The HEP Software Foundation

Graeme Stewart, for the HSF

Geant4 Collaboration Meeting, 2020-09-21

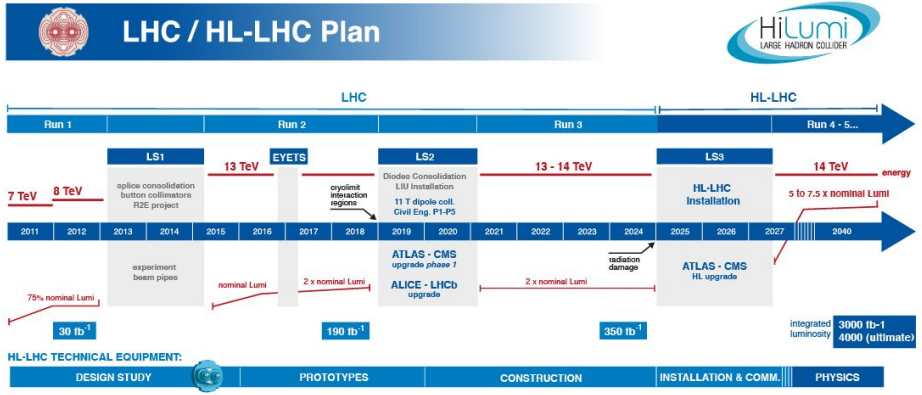
HL-LHC, the Intensity Frontier, and beyond

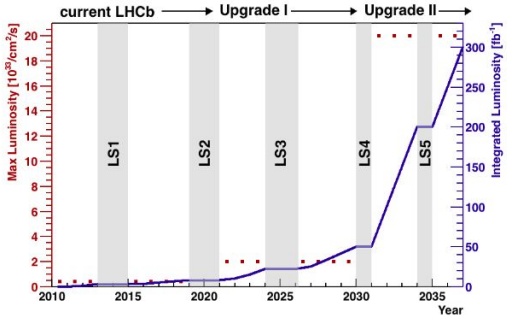
Our mission:

- Exploit the Higgs for SM and BSM physics
- b, c, tau physics to study BSM and matter/antimatter
- Dark matter
- QGP in heavy ion collisions
- Neutrino oscillations and mass
- Explore the unknown

Our Tools:

- (HL-)LHC, DUNE, Belle II
- Future Colliders: ILC, FCC

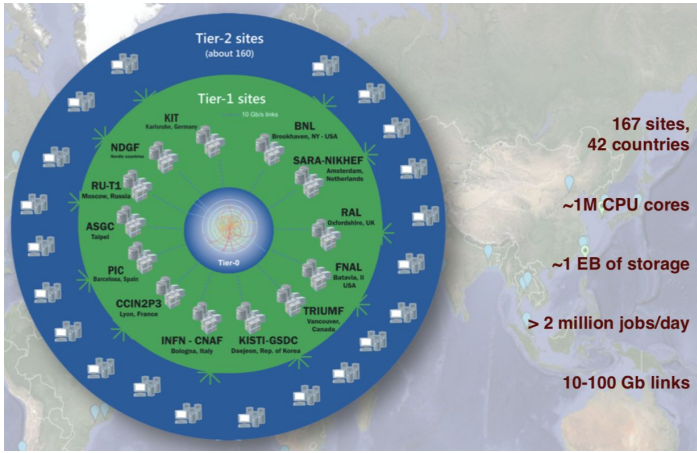




LHCb luminosity
arXiv:1808.08865

HEP Software and Computing

- 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
- It is a critical part of our physics production pipeline, from triggering all the way to analysis and final plots as well as simulation
- LHC experiments use about 1M CPU cores every hour of every day, we have around 1000PB of data with 100PB of data transfers per year (10-100Gb links)
  - We are in the exabyte era already
- This is a *huge* and *ongoing* cost in hardware and human effort
- With significant challenges ahead of us to support our ongoing physics programme



ROOT
Data Analysis Framework

athena
ATLAS Experiment main repository for Athena

Geant 4

DD4hep

Gaudi
Project ID: 38 | Request Access

a:ts

Unstar 39 Fork 58

7,007 Commits 36 Branches 459 Tags

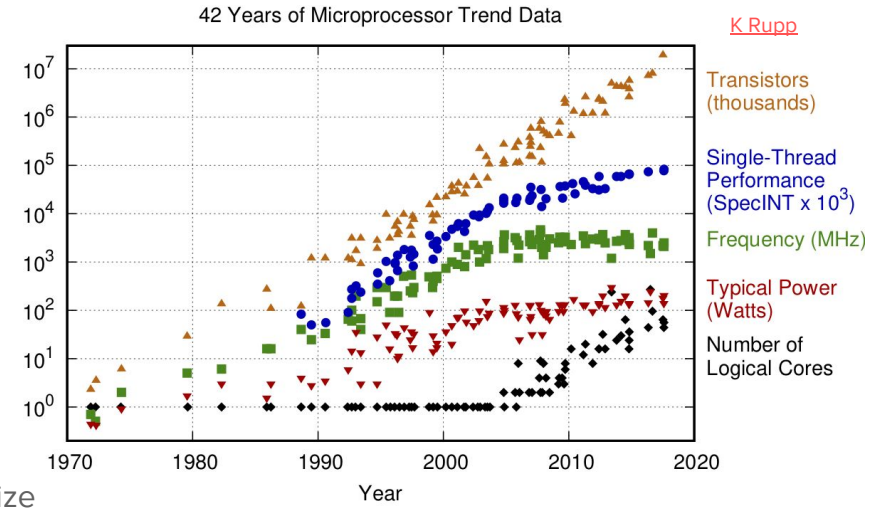
46.3 MB Files 10.9 GB Storage 26 Releases

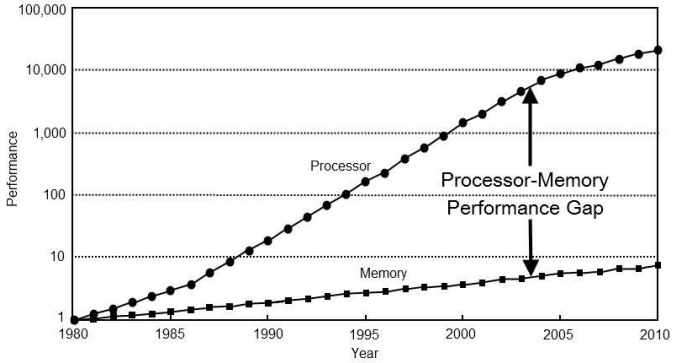
CMS and CMSSW

Scikit-HEP

Technology Evolution

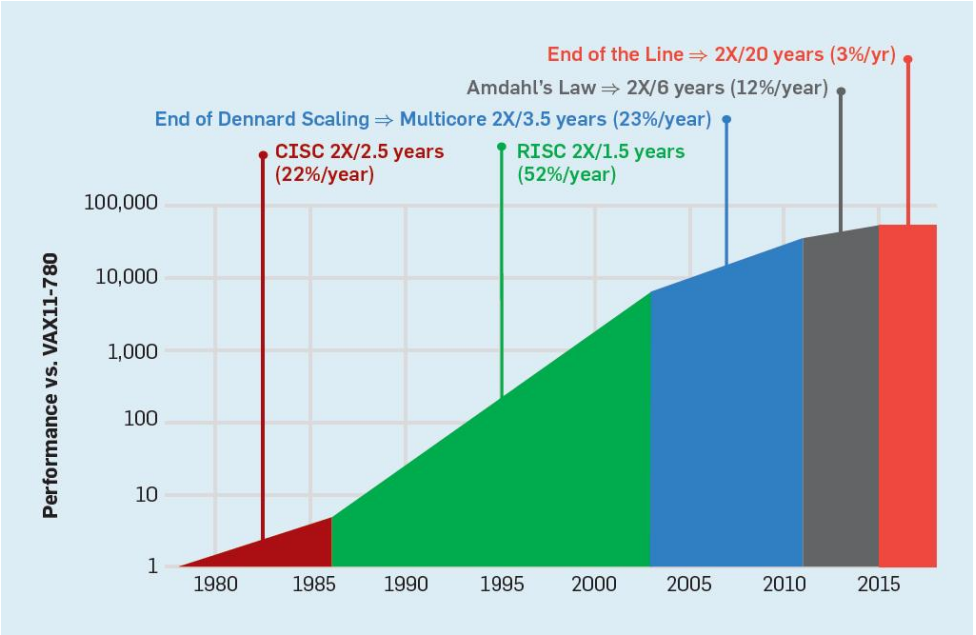
- Moore's Law continues to deliver increases in transistor density
  - But, doubling time is lengthening
- Clock speed scaling failed around 2006
  - No longer possible to ramp the clock speed as process size shrinks
  - Leak currents become important source of power consumption
- So we are basically stuck at ~3GHz clocks from the underlying Wm^{-2} limit
  - This is the *Power Wall*
  - Limits the capabilities of serial processing
- Memory access times are now ~100s of clock cycles

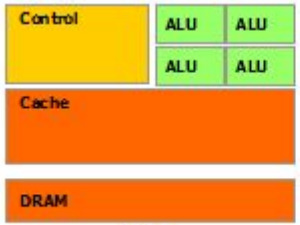




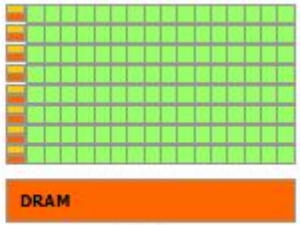
Decreasing Returns over Time

- ACM Conclusion: diversity of new architectures will only grow
- Best known example is of GPUs
  - Also FPGAs, TPUs, ...
- *A64FX ARM CPU recently took the #1 crown by fixing the memory latency issue*



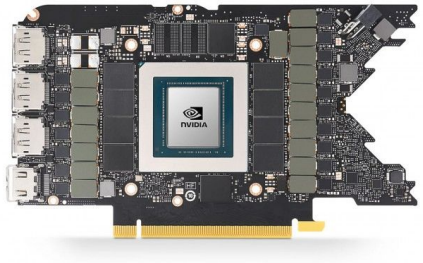


CPU



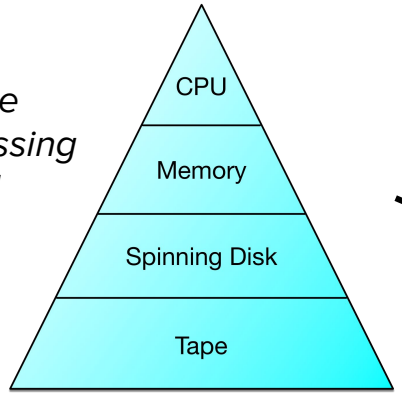
GPU

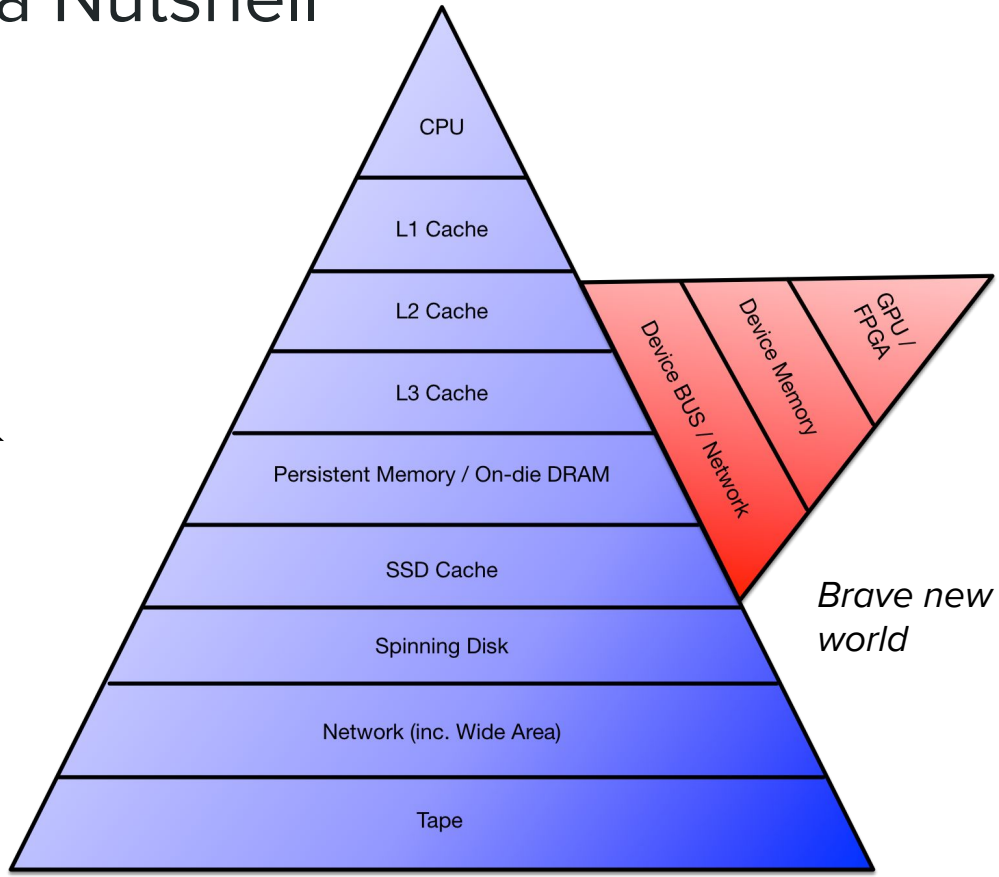
GPUs dedicate far more transistors to arithmetic



Hardware Evolution in a Nutshell

Ye olde processing model





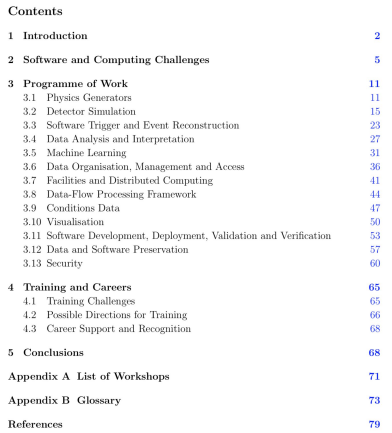
Brave new world

"We're approaching the limits of computer power – we need new programmers now"
John Naughton, Guardian

HEP Software Foundation (HSF)

- The LHC experiments, Belle II and DUNE face the same challenges
  - HEP software must evolve to meet these challenges
  - Need to exploit all the expertise available, inside and outside our community, for parallelisation
  - New approaches needed to overcome limitations in today's code
- Cannot afford any more duplicated efforts
  - Each experiment has its own solution for almost everything (framework, reconstruction algorithms, ...)
  - New experiments should not be starting from scratch, but building on best-of-breed
- HSF started with a number of workshops and working groups on common topics (packaging, licensing)
- 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*

HSF Vision: Community White Paper, European Strategy, Snowmass and HL-LHC


Contents

- We wanted to describe a **global vision for software and computing** for the HL-LHC era and HEP in the 2020s
  - This was the *Community White Paper* with 310 authors from 124 institutes, 14 chapters
  - Published in *Computing and Software for Big Science*, https://doi.org/10.1007/s41781-018-0018-8 (and on arXiv)
- We have prepared additional input for European Strategy Update (talk), LHCC review of HL-LHC and US Snowmass process
- We were engaged in both **projecting a voice on the importance of software** to our field and in **building a community** committed to the open and collaborative development
- Geant4 colleagues have been involved in this process from the beginning

Thank you!

HSF Organisation

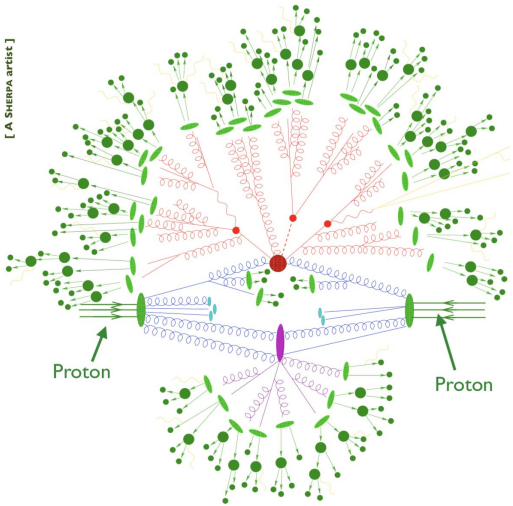
- As a do-ocractic inspired organisation we try to have as lightweight as possible structures to support activities
- Coordination Team for oversight and driving overall engagement, organising workshops
  - Modest sized group of motivated individuals who contribute to general running of HSF
  - Ex-officio members from experiments and WLCG as stakeholders
- Working Groups for key areas of HEP activity
  - Event generation, detector simulation, reconstruction, analysis, frameworks, tools and packaging, education and training, Python in HEP
- The HSF's role here is one of an information conduit and meeting point
  - Report on interesting and common work being done
  - Forum for technical comments and discussion
  - Encourage cooperation across experiments and regions
  - Sustainable training for software developers

Laying Good Foundations for Open Source Projects

- Best Practices for Open Source Projects
  - We wrote a guide for how to best setup and run open source science projects
  - Practical matters are those concerning
    - Copyright and Licensing
    - Contribution guide and code of conduct
    - Continuous integration and managing pull requests
  - One of the most important lessons is *how to build a community* around your project
- Software Tools and Packaging
  - Active group promoting best practice for correctness and performance
  - We don't build our experiment software in isolation
  - Need a software stack, incorporating many components from the open source world and HEP community
    - Actively developed in EP R&D Key4hep project and SFT SPI project
  - There are many points of best practice here that projects should follow!

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

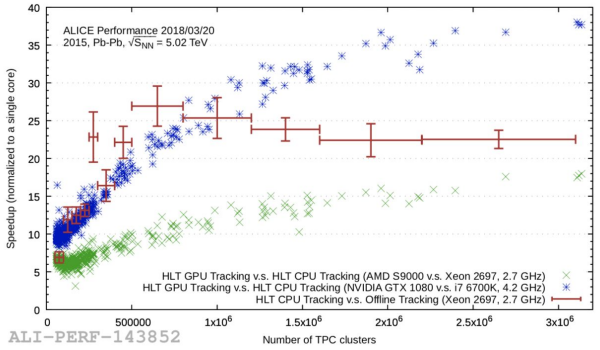
Event Generators



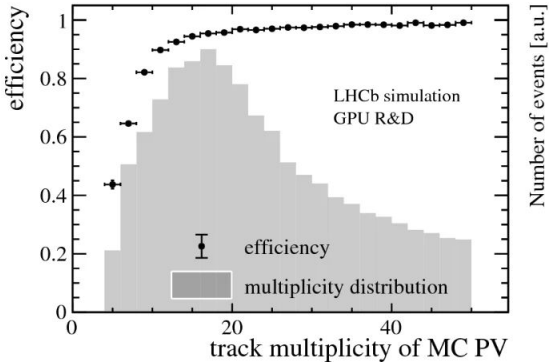
- Base of all simulation
  - In Run-1 leading order generators and little contribution to overall CPU budgets
- Increasing importance for LHC precision measurements
  - ATLAS and CMS use will use higher order generators like Madgraph and Sherpa
  - Technical and physics challenges arise particularly 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 different event generation choices
  - As well as raising the issue of generators more widely (LHCC talk and related paper), highlighting proper career incentives
  - Involved in porting efforts for running event generation on GPUs
    - Madgraph making good progress

Reconstruction and Software Triggers

- Hardware triggers no longer sufficient for many modern experiments
  - More and more initial reconstruction needs to happen in software
  - LHCb and ALICE strategies for Run3 embody this par excellence
- Close to the machine, need to deal with tremendous rates and get sufficient discrimination
  - Pressure to break with legacy code is high
  - Lots of developments rewriting code for GPUs
    - Physics can get better!
  - Lessons learned: keep data model simple, bulk data, be asynchronous, minimise data transfers
- This work is driving more and more interest in GPUs in HEP
  - ALICE have made effective use of GPUs since Run-2
  - Choice of LHCb to use Allen for HLT1 is a boost for this R&D line and a general retooling of HEP software
- Reconstructing at sufficient accuracy close to the detector brings the promise of *Real Time Analysis*



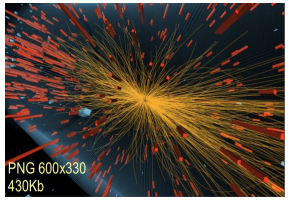
arXiv:1905.05515

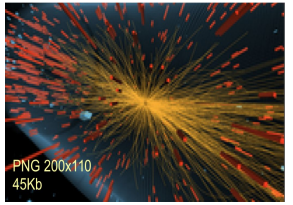


Allen: A High-Level Trigger on GPUs for LHCb, doi:10.1007/s41781-020-00039-7

Analysis and PyHEP

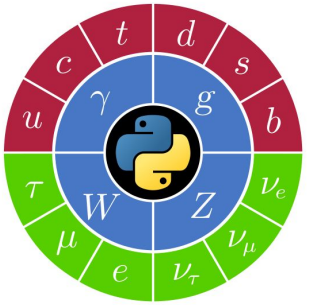
- Scaling for analysis level data also a huge challenge for all LHC experiments
- Data processing trains reduce staging costs and reducing the data volume (CMS nanoAOD ~1kB) helps enormously
  - Interest in analysis clusters, specialised for analysis operations over the generic grid resources (WLCG/HSF pre-CHEP workshop 2-3 November 2019)
    - Columnar analysis and declarative models
- Python more and more considered the language of choice here
  - HSF PyHEP working group attracted 1000 registrants to the 2020 workshop
  - Data science and machine learning toolkits with integration with particle physics tools (Coffea, pyhf, PyROOT, Scikit-HEP, SWAN, zfit)
  - Videos on YouTube



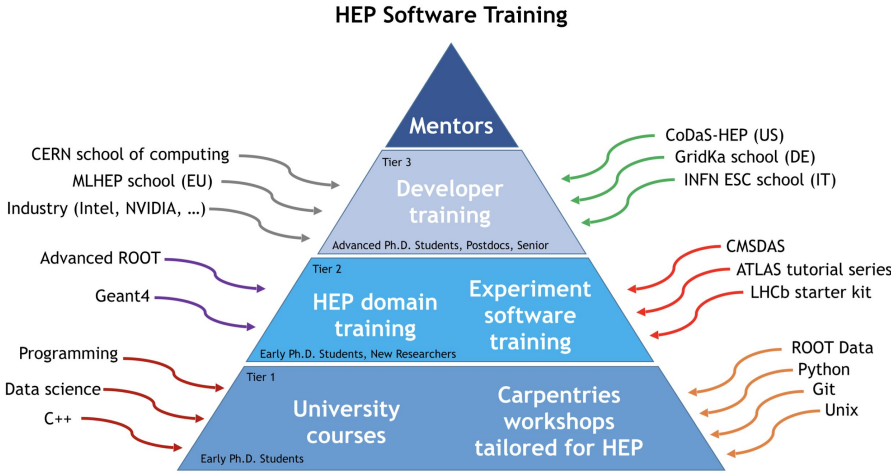


PyHEP 2020 Workshop

32 videos • 631 views • Last updated on 19 Jul 2020



Training



- Many new skills are needed for today's software developers and users
- Base has relatively uniform demands
  - Any common components help us
- HSF Training Group ran a Software Carpentries tutorial at CERN last year (co-organised between the HSF IRIS-HEP)
  - COVID adapted remote training in Virtual Pipelines and Docker also ran this year
- Now assembling a complete curriculum for training in HEP, using Carpentries templates

Intermediate

Module	Description	Status	Authors	Repo	Site/Material
Parallel programming					
Docker	Introduction to the docker container image system	✓	M. Feickert		
Workflows & reproducibility	E.g. `yadage` and `reana`				
Machine learning			L. Polson		
CI/CD	Continous integration and deployment with gitlab	✓	G. Stark		

Training and Careers

- New areas of challenge concurrency, accelerators, data science, ...
- Here we want to gather the knowledge that the community has, reach out to experts in software engineering, and to other data intensive sciences
- Software Institute for Data-Intensive Science
  - Alpaka hackathon organised by HSF, openab and SIDIS.
  - Need to foster new C++ expertise in a modern style - thanks to Sebastien Ponce for bolstering our training efforts here
- New Compute Accelerator Forum
  - Kick off meeting next month (1 October)
  - Fundamental aspects of programming compute accelerators and heterogeneous systems
    - Introduce basic concepts, infrastructure and tools
    - Discuss advanced topics of compute accelerators
- Careers area for HEP software experts is an area of great concern
  - Need a functioning career path that retains skills and rewards passing them on...

And last but not least… Detector Simulation!

- Future experiment needs call for a massive increase in our simulation capacity
  - Match the data rates expected from high luminosities and event rates
  - So future simulation will have to get *faster*, but still give us the necessary *physics accuracy*
- Three pronged approach
  - Technical improvements to Geant4 as the LHC simulation workhorse
    - GeantV R&D showed there are gains to be made here
    - Accurate measurements of detailed performance are the starting point
  - Improve fast simulation
    - Parametric approaches
    - Machine learning (not as easy as some of us hoped!)
    - Hope to be able to improve generality and integration of approach
  - R&D into compute accelerators for particle transport simulation
    - This brings unique challenges, due to the stochastic nature of the problem, so it will be hard
    - But with the spread of GPUs and other devices it's essential R&D for the field

HSF Directions and International Efforts

- Particle physics is in inherently international effort, with an excellent tradition of cooperation in many different domains
  - Detector R&D, Experiments, WLCG, Common Software
- HEP Software Foundation tries to foster a shared vision
  - This encourages diverse R&D!
  - Success in attracting funding to this area helps reinvigorate the developer community with young people
  - Recognised links to other main players (WLCG, LHCC, IRIS-HEP, EPPSU, Experiments, Snowmass Process)
    - And we want to work with Geant4!
- Next HSF Workshop will be a virtual event in November (19-20 + 23-24)
  - One session we will dedicate to Detector Simulation R&D, co-organised with Geant4 colleagues

Backup

Getting Involved with HSF...

- Join the HSF Forum, hsf-forum@gmail.com
  - Few messages a week with updates, jobs, items of interest
  - Owned by the community - please just post items of relevance
- Join a working group, https://hepsoftwarefoundation.org/what_are_WGs.html
  - Follow the group's meetings and discussions
  - Suggest a meeting topic
- Indico Main Page
- Annual meetings
  - Established tradition of a joint meeting with WLCG each Year
  - Next one (virtual) in November
- Propose a new activity area
  - The HSF is there to help gather interest

- Data Analysis
- Detector Simulation
- Frameworks
- Physics Generators
- PyHEP - Python in HEP
- Reconstruction and Software Triggers
- Software Developer Tools and Packaging
- Training