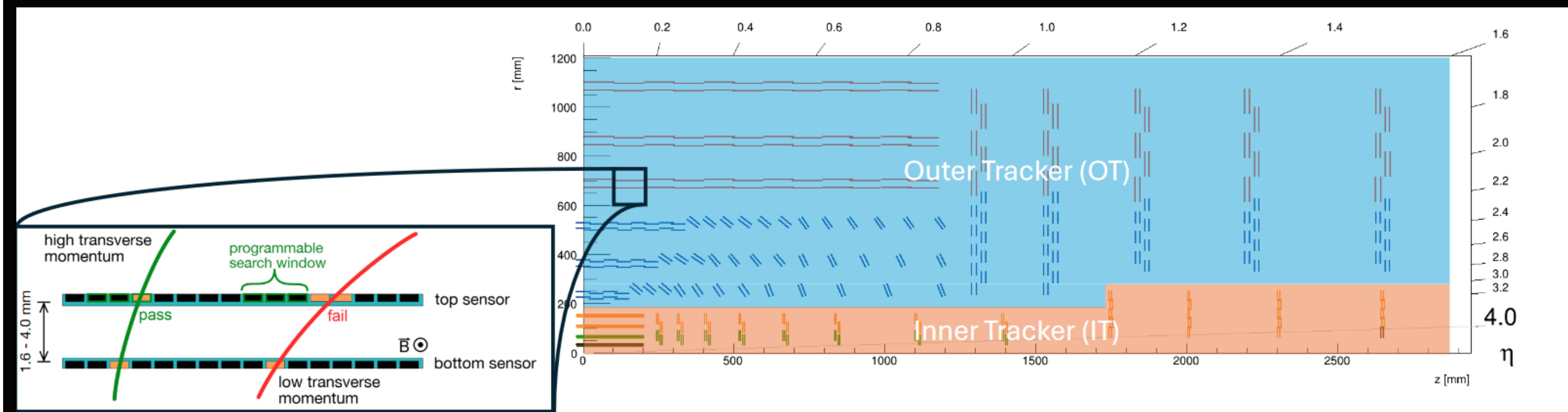
http://hepsoftwarefoundation.org

Collaboration via the HSF led to national and regional initiatives. For example, in the US, 17 universities collaborate on the IRIS-HEP software institute.

- Innovative algorithms for data reconstruction and triggering
- Highly performant analysis systems that reduce `time-to-insight' and maximize the HL-LHC physics potential; and
- Data organization, management and access systems for the community's upcoming Exabyte era.

Example outcomes

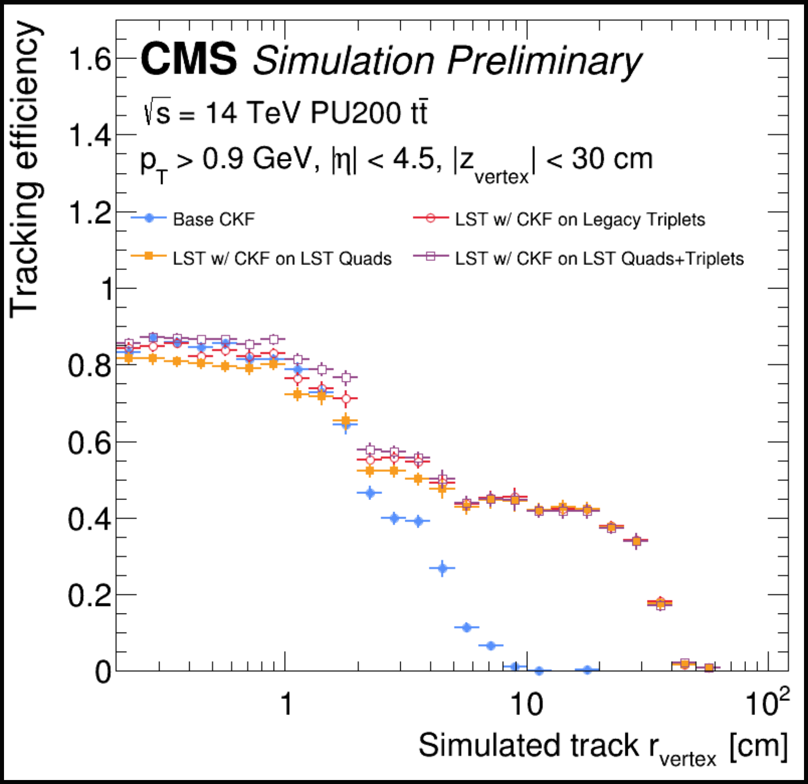
Line segment tracking exploits unique geometry of upgraded CMS tracker detector



- Key characteristic of the CMS Phase 2 Outer Tracker (OT): Each layer comprises 2 closely-spaced silicon sensors.
- Algorithmically link pairs of hits in sensors of the same layer and build up tracks
  - Reduce combinatorics.
  - Can be locally reconstructed ➔ Allow for parallelization.
  - Elementary building block for tracks.

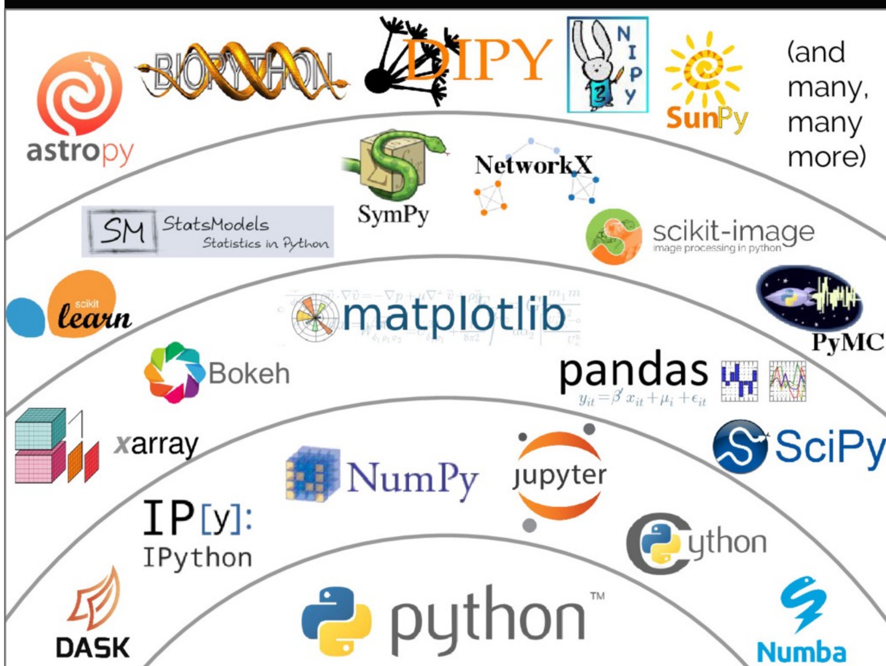
Line segment tracking for CMS

- LST algorithm idea
  - Start small, build up tracks by making doublets, then triplets out of doublets, etc
  - Apply ML to final stages where combinatorics are greater. A DNN works very well at reducing fake tracks while retaining efficiency
- Being able to quickly and efficiently find tracks that are very displaced from the primary vertex, LST adds a completely new physics capability for the CMS trigger (HLT)



Leveraging data science for HEP analysis

Scientific Python / PyData vision/ecosystem



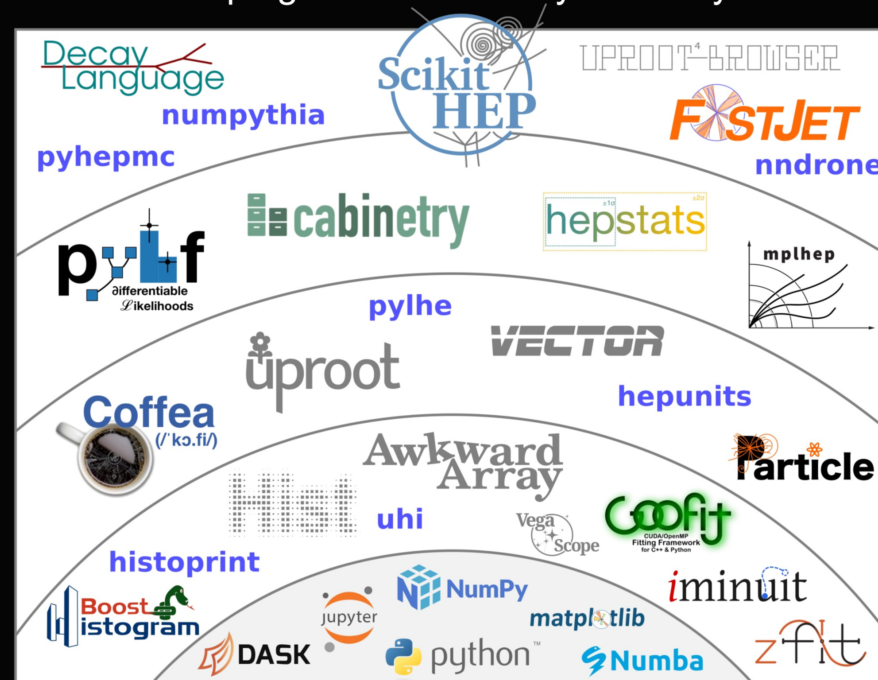
Application Specific

Domain Specific

Technique specific

Foundational

Developing HEP data analysis ecosystem



Awkward Array – numpy for HEP data

```python
array = ak.Array([
    [{"x": 1.1, "y": [1]}, {"x": 2.2, "y": [1, 2]}, {"x": 3.3, "y": [1, 2, 3]}],
    [],
    [{"x": 4.4, "y": [1, 2, 3, 4]}, {"x": 5.5, "y": [1, 2, 3, 4, 5]}]
])
```

```python
output = []
for sublist in python_objects:
    tmp1 = []
    for record in sublist:
        tmp2 = []
        for number in record["y"][1:]:
            tmp2.append(np.square(number))
        tmp1.append(tmp2)
    output.append(tmp1)
```

2.3 minutes to run (22 GB footprint)

```python
output = np.square(array["y", ..., 1:])
```

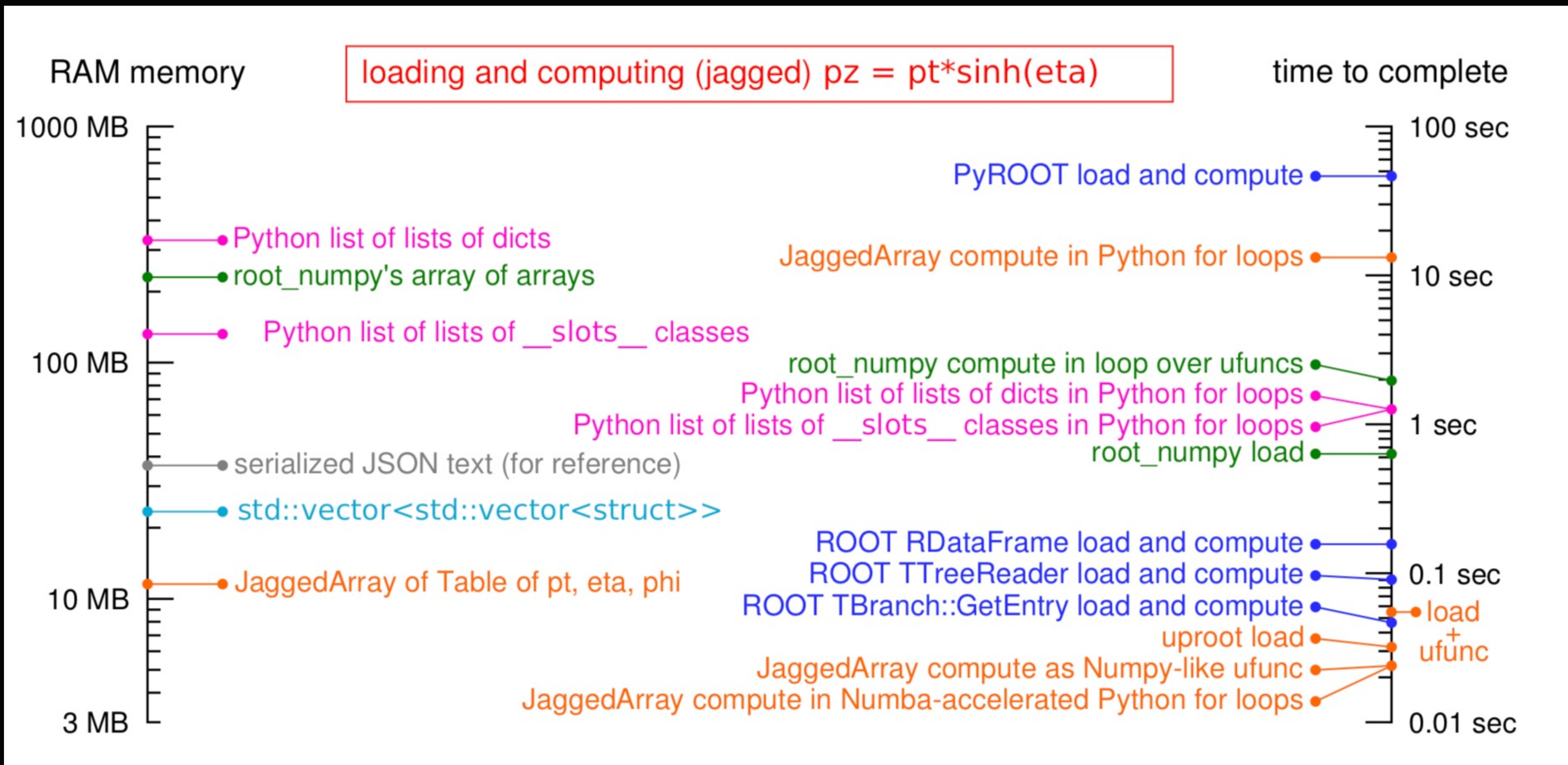
```
[
    [[], [4], [4, 9]],
    [],
    [[4, 9, 16], [4, 9, 16, 25]]
]
```

4.6 seconds to run (2.1 GB footprint)

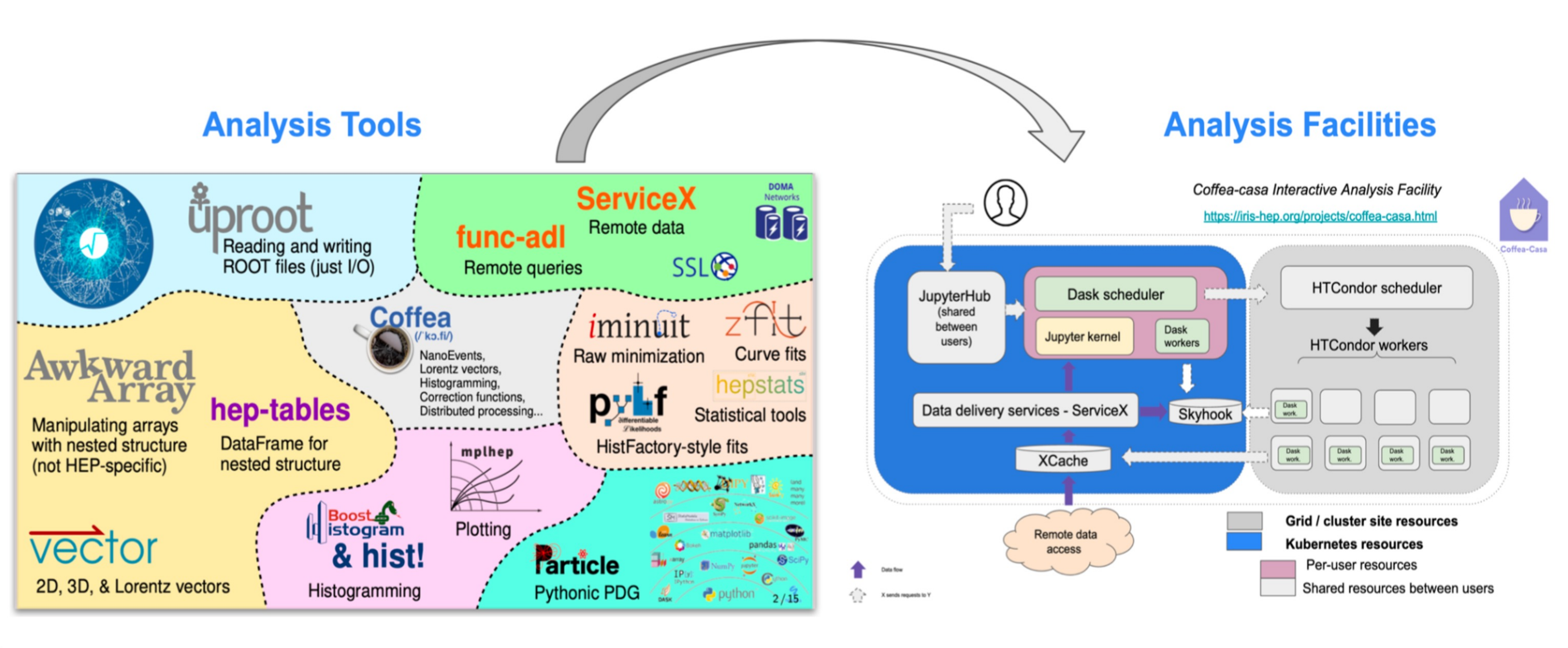
(single-threaded on a 2.2 GHz processor with a dataset 10 million times larger than the one shown)

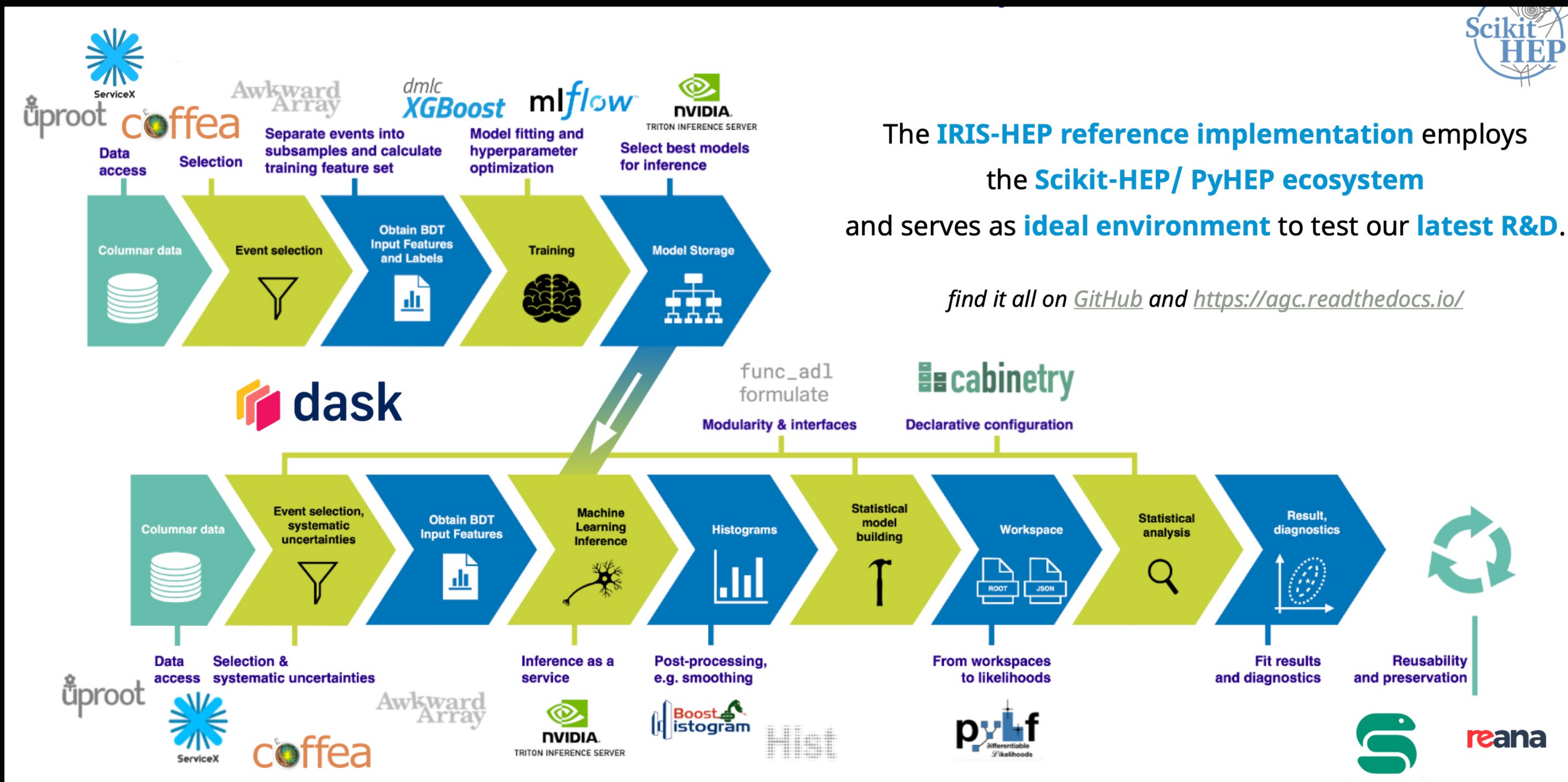
- General tool for manipulating JSON-like structures in a NumPy-like way
- Motivated by problems in HEP which commonly include irregular, “jagged” data

Exciting results are possible: Orders of magnitude speed ups



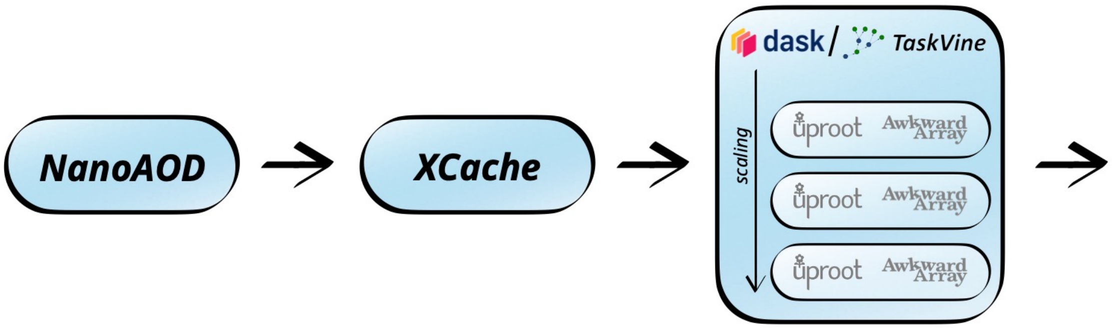
What comes next? Analysis tool chains and facilities rather than just tools

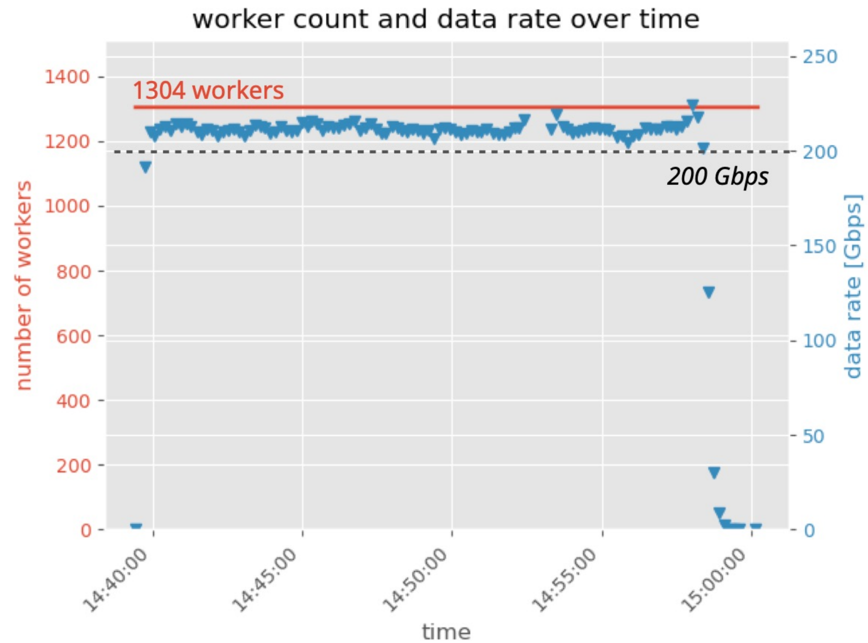




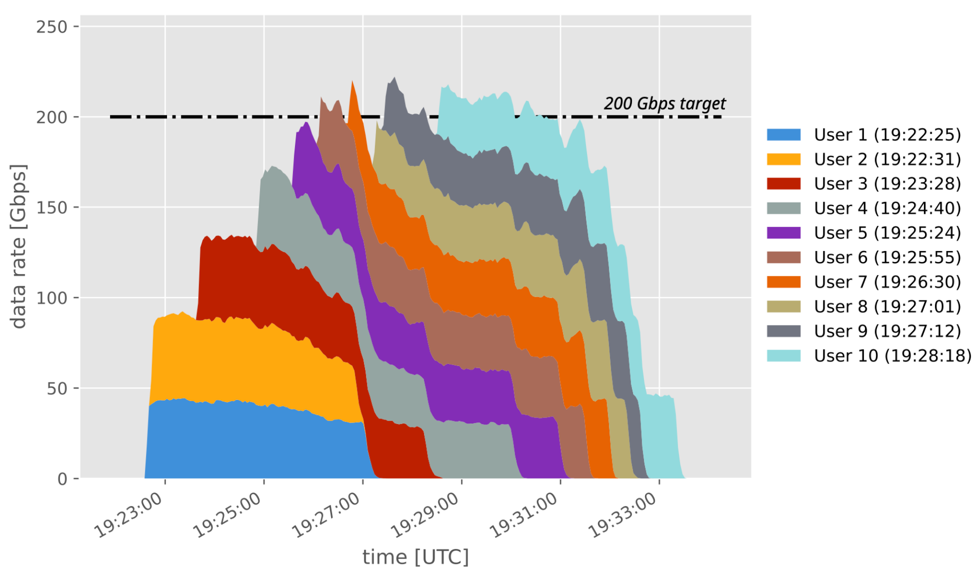
ServiceX
uproot
coffea
Awkward Array
dmlc XGBoost
mlflow
NVIDIA
TRITON INFERENCE SERVER
Data access
Selection
Separate events into subsamples and calculate training feature set
Model fitting and hyperparameter optimization
Select best models for inference
Columnar data
Event selection
Obtain BDT Input Features and Labels
Training
Model Storage
dask
func_adl
formulate
cabinetry
Modularity & interfaces
Declarative configuration
Columnar data
Event selection, systematic uncertainties
Obtain BDT Input Features
Machine Learning Inference
Histograms
Statistical model building
Workspace
ROOT
JSON
Statistical analysis
Result, diagnostics
Data access
Selection & systematic uncertainties
Inference as a service
Post-processing, e.g. smoothing
From workspaces to likelihoods
Fit results and diagnostics
Reusability and preservation
Boost histogram
Hist
pyhf
reana
Scikit HEP
The IRIS-HEP reference implementation employs the Scikit-HEP/ PyHEP ecosystem and serves as ideal environment to test our latest R&D.
find it all on GitHub and https://agc.readthedocs.io/

Reaching data rates matching HL-LHC scale in multiuser prototype facilities





200+ Gbps with Dask + HTCondor



Larger data samples require more sophisticated statistical analysis

Likelihoods are central for High Energy Physics

$$L(\vec{n},\vec{a}|\vec{\eta},\vec{\chi}) = \prod_{c\in unbinned\ ch}\ \prod_{i\in obs} \frac{f_c(\vec{x}_{ci}|\vec{\eta},\vec{\chi})}{\int f_c(\vec{x}_{ci}|\vec{\eta},\vec{\chi})\, d\vec{x}_c} \cdot$$

$$\prod_{c\in binned\ ch(analytical)}\ \prod_{b\in obs} Pois(n_{cb}|\nu(\vec{\eta},\vec{\chi})) \cdot \prod_{\chi\in\vec{\chi}} c_\chi(a_\chi|\chi)$$

Numerical and analytic integrals

$\vec{n}$: data, $\vec{a}$: auxilary data, $\vec{\eta}$: unconstrained parameters, $\vec{\chi}$: constrained parameters

Recently automatic differentiation techniques work well enough to replace these numerical calculations with faster and more robust methods

Automatic Differentiation in a nutshell

$$f(x) = e^{e^{e^{e^{e^{x}}}}}$$

Symbolic via Wolfram Alpha →

$$\frac{d}{dx}\left(e^{e^{e^{e^{e^{x}}}}}\right) = e^{x+e^{e^{e^{e^{x}}}}+e^{e^{e^{x}}}+e^{e^{x}}+e^{x}}$$

Figure out the analytical fn ↑ ↓ *Handcode*

Handcode, optimized by expert ↓

```
// f(x)=e^(e^(e^(e^(e^x))))
#include <cmath>
double f (double x) {
  double result = x;
  for (unsigned i = 0; i < 5; i++)
    result = std::exp(result);
  return result;
}
```

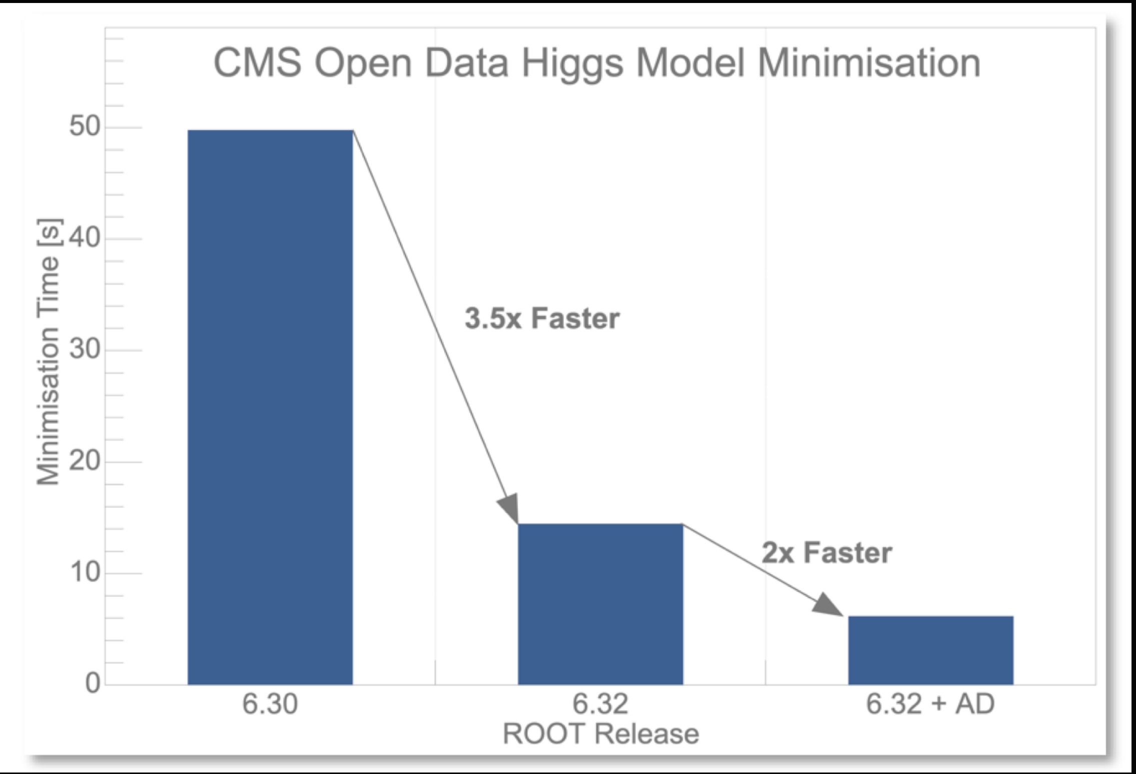
AD →

```
double f_dx(double x) {
  double result = x;
  double d_result = 1;
  for (unsigned i = 0; i < 5; i++) {
     result = std::exp(result);
     d_result *= result;
  }
  return d_result;
}
```

```
auto f_dx = clad::differentiate(f, "x");
```

AD now used by "RooFit" instead of numerical derivatives

- Use of AD in RooFit brings two benefits:
  - Faster answers
  - More robust minimization: Removes the need for numerical stability tricks
- Note: Not all functions in RooFit supported by AD techniques (yet)…



The HSF-India Project

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

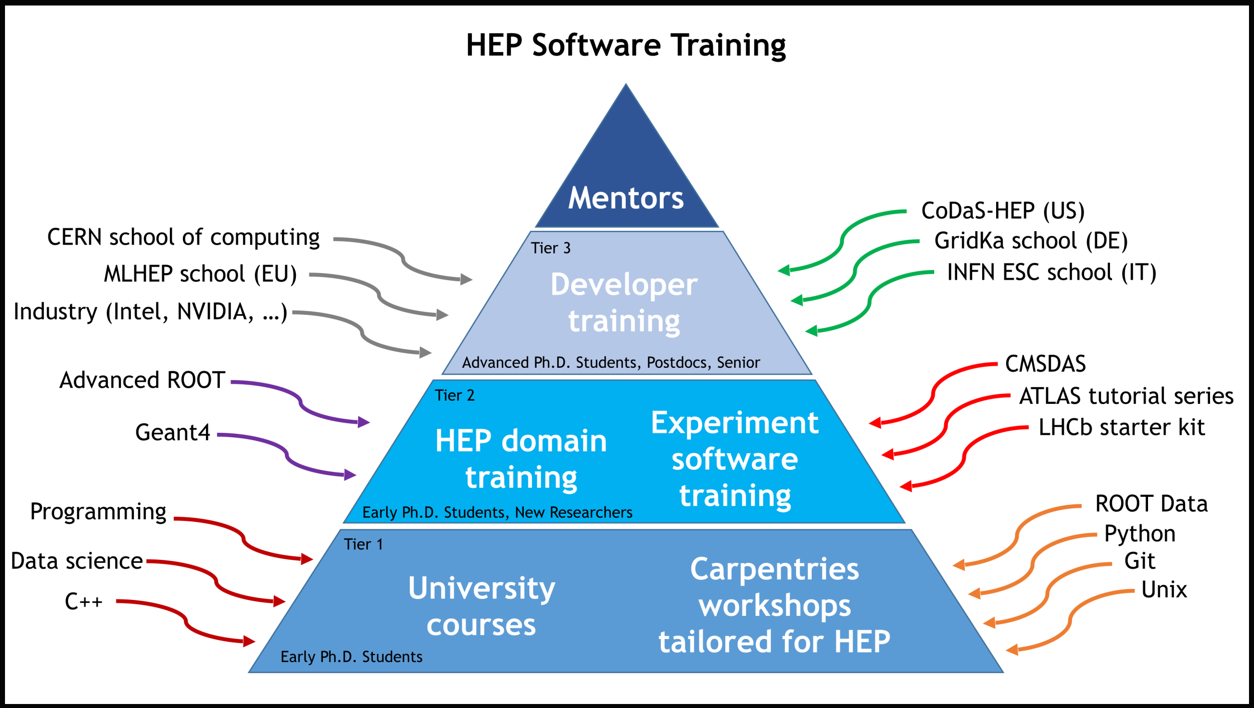
- Given the growing complexity of our scientific data and collaborations, building and fostering collaborations are increasingly important to raise the collective productivity of our research community.
- HSF-India project aims to build international research software collaborations between US, European, 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), and Delhi (University of Delhi)

- 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

Scientific Python

Marshmallow Challenge

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

Upcoming events

Scientific Computing Workshop for High Energy Physics and Tomography

16–20 Dec 2024
Asia/Kolkata timezone

Enter your search term

- Overview
- Timetable
- Expression of Interest Form

Contact us

vikas@vecc.gov.in
+91 33 2318 2424

Software and data analysis skills are essential for doing research. Now a days, Parallel and GPU programming, machine learning, Scientific Python and simulation techniques are crucial for advancing the scientific objectives of next-generation particle, nuclear, astroparticle, High-energy physics and Muon Tomography. We are organizing a 5-day workshop entitled "Scientific Computing Workshop for High Energy Physics and Tomography" during December 16-20, 2024 at VECC, Kolkata. The workshop will consist of topical sessions that target specific data analysis and software skills critical for research in physics with application for Tomography. National and International subject matter experts will lead interactive lectures on scientific Python, data analytics using machine-learning tools, the opportunistic utilisation of GPU computing and other topics. These hands-on sessions will facilitate practical learning and collaborative experiences.

This workshop covers

- Scientific Python
- Parallel programming and GPUs
- Basics of machine learning
- Techniques and applications for muon tomography

This event is organized primarily for M. Sc. and PhD students. Number of seats is limited. Researchers interested to participate are requested to provide their details in the linked form (found on the left toolbar once the application window opens). latest by November 8, 2024. Registration link will be sent to the selected candidates - participation is by invitation only.

This event is organized by Variable Energy Cyclotron Centre, Kolkata - A R&D unit under Department of Atomic Energy, Government of India . Expenses towards participation of HEP (High Energy Physics) Software Foundation experts is provided by the U.S. National Science Foundation (grant OISE-2201990).

HSF-INDIA HEP SOFTWARE WORKSHOP

January 13th to 17th, 2025
Centre for Advanced Studies in Electronics Science and Engineering
School of Physics
University of Hyderabad, Hyderabad, India

Chief Patron
Prof. B. J. Rao
Hon. Vice Chancellor
University of Hyderabad

Special Invitee
Prof. M. Ghanashyam Krishna
IOE Director
University of Hyderabad

Organizing Committee
Prof. M. Ghanashyam Krishna, UoH
Prof. James Raju , UoH
Prof. Samrat Sabat, UoH
Prof. Nageswara Rao, UoH
Prof. Rukmani Mohanta, UoH
Prof. Soma Sanyal, UoH
Dr. Bhawna Gomber , UoH
Dr. Pratap Kollu, UoH
Dr. Anjali Priya, UoH
Dr. David Lange, Princeton University, USA
Dr. Peter Elmer, Princeton University, USA
Prof. Rafael Coelho Lopes de Sa, UMass-Amherst, USA
Prof. Verena Martinez Outschoorn, UMass-Amherst, USA

Topics

Scientific Python
Parallel Programming & GPUs
Basics of Machine Learning
Real-time triggering software

The workshop primarily targets masters & early stage PhD students

Registration
https://indico.cern.ch/event/1394564/
Deadline: November 1, 2024

HEP Software Foundation

The HSF-India project aims to promote the development of international research software collaborations. This is the fifth in a series of workshops for software and data analysis skills essential for doing research software in physics.

Conveners
Dr. Bhawna Gomber (bhawna.gomber@cern.ch)
Dr. David Lange (David.lange@cern.ch)

Sponsored by IOE, University of Hyderabad and HSF-India (NSF/USA)

OISE-2201990

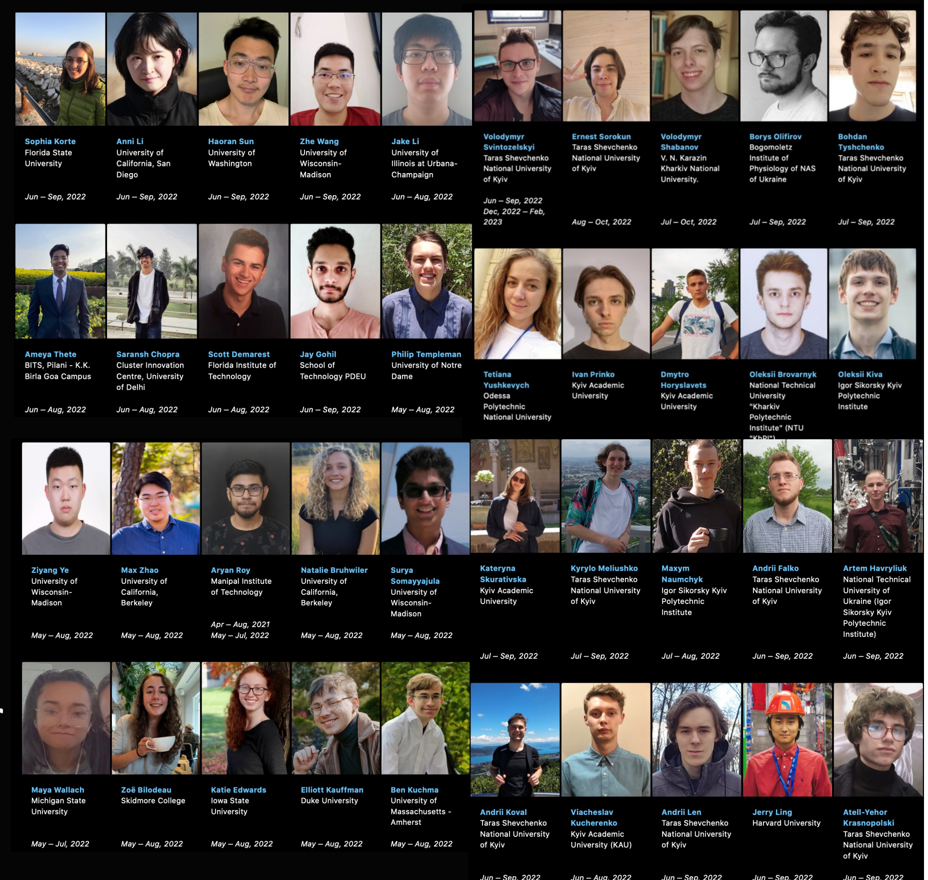
https://indico.cern.ch/event/1461967/

https://indico.cern.ch/event/1394564/

3-6 month project Fellows Program

IRIS-HEPs fellow program.
https://iris-hep.org/fellows.html

- **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



Bidirectional Research Exchange Program

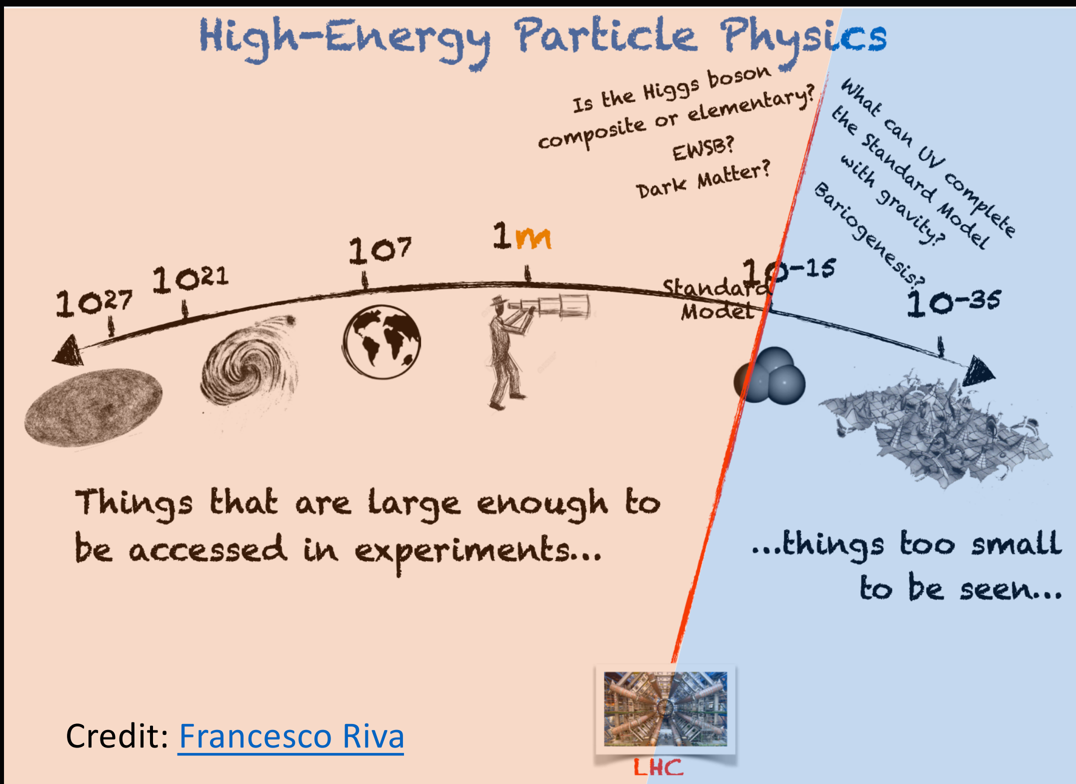
We also have funding for “research exchanges” that support travel costs for 1-3 months. These are meant for very senior PhD students and more senior researchers that have already

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

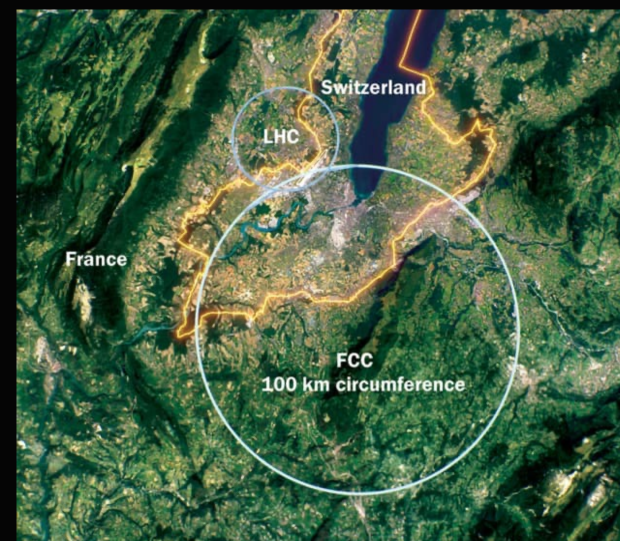
We are looking for either project/host ideas or those interested in doing an exchange. If you have ideas for projects that interest you, we can help identify matches with US researchers

We know that physics beyond our current model exists. However, we do not yet know the energy scale to probe.



Credit: Francesco Riva

Next-generation colliders will bring precision science with Higgs



CERN feasibility study of next generation ee and hadron collider

Conclusion and Opportunities

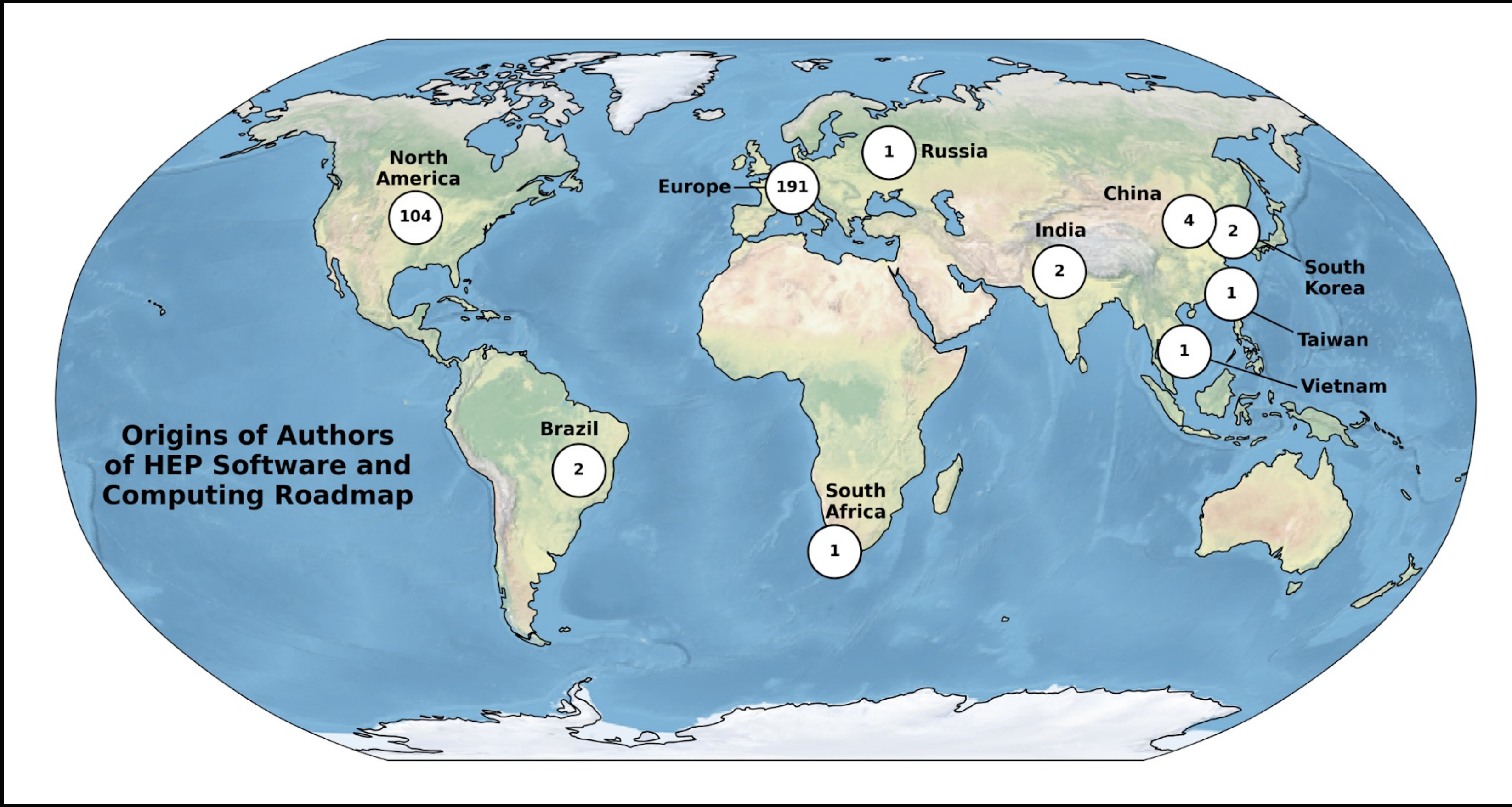
- Large international teams of scientists allow experimental endeavors to exploit regional strengths. Just as with other aspects of our science, software teams that are inherently international are most likely to develop performant, highly usable, and sustainable research software ecosystems.

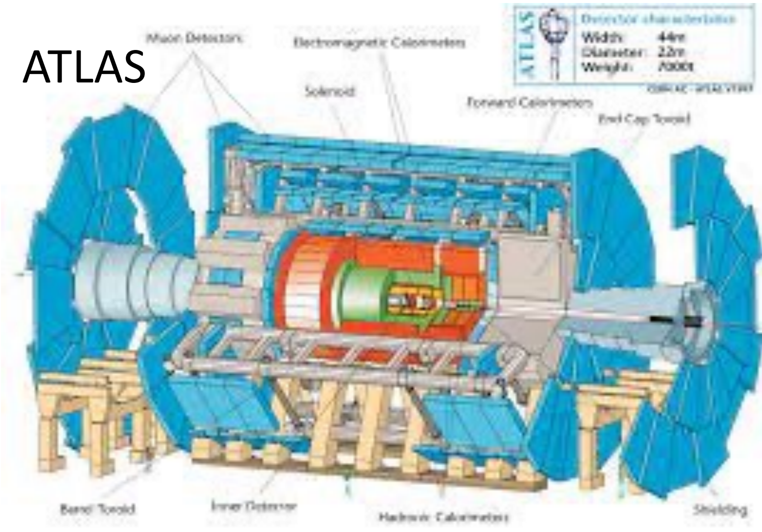
- Long experimental lifetimes with an evolving developer community, large code bases, and distributed computing systems push our community to treat software as infrastructure

- HSF-India is a new project that we hope can catalyze global collaboration in research software in Physics.
  - http://research-software-collaborations.org

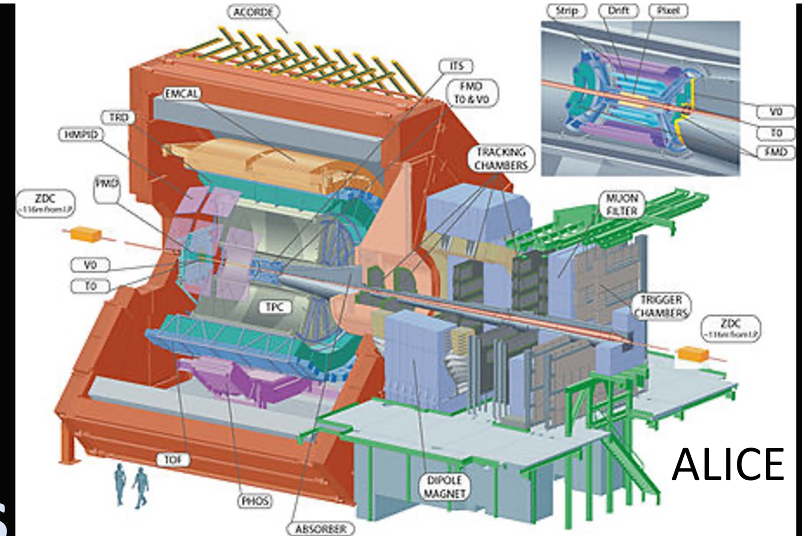
Backup

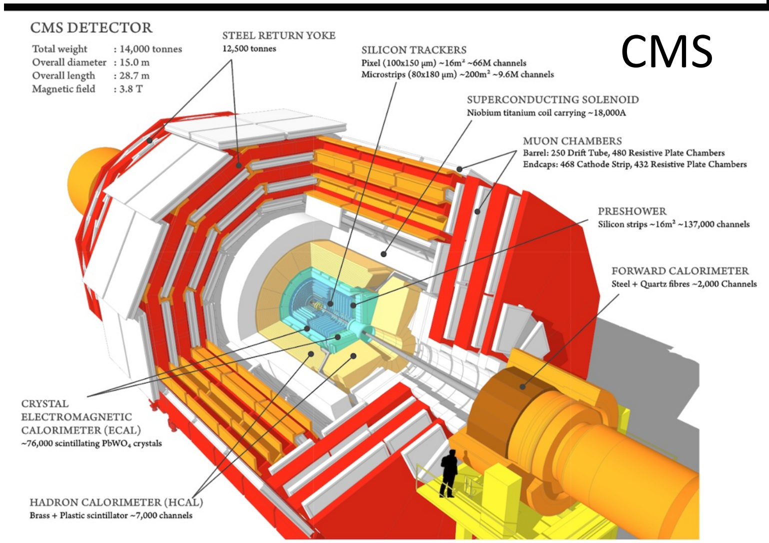
However, nearly all authors of the HSF Community Roadmap were from institutions in Europe and the US

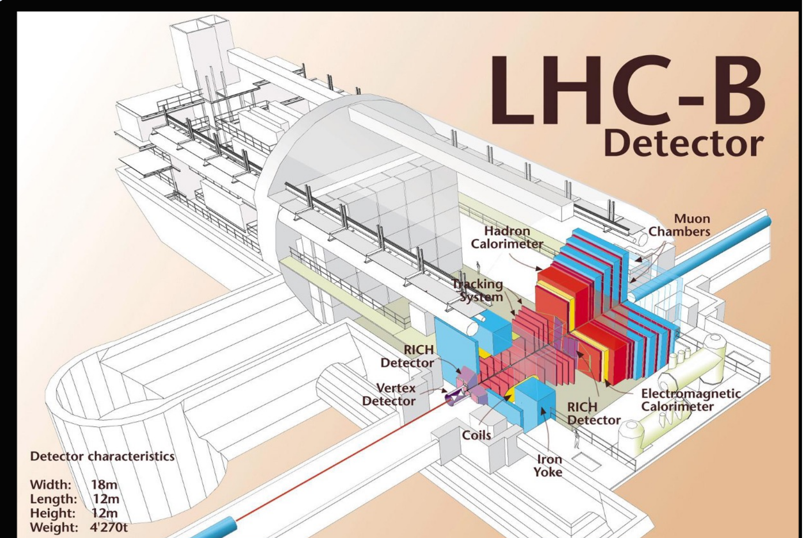


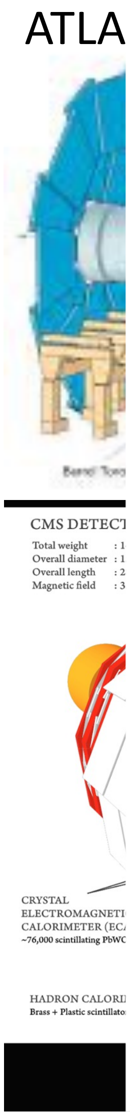


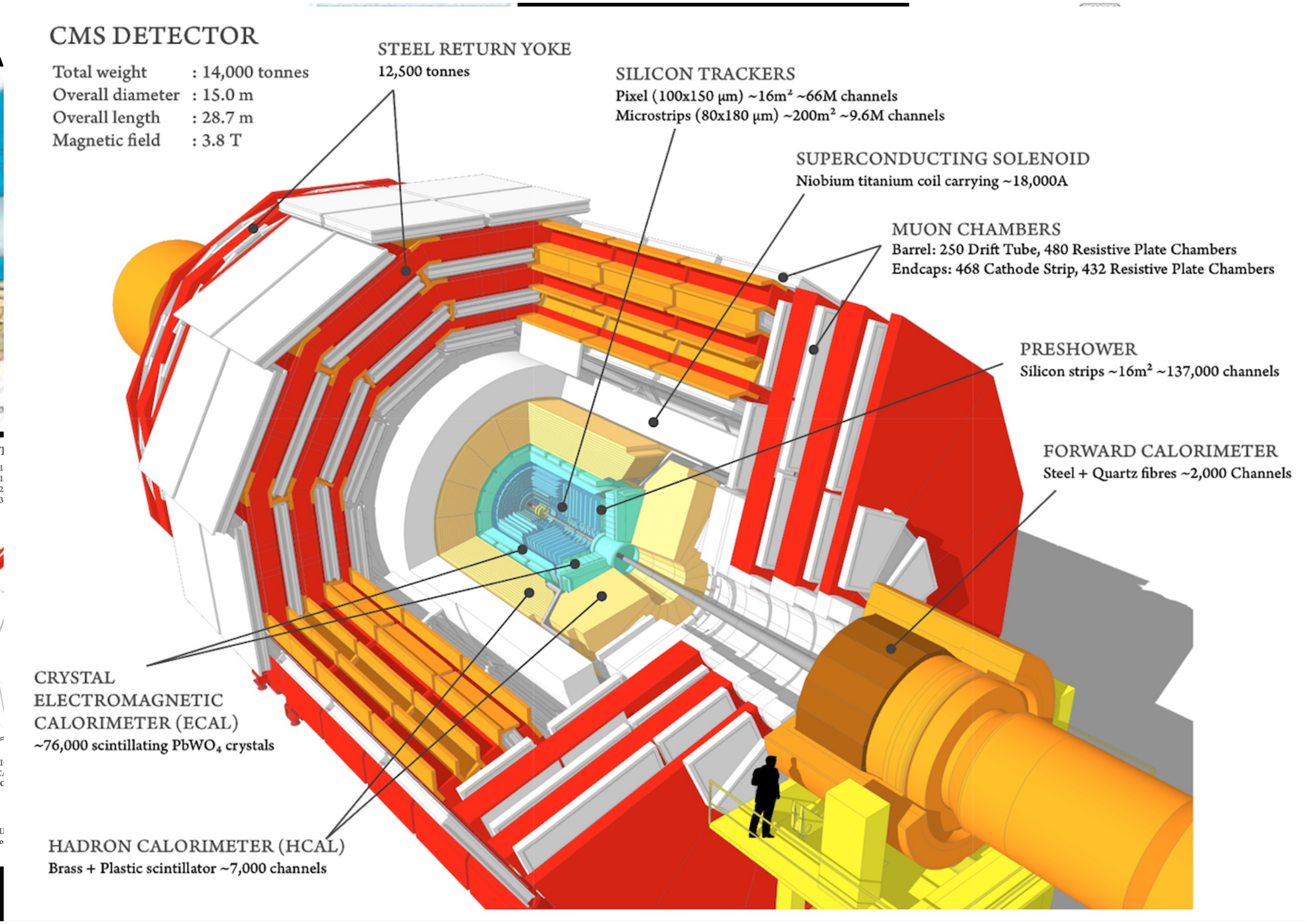
The 4 LHC Experiments

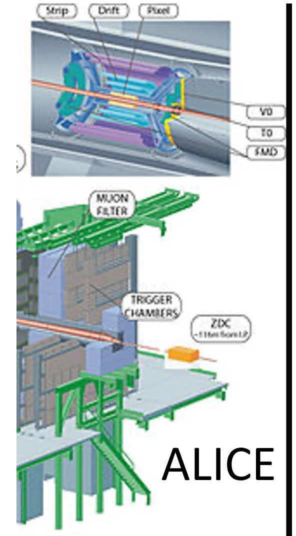


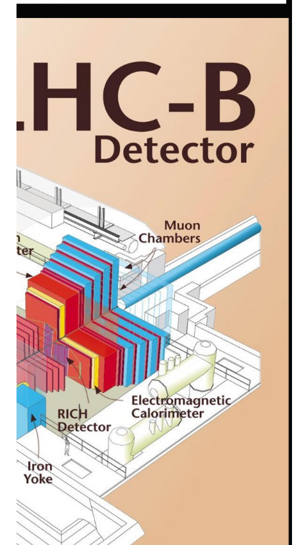





ATLA


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 PbWO4 crystals
HADRON CALORIMETER (HCAL)
Brass + Plastic scintillator ~7,000 channels


Strip
Drift
Pixel
V0
T0
FMD
MUON FILTER
TRIGGER CHAMBERS
ZDC
ALICE


HC-B
Detector
Muon Chambers
RICH Detector
Electromagnetic Calorimeter
Iron Yoke