HEP Computing Challenges and the HSF

Graeme Stewart, for the HSF

ILCX Workshop, 2021-10-28

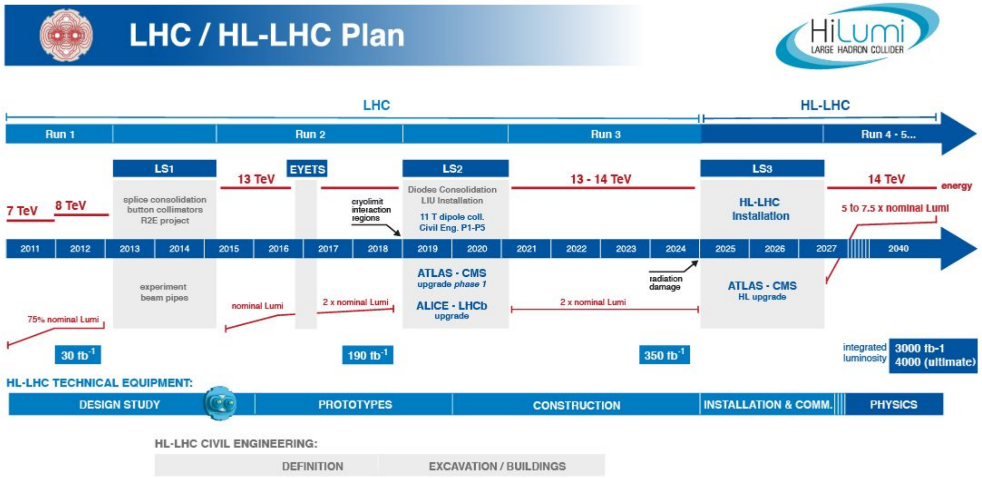
HL-LHC, the Intensity Frontier, and beyond

Our mission:

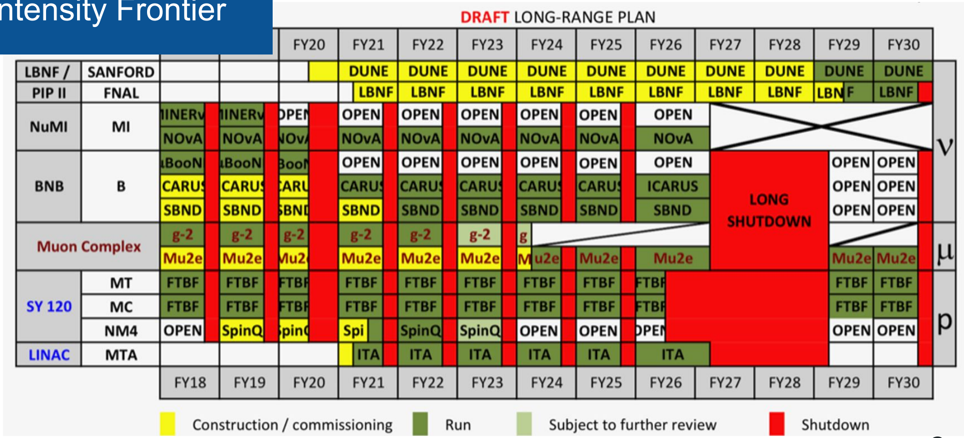
- Exploit the Higgs for SM and BSM physics
- b, c, tau physics to study BSM and matter/antimatter
- Dark matter
- QGP in heavy ion collisions
- Neutrino oscillations and mass
- Explore the unknown

Our Tools:

- [present] (HL-)LHC, DUNE, Belle II
- [longer-term] ILC, FCC, CEPC, BEPC
- also connected to HSF:
  - smaller LHC-adjacent experiments (e.g. FASER)
  - nuclear physics experiments (e.g. FAIR, EIC)

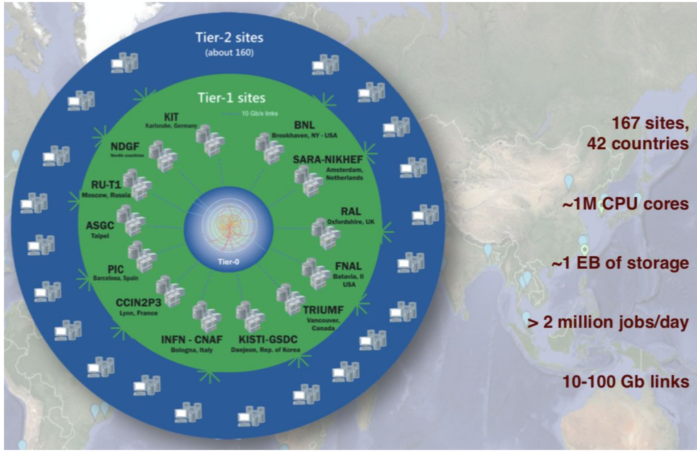


FNAL Intensity Frontier



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 1000PB 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

iLCSoft DD4hep acts

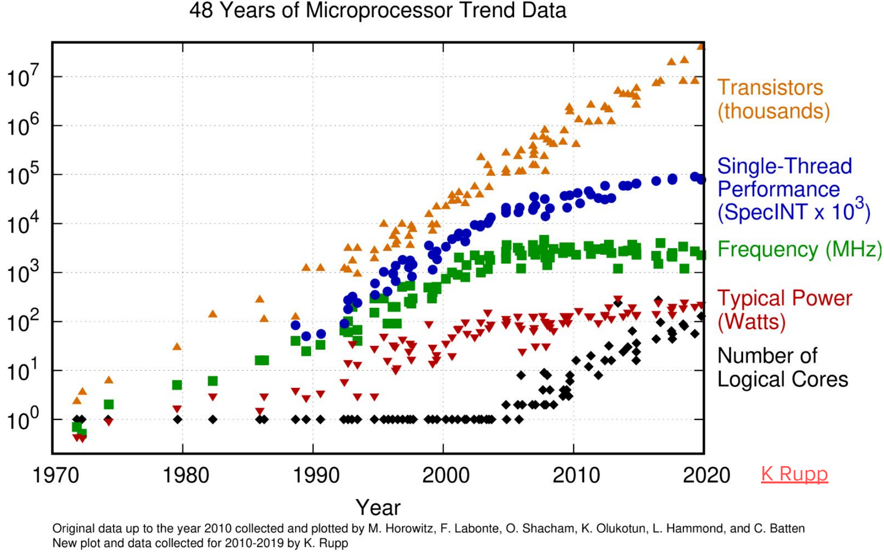
HEP.TrkX GEANT4 A SIMULATION TOOLKIT

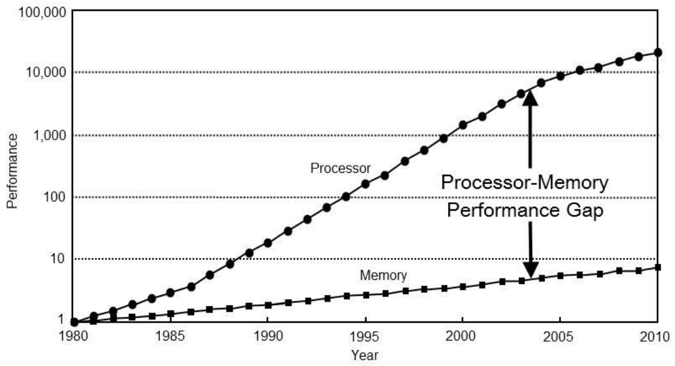
Scikit HEP

KEY4HEP

Technology Evolution

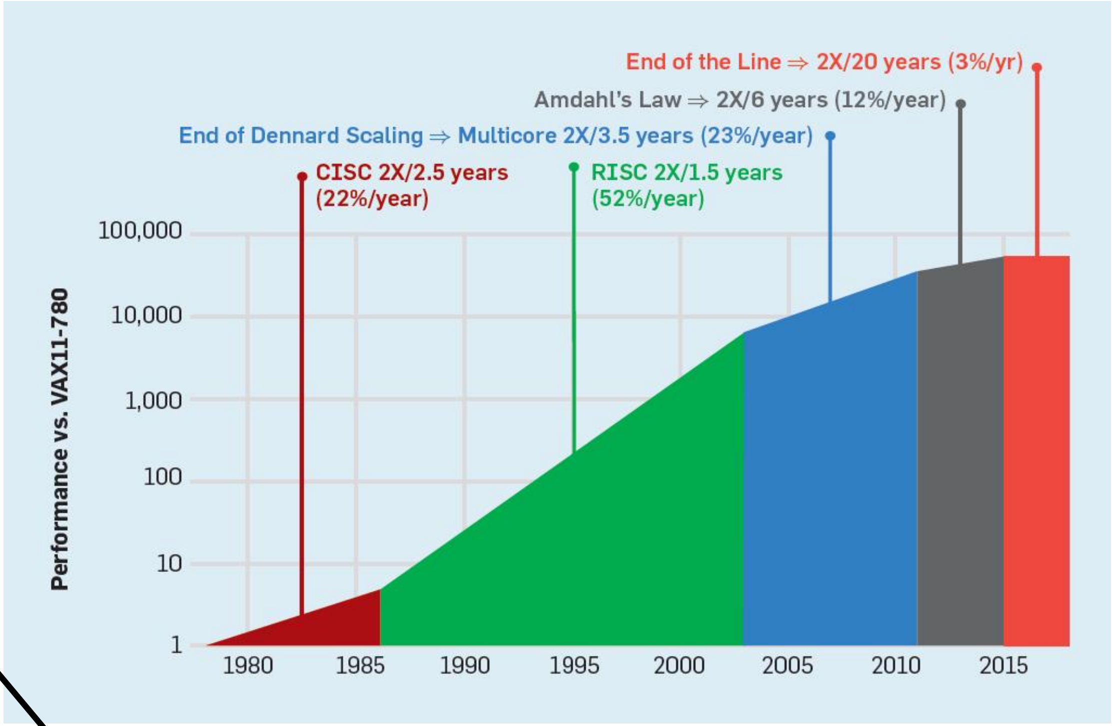
- Moore's Law continues to deliver increases in transistor density
  - But, doubling time is lengthening
- Clock speed scaling failed around 2006
  - No longer possible to ramp the clock speed as process size shrinks
  - Leak currents become important source of power consumption
- So we are basically 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

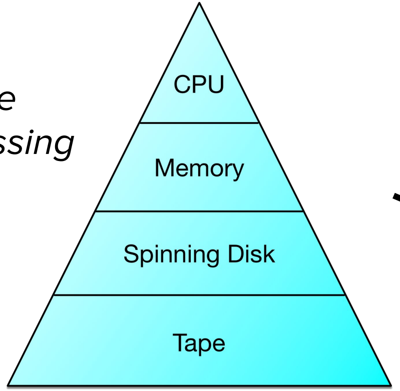
- Conclusion: diversity of new architectures will only grow
- Best known example is of GPUs
- Also FPGAs, TPUs
- As well as non-trivial innovations for CPUs
    - Apple M1
    - Fujitsu A64FX
    - Google Tensor

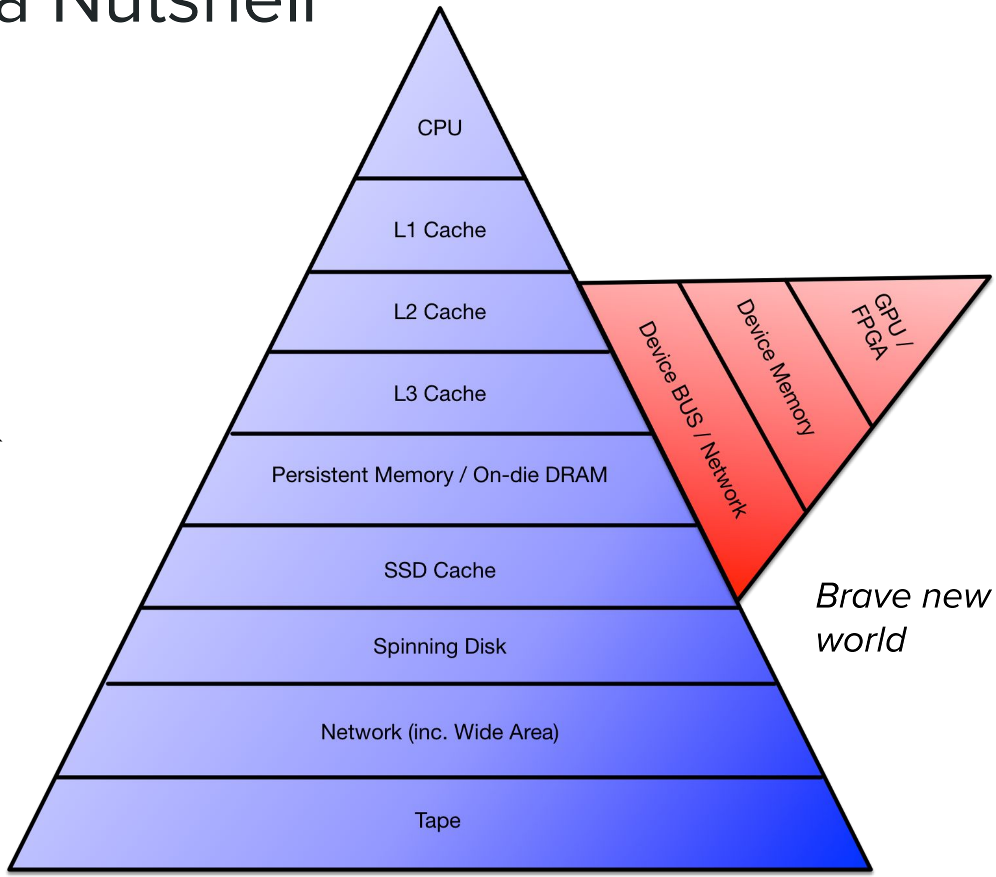


GPUs dedicate far more transistors to arithmetic

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, DUNE and future experiments 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
- The established role of the HSF, since 2015, 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*

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)
  - Community engagement: 310 authors from 124 institutes, 14 chapters
- Last year we ‘updated’ some of the CWP chapters, specifically focused on HL-LHC
  - Analysis, Reconstruction, Detector Simulation and Event Generation
  - HL-LHC Computing Review: Common Tools and Community Software
- This review process continues with the LHCC (Large Hadron Collider Committee) which has on ongoing review of HL-LHC preparations and in which the HSF had a coordinating role
- HSF also had a significant engagement with European Strategy Update and with Snowmass process

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.*

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

Software and Computing International R&D Projects

Lobbying pays off! HSF does not itself seek funding, but supports bids to funding agencies. Many of these projects received a letter of support or collaboration from the HSF:

- IRIS-HEP, NSF USA
  - Analysis systems, innovative algorithms, DOMA
- ErUM-DATA, Helmholtz Institute DE
  - Heterogeneous computing and virtualized environments, machine learning for reconstruction and simulation
- EP R&D, CERN
  - Turnkey software systems, faster simulation, track and calo reconstruction, efficient analysis
- HEP-CCE, DOE USA
  - Portable Parallelization Strategies, I/O Strategy on HPC, Event generators
- AIDAInnova, European Commission EU
  - Turnkey software, track reconstruction, particle flow, ML simulation
- SWIFT-HEP, STFC and ExCALIBUR-HEP, UKRI UK
  - Exascale data management, Event generators, detector simulation on GPUs, FPGA tracking for HLT

```
int main {
  cout << “write software” << endl;
  return 0;
}
```

Key4hep

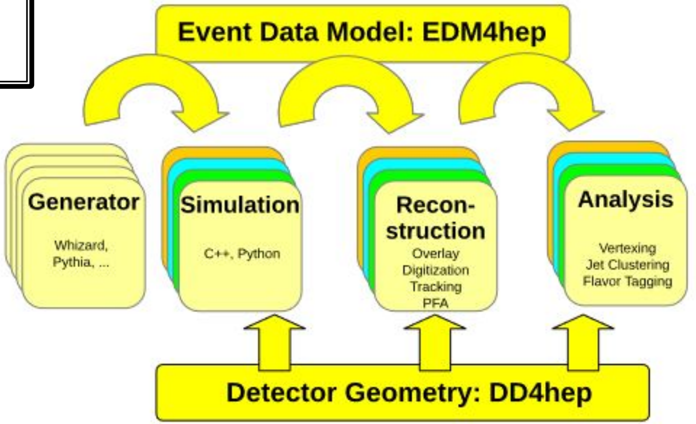
More details in Key4hep and EDM4hep talks earlier today

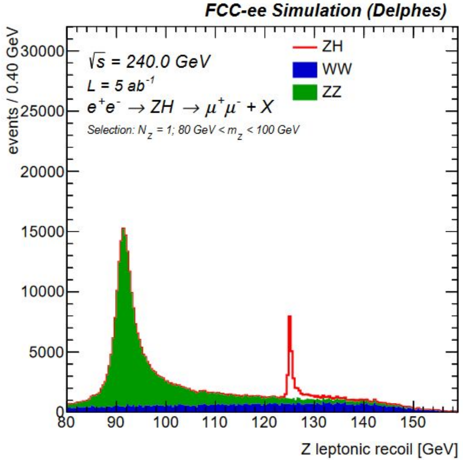
- HEP software stacks are wide and deep - many dependencies
- Want to be able to run full chains for detector design studies easily and in a validated setup
- Ingredients
  - Event data model, EDM4hep based on LCIO and FCC-EDM
  - DD4hep for geometry
  - Gaudi event processing framework (with *Marlin wrappers*)
  - Packaged and deployed using Spack
  - Fast (Delphes) and full (Geant4) simulation available
- Contributions from ILC, CLIC, FCC and CEPC Communities
  - ILC community is investigating a smooth transition to Key4hep (keeping existing tools/algorithms working)

Allied with activities like HSF packaging group

- Feature and performance evaluations favoured Spack (LLNL) as a solution

And discussions on best practices for copyright and licensing





courtesy of C. Helsens

Event Generators

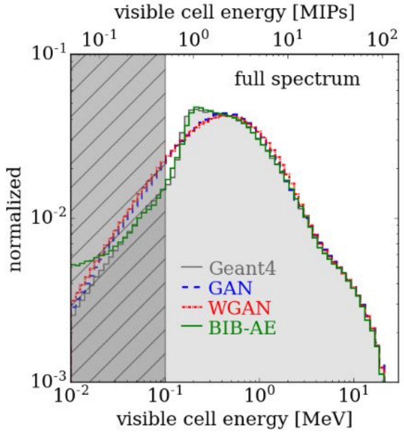
- Base of all simulation
  - LHC Run-1 leading order generators and little contribution to overall CPU budgets
- Increasing importance for LHC precision measurements
  - ATLAS and CMS now use higher order generators like Madgraph and Sherpa
  - Technical and physics challenges arise particularly from 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, CSBS paper)
  - Involved in porting efforts for running event generation on GPUs (Madgraph making good progress)

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)**

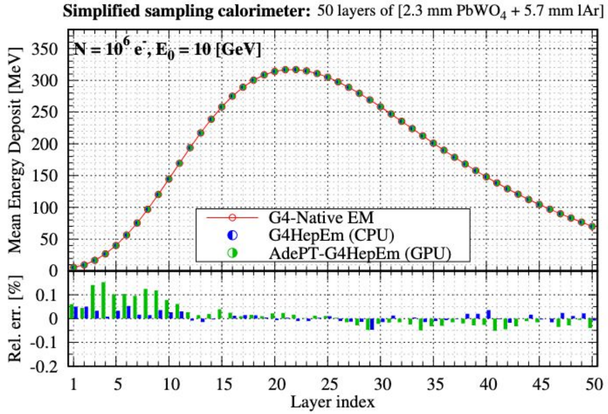
A Valassi et al.

Detector Simulation

Performance of different ML architectures for photons in ILC Calorimeter, S Diefenbacher et al.



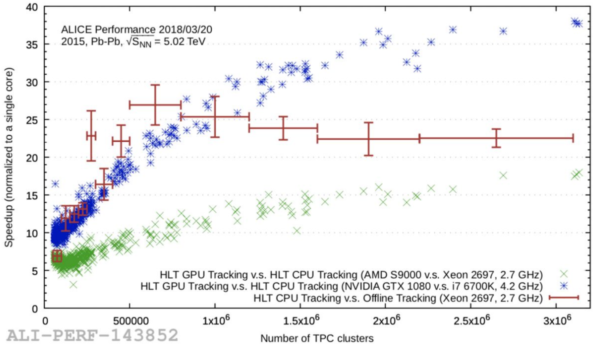
- A major consumer of LHC grid resources today
  - Experiments with higher data rates will need *more simulation*
- Faster simulation, with minimal loss of accuracy, is the goal
  - Range of techniques have been used for a long time (frozen showers, parametric response)
  - Key point is deciding when it's good enough for physics
- Machine learning lends itself to problems like this
  - Calorimeter simulations usually targeted
  - Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs)
    - This is probably *not as easy as we thought* - traditional parametric approaches are hard to beat
- Particle tracking on GPUs is very challenging, but work has started
  - AdePT demonstrator (CERN EP-SFT and SWIFT-HEP)
  - Celeritas (DOE)
  - Opticks, using ray tracing on GPUs for optical photons



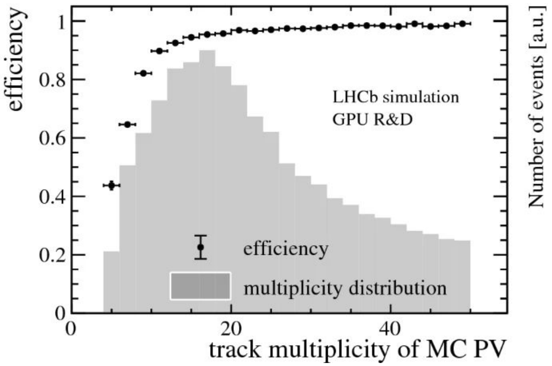
New GPU friendly EM Physics library integrated into the AdePT prototype (M Novak, J Hanfeld) - per mil agreement with Geant4

Reconstruction and Software Triggers

- Hardware triggers no longer sufficient for modern experiments
  - More and more initial reconstruction needs to happen in software
- 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 pioneered this in Run-2 - other LHC experiments also active
  - Choice of LHCb to use Allen for HLT1 is a boost for this R&D line and a general retooling of HEP software

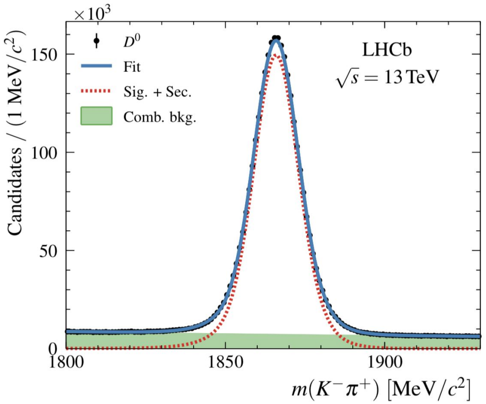


arXiv:1905.05515



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

Reconstruction and Software Triggers

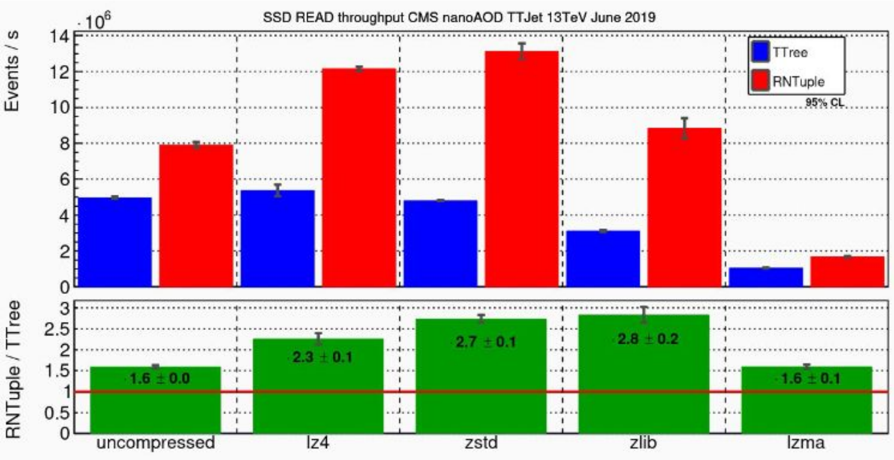


LHCb charm physics analysis using Turbo Stream (arXiv:1510.01707)

- Real Time Analysis (HEP Version)
  - Design a system that can produce analysis useful outputs as part of the trigger decision
    - If this captures the most useful information from the event, can dispense with raw information
  - *This is a way to fit more physics into the budget*
- LHCb Turbo Stream was a radical way to do this and its use in Run3 is a vindication of the approach (see backup)
- Whole ALICE data reduction scheme is based around keeping ‘useful’ parts of events (no more binary trigger)
  - O2 ➔ Online/Offline Data Reduction Farm
- ATLAS and CMS have schemes under development for special handling of samples for which full raw data is unaffordable (e.g., low P_T di-jets)
- For future detectors handling timing information becomes more and more important

Analysis

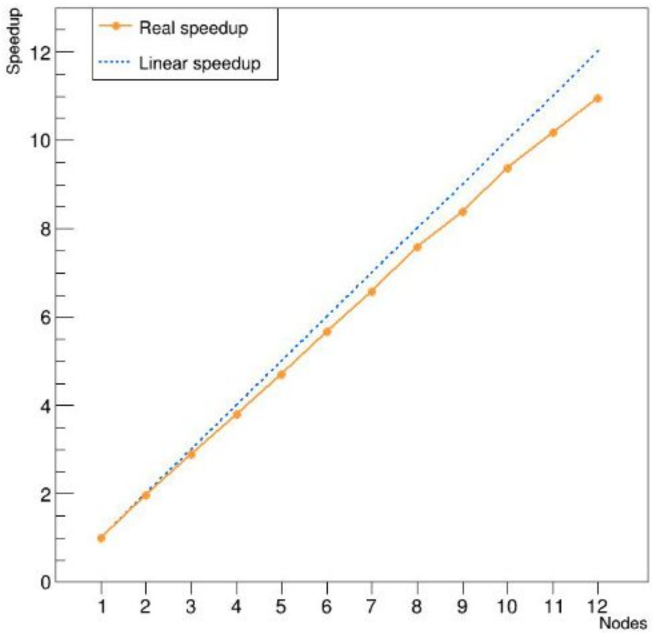
- Scaling for analysis level data also a huge challenge for experiments
- Efficient use of analysis data can come with combining many analyses as carriages in a train like model (PHENIX, ALICE, ATLAS)
- Reducing volume of data needed helps hugely
  - CMS ~1kB nanoAOD makes a vast difference to analysis efficiency and "papers per petabyte"
- Re-inventing data formats for modern devices is a key piece of re-engineering by ROOT to scale up in speed (and down in size!)
  - Also adapt for object stores



Comparison of RNTuple vs. TTree on CMS NanoAOD, J Blomer

Analysis

- Declarative models: RDataFrame (and others)
  - Say what, not how and let the backend optimise
  - E.g. split and merge, GPU execution, cluster-wide distribution
- Analysis facilities may offer specialist solutions to the different working point of analysis vs. other workflows
  - E.g. coffea-casa prototype with columnar backend
- Standard candle analysis benchmarks help compare approaches and Analysis Grand Challenges (IRIS-HEP) test solutions end-to-end

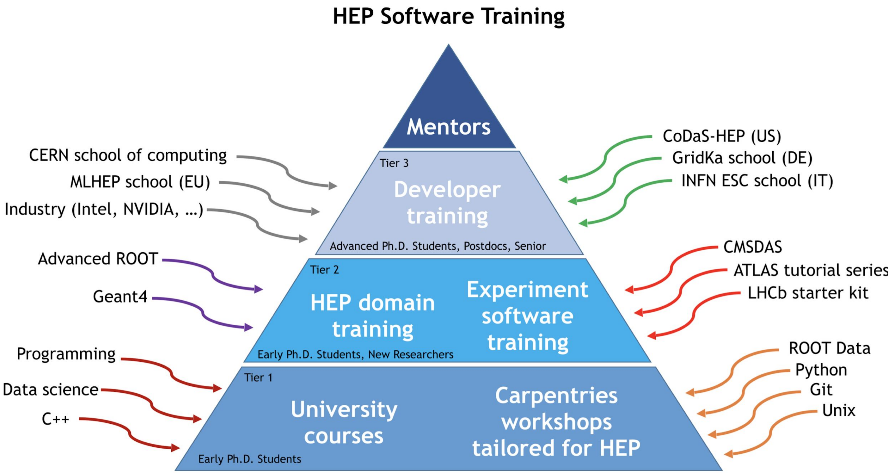


Distributed RDataFrame scaling (Dask backend), V Padulano et al

PyHEP Workshops

- Started in 2018 at CHEP Sofia, blossomed into two hugely successful virtual workshops in 2020 and 2021
- More than 1300 people registered, demonstrating a huge interest in this area because of:
  - 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
  - New HSF activity area started just before the summer
- 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
  - Captions thanks to support from Python Software Foundation amongst others

Training and Careers



- Many new skills are needed for today's software developers and users
- Base has relatively uniform demands
  - Any common components help us
- LHCb StarterKit initiative taken up by several experiments, sharing training material
- HSF Training Group runs Software Carpentries and other tutorials (co-organised between the HSF IRIS-HEP)
- Highly successful C++ training courses (3 this year from SIDIS and HSF)
  - Inspires continued curriculum development and sharing material
- Assembling a complete curriculum for training in HEP, using Carpentries templates
- Recent paper published on HEP Software Training Challenges

HSF and directions for International Efforts and Cooperation

- Particle physics is an inherently international effort, with an excellent tradition of cooperation in many different domains
    - Detector R&D, Experiments, WLCG, Common Software
- But we have also had incoherent approaches and duplication
- HEP Software Foundation has been fostering much more the shared vision
    - This encourages diverse R&D!
    - There is now real success in attracting funding to this area
    - Recognised links to other main players (WLCG, LHCC, IRIS-HEP, EPPSU, SWIFT-HEP, Experiments)
- There is a lot of work we need to do for upgrades and future projects
    - HSF offers an excellent place to present work, discuss successes (and disappointments!) and to help lead community activities
        - WG convener nominations welcome for 2022

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
- Workshops and Annual meetings
  - Not quite sure what our 2022 plans are, but we will do something!
- Propose a new activity area
  - The HSF is there to help gather interest

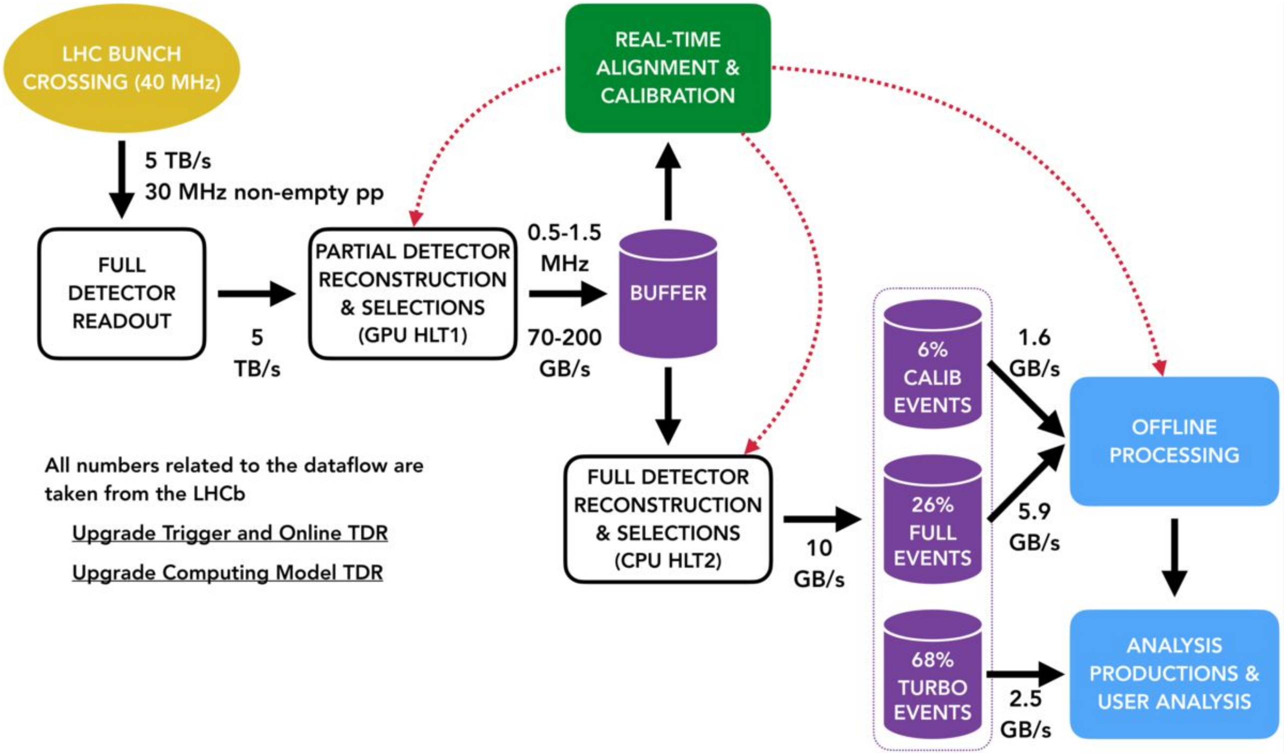
HEP Software Foundation

- 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

Backup

LHCb Real-time Dataflow



https://lhcbproject.web.cern.ch/lhcbproject/Publications/f/p/LHCb-FIGURE-2020-016.html