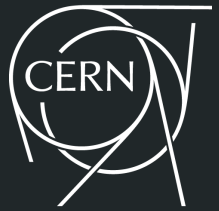


HEP Software Foundation

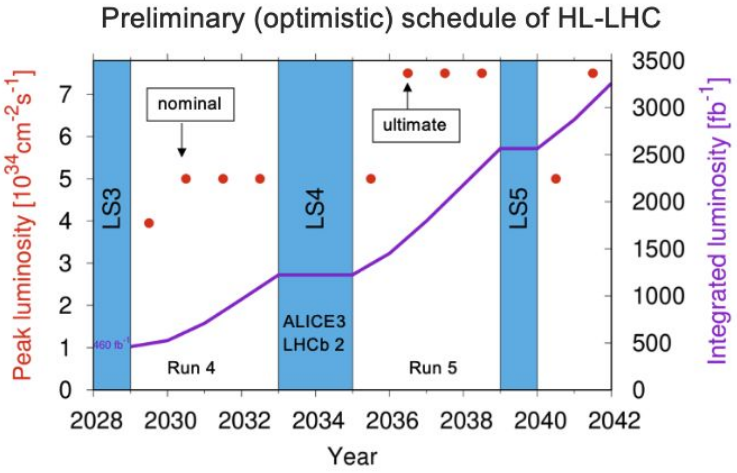
Benedikt Hegner (CERN EP-SFT)

SMARTHEP Kick-Off Workshop, 2022-11-23

Overview

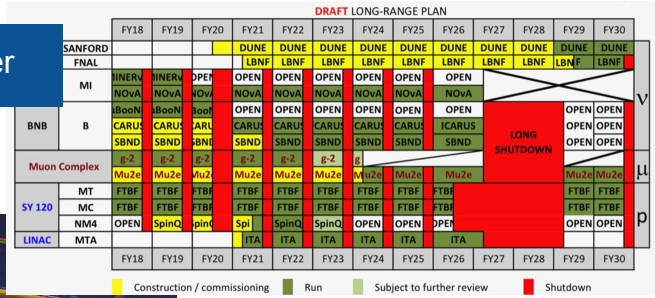
- HSF promotes and encourages common software developments
- Serving the needs of high-energy and nuclear physics
  - Primary motivation is the physics programme
  - (HL)-LHC, but also intensity frontier and nuclear physics programme
  - Software is a key component of our physics exploitation

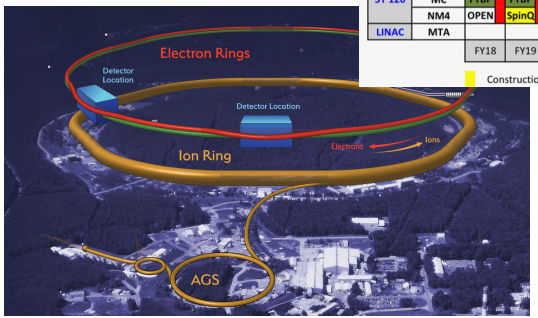
- There is always space for fresh and innovative people and ideas
  - And there is as well lots to learn from experts if you are new to the field!



HL-LHC Schedule

FNAL Intensity Frontier

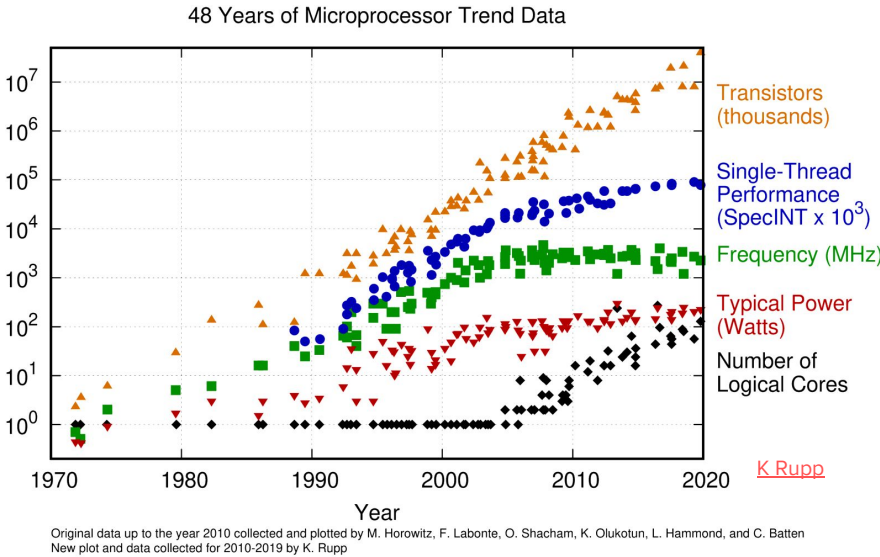




Electron-Ion Collider Early 2030s

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*

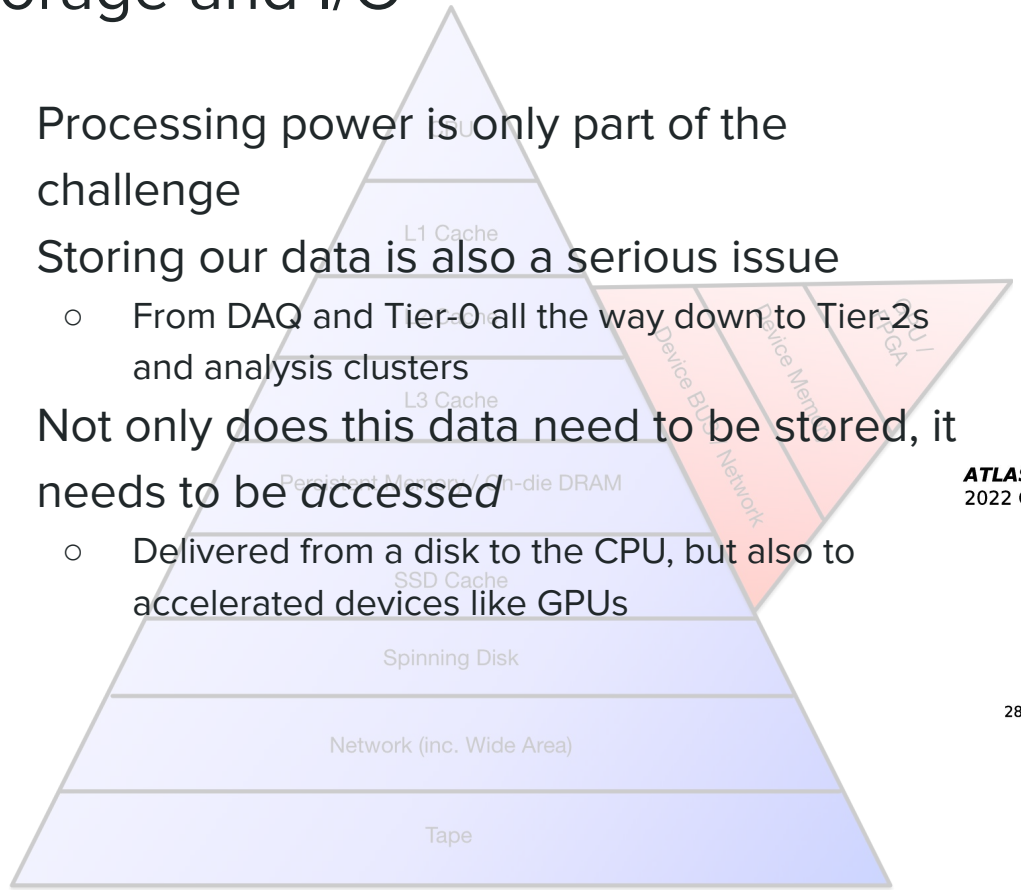


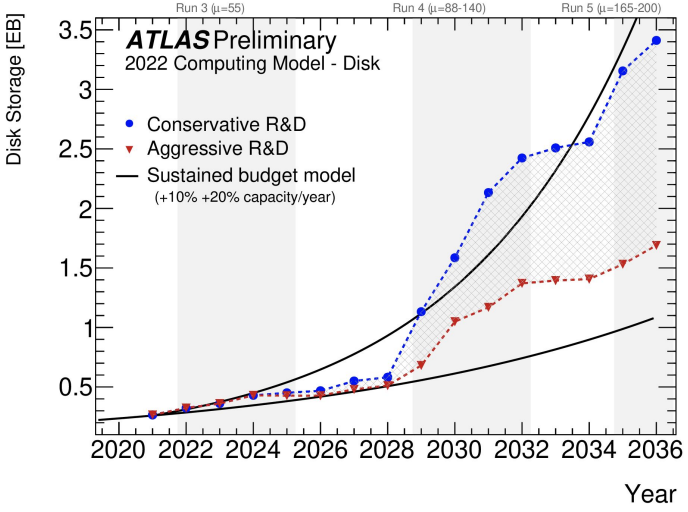
New process schedules for future chips

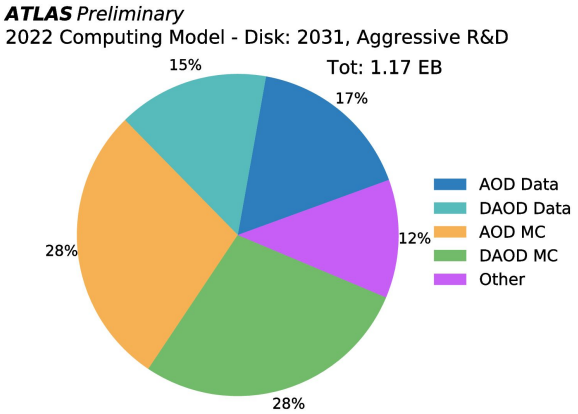
	2018	2020	2021	2022	2023	2024	2025
TSMC	7nm	5nm/6nm		3nm	3nm+		2nm
Samsung	8nm	5nm/4nm			3nm		2nm
Intel	Intel 10		Intel 7		Intel 4/Intel 3	Intel 20A	Intel 18A

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







General Activities

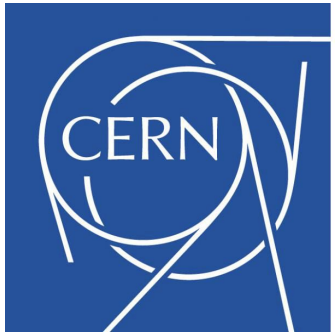
HSF Activities

- Plenty of working groups tackling topics community-wide:
  - Data Analysis
  - Detector Simulation
  - Frameworks
  - Physics Generators
  - PyHEP
  - Reconstruction and Software Triggers
  - Software Developer Tools and Packaging
  - Training

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
    - Experiments Starting Up
    - Workflow Management
    - Streaming Readout
    - Analysis - techniques and tools

GSoC / GSoD



- CERN/HSF has been accepted again in 2022 as an organisation
- Project proposals this year
  - 18 projects
  - 25 organisations
  - 39 proposals
  - 27 student projects accepted in the end
- Always a nice source of ideas from outside the field
  - European participation is rather weak though
- 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 was in March
  - Many more meetings since then
- Conditions Databases
  - Early HSF activity
  - Renewed interest from experts in ATLAS, CMS, Belle II, DUNE
  - Still needs to concretize for work to start

Working Group Roundup

Training

- Software Carpentry Training
  - Hosting 3-4 times a year
- C++ Training
  - So far 5 HEP C++ courses and hands-on training
  - Split course into 2: *Essentials* and *Advanced*
  - New course version taught last week Essentials, 15-17 March
    - Initial feedback seems very positive
- Diverse set of other trainings
  - This year new Matplotlib course
- Monthly training hackathon
- Full list of community trainings here

Generators

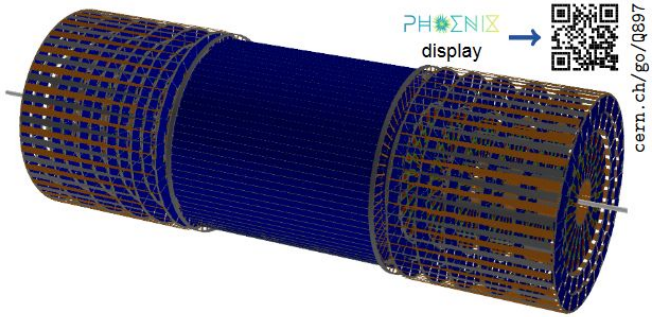
- Worked on a Snowmass input document, [arXiv:2203.11110]
- Established better links with the nuclear (EIC) and neutrino community
- Continue to work on highlight topics
  - Negative weights at NLO and NNLO
    - Lots of recent papers to discuss
  - Porting to GPU and better use of modern CPUs (with IT-SC-RD, Madgraph and Sherpa teams)
  - Optimising generator use by the experiments (e.g., arXiv:2112.09588)
  - On the more disruptive side looked into ML and Julia

Simulation R&D

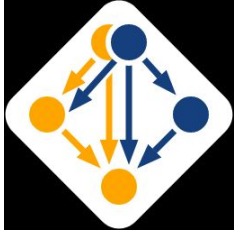
- HSF Working Group
  - Use of FPGAs for simulation looked at last year
  - This year
    - Non-Geant4 simulation (MARS)
    - Geometry topics
    - ML for simulation
    - Differentiable simulation

Reconstruction and Software Triggers

- Discussions on Patatrack and ACTS
  - Joint meeting with the Compute Accelerator Forum
- 4D Reconstruction Algorithms
  - For detectors with timing information
- Lots of work on the Open Data Detector
  - Development of TrackML detector
  - To be a testbed for experiment independent datasets for algorithm development
- However, most experts absorbed by LHC run



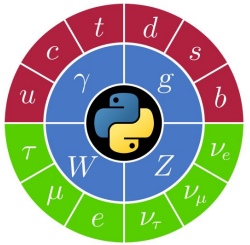
Frameworks and Tools



- 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
  - Looking more into non-LHC communities
    - Mu2e
    - sPHENIX / EIC
- Tools and Infrastructure
  - Continued effort on Spack (becoming the packaging tool of choice in HEP)
  - C++20 features and training topics
    - Joint interest with Frameworks and Training

Analysis Topics

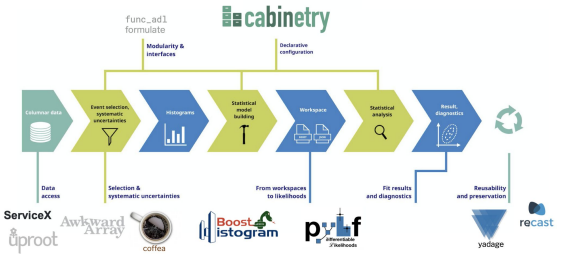
Analysis and PyHEP WG



- 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
  - Established strong program with Analysis Facilities
- PyHEP
  - Restarted topical meetings, with a look at Boost-Histogram / Hist and Awkward Array Updates
  - PyHEP Workshop 2022 very well attended (>1k registrants)
  - Discussion/outlook for Julia in HEP
  - LHCC input document is arXiv:2202.02194

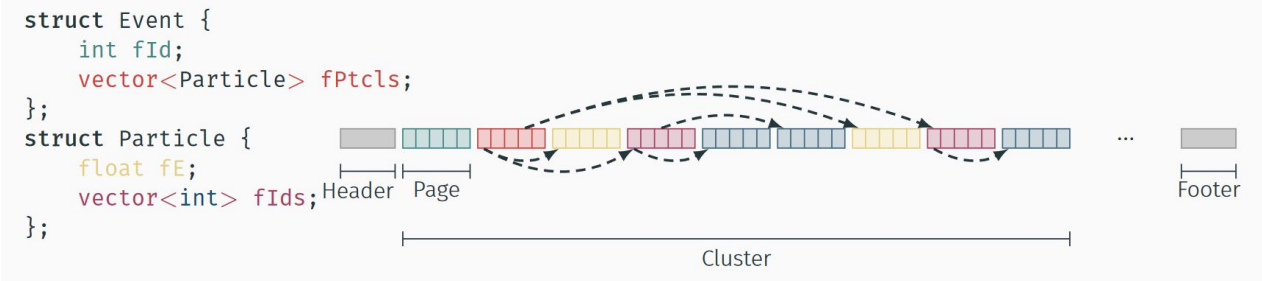
Analysis Grand Challenge

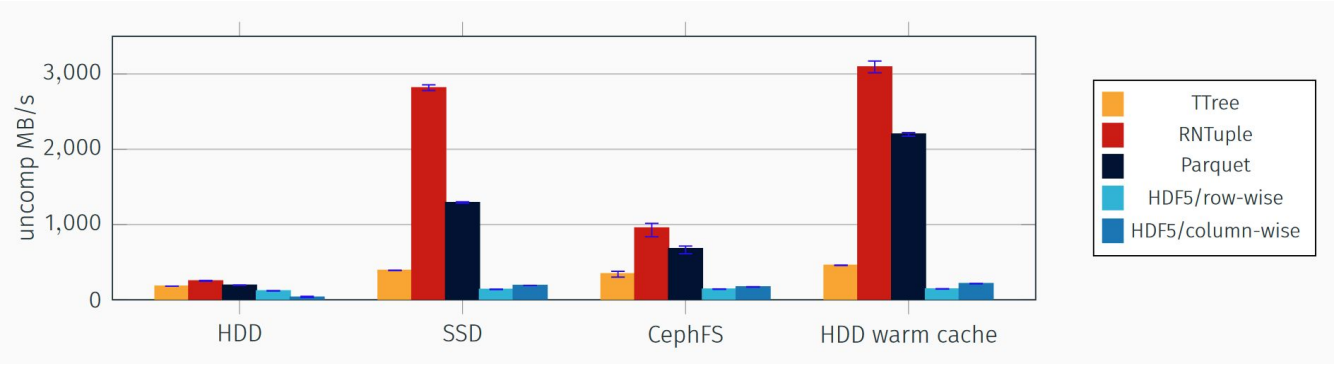
Sketch from IRIS-HEP



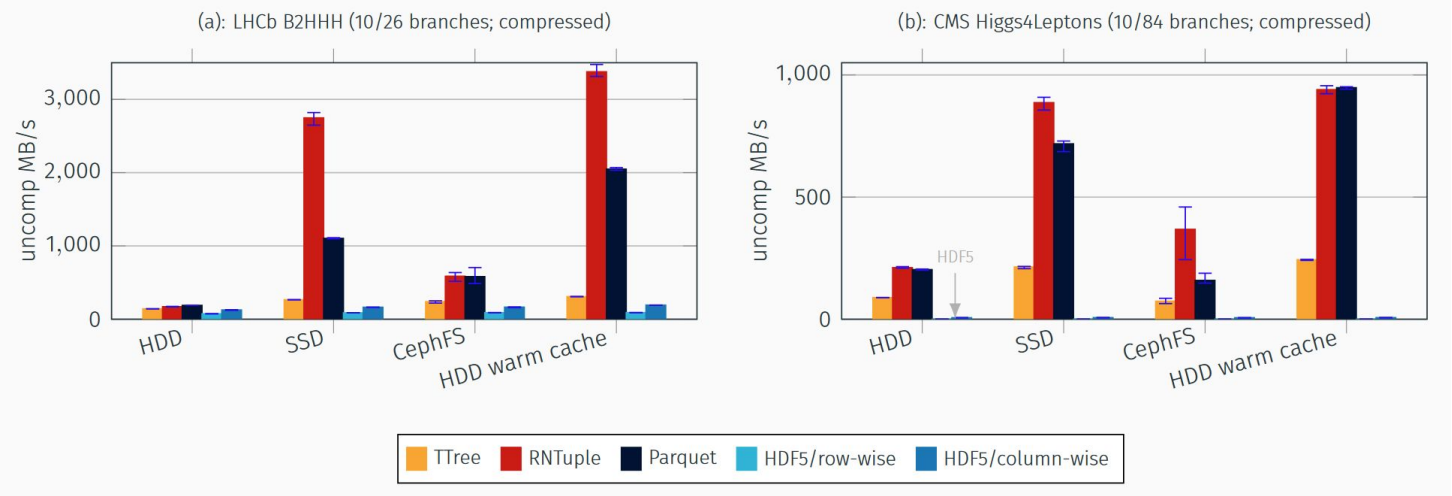
RNTuple

- 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

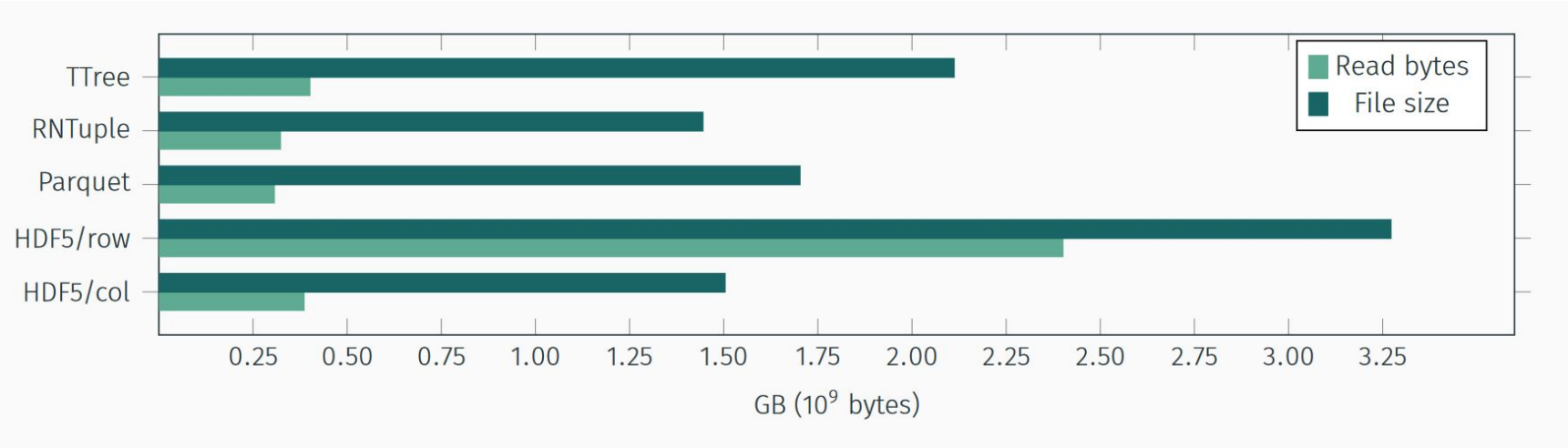




RNTuple performance - read speeds



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



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

Summary

- Plenty of successful software projects and contributions by the experiments
  - In 2021 an LHCC review gave us similar feedback
  - But there is still a lot to do from now to HL-LHC!
- Despite the busy LHC times, the HSF Activities continued to prosper in 2022
  - New areas of interest are identified and the HSF can provide a home
- Analysis Facilities Forum and Analysis Ecosystems Workshop were particularly important this year
  - Lots of development in ROOT to help provide new tools and infrastructure here
  - Better exchange with projects outside the field to see what we can learn from each other