

Common Software Updates

Graeme Stewart (CERN EP-SFT) and Liz Sexton-Kennedy (FNAL)

LHCC Referees, 2022-03-08



GSoC/GSoD Update



- Summer of Code, 41 projects were proposed by mentors
- Interest from prospective students varied a lot
 - From =0 to >50
 - Machine learning is very much the hot topic
- We finally have **27 students** who will start GSoC projects
 - To first order, these are all projects where there was a credible student
 - This is up 2 from last year
- Projects can run for as little as 10 to up to 22 weeks
 - 175 or 350 total coding hours
- Students writing a **blog on the HSF website** is now mandatory
- Season of Docs project proposed (it's only possible to have 1, but this can be multi-part)
 - Modern C++ interpreter (CLING in ROOT, evolving to CLANG-REPL in LLVM)
 - Reinterpretation packages, Gambit and Rivet

Physics Generators

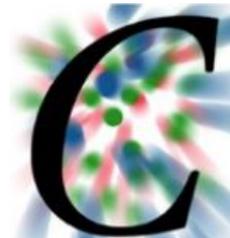
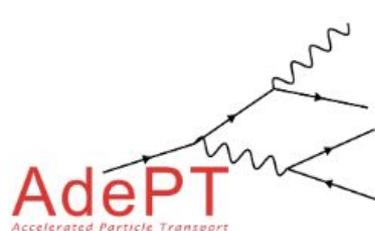
- Expanding scope of the HSF WG to include neutrino physics, nuclear physics, and more specific emphasis on the role of AI/ML
 - Very productive [meeting on neutrino physics](#) in early May [arXiv:[2203.09030](#)]
 - To come: dedicated EIC meeting end of June, ML-focused discussion in September [arXiv:[2203.07460](#)]
- Covering new technologies, e.g., event generation on GPUs or utilising the Julia language
- Work on software developments continues at pace
 - **Significant speedups in Sherpa**, covered in two meetings [[1](#), [2](#)] [arXiv:[2112.09588](#)]
 - Some of the LHAPDF work supported by SWIFT-HEP
 - **MadGraph GPU port** getting close to being tested by experiments
- Provided contribution to **community white paper for Snowmass: “Event Generators for High-Energy Physics Experiments”** [arXiv:2203.11110](#)
 - Emphasis on physics models and algorithms and cross-cutting aspects between experiments
- HSF discussed licencing of HepMC3 with the authors
 - Now [relicenced as LGPL](#) (from GPL), which is much better for non-GPL experiment codes

Simulation



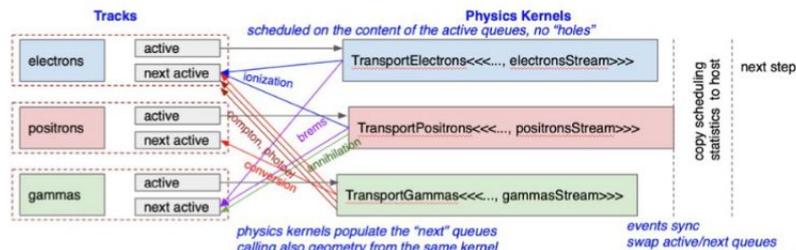
- **Geant4 11.1-Beta in preparation** (end of June 2022)
 - Updated tasking system based on PTL v2.3.3 (or Intel/TBB)
 - Robustness fixes and porting on new compilers
 - New specialized transportation for EM multiple-scattering
 - New integral approach for hadronic processes
 - Improved isotope production for spallation fragments
 - Updated thermal scattering data
 - Simplified configuration for data sets
 - Improved graphics representations for meshes support in visualization
- **New release VecGeom v1.2.0**
 - Major revision of the build system with enhanced CUDA support; C++17 default Standard
 - Enhanced BVH navigation system on GPU
 - Bug fixes and robustness improvements
- [Virtual] **Geant4 Beginners Course** – 9-13 May 2022
 - Full attendance; fully virtual on Zoom with hands-on sessions
- HSF WG meetings have covered **Fast Calorimeter ML Simulation challenge**; the **DD4hep** geometry package; **MARS** code (also interesting discussion on licenses)

Simulation - HSF Community Meeting



- Interest in **porting particle transport simulation to GPUs** has been growing in the last few years
 - Motivated by the large consumption of compute cycles today and the HL-LHC needs for the future
 - **AdePT and Celeritas projects** have been in development now for about 2 years
- HSF co-organised a [community meeting](#) on this important topic (3-6 May - virtual meeting)
 - Can we transform HEP particle transport to be efficient-enough on GPU?
 - How much effort will it take to create a production-level tool ?
 - What level of changes would be required to port key elements of the user code of production experiment simulation?

Problem Statement



- It's a challenge!
 - GPUs want every thread doing the same thing, with contiguous memory
 - **MC simulation has every particle doing something a bit different**
 - Energy, geometry, stochastic physics processes, ...
- **Electromagnetic physics in calorimeters is the focused problem to tackle at the moment**
 - Restricted problem (e+, e-, gamma) with limited physics processes of interest
 - Calorimeters restrict geometry and material
 - But still significant in resource consumption
- Meeting organised to look at key pieces of the problem
 - Project Overviews
 - Geometry and Physics
 - Integration

Main Conclusions

Wall time per primary (ms)

	geo	arch	mean	σ
Geant4 10.7.1	Geant4	CPU	2.9	0.1170
AdePT 68508ef7 (sethryj/adept/summit, 2 May 2022)	VecGeom	GPU	0.0850 <small>Latest optimization, 0.063</small>	0.0005
Celeritas 8d83ebab (29 Apr 2022)	ORANGE	CPU	2.09	0.0192
		GPU	0.046	0.0012
	VecGeom	CPU	1.95	0.0352
		GPU	0.0627	0.0004

- Performance

- Both projects **measuring ~x40 speed-up on Summit style hardware using GPUs**

- 2 CPUs (IBM Power 9),
- 6 GPUs (Nvidia Tesla T100s)

- Caveat Emptor: This on a **very simplified example (TestEM3)**, sampling box calorimeter, no magnetic field, no multiple scattering

- Physics

- **Per-mil agreement with Geant4** in both projects

- Geometry

- This is a **very serious bottleneck today**, sub-optimal and realistic geometries run ~x3 slower
- Both projects looking at surface based alternative geometries to VecGeom

- Integration

- Two approaches possible ‘fast simulation’ (AdePT) or ‘tasking model’ (Celeritas)
- User code progress needs engagement with users...

- Projects have engaged in dialogue to date and **Celeritas team should visit CERN** in June to carry this discussion forward

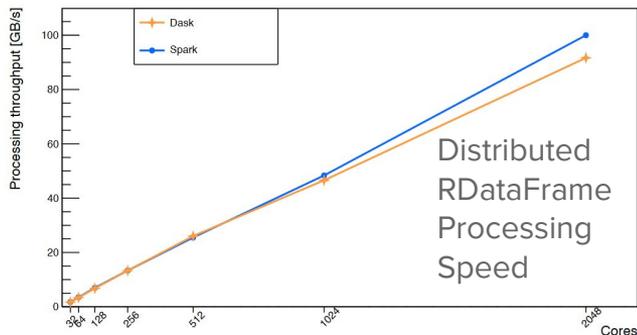


ROOT* Users Workshop

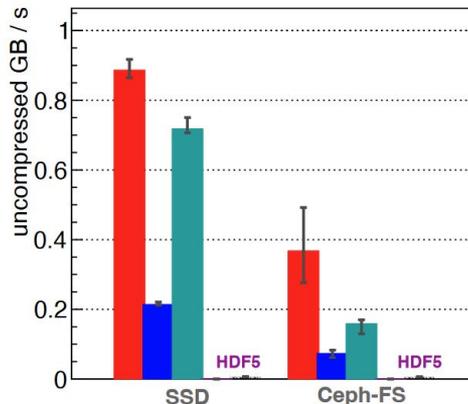
- Workshop held 9-11 May, virtually, 150 attendees
 - 12th in the Users Workshop series, first since Sarajevo 3½ years ago
- **38 talks over 3 days**
 - Packed into European afternoon slot / US mornings - very dense
 - Continued to meet the informal discussion time over coffee, etc.
- High level view
 - **News and outlook** from ROOT developers
 - **Novice users' experiences** and feedback on problems
 - **Expert users' use cases** and desiderata
 - **Experiment presentations** on workflows and analysis models
 - All LHC experiments, DUNE, mu2e, COMPASS
- Plus a Train the Trainer event

ROOT Development Highlights

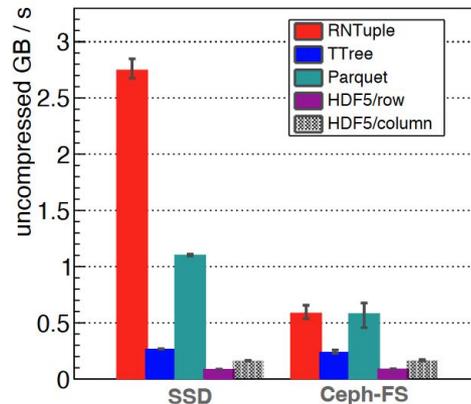
- **RNTuple is world beating**
 - The I/O format for the next 25 years
- Huge **Roofit speed-ups** on CPU and GPU
- **Distributed RDataFrame** can read and process data at up to **100 GB/s on 2048 cores**



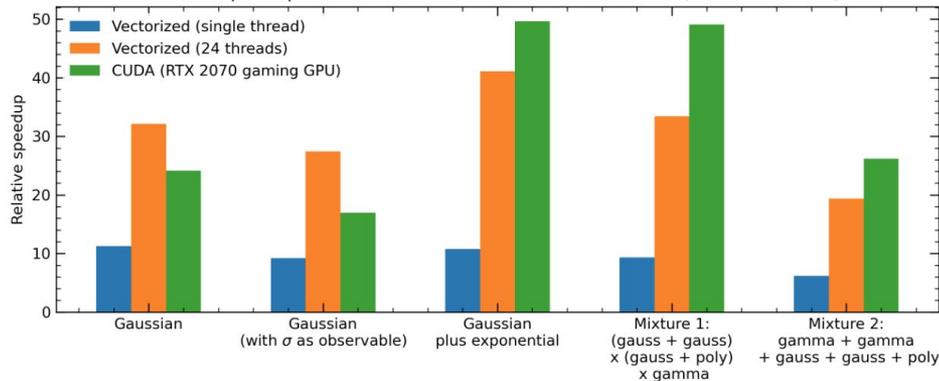
CMS Higgs4Leptons (10/84 branches)



LHCb B2HHH (10/26 branches)



Roofit: speedup in benchmark fits relative to scalar mode (1 million events)



ROOT Future

- First class **integration with machine learning** tools
 - SOPHIE provides lightweight, extremely fast inference
- **Python integration** got much better and keeps improving
 - `import ROOT` is a gateway to performant C++ under the hood
- ROOT has never been **easier to install**
 - Conda, homebrew, pip, snap...
 - Modularity - grab only what you need for simpler installation
- C++ reflection instead of dictionaries
- Auto-differentiation (CLAD) and GPU use on the rise
- Worries: general decline in the HEP level of C++, *which remains vital for performance*, ROOT becomes less accessible as a code base to physicists

HSF and IRIS-HEP Analysis Ecosystems Workshop

- Workshop held in hybrid mode at IJCLab
 - More than **70 people attended in person**
 - Held 5 years after the first workshop in Amsterdam
- Focused on 6 key topics for analysis
 - Analysis Facilities
 - ML tools and differentiable computing workflows
 - “Real-time” trigger-level analysis
 - Analysis User Experience and Declarative Languages
 - Analysis on reduced formats or specialist inputs
 - Bookkeeping and systematics handling
- HL-LHC was one focus, but not the only one
 - Run 3, Belle II, DUNE, ...



Workshop Outcomes

- We aimed for an outcomes focused workshop
 - The direction that we should be moving in the next 5 years
- The topic conveners are drafting their conclusions now and they should be available in a few weeks
 - Summaries in the [last session](#) give a good view of what was discussed
- A few personal observations
 - Having an **in-person event was extremely productive**
 - Lots of opportunity for follow-on discussions and making contact with new people
 - We agreed that there is **one HEP analysis ecosystem**
 - ROOT and Scikit-HEP are both there and both highly engaged

Topic Summaries

- **Analysis facility prototypes look fast enough now** (μs - ms per event)
 - AF focus now has to be on **ease of use** - for users and sites that deploy them
 - Many questions: scale-out, authentication, deployment complexity, user feedback, ...
 - Topics to be taken up in the HSF Analysis Facilities Forum
- **ML is much more widespread, becoming easier**, but still very dynamic; Autodiff is extremely interesting, but utility not yet established clearly
 - **Standard benchmarks** for performance will help
- User experience (UX) aims at reducing boilerplate and error prone/inefficient code
 - Do physicists need to do software engineering (and should they)? **There is training needed!**
 - **Bookkeeping and systematics remain pain points, as well as scale-out**
 - ROOT's `.Vary()` points the correct way
 - **Interoperability** between different ecosystem pieces is inconsistent
- **Reduced formats must to be used** to scale (NanoAOD, PHYSLITE)
 - Also need to **support the other analyses** - custom formats, dedicated skims?
 - Augmentation can be improved to only add for selected events
- Bookkeeping and systematics was discussed a lot in the UX context
 - Metadata paper reviewers suggested follow-ups, to be discussed in HSF
 - **Systematics challenge proposed**



CernVM File System

- **Major performance improvement** for many core nodes (64+)
 - Affects mostly building software with dependencies coming from cvmfs (e.g., compiler)
 - Increased node utilization for ATLAS builds from ~60% to ~90%
<https://github.com/cvmfs/cvmfs/issues/2879>
- Ongoing work on the 2.10 release, ETA August 2022
 - Several valuable **code contributions from industry** (>30 pull requests)
 - ongoing work on triaging and (where relevant) merging
 - Many bug fixes and performance improvements
- Upcoming **workshop at NIKHEF** in September 2022:
<https://indico.cern.ch/e/cvm22>



Software Releases and SPI

- Several **layers for ATLAS** LCG_101_ATLAS_[15,16,18,20,21,22]
 - Many updates for MCGenerators
- Collected more updates for **LCG_102 (ROOT 6.26/04*, Geant4 11.0.3)**
 - Updated creation of RPMs
- Added nightly builds for Apple M1 on macOS 12
- Continued work on the **Spack built stack:**
 - Setting up buildcache for sharing build results
 - Started discussing spack use with LHCb

Training

- **Software Carpentry Training** (58 students)
 - [March 2021](#) SC edition was held, taught by The Carpentries, ROOT and IRIS-HEP
- **Matplotlib Training** (100 students)
 - [New course](#) developed and held 21-22 April
- C++ Training
 - [Essentials course, 15-17 March](#)
 - Two new teachers from EP-SFT and IT-SC-RD
 - Good feedback from students
 - **Advanced course** planned for after the summer

Conclusions and Outlook

- May turned out to be the workshop month!
 - GPU Simulation, ROOT Users Workshop, Analysis Ecosystems, ...
- Lots of progress in many areas of software development
 - Generators, Simulation, Foundation and Analysis
- Upcoming events
 - ICHEP in Bologna (Expanded [Computing and Data Handling](#) track), July
 - [CERN VM Workshop](#), September
 - [PyHEP 2022](#), September
 - [ACAT2022](#), October
- And for posterity, we wrote a short paper on the history of the HSF
 - The HEP Software Foundation Community [[arXiv:2205.08193](#)]