Community Summer Study

SNOWMASS

July 17-26 2022, Seattle

Seattle Snowmass Summer Meeting 2022

Collaborative Initiatives Across Experiment Boundaries

Liz Sexton-Kennedy

Fermilab

HSF

HEP Software Foundation

Introduction - WLCG

- Reflecting the global nature of HEP, we have a global umbrella computing organization in the World wide LHC Computing Grid
- The WLCG is an MOU-governed computing services organization. Originally (2006) for the LHC but recently has been expanded to service BELL2, DUNE, and Rubin

WLCG
Worldwide LHC Computing Grid

OSG

Data Transfer Video

Introduction - HSF

- Reflecting the global nature of HEP, we have a global umbrella software organization in the HEP Software Foundation, HSF.
- HSF does not itself seek funding, but supports bids to funding agencies for projects that align with the goals documented in the Community White Paper, CWP:

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

HEP Software Foundation Collaboration • Johannes Albrecht (Dortmund U.) Show All(310)

Dec 18, 2017

49 pages
Published in: *Comput.Softw.Big Sci.* 3 (2019) 1, 7
Published: Mar 20, 2019
e-Print: 1712.06982 [physics.comp-ph]
DOI: 10.1007/s41781-018-0018-8 (publication)
Report number: HSF-CWP-2017-01, HSF-CWP-2017-001, FERMILAB-PUB-17-607-CD
View in: OSTI Information Bridge Server, HAL Archives Ouvertes, CERN Document Server, ADS Abstract Service

pdf links cite 141 citations

arXiv:1712.06982v5 [physics.comp-ph]

HSF-DOC-2022-01
May 17, 2022

arXiv:2205.08193v1 [physics.comp-ph] 17 May 2022

The HEP Software Foundation Community

The HEP Software Foundation

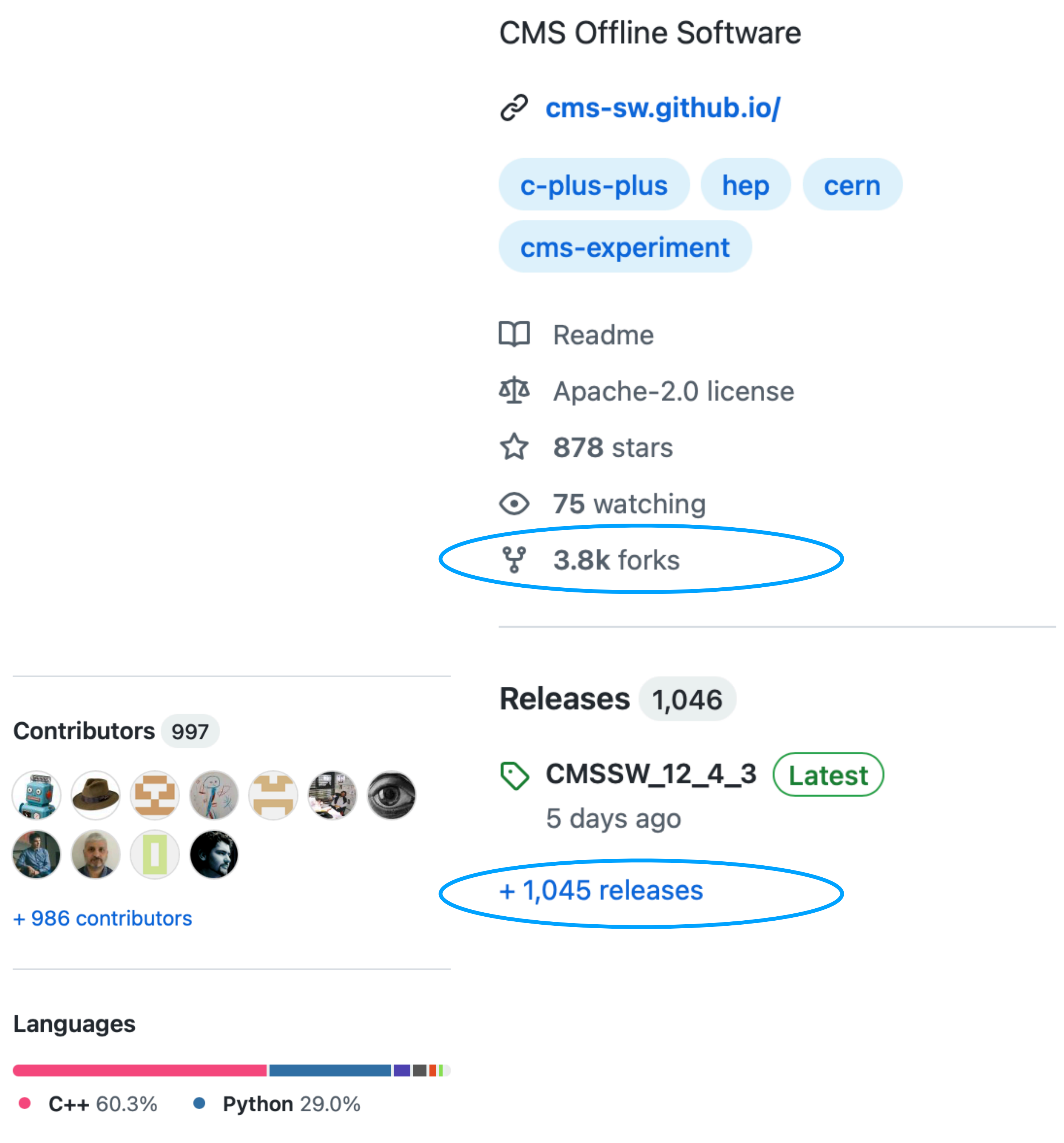
Contact editors:
Graeme A Stewart, CERN (graeme.andrew.stewart@cern.ch)
Peter Elmer, Princeton University (Peter.Elmer@cern.ch)
Elizabeth Sexton-Kennedy, Fermilab (sexton@fnal.gov)

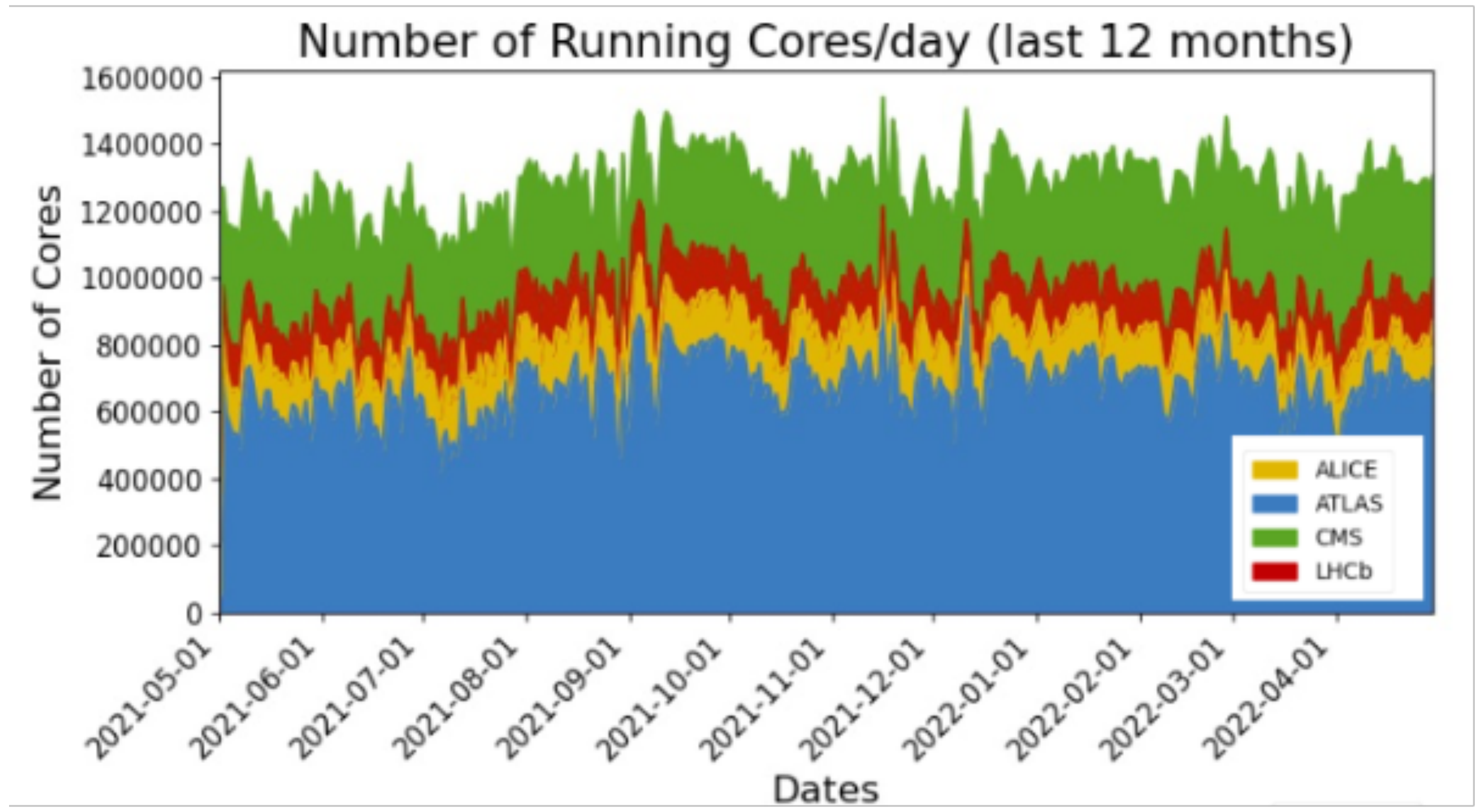
Abstract: The HEP Software Foundation was founded in 2014 to tackle common problems of software development and sustainability for high-energy physics. In this paper we outline the motivation for the founding of the organisation and give a brief history of its development. We describe how the organisation functions today and what challenges remain to be faced in the future.

arXiv:2205.08193v1 [physics.comp-ph]

Introduction - Scale

- High Energy Physics has a vast investment in software
  - Estimated to be around 50M lines of C++
  - Which would cost more than 500M$ to develop commercially
- S&C is a critical part of our physics production pipeline, from experimental design, to data collection (e.g. triggering), to data analysis/interpretation
- LHC experiments use about 1.4M CPU cores every hour of every day, we have around 1000PB of data with 1000PB of data transfers per year (10-100Gb links)
- This is a huge and ongoing cost in hardware and human effort and it should have broader impacts!





The Big Picture in the 10 Year Timescale

- These initiatives are driven by:
  - The physics objectives of our experiments and the timeline of their data-taking
  - Evolution of technologies we use: hardware (including GPUs, FPGAs, and TPUs), operating systems, Grid/Cloud, compilers, standard foundation libraries, and more
  - Resource limitations in both people and services
- Adds up to an overwhelming challenge, hence the need to ban together and eliminate duplicate efforts
  - The established role of the HSF, since 2015, is to facilitate coordination and common efforts in software and computing across HEP in general

Outline

- Who has taken up the call of the CWP?
- What have they proposed to do?
- What are the opportunities?
- How are they coordinating with each other?
- Examples of community collaborative initiatives
- Conclusions

Who?

17 universities

IRIS-HEP, NSF USA - 2018

ErUM-DATA, Helmholtz Institute Germany - 2019

HEP-CCE, DOE USA - 2019

2 hardware projects (HEP&NP)
SciDAC, ECP (ASCR)

170 collaboration members - 2001

7 universities

PATH, NSF USA - 2020

EP R&D, CERN - 2020

19 universities + RAL

SWIFT-HEP, STFC 2020 and
ExCALIBUR-HEP, UKRI UK - 2021

Celeritas - ECP, DOE USA - 2021

37 universities + 8 industry participants

AIDAInnova, European Commission EU -2021

Plus lots of single PI level programs and:
DOE SciDAC
NSF Institutes
DOE Quantum Centers

What?

- OSG-LHC, Data Organization Management & Access, Analysis Systems & Facilities, Facilities R&D, Innovative Algorithms, Training
- Heterogeneous computing and virtualized environments, Machine Learning for reconstruction and simulation
- Portable Parallelization Strategies, I/O Strategy on HPC, Event Generators
- Turnkey software systems, faster simulation, track and calorimeter reconstruction, efficient analysis
- Exascale data management, Event generators, Detector Simulation on GPUs, FPGA tracking for HLT
- GPU offloading for Geant EM showers, and LQCD software/algorithms, Cosmology and Accelerator modeling software
- Turnkey software systems, Track Reconstruction, Particle Flow, ML Simulation

Opportunities

- There is still plenty of scope left as laid out in the CWP, the field needs these efforts to grow.
- Some areas are more well covered then others
- Example: Portability Frameworks are a small part of the scope of the Frameworks topic, yet well covered… Integration with ML frameworks is lacking
- Example: GPU acceleration of Geant is covered; physics models are not
- When there are 2 research groups working on the same deliverable the HSF provides a forum to get them talking to each other.
    - Example is Adept and Celeritas have been fruitfully collaborating

Coordination

- After the CWP many of the groups that collaborated on writing the paper continued on as HSF & WLCG working groups.

WLCG Archival Storage Group

- WLCG Archival Storage Group

WLCG Network Throughput WG

- WLCG Network Throughput WG

WLCG Security Working groups

- WLCG Traceability Working Group
- WLCG Authorisation Working Group
- WLCG Federated Operations Security Working Group

WLCG Containers Working Group

- WLCG Containers Working Group

Kubernetes WG

- Kubernetes

WG for Transition to Tokens and Globus Retirement

- WG for Transition to Tokens and Globus Retirement

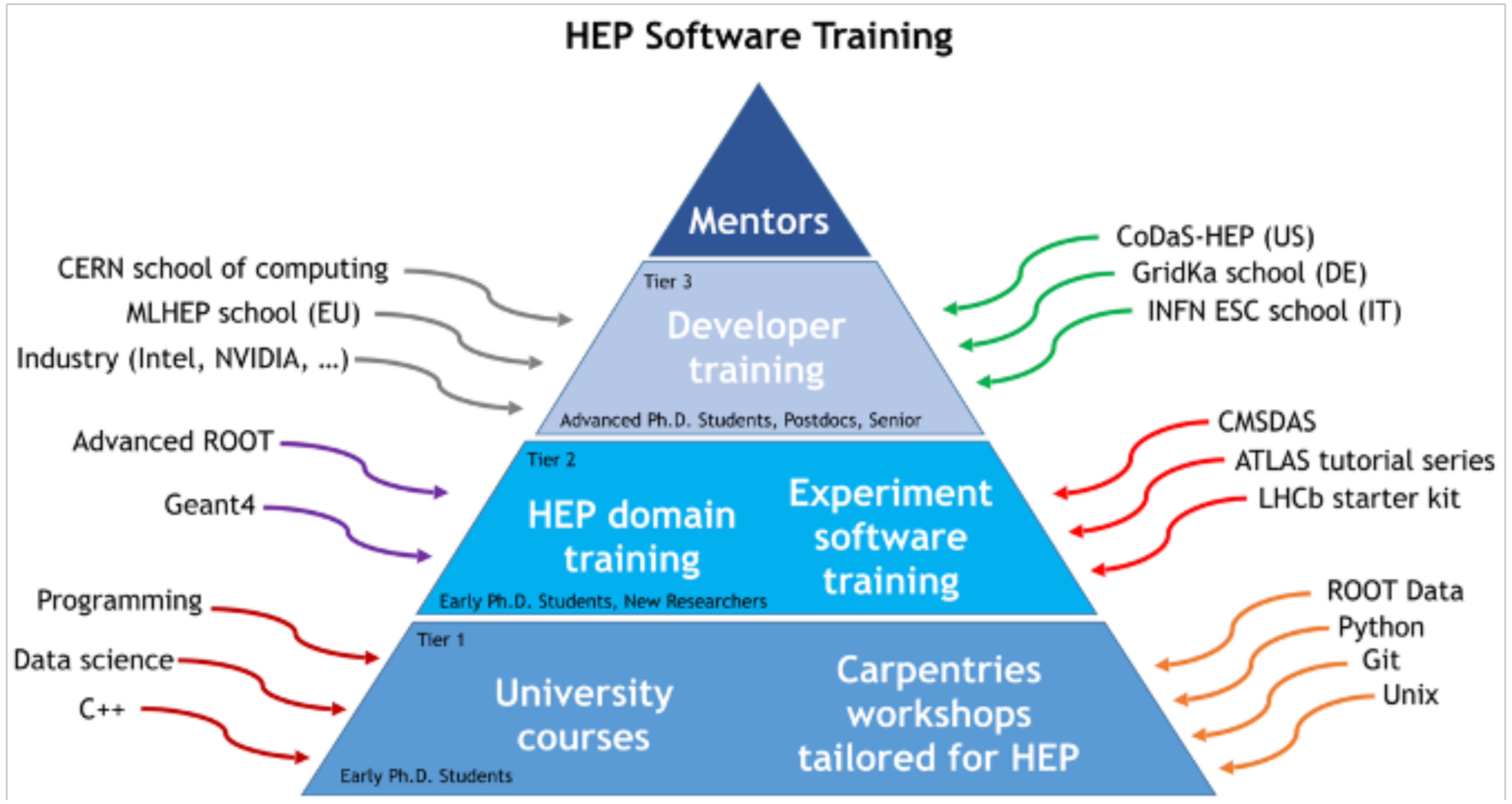
WLCG
Worldwide LHC Computing Grid

HSF
HEP Software Foundation

- Data Analysis
- Detector Simulation
- Frameworks
- Physics Generators
- PyHEP - Python in HEP
- Reconstruction and Software Triggers
- Software Developer Tools and Packaging
- Training
- Differentiable Computing
- Season of Docs
- Google Summer of Code
- intelligent Data Delivery Service
- Licensing
- Quantum Computing
- Reviews
- Visualisation
- Analysis Facilities

Example: Training

- Many new skills are needed for today's software developers and users
- Base of the pyramid has relatively common demands
  - Common software components help in training
- HSF Training Group runs Software Carpentries and other tutorials (co-organised between the HSF IRIS-HEP)
- Highly successful C++ training courses (from SIDIS and HSF)
  - Inspires continued curriculum development and sharing material
- Assembling a complete curriculum for training in HEP, using Carpentries templates
- Paper published on HEP Software Training Challenges



Ex: Event Generators

- Base of all simulation
  - LHC Run-1 leading order generators used and contributed little to overall CPU budgets
- Increasing importance for LHC precision measurements
  - ATLAS and CMS now use higher order generators like MG5_aMC and Sherpa
  - Technical and physics challenges arise particularly from negative event weights
- HSF Working Group formed after the 2018 computing for event generators workshop
  - Active in a number of areas such as understanding costs and the physics impact of event generation choices
  - Raising the issue of generators more widely (LHCC talk, CSBS paper)
  - Involved in porting efforts for running event generation on GPUs

Implementation ($e^+e^- \to \mu^+\mu^-$)	MEs / second Double
1-core MadEvent Fortran scalar	1.50E6 (x1.15)
1-core Standalone C++ scalar	1.31E6 **(x1.00)**
1-core Standalone C++ 128-bit SSE4.2 (x2 doubles, x4 floats)	2.52E6 (x1.9)
1-core Standalone C++ 256-bit AVX2 (x4 doubles, x8 floats)	4.58E6 (x3.5)
1-core Standalone C++ "256-bit" AVX512 (x4 doubles, x8 floats)	4.91E6 (x3.7)
1-core Standalone C++ 512-bit AVX512 (x8 doubles, x16 floats)	3.74E6 (x2.9)
Standalone CUDA NVidia V100S-PCIE-32GB (2560 FP64 cores*)	7.25E8 **(x550)**

Ex: PyHEP Workshops

- Started in 2018 at Sofia CHEP and grew into two very successful virtual workshops in 2020 and 2021
- More that 1300 people registered, demonstrating huge interest in this area driven by
    - Data science and machine learning toolkits
    - Integration with particle physics tools (Coffea, pyhf, PyROOT, Scikit-HEP, SWAN, zfit)
- Trends and hot topics included automatic differentiation
    - An HSF activity area started just before the summer 2021
- Many talks and tutorials done as notebooks
    - Participants could follow live or use them as offline resources
    - Integrated into Binder

- Everything uploaded to the HSF's YouTube channel - supported by Python Software Foundation

Recommendations

European Strategy Update

2020 Strategy Statements

4. Other essential scientific activities for particle physics

Computing and software infrastructure

- There is a need for strong community-wide coordination for computing and software R&D activities, and for the development of common coordinating structures that will promote coherence in these activities, long-term planning and effective means of exploiting synergies with other disciplines and industry
- A significant role for artificial intelligence is emerging in detector design, detector operation, online data processing and data analysis
- Computing and software are profound R&D topics in their own right and are essential to sustain and enhance particle physics research capabilities
- More experts need to be trained to address the essential needs, especially with the increased data volume and complexity in the upcoming HL-LHC era, and will also help in experiments in adjacent fields.

d) Large-scale data-intensive software and computing infrastructures are an essential ingredient to particle physics research programmes. The community faces major challenges in this area, notably with a view to the HL-LHC. As a result, the software and computing models used in particle physics research must evolve to meet the future needs of the field. *The community must vigorously pursue common, coordinated R&D efforts in collaboration with other fields of science and industry to develop software and computing infrastructures that exploit recent advances in information technology and data science. Further development of internal policies on open data and data preservation should be encouraged, and an adequate level of resources invested in their implementation.*

Summary

- Particle physics is an inherently international effort, with an excellent tradition of cooperation in many different S&C domains
- There has been many funding opportunities for cross experiment R&D in the S&C domain.
  - More are needed!
  - Operations programs for S&C can provide sustainability but can also lead to duplication AND small experiments may not have them!
    - Usually cross experiment sustainability and enhancement is not supported (Geant, Event Generators) See recommendation 1 from CompF.
- HSF and WLCG umbrella organizations offer an excellent place to present work discussing successes (and disappointments!), to help avoid duplication, and to lead community workshops & other activities