The HSF landscape

Recent activities & plans

Eduardo Rodrigues - University of Liverpool

On behalf of the HSF coordination team

SWIFT-HEP Workshop, All@Home, 14-15 January 2020

The HSF and SWIFT-HEP share activity areas!

- HSF very happy to be involved with ECHEP / ExCALIBUR and now SWIFT-HEP from the onset !

The goal of the HSF is to facilitate coordination and common efforts in software and computing across HEP in general
- *Our philosophy is bottom up, a.k.a. do-ocracy*

- DISCLAIMER
  - What follows is not an introduction to the HSF
  - Assumed unnecessary for this community at this stage ;-)
  - Rather an overview of HSF community activities in the last year

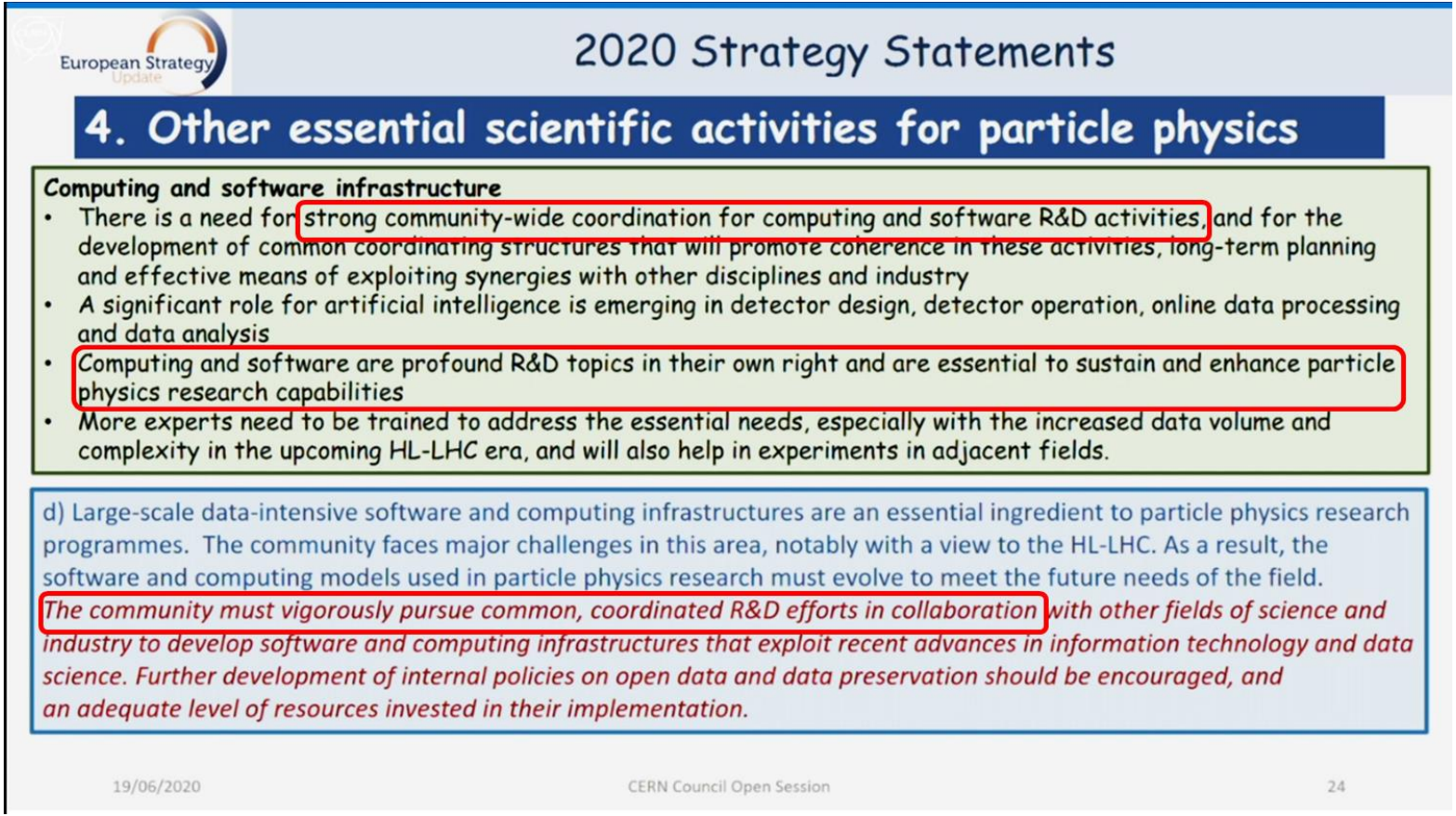
HEP Software & Computing – recap of core challenges

- High Energy Physics has a vast investment in software & computing
  - Demands a *huge* and *ongoing* cost in hardware and human effort
  - Demands ever more software skills and expertise

- As diversity of new architectures will only grow (e.g. see ACM Conclusion), the challenges will only increase (at least in the medium term)
- Computational *power efficiency* is a driver to exploit HPCs (+ pressure from funding bodies)

- Demands in data storage and data transfers are also increasing considerably
  - We are in the exabyte era already

HEP Software & Computing – we are being listened to!



Community building

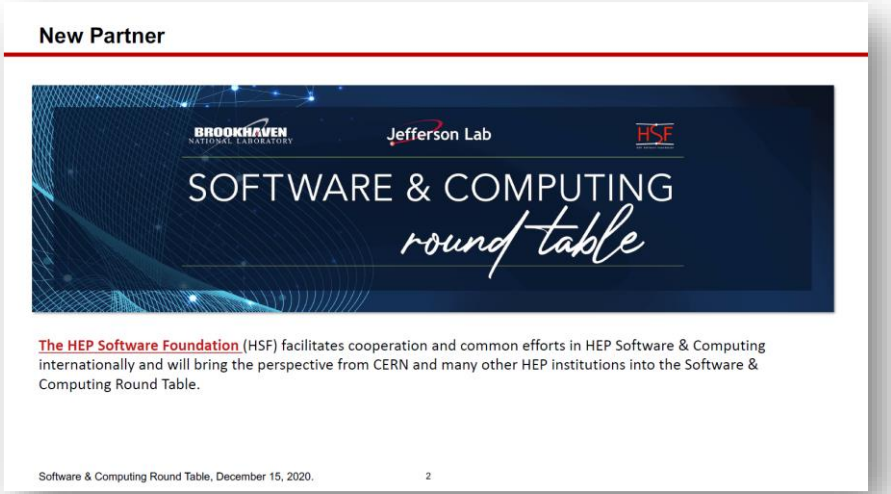
Organisational Engagement

- Not withstanding that the HSF is an organisation of people in HEP and mostly in experiments and often in other projects...
- Who does the HSF engage with?
  - Experiments
  - CERN Openlab & SIDIS (Software Institute for Data Intensive Science)
  - WLCG – HSF has a point of contact
  - Funded R&D projects
    - IRIS-HEP (HSF also member of its steering board)
    - HEP-CCE (Center for Computational Excellence)
    - SWIFT-HEP & ExCALIBUR
    - CERN-EP R&D
  - ESCAPE (future research infrastructures for astronomy and acc.-based particle physics)
  - ECFA and European Particle Physics Strategy Update (https://europeanstrategy.cern/)
  - Snowmass
  - Nuclear physics

Engaging with other communities – Nuclear Physics

- Nuclear Physics SW&C round-table organised since 2016
- Now jointly organised by BNL and JLab *and* the HSF, who joined in late 2020
  - Encourage cross-promotion with and participation in the HSF

- HSF presented this week at the 1st 2021 round-table:
  - General HSF presentation
  - Python and PyHEP
  - Training activities

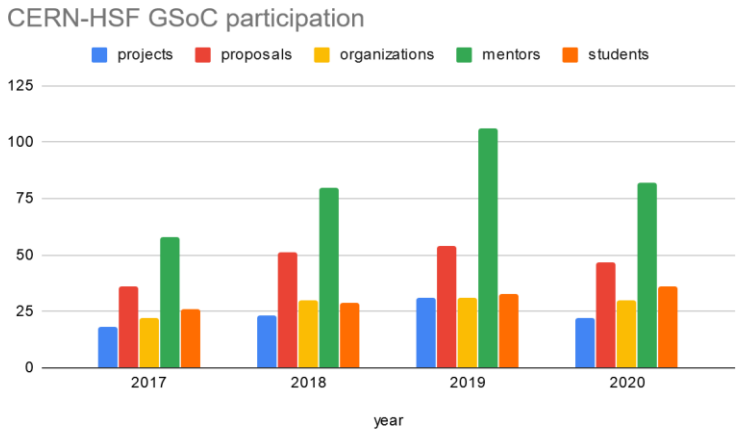


Google Summer of Code & Season of Docs in 2020

- HSF acts as an umbrella organisation for HEP

Developments related to SWIFT-HEP/ExCALIBUR would be very welcome this year

- Google Summer of Code (GSoC)
  - 36 slots awarded from Google
  - 34 students were successful in their projects
  - Google has announced that the coding time awarded in GSoC 2021 will be about *half* of previous years, which is not good
- Google Season of Docs
  - 4 projects in Season of Docs
    - AllPix
    - Rucio
    - ROOT (x2 – general + Python documentation)



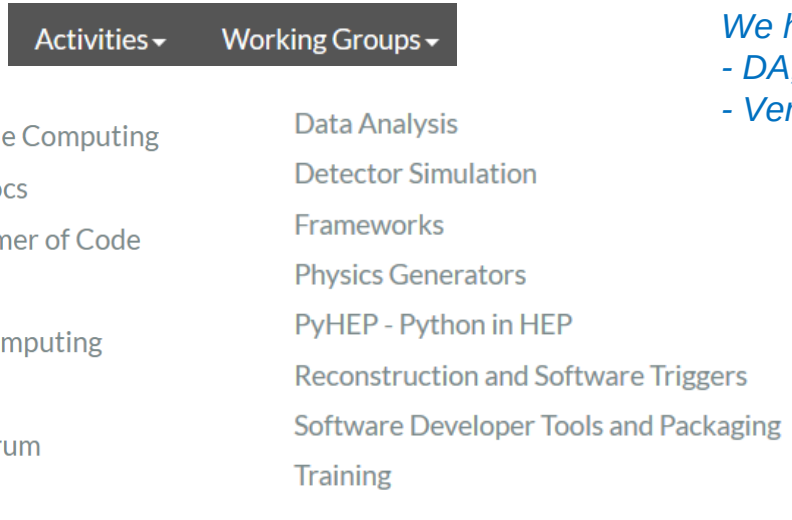
Community White Paper, HL-LHC Review and Software Advocacy...

- Early HSF goal to describe a *global vision for software and computing* for the HL-LHC era and HEP in the 2020s
  - **Community White Paper** published in *Computing and Software for Big Science*, https://doi.org/10.1007/s41781-018-0018-8 (and on arXiv) with a strong community engagement
- Last year we 'updated' some of the CWP chapters, specifically focused on HL-LHC
  - Analysis, Reconstruction, Detector Simulation and Event Generation (SWIFT-HEP themes, BTW)
  - HL-LHC Computing Review: Common Tools and Community Software
- This *review process continues* with the LHCC (Large Hadron Collider Committee), which has an ongoing review of HL-LHC preparations
  - LHCC feedback has been very positive
- HSF also had a significant engagement with the European Strategy Update and with the Snowmass process

HSF Working Groups

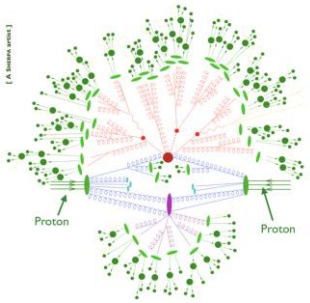
HSF Working Groups and Activity Groups

- Workforce of the HSF & means to engage and exchange with the community
- Groups organise around a particular area of interest
  - These can be topical and cross cutting



We have come a long way, e.g.,
- DA, DS and R&STs were new WGs in 2018
- Very active groups in 2020

Physics Generators



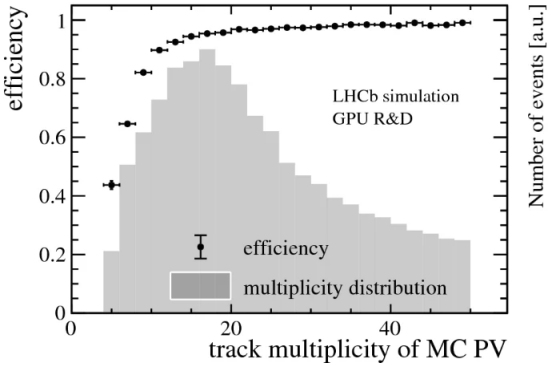
- Base of all simulation
- Increasing importance (CPU budget wise, from viewpoint of software) for LHC precision measurements
  - ATLAS and CMS are relying on higher order generators and more complex calculations more and more
    - E.g., Madgraph and Sherpa
  - Technical and physics challenges arise particularly from negative event weights
- Technical work on leading-order generators and hadronisation, filtering and shower step is important
  - Non-thread safe generators (EvtGen)
  - Technical improvements in, e.g., Pythia and LHAPDF
- Activities (not comprehensive):
  - In a number of areas, such as understanding costs and the physics impact of different event generation choices
  - As well as raising the issue of generators more widely (LHCC Sep.'20 talk and related paper), highlighting proper career incentives
  - Involved in porting efforts for running event generation on GPUs (Madgraph5_aMC@NLO making good progress)
- Several WG meetings and contributions to workshops
- Dedicated session in Nov.'20 HSF-WLCG workshop highlighted progress

Detector Simulation

- Crucial area for experiments as several spend significant fraction of processing power on simulation
- A major consumer of LHC grid resources today
  - Experiments with higher data rates will need more simulated samples
  - Faster simulation, with minimal loss of accuracy, is the goal
- Machine learning lends itself to problems like this
  - Calorimeter simulations usually targeted
  - Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs), etc.
  - Key point is when is it good enough for physics
- Increasingly, GPUs are seen as the real target (some architecture convergence)
  - Opticks for optical photon simulation use in Geant4
  - US Celeritas Project (exascale simulation of HEP for detector modeling)
  - CERN EP-SFT AdePT Project (Accelerated demonstrator of electromagnetic Particle Transport; UK involvement)
- Dedicated session in Nov.'20 HSF-WLCG workshop highlighted progress
  - Good summary of recent R&D work on GPUs
  - Review talks Geant4, ML-based fast simulations and experiment [technical] requirements

Reconstruction and Software Triggers

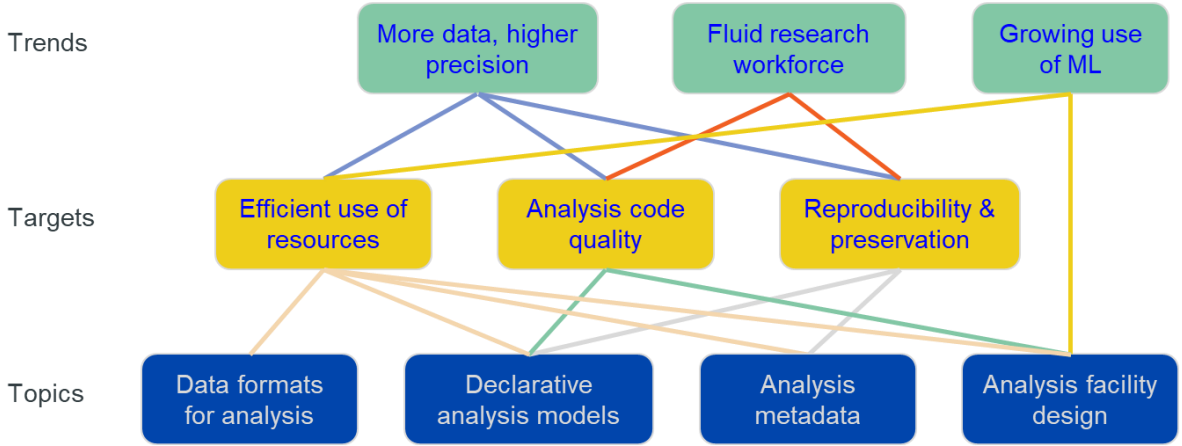
- Hardware triggers no longer sufficient for modern experiments
  - More and more initial reco. needs to happen in software, c.f. ALICE and LHCb strategies for Run 3
- Close to machine, need to deal with tremendous rates & get sufficient discrimination
  - Lots of developments rewriting code for GPUs – physics can get better!
  - Lessons learned: keep data model simple, bulk data,be asynchronous, minimise data transfers
  - High quality reconstruction close to the machine is *Real Time Analysis*
- This work is driving more and more interest in GPUs in HEP
  - Choice of LHCb to use Allen for HLT1 is a boost for this R&D line and a general retooling of HEP software
    - Allen software framework could be of interest to other experiments
- Public GitHub organization for common trigger & reco code
  - Currently hosting code for tracking with graph ML Exa.trkX
- Several WG meetings and contributions to workshops
  - E.g., Nov.'20 joint meeting with Long-Lived Particle Community on use of reco. software in non-LHC experiments
  - Many of those experiments share hardware/software with LHC experiments
    ⇒ showcase return-on-investment on common software development



Allen: A High-Level Trigger on GPUs for LHCb, doi:10.1007/s41781-020-00039-7

Analysis

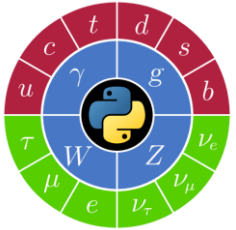
- Aims:
  - Reduce monotonous and laborious tasks in physics analysis
  - Optimise human and computing costs of publishing physics results
- Priorities:
  - Define problems by identifying the needs of physicists and the requirements of analyses across experiments
  - Find solutions by connecting physics analysis experts and technological innovators within and beyond the HEP community
- Development targets:



Analysis

- LHCC review priorities set the themes for 2021 activities, planning in progress
  - Analysis resource usage & computing model
  - Declarative analysis -- interfaces & backends
  - Metadata handling for analysis -- calibration & book-keeping
  - Quality assurance: job curation & code testing
- Aim to increase cooperation with other WGs
  - Training: improve resources for new analysts, introduce tools for effective SW development
  - PyHEP: Python ecosystem offers very interesting tools. Envision a Python stack for analysis?
- Participated in Snowmass Computing Frontier kick-off
- Scaling for analysis-level data is a huge challenge for all LHC experiments (and others)
  - Reducing volume of data needed helps hugely
    - CMS ~1kB nanoAOD makes a vast difference to analysis efficiency and “papers per petabyte”
- Efficient use of analysis data can come with combining many analyses as carriages in a train-like model (pioneered by PHENIX then ALICE)
  - Interest in analysis clusters, specialised for analysis operations over the generic grid resources
- Declarative models more popular than ever
  - E.g., ROOT’s RDataFrame and friends for efficient data representations

PyHEP – Python in HEP



- The “Python in HEP” WG started in early 2018 as an activity group
- It became “formally” a WG last year
- Main activity – the PyHEP workshops
  - There will be a PyHEP 2021
- But topical meetings being organised monthly in 2021

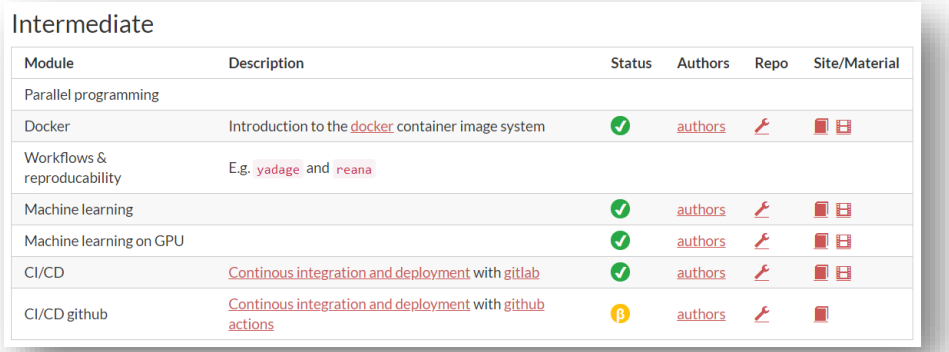
- Interest in these was one of the PyHEP 2020 workshop feedbacks
- Idea came up to follow somewhat the theme of *Python 3 Module of the Week*, but with a spirit adapted to our needs, hence rather a “Python Module of the Month”, presentations with a focus on libraries relevant to data analysis in Particle Physics
  - First tutorial-like presentation will be Feb. 3rd on Numba, see Indico
- Interest from a growing community, with several experiment-agnostic projects:

- Lots of ways to communicate !
  - The main (Gitter) channel now has > 160 people registered

- https://github.com/CoffeaTeam
- https://github.com/FAST-HEP
- https://github.com/root-project/
- https://scikit-hep.org/
- https://github.com/zfit

Training

- "Create a full HEP software curriculum built from standardised modules of hands-on training sessions"
- Impressive work done last year to prepare and deliver training (in spite of COVID-19!)
  - Modules are hosted at https://github.com/hsf-training
- Overview of modules relevant for HEP at https://hepsoftwarefoundation.org/training/curriculum
  - Example:



Intermediate

Module	Description	Status	Authors	Repo	Site/Material
Parallel programming					
Docker	Introduction to the docker container image system	✓	authors		
Workflows & reproducibility	E.g. `yadage` and `reana`				
Machine learning		✓	authors		
Machine learning on GPU		✓	authors		
CI/CD	Continous integration and deployment with gitlab	✓	authors		
CI/CD github	Continous integration and deployment with github actions	β	authors		

- Challenge is to put all of this training on a **sustainable basis**, with suitable rewards for tutors and helpers

Training

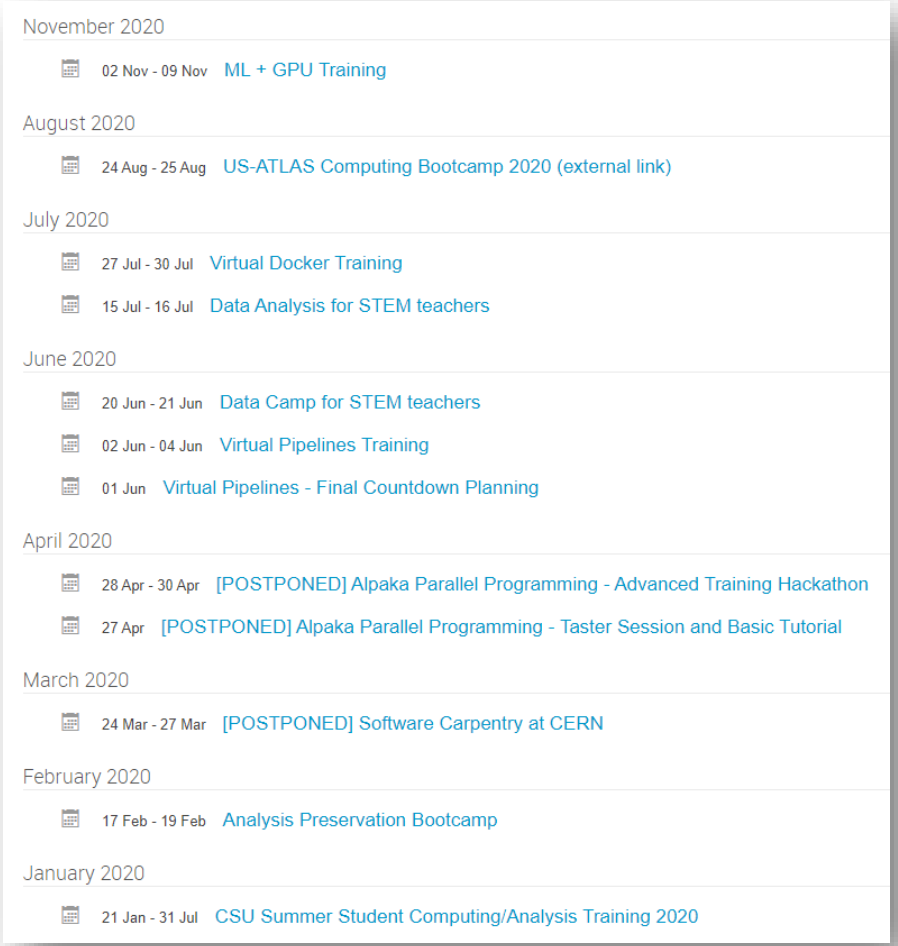
- Lots of training events in 2020 !
  - Help from community as material builder/tutor/helper has been impressive
  - Events often over-subscribed !
- More tutors/helpers always welcome!

- Some trainings organised as series, e.g., Analysis preservation training:

  [1] GitLab Pipelines - June

  [2] Docker - July

  [3] Workflows/REANA - August



HSF Events & Workshops

HSF Events & Workshops – overview

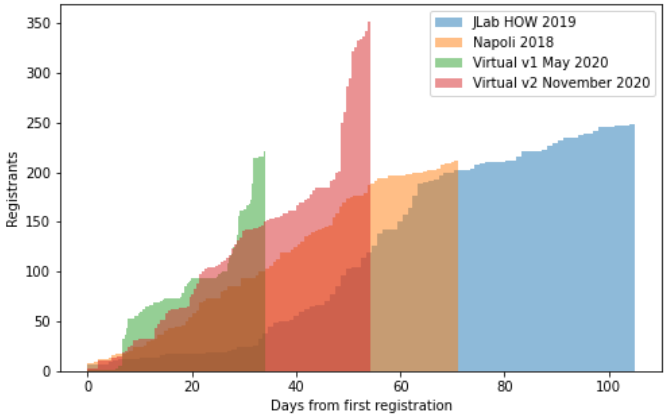
- Weekly coordination meetings
  - See https://indico.cern.ch/category/7970/ for details
- HSF organises workshops and WG meetings
- New – co-organises the Compute Accelerator Forum

Compute Accelerator Forum

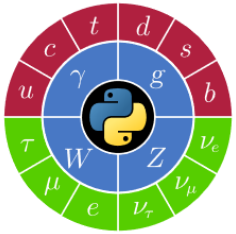
- We have noted in the software development community that there is a move away from pure CPU processing
  - This brings many opportunities, but many new challenges to development teams
- Forum co-organised by the HSF, SIDIS and openlab, who identified the need to discuss *fundamental aspects* of programming on these new devices
  - Foster discussion and teaching
  - Discuss fundamental aspects of software engineering for compute accelerators and heterogeneous computing platforms
- Compute Accelerator Forum started in October last year, with monthly meetings
  - Cross-cutting on this important topic
  - Schedule being populated for this year, https://indico.cern.ch/category/12741/

HSF Workshops in 2020

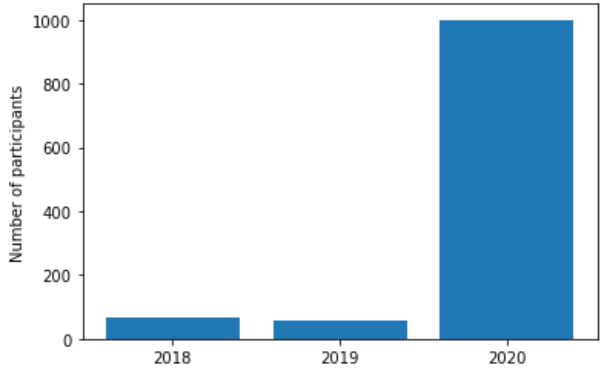
- Our workshops bring together the community to discuss areas of significant interest and, particularly, to encourage cross-talk between people in different areas
  - Commonly organised with WLCG
- In 2020, due to the pandemic, we had to postpone our intended face-to-face workshop in Lund
- May 2020 - New Architectures, Portability, and Sustainability
  - Held jointly with WLCG
- July 2020 – PyHEP 2020 workshop
- November 2020 - HSF Software Focused Workshop
  - Training, Event Generators, Detector Simulation, General R&D Session + NuclPhys Trends Report
- In 2021 we may focus on smaller ‘one shot’ events to combat Zoom-fatigue



PyHEP 2020 workshop



- This 3rd edition was meant to be in the US for the first time, co-locating with the important SciPy 2020 conference
  - We engaged with this very large scientific community
    - Had several talks from HEP colleagues @ SciPy 2020
  - Both had to be organised as virtual events (blame COVID-19)
- PyHEP 2020 (July'20) a truly global event with participants from all over the world (benefit from running virtual)
- Agenda organised in 2 time zones to accommodate Asia, Europe and Americas
  - Remarkable level of interest - *we limited at 1000 registrations!*



Organising Committee

Eduardo Rodrigues - University of Liverpool (Chair)
Ben Krikler - University of Bristol (Co-chair)
Jim Pivarski - Princeton University (Co-chair)
Matthew Feickert - University of Illinois at Urbana-Champaign

Local organisation

Chris Tunnell - Rice University
Peter Onyisi - The University of Texas at Austin

Sponsors

The event is kindly sponsored by

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

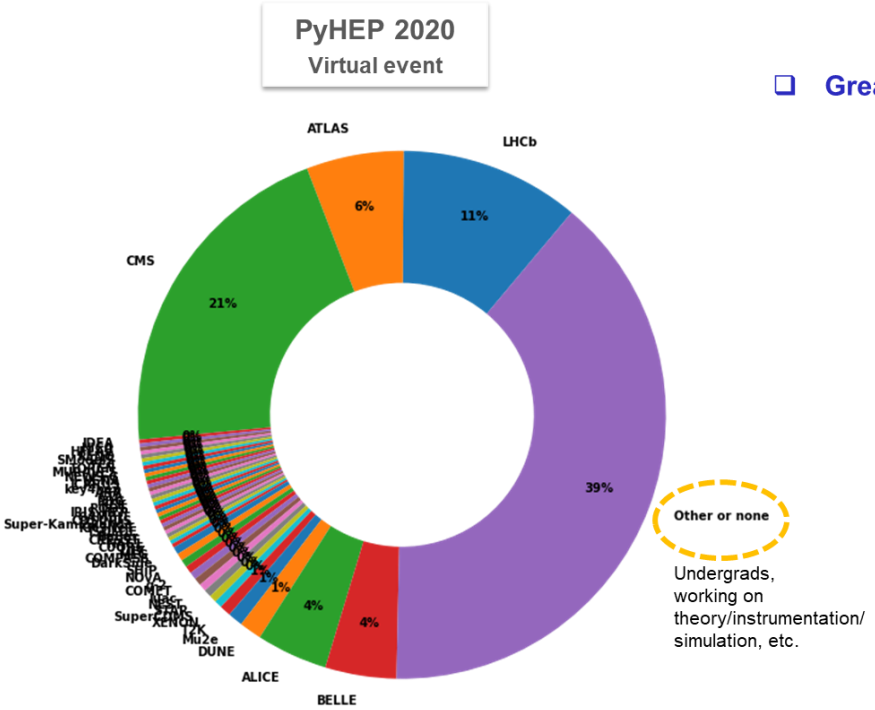
UNIVERSITY OF LIVERPOOL

python SOFTWARE FOUNDATION

Software Sustainability Institute

Fermilab

PyHEP 2020 – diversity and inclusion



- **Great to see such a diverse set of participants !**

(Pie chart and “logo art” with information taken from the pre-workshop questionnaire)

PyHEP 2020 – organisational aspects overview

- **Sessions & presentations**
  - Spread in sessions for “Atlantic”- and “Pacific”-friendly time zones
  - We strongly encouraged notebook presentations, available in public Github repositories with a Binder launch button
  - All presentational material posted on workshop agenda and later given a DOI with Zenodo, in a dedicated “pyhep2020 community” – formal citation, replaces proceedings
  - All talks got recorded, captioned and later uploaded to the HSF YouTube channel – dedicated playlist “PyHEP 2020 Workshop”

- **Zoom video conferencing system**
  - With capacity for 1000 participants
  - Public room but PIN provided via email

- **Slack channels**
  - Various channels:
    - By topic, mapping to sessions, discussions encouraged here
    - Announcements, for actual announcements
    - Random, used to encourage community spirit and add social context

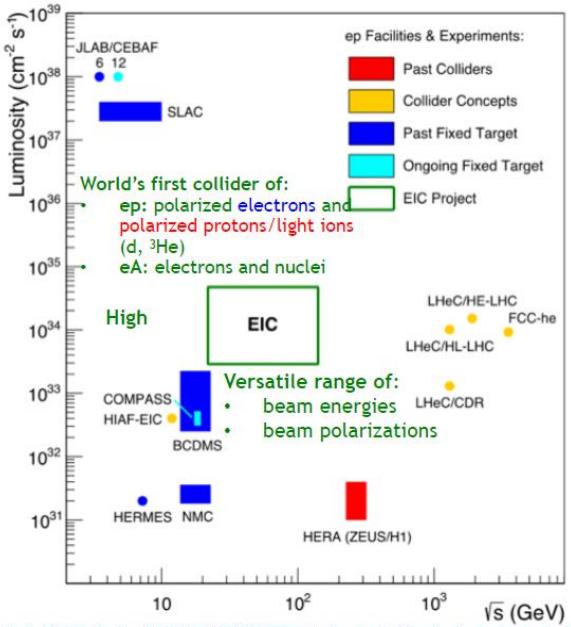
- **Questions & answers with slido**
  - Used *slido* to crowd-source questions, to prioritise the most popular ones upvoted by participants
  - Session chair shares link to questions at end of presentation
  - Most popular ones get answered/discussed
  - At end of Q&A all questions are copied to Slack in the appropriate topical channel
    ⇒ participants can continue to discuss and exchange
  - A few polls also run via slido

- **Communication also on**

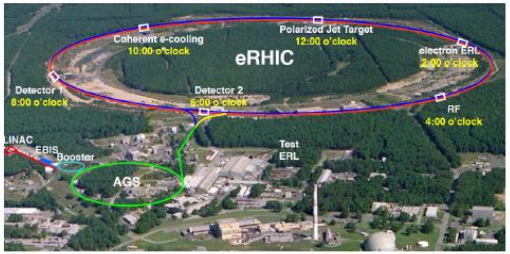
Thank you !

The Nuclear Physics Frontier

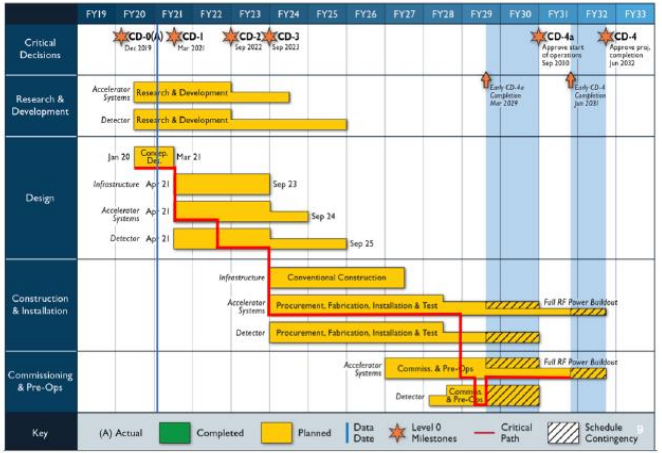
The mission of the Nuclear Physics program in the U.S. is to discover, explore, and understand all forms of nuclear matter.



The Electron-Ion Collider will operate at BNL with JLAB as a major partner in realizing the project



In the meantime, a robust NP program continues at BNL (sPHENIX, STAR) and at JLAB (12 GeV) with its own evolving computing challenges



Ofer Rind, Workshop "Future trends in Nuclear Physics computing" summary, GDB Meeting, Dec. 9th, 2020

Software Tools, Packaging and Licensing

- Copyright and Licensing
  - Long neglected inside collaborations
  - Essential to be able to
    - Open source our software
    - Combine with other open source projects
- Copyright
  - Keep as low a number as practicable
  - E.g. © CERN for the benefit of collaboration X
- License
  - Favour liberal licenses for industry collaboration: LGPL, Apache, MIT
  - Avoid GPL for libraries you want non-GPL projects to use
- Software Tools
  - Active group promoting best practice for correctness and performance
- Packaging
  - We don't build our experiment software in isolation
  - Need a software stack, incorporating many components from the open source world and HEP community
  - Preference for tools that are not home grown and have a wider support base
  - Spack actively being prototyped (link to Key4hep project in EP R&D)

[HSF-TN-2016-01; HSF-TN-2016-03; HSF-TN-2020-01]

HL-LHC Computing Review

LHCC commissioned review by HSF: “Common Tools and Community Software”

Analysis highlights:

- Analysis data formats -- centralised production, disk costs, data access patterns, systematic uncertainties
- Metadata handling -- bookkeeping analysed data (does processing 100% of data scale to HL-LHC?), validity & retrieval of calibrations, cross-sections, ...
- Quality assurance -- code testing for accuracy & efficiency
- Analysis interfaces -- declarative configuration, transparency, preservation