Workshop Notebook

SwiftHEP/ExcaliburHEP Virtual Workshop, January 14-15th 2021

https://indico.cern.ch/event/976081/

The notebook here is for comments, discussions as well as any issues that should be followed up after the workshop. There are assigned note-takers for each session, but anyone is welcome to add their comments and observations.

Feel free to also write questions and discuss matters here before the presentations begin (questions will then be read by the chair after the presentation).

- Keep comments in the right section for each talk
- Don't delete anything anyone else wrote
- Identify yourself in the comment(s) so we know who made the comment
    - We suggest your name in []s, e.g. [Graeme S]

Day 1 - Thursday 14 January

Introduction

- Can we record the sessions? [Akram Khan]
    - Started from Eduardo's talk [Davide Costanzo]
- Slide 3, some sign of optimism. Several working hard to ensure SwiftHEP-2. Relevant contacts in STFC fully onboard. Don't want to overstate, but noises are good. Asked for £20M over 4yrs for PPAN [Pete Clark]
    - Could it overlap with phase 1? [Philip Clark]
    - Will ensure link up/contacts when/if happens [Pete Clark]
- (From Zoom Chat) For the website, would a wiki approach will be the most effective (otherwise it may end up quite static) [Philip Clark]
    - Downside is the typical Wiki decay (though true for any site) [Ben Morgan]
    - I think we need a "fancy" static site for publicity. Then we can have a wiki underneath - like e.g. the atlas website [Davide]
    - That makes sense, perhaps you can also have recordings of some public "highlight" talks and so on as well. To keep it a bit dynamic. [Philip Clark]
    - For publicity is something like a twitter account also useful for advertising? [Mark Hodgkinson]

The HSF Landscape

- On the engagement with the Nuclear Physics community, do we know how many UK people/institutes are participating [Ben Morgan]
  - As far as is known, not yet, likely due to JLAB+BNL organisation at startup [Eduardo R]
- On the training. Hopefully our data parallel school will provide material that can be used [Davide]
  - Material provided from across community, but need to work together to avoid reproduction and sustainability [Eduardo R]
- Would the material need to be in some specific format to be (the most) useful? If so, we should ask early on so that could be taken into consideration in the preparation [Mark Hodgkinson]
  - Defer to dedicated talk tomorrow [Davide]
  - The material largely follows the Carpentry style, hence Markdown with Jekyll, but let's discuss tomorrow … [Eduardo R]
- Given the popularity with Python, would cross-WG links/meetings be worthwhile, e.g. "Python for Simulation" ?[Ben Morgan]
  - Yes, by all means, WG encouraged to talk to each other and link up where possible [Eduardo R]
- Is there a specific ability level one should bear in mind when thinking about projects for GSOC? What level of students typically take part (e.g. final year undergrads or first year undergrads etc)? [Mark Hodgkinson]
  - Mostly later years or technically experienced people [Eduardo R]
  - Had one at RAL last year, very experienced so good. But lots (30+) of applications to sift [Stewart M]
    - HSF can help here with documentation, templates, etc [Eduardo R]
- Is there any activity around FPGA acceleration? (Claire Shepherd-T)
  - Not as widely discussed as GPUs, but large amount of UK expertise [Eduardo R]
  - Will be Compute Accelerate Forum meeting(s) in future of FPGA [Davide C][Ben M]
- Looking at the future architectures workshop from May 2020 it was before all the ARM related improvements really took off with Apple silicon, this could be interesting to be updated. [Philip Clark]
  - LHCb has put some effort in here, but not sure of status [Eduardo R]
  - Geant4 want to look at this during 2021, both Apple M1 and HPC ARM [Ben M]

Event Generators

- Just a comment: similarly heavy LHAPDF usage was observed in recent MG5_aMC profiling done in ATLAS [Josh M]
- You mentioned using a more powerful GPU would be interesting - we setup a VM in the STFC cloud with a GPU for Excalibur work, so you could try that if you think it would be useful [Mark Hodgkinson]

- - We also have some at Warwick which I would have already tested on if it not for certificate difficulties. But I may ping if we would like to try more. [Tim Martin]
  - What was your driver of using SYCL as an abstraction layer? Did you have the chance to look into other tools (Alpaka, Kokkos, etc)? [Stefan Roiser]
    - Due to the size of the companies involved: Khronos (spec), CodePlay (Compute++) (also: UK-based), Intel (OneAPI), Xilinx (trySYCL). It looked like the best framework to target, i.e. most likely to be supported into the future. And to have the most mature tooling. I did not experiment with any of the other players in this area so far, mainly due to limited time [Tim Martin]
  - Since you are doing matrix operations, could this be done on an FPGA as well? In principle SYCL should allow for a transparent porting to it. [Davide]
    - Yes. In my Pythia8 investigation I have only needed simple matrix operations (4-vecs, so only 4x4). However both BLAS and Eigen have SYCL implementation for more complicated cases. I haven't yet attempted to deploy on an FPGA, which would be possible if I had one available. [Tim Martin]
    - There should be some on STFC cloud. Alastair/MarkH would know the details. I'm not sure they are online yet (though they were ordered) [Davide]

Simulation

- Slide 12: moving the tracks to the relevant thread looks like a very dynamic operation in terms of data movement. IIRC this was hard to do for GeantV on CPUs. Are GPUs up to the task? [Paolo]
  - Yes, this is under investigation by both projects [Ben M]
  - One difference from GeantV is that it needed to reorganise/restructure data at each iteration. On GPU, can avoid this by masking out tracks that don't meet the criteria for the computation. To be determined how well this can work, but it does allow for more flexibility to balance reduced warp efficiency with the costs of a data reorganisation. [Graeme S]

Invited external talk: The DOE Center for computational excellence (CCE)

- Slide 7:"Runtime memory mapping of data" - isn't this what awkward-array (from IRIS-HEP) addresses (for local nodes) [Luke K]
- We should organise a technical discussion on event generators, as that is one area where we could collaborate [Davide]
  - Would be fantastic, at the point where having collaboration would push us over development into concrete outputs [Paolo C]
- Slide 10: A "Complete rewrite of matrix-element generator for CPUs & GPUs" sounds rather ambitious. From recent discussions with current generator authors this was not thought to be possible/useful/optimal… So is this considered a realistic goal/high priority? [Josh M]

- - Fair to say we came to same conclusion, extremely hard to do. Also looking at more efficient matrix-element calcs. [Paolo C]
    - Have had hackathon with Intel engineers on Madgraph code generation [Vincent P]
      - @Vincent, was this specifically for OneAPI? I would be keen to be involved with something like this in future if they are ~open invite… :-)
- Out of curiosity did you consider at all ASDF (asdf · PyPI) instead of HDF5? I know nothing about it but see some interest and traction, so just being curious. [Eduardo R]
  - Advanced Scientific Data Format (ASDF)
  - https://asdf-standard.readthedocs.io/en/1.5.0/
  - Thanks for the pointer, Eduardo: from a little reading, ASDF seems to mainly be interesting as a semi-human-readable (via YAML) take on HDF5. They (developers for James Webb Space Telescope) have a couple of other gripes that are summarised & echoed here: https://cyrille.rossant.net/should-you-use-hdf5/ Maybe a bit of NIH or grass-always-greener syndrome! The issues don't seem to include any performance concerns, which would be our dominant use-case for HDF5 at least while ROOT still has to be converted to as an endpoint format for experimentalists, and HDF5 is obviously much more widespread, but it's good to know about. [Andy B]
- Just a comment: the work on I/O and HDF5 would be of interest in simulation for physics model data, both raw and calculated, so maybe another point of contact, and I think ExaLAT have mentioned large parallel I/O as well[Ben M]
  - All welcome to attend meetings. Lots of experience to provide feedback/help [Paolo C]
- Slide 2: Not a criticism, are GPUs completely suitable for all workloads, e.g. in cost terms per cycle and subscriptions? [Pete Clark]
  - Basically agree, differentiator cf UK is that these machines (US) are being built based on government investments, so encouraged to use them [Paolo C]
  - Limitations on some systems with communication between nodes, e.g. by file rather than network [Pete C].

Day 2 - Friday 15 January

Analysis [Luke]

- On the RSE career. It is also important to engage with national initiatives, like software sustainability institute, etc [Davide]
  - There is a similar discussion within ExCALIBUR [Ben M]
- How visible is caching to the user? [Claire]
  - When doing caching, need enough provenance info to enable user to trace back exactly what was done [Luke K]
- Are there any examples of declarative code/scripts ? [Pete C]
  - Simplest example is cuts in ROOT's TTrees. Declare, as a text string, what to do [Luke K]

- - Also ROOT's RDataFrame, Scikit-HEP project awkward-array's package and Coffea project packages (the non-ROOT examples being from the PyHEP big data ecosystem) [Eduardo R]
  - Is this like Mathematica ? [Claire S]
  - Yes, to the extent of the symbolism [Luke K]
  - Another important aspect of declarative analysis is that there's no event loop. Just saying what should happen, not how. The how is left up to the backend programmer [Graeme S]

Data Management [Alastair]

- How does the UK data lake prototype compare with what is planned (or exists) elsewhere? [Davide]
  - UK will follow WLCG model. Advantage to our setup is closer connections between sites [Alastair D].
- In the near to long term, how will this change the balance between institute level and national resources for Disk/CPU, and how we bid for these in project/consolidated grants? Similarly, how will we link up with EPSRC etc ? [Ben M]
  - It will shift more towards someone running a large facility. It's easier now to provide a large variety in fewer places but that is likely to shift. [Alastair]
  - How is this linked up with agencies/funding providers? (STFC) Have gotten better at this (coordinated purchasing between councils). [Alastair]
  - When we put in the excalibur we said we might need to manage data across many sites in many countries, ut councils don't have that yet- usually local/single sites. We (EXCALIBUR) have the lead for something others may be doing later on [Pete]
- S4: A data lake should be a good thing, but what is it that has moved on from what we are doing anyway? ESCAPE showed that VOs are moving stuff in and out of a data lake. [Pete]
  - From a big VO point of view they will care about managing their own catalog, but what this configuration allows is for sites to configure how their storage gets used. That extra detail gets put in by the infrastructure rather than the expt, so they don't need to repeat this across each expt. [Alastair]
  - Pointer to escape talk https://indico.cern.ch/event/876772/#3-escape-full-dress-rehearsal
  - Perhaps Datalake is sort of "declarative programming" for infrastructure. That is, the VOs would then declare what they wanted to do but the infrastructure would decide how that is best done. [Dave Britton, in zoom chat]

Data Parallel School [Mark]

- When does registration close? [Davide]
  - End of January. If places aren't filled we can open it to the full HSF list.
  - Would be good to record the material and determine what can be reused.

Invited external talk: The ExaLat Project

- Jewels booster: 3700 A100 cards is a massive amount, no? [Pete]
  - Followed up on S6. It's 3700 of the cards noted in blue.
  - https://en.wikipedia.org/wiki/TOP500 Number 7 [Davide]
- On the data curation, is this just management or does it also cover I/O on HPC systems (e.g. data I/O over MPI etc) ?[Ben M].
  - It's the entire process. We're advanced on the I/O side but have a lot to learn on the lake side of things [Luigi]
- I think it would be interesting to at least sit in on the ExaLat schools to see what material is covered to find the commonalities [Ben M]
- Rucio is becoming the standard across PP and SKA. If lattice found it worked, there is a lot of shared support and development for it. [Pete]
  - We have been talking to Alastair already about it [Luigi]
- On GPU benchmarking, obviously NVidia has dominated, but with others coming online, are there HEP/Lattice cross-vendor benchmarks to compare? [Alastair]
  - Current ones can work on many architectures, but it does require work. Competition in this area between vendors would be healthy here, so need this type of systems [Luigi].
  - Would be good to link up here with LHC/HEP to compare/share [Alastair]
  - Code/etc all publicly available (GitHub) so happy to share/discuss further [Luigi]
    - The multi architecture benchmark Luigi was referring to is available at this address https://github.com/paboyle/Grid/wiki/Dirac-ITT-2020-Benchmarks [Luigi, on chat]
  - New "HepScore" benchmarks for CPU, but would/should allow other workflows such as GPUs. Important thing is to use actual production code. See update at Jan GDB [Graeme]
    - Hepscore talk https://indico.cern.ch/event/876772/contributions/4177112/attachments/2170154/3663847/2021-01-13-GDB-HEPScoreTFReport.pdf [Pete on chat]
    - The GDB talk was concerning to me. The current benchmark is ~25% off when comparing Intel and AMD CPUs. The HEPScore is not ready to use for procurement yet. It would be nice to have GPU benchmarks in place before we have to buy at scale.[Alastair]
  - Agree - Lattice benchmarks use simplified but realistic set of production code [Luigi]

- Have/are you looking at Arm, such as A64FX [Stewart]?
  - Yes, Japanese colleagues are looking at this and we have benchmarks that can run on this architecture. However, Fujitsu didn't bid on the Dirac procurement. Most likely short term is GPUs for the main machines [Luigi]
- Luigi will upload slides with info/mail to contact about the upcoming Schools.
- Can Lattice codes be considered "trivially easy" on GPU? [Pete]
  - Issue is that lattice cannot fit on single node, challenge is optimization for communication between nodes [Luigi]

Reconstruction and Trigger [Stewart]

- Largest work package but also the most difficult one to organise given expt specialisation. Need to think of how to go forward with this in mind [Davide]