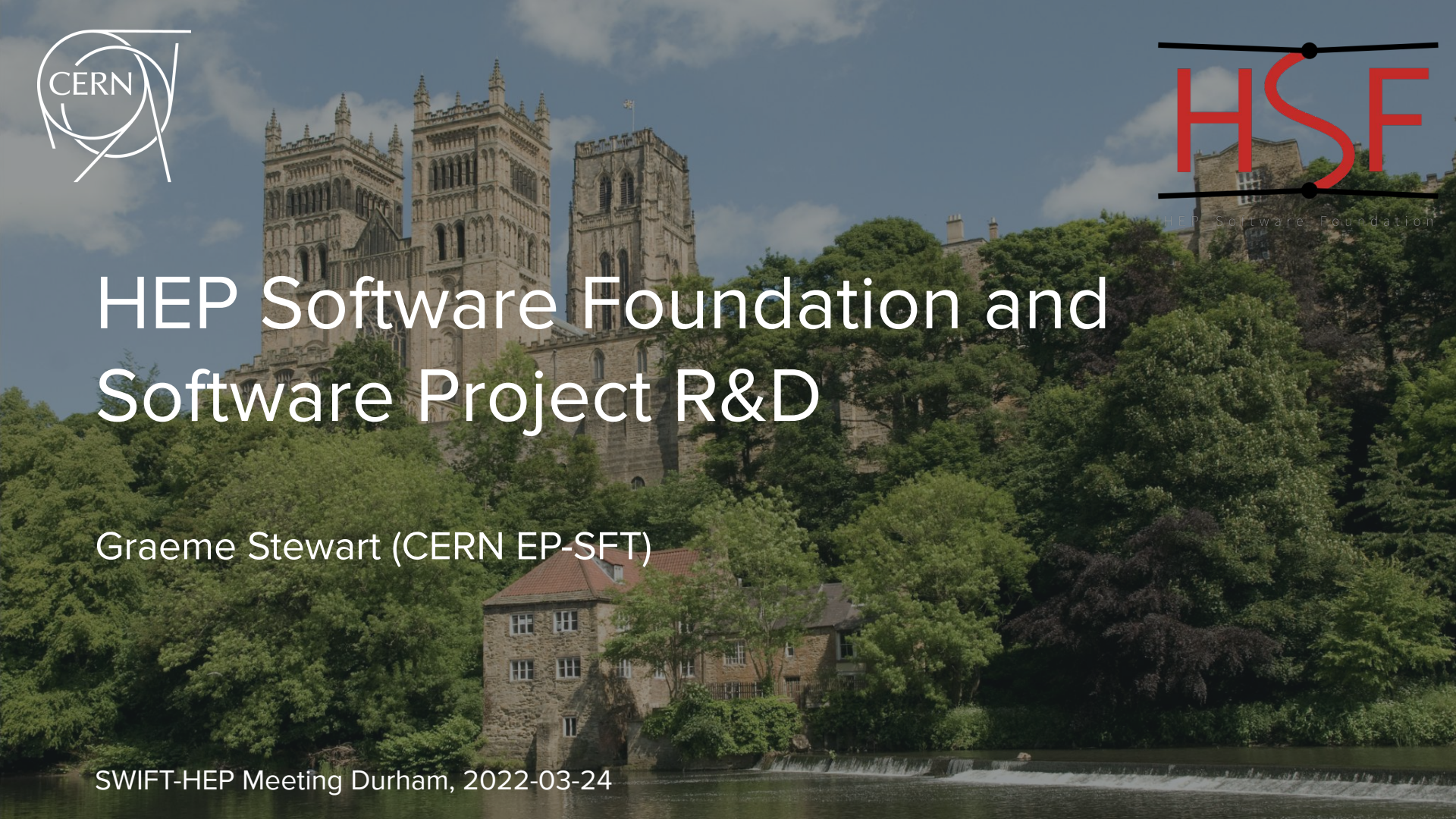


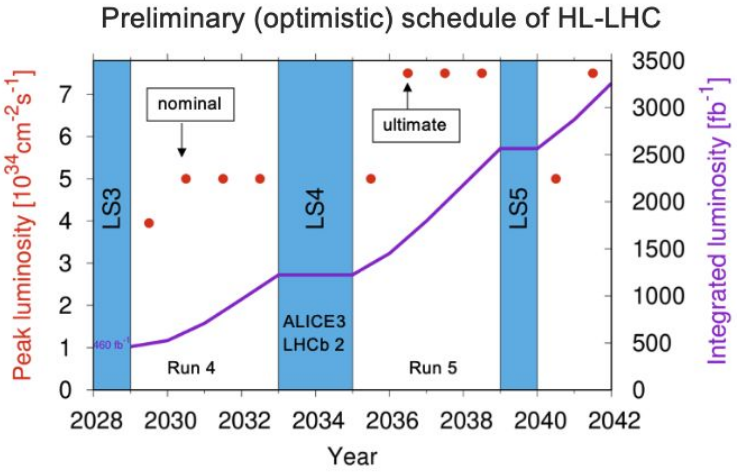
HEP Software Foundation and Software Project R&D

Graeme Stewart (CERN EP-SFT)

SWIFT-HEP Meeting Durham, 2022-03-24

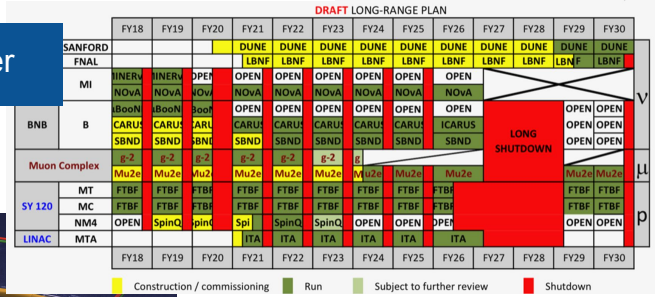
Overview

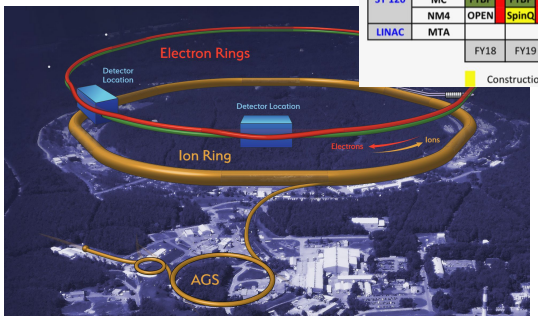
- HSF promotes and encourages common software developments
- Serving the needs of high-energy and nuclear physics
  - Primary motivation is the physics programme
  - (HL)-LHC, but also intensity frontier and nuclear physics programme
  - Software is a key component of our physics exploitation
  - Consistent with EPPSU
- Involvement from the UK has been great in the HSF
  - We can always foresee additional points of contact and interactions



HL-LHC Schedule

FNAL Intensity Frontier

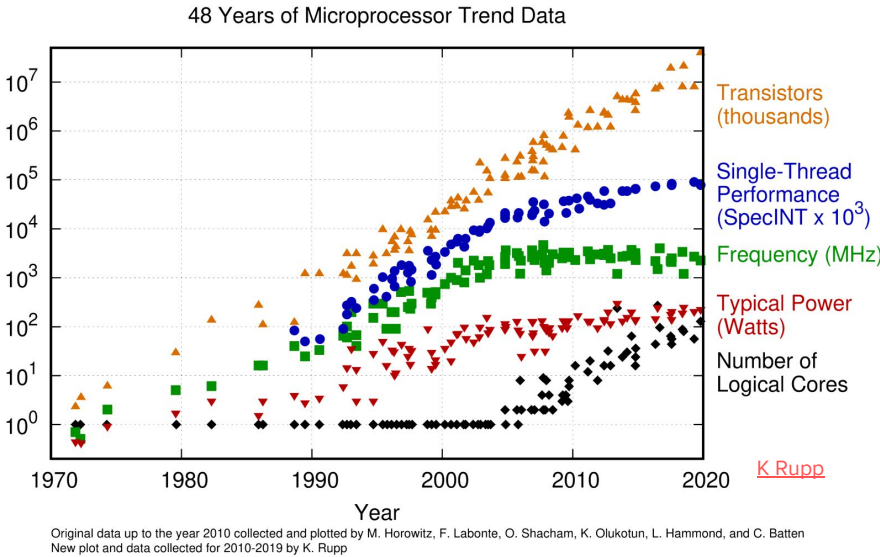




Electron-Ion Collider Early 2030s

Technology Evolution

- Moore's Law continues to deliver increases in transistor density
  - Increasingly challenging technical issues, but there is a roadmap to 2nm by 2025
- Clock speed scaling failed around 2006
  - Limits the capabilities of serial processing
- Memory access times are now ~100s of clock cycles
- Processing technology evolves from being solely based on CPUs towards heterogeneity
  - GPUs, FPGAs, TPUs, ...
  - This hardware is very much adapted towards different computing paradigms, *machine learning* and now towards *differentiable computing*

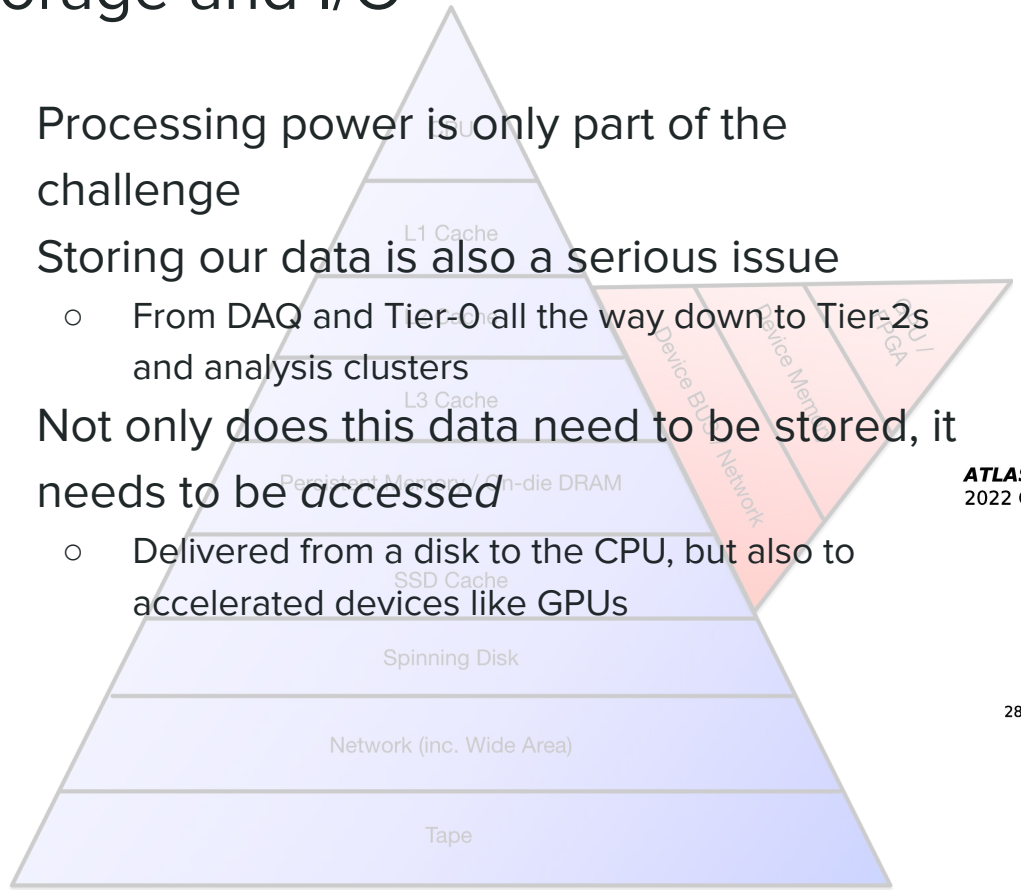


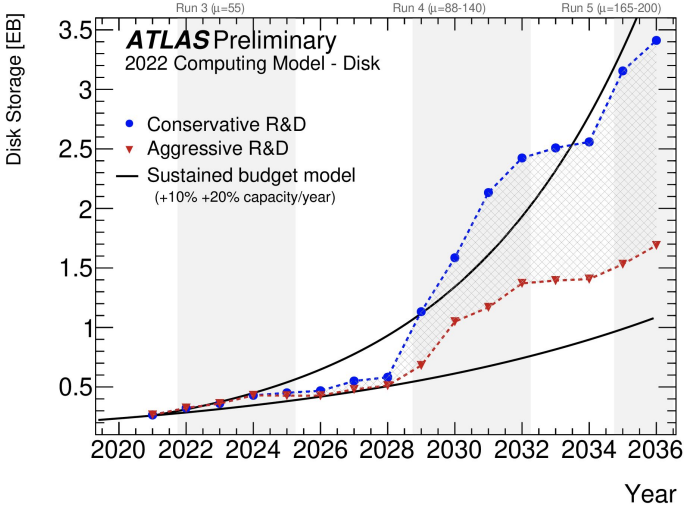
New process schedules for future chips

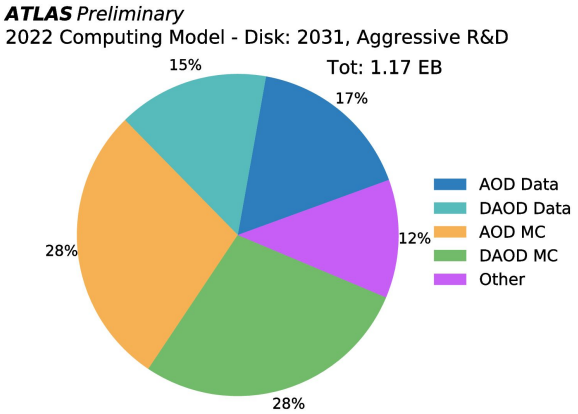
	2018	2020	2021	2022	2023	2024	2025
TSMC	7nm	5nm/6nm		3nm	3nm+		2nm
Samsung	8nm	5nm/4nm			3nm		2nm
Intel	Intel 10		Intel 7		Intel 4/Intel 3	Intel 20A	Intel 18A

Storage and I/O

- Processing power is only part of the challenge
- Storing our data is also a serious issue
  - From DAQ and Tier-0 all the way down to Tier-2s and analysis clusters
- Not only does this data need to be stored, it needs to be *accessed*
  - Delivered from a disk to the CPU, but also to accelerated devices like GPUs

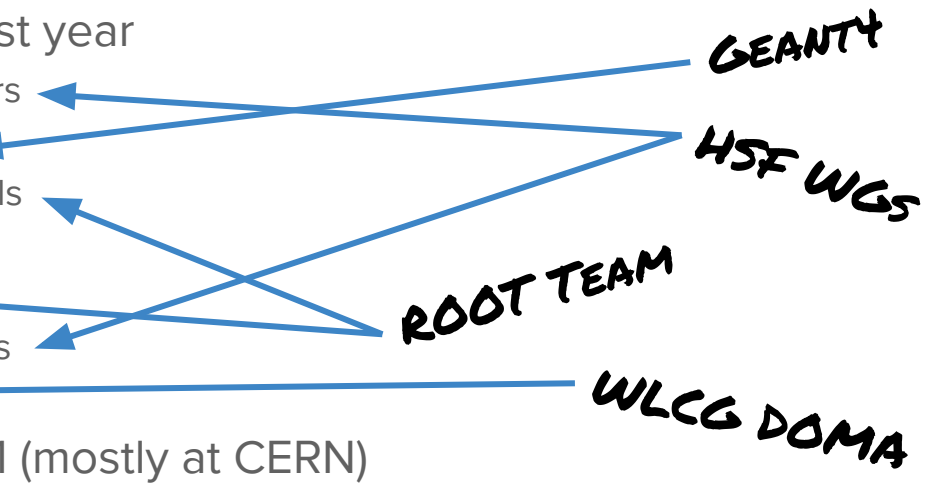






LHCC Review

- LHCC review of HL-LHC Software and Computing, Part 2
  - Standing review committee, chaired by Amber Boehnlein (JLab)
- Preparation of inputs last year
  - Physics Event Generators
  - Detector Simulation
  - ROOT Foundational Tools
  - Data Analysis
    - ROOT
    - Data Science Tools
  - DOMA
- Review November 2021 (mostly at CERN)
- Report published March 2022, https://cds.cern.ch/record/2803119



Review Outcomes

- *From a technical perspective, all activities and the experiments are showing excellent progress*
- HSF: discussion forums in generators and data science tools *positive collaboration*
- Simulation
  - More coordination on fast simulation would be welcome
  - R&D for GPUs with the G4HepEm library shows promising results
- Analysis
  - ROOT: *incredible level of technical expertise*; importance of RNTuple to reduce size and improve access speeds
  - PyHEP: more interactions with ROOT (best of both worlds), experiments should be involved
- Generators
  - Despite mismatch in incentives (theory advances != computational improvements) impressive progress in performance
  - Negative weight events (a major pain point) showing progress (e..g, MC@NLO-Δ)

Still discussing how to take the observation and recommendations on board

General Activities

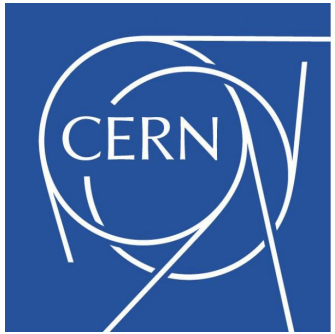
Activities and Planning

- Much activity at the start of 2022 devoted to review of 2021 accomplishments and plans of work
- For details please see:
  - HSF Planning
    - Coordination Meeting on 20 January
    - 2022 Planning Page
  - SFT Group Meetings
    - 24 January: SPI, Key4hep, HSF
    - 31 January: CVMFS
    - 7 February: Geant4
    - 28 February: ROOT

CAF and Roundtable

- Compute Accelerator Forum
  - General series of meetings on the use of *accelerators* in HEP
  - Topics range from APIs to application software (Patatrack and ACTS) to available facilities
- Software and Computing Roundtable
  - General series of meetings on interesting topics in NHEP
  - Organised with JLab and BNL
    - Strong links to the EIC community
  - Meetings in 2022
    - Different languages in HEP: C++, Python and Julia
    - Data management, in HEP and SKA

GSoC / GSoD



- It's Google Summer of Code time again!
- CERN/HSF has just been accepted again as an organisation
- Project proposals this year
  - 18 projects
  - 25 organisations
  - 39 proposals
- Note that this year Google are allowing again longer projects of 350 hours, which generally works better for us
- N.B. We find that we have few applications from UK institutes, so please help advertise this programme
  - We even have people today working in the field who started in GSoC!
- Less well known is Google Season of Docs, which supports a technical writer to work on documentation for open source projects

New Activities

- Analysis Facilities Forum
  - Community platform for those interested in contributing to the development of analysis facilities for use by HEP experiments
  - Kick-off meeting 25 March
  - Alessandra will discuss this in much more detail in the next talk...
- Conditions Databases
  - Early HSF activity
  - Renewed interest from experts in ATLAS, CMS, Belle II, DUNE
  - Work to defining an API embodying best practice

Working Group Roundup

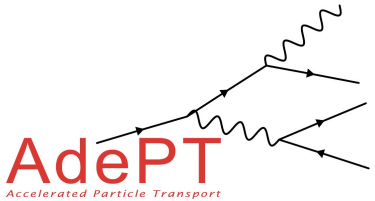
Training

- Software Carpentry Training
  - December 2021 edition was held (with The Carpentries, ROOT and IRIS-HEP)
  - March 2022 edition to come
  - Plan to host this 3-4 times a year
  - New: will hold Carpentry Trainer Event for our community (accreditation)
- Matplotlib Training close to first run
- C++ Training
  - Regular material development meetings
  - Split course into 2: *Essentials* and *Advanced*
  - New course version taught last week Essentials, 15-17 March
    - Initial feedback seems very positive

Generators

- Working on a Snowmass input document, [arXiv:2203.11110]
  - Topical meeting two weeks ago to discuss that
- Establishing better links with the nuclear (EIC) and neutrino community
- Continue to work on highlight topics
  - Negative weights at NLO and NNLO
    - Lots of recent papers to discuss
  - Porting to GPU and better use of modern CPUs (with IT-SC-RD, Madgraph and Sherpa teams)
  - Optimising generator use by the experiments (e.g., arXiv:2112.09588)
    - Would like to review this in the context of new generator releases with features that improve physics and performance

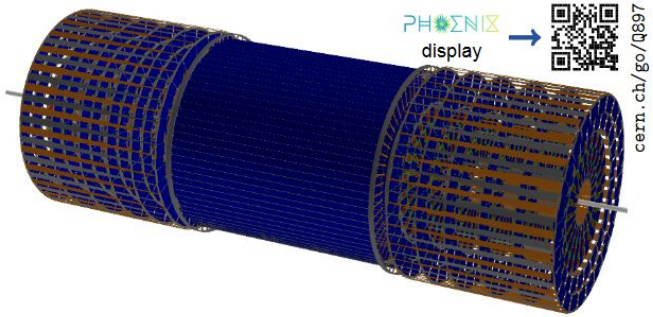
Simulation R&D and HSF Activity



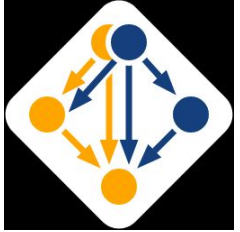
- New extended example in Geant4.11 demonstrating the incorporation of inference libraries (ONNX, LWTNN) into Geant4 for Machine-Learning based fast simulation
  - The current example is uses a VAE network, trained on some generalised calorimeter geometries, conditioned on the energy and angle of the incoming particle
- LPCC Fast Sim Workshop was held 22-23 November
  - Now to be held more regularly, based on feedback from participants and experiments
- Completion of the new AdePT example demonstrating the interfacing to Geant4 through the fast simulation hooks
  - This examples allows one to run Geant4 simulation with the GPU based transport (AdePT) enabled in specific parts of the geometry (calorimeter) for the EM physics
  - Energy depositions are collected in selected volumes and transferred back to the host
  - Full validation for CMS geometry, benchmarking and optimisation is ongoing
- HSF WG
  - Use of FPGAs for simulation looked at last year
  - This year
    - Non-Geant4 simulation (MARS, 28 March)
    - Geometry topics (links to reconstruction geometry to be examined)
    - ML for simulation

Reconstruction and Software Triggers

- Discussions on Patatrack and ACTS
  - Joint meeting with the Compute Accelerator Forum
- 4D Reconstruction Algorithms, 16 March
  - For detectors with timing information
- Open Data Detector
  - Development of TrackML detector
  - To be a testbed for experiment independent datasets for algorithm development
- Machine learning topics to be addressed with special ‘learning to discover’ (Institut Pascal Workshop) discussion



Frameworks and Tools



- Frameworks
  - Started to look at C++ features
    - First of a series of topics on C++20, also to look at modules, reflection, … with tools group
  - Efficient I/O for parallel data processing
  - Common topics with other groups
    - Scheduling accelerator workloads (espc. for reconstruction)
    - Conditions database interactions
- Tools and Infrastructure
  - Continue to look at Spack (becoming the packaging tool of choice in HEP)
  - C++20 features and training topics
    - Joint interest with Frameworks and Training

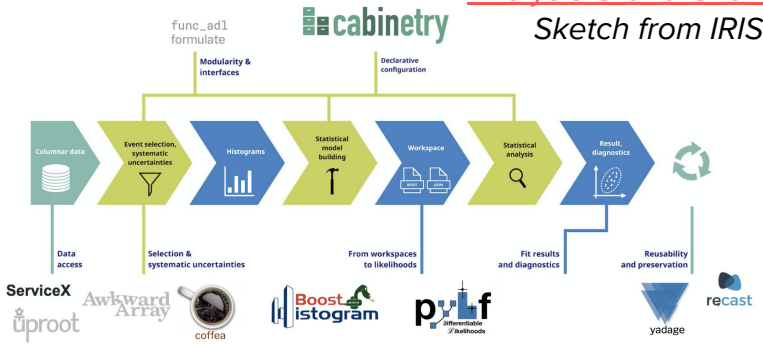
Analysis Topics

Analysis and PyHEP WG

- Analysis
  - Analysis working group paper on Metadata is now on arXiv:2203.00463
  - Continue to look at workflow management tools
    - Very popular topic in December
  - Implementation of systematic uncertainty tools
  - Meetings on Analysis Facilities at the GDB in December and January
- PyHEP
  - Restarted topical meetings, with a look at Boost-Histogram and Hist
  - Next meeting will be on Awkward Array developments
  - Planning the 2022 PyHEP Workshop for September - primarily virtual event
  - LHCC input document is arXiv:2202.02194

Analysis Grand Challenge

Sketch from IRIS-HEP

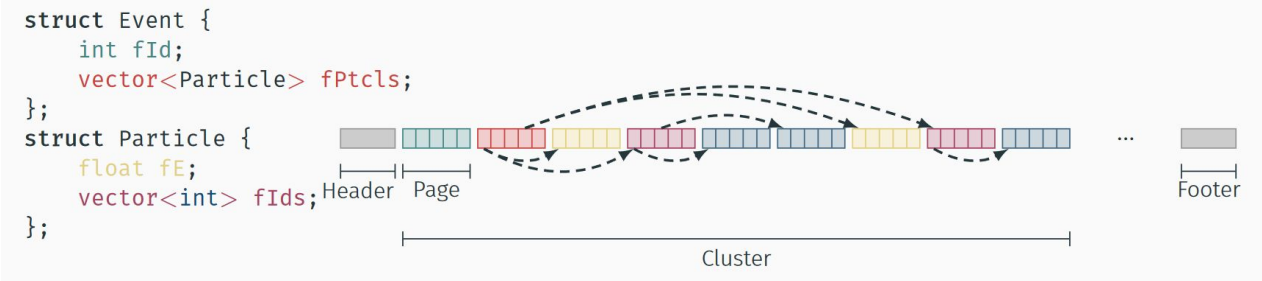


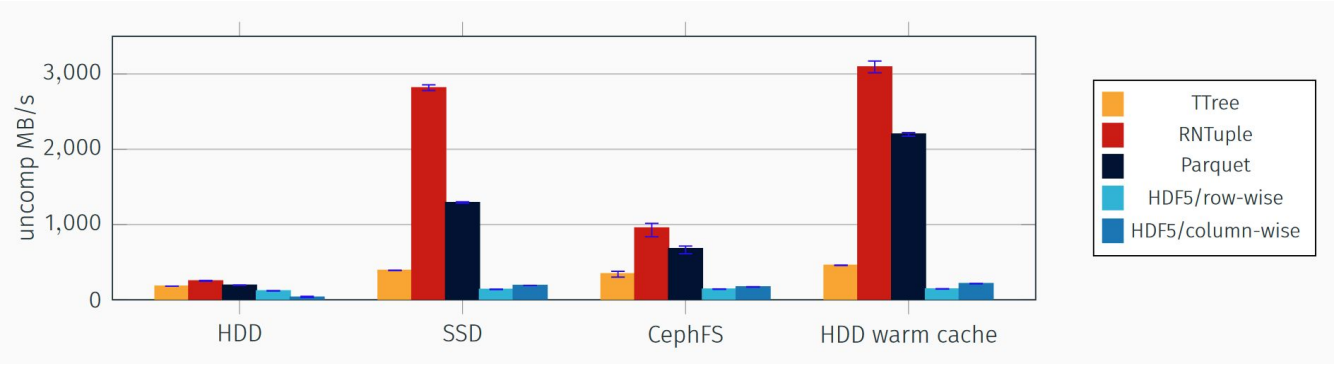
EP R&D - Efficient Analysis

- EP R&D is 5 year funded project on next generation detector technologies
- Including software, supporting projects in reconstruction, simulation, software stacks and analysis
- Analysis is helping with the development of ROOT's **RNTuple** and **Distributed RDataFrame**
    - We believe that these would be key components of any future Analysis Facility

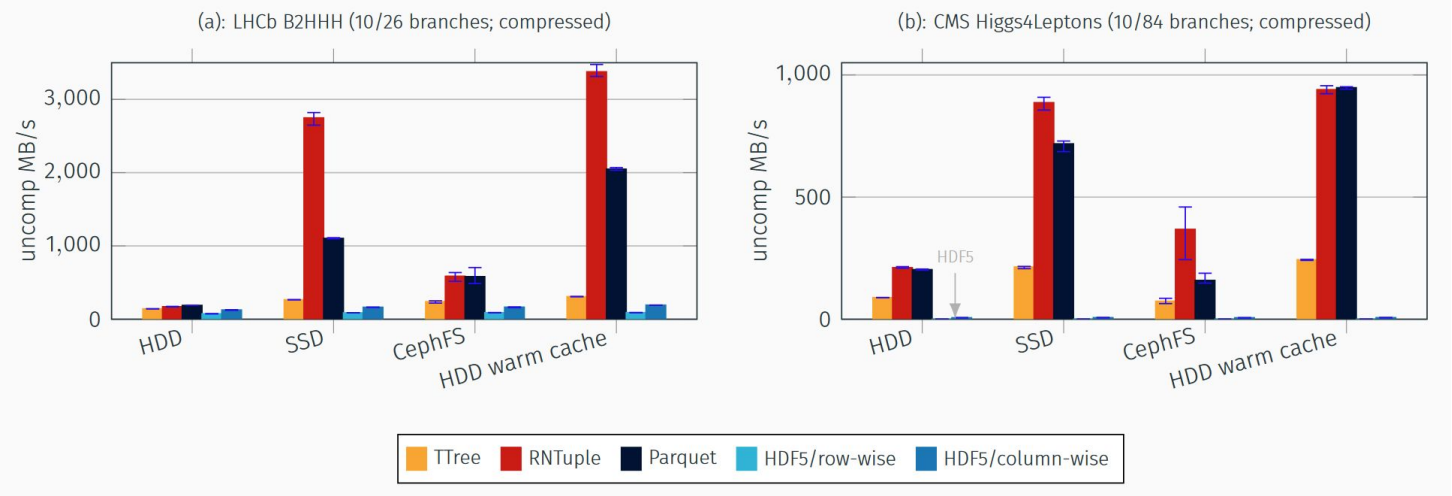
RNTuple

- Development of tailor-made I/O subsystem for HEP
  - We believe the volumes and costs of HEP data storage merit this investment
- Goals
  - Faster storage access for the same content as TTree
  - Smaller file sizes than the same data in TTree
  - Native support for object stores
  - Better support for error handling than TTree
  - Schema evolution for C++ (cling) and Python (cling + PyROOT)
  - Long term maintenance and support

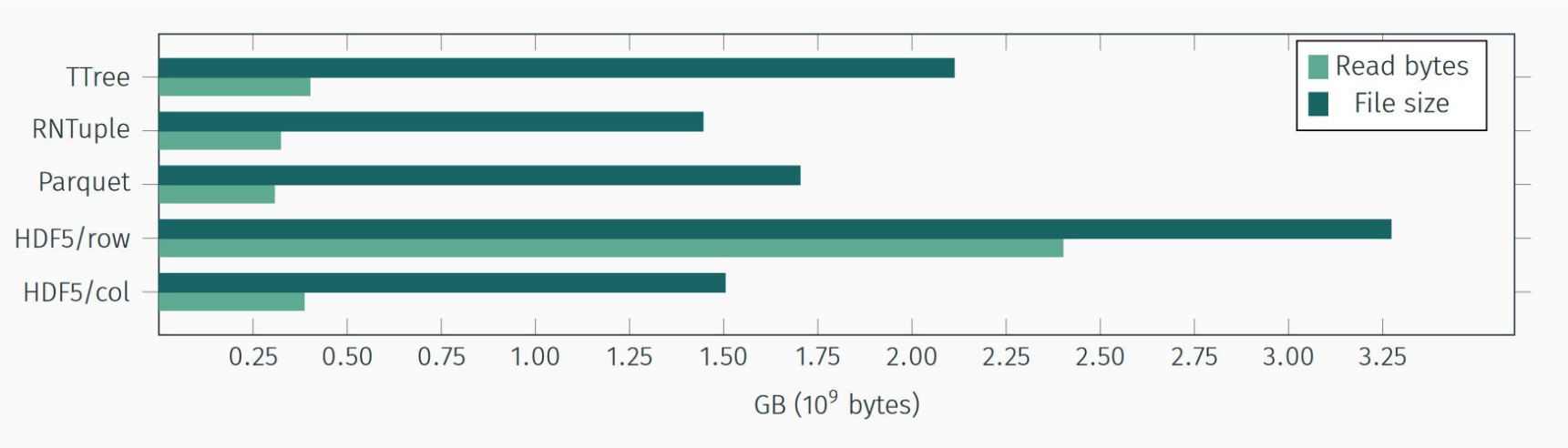




RNTuple performance - read speeds



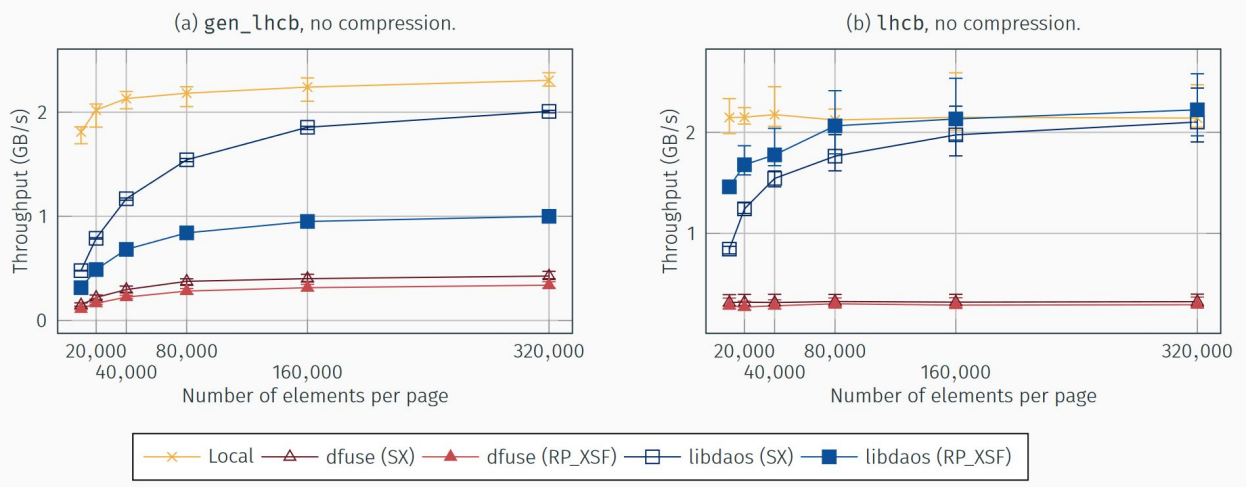
- RNTuple storage size and bytes-read for CMS Higgs to 4 leptons analysis



RNTuple on DAOS (Intel Object Store)

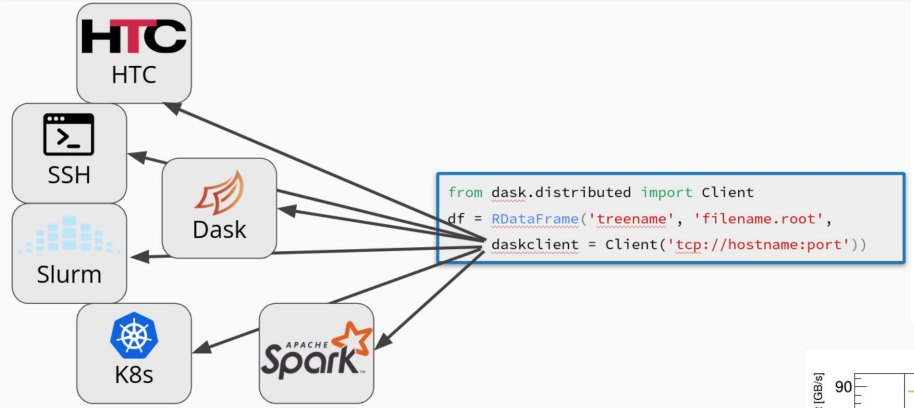
Data import (write)

Run analysis (read)



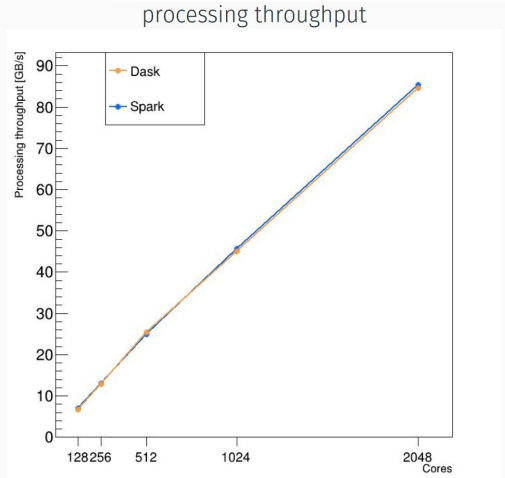
- Take away message is that access via object store interfaces is much faster than POSIX
- Promises event to be faster than local disk

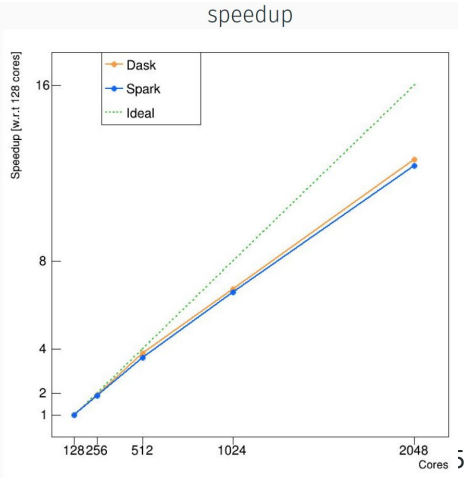
Distributed RDataFrame



- Programming model is a simple extension to an existing RDataFrame

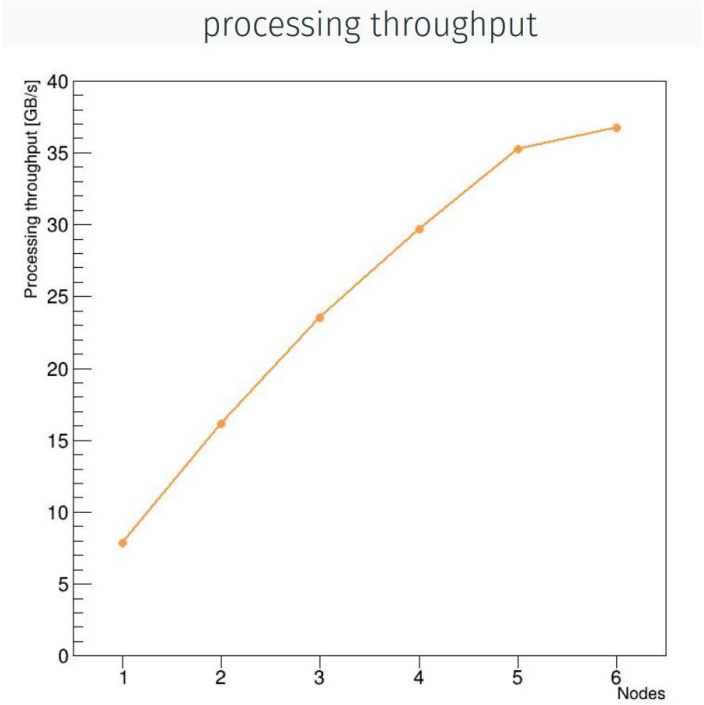
- Excellent scaling on CMS Dimuon analysis
  - 9TB read, 100% branches processed
- N.B. In this case data is *local*





Distributed RDataFrame on DAOS

- More realistic use case is to read the input data from a high-performance mass storage system
- Promising tests from the HPE DAOS test cluster
  - 1.2TB dataset, 70% branches read (900GB)



RDataFrame Systematics

- Systematic variations give rise to a lot of loops and re-runs of analysis code
- Since ROOT 6.26/00 there is a new interface to handle this directly in RDataFrame, using the `Vary()`/`VariationsFor()` interface

```
auto nominal_hx =
   df.Vary("pt", "ROOT::RVecD{pt*0.9f, pt*1.1f}", {"down", "up"})
     .Filter("pt > pt_cut")
     .Define("x", someFunc, {"pt"})
     .Histo1D<float>("x");

// request the generation of varied results from the nominal
ROOT::RDF::Experimental::RResultMap<TH1D> hx = ROOT::RDF::Experimental::VariationsFor(nominal_hx);

// the event loop runs here, upon first access to any of the results or varied results:
hx["nominal"].Draw(); // same effect as nominal_hx->Draw()
hx["pt:down"].Draw("SAME");
hx["pt:up"].Draw("SAME");
```

- Execution is made *efficient* by reusing as many pieces as possible
- Will be available in distributed mode for 6.28

Summary

Important Workshops in 2022

- Detector Simulation on GPU Community Meeting
  - Discuss and review progress on the GPU detector simulation projects
  - AdePT and Celeritas will both participate
  - Virtual format: 4 European afternoons, 3-6 May
- ROOT Users Workshop
  - ROOT roadmap and new ROOT features
  - Updated RNTuple performance studies
  - User feedback
- Analysis Ecosystems Workshop 2
  - Continuing the quest for reduced ‘time-to-insight’, 5 years after the Amsterdam workshop
  - Looking at the latest developments in ROOT, data science tools, analysis facilities
  - Discussion on next steps and missing pieces for end-to-end analysis - to produce a written report
  - In person (!) event at IJCLab, 23-25 May
- PyHEP 2022
  - Mostly virtual, anticipated for early September 2022
- ACAT2022
  - Our big HEP software conference this year
  - In person, Bari 24-28 October

Summary

- LHCC review has given praise to software projects and the experiments
  - But there is still a lot to do from now to HL-LHC!
- HSF Activities starting up again for 2022, continuing to provide a focus for exchange of ideas and common projects
  - New areas of interest are identified and the HSF can provide a home
- Analysis Facilities Forum and Analysis Ecosystems Workshop will provide a focal point
  - Lots of development in ROOT to help provide new tools and infrastructure here
  - Plus following talks from Alessandra and Luke covering IRIS-HEP Analysis Challenges and Analysis Facilities Forum