CERN EP R&D Software Working Group - Notes on Lightning Talks 5th February

This document is open to anyone to contributed their notes that thoughts from the first lightning talks meeting of the Software R&D WG.

Some comments after the meeting added by the core group are in red.

Introduction - Jakob

We do R&D already in software. This is an opportunity to do something that might not otherwise get done.
Not clear how funding will happen - may have to fit into the existing EP budget.
Our goal to define as compelling a case as possible for software.

New Approach for ATLAS Detector Geometry Modelling - Alexander

What about geometry for detector design studies? E.g., fast prototypes where CAD is too heavy. A. Not clear, need to study other use cases.
Is this R&D at all as all products seem to exist?
A.: product doesn't exist. We need research study of methods and tools for development
Reference Geometry (RF) and transformation flows from RF into applications
Can this be extended to electronics simulation and emulation (same concepts of online/offline synchronisation exist and one single truth is probably also a useful concept)? In general, multi-physics and systems modelling in HEPP is not tightly integrated as in industry.
Not really R&D, also approach has been tried before and was not successful.

Software Defined and Named Data Networking, and the Consistent Operations Paradigm - Harvey

Can this be made appealing to non-experts? Yes, we need to use the network well, as a good customer and we are recognised by the network operators.
Why not just centralise the computing? Funding agencies like contributions from all places. Need to integrate other resources too, e.g., HPCs and Clouds, so WAN data delivery is not going away. (Graeme: Tier-2's often over deliver a lot on CPU).
Not well aligned with EP expertise (more IT), but should consider the impact on software (e.g., latency hiding) and data management areas, watching brief.

VectorFlow, a subscription-based vector data-shipping service - Andrei

Looks like a very important project even to help with shorter term problems that exist for Run 3. Good and concrete.

Beyond vectorisation - need to buffer data and care about memory bound processing (is 3D Xpoint interesting here?).

Gets a positive reaction and aligns well with challenges. Relationship to projects that have already happened (GeantV) should be clarified. Important impact on frameworks.

Generally we feel that concurrency needs continued investment - inhomogeneous future.

Future technologies for efficient and light-weight handling of non event data - Gianluca

What's the killer argument? Analysis users.

Is it R&D? Yes, not all problems are really solved, even if there are developments in progress.

Experiments have evolved in this area (CMS, Belle II). Overall direction is relatively clear now, so more in implementation phase. Different phase space points mean different technical solutions. Not much prospect of one system to rule them all. Calibration is more of the challenge for the future.

Filesystem-less high performance I/O of HEP Data - Danilo

Independent of a namespace, so would need to be in concert with evolution in this area. ATLAS very keen on this - meshes with evolution of data management system that is being developed to be file agnostic.

Important topic because it addresses the challenge between how we persistify and what underlying object type stores can deliver. Does this adapt to existing solutions or develop new storage layers? Probably the former. Open question - can we live without namespaces? Needs strong relationship with storage experts. There is a real need to evaluate the performance of our vectored reads against both filesystems (~storage systems+associated protocols) and objectstores (essentially ceph-native or S3). And to make sure we can do it not only on the LAN but also over the WAN. Additionally, programmable filesystems promised to help here (e.g., automatic separation of big files into their constituents (think root files -> separate branches)), also unclear how this progressed in industry and if that is still a thing.

Track Reconstruction in a Concurrent world - Andreas

Real R&D work here to actually do the track reconstruction in parallel.
Multi-dimensions to problem - too much data to process a whole event efficiently; complementary to Andrei's talk on bunching data for efficient processing.
Offloading is important for throughput.
Investment in ML hardware is desirable.
This is an unsolved issue - current approaches are very sequential. Not yet tackling the algorithmic challenges. CMS doing interesting work here too. Is convergence really possible? Links outside CERN as AIDA2020 taking ACTS developments.
Similar issues would arise for HGCAL.

Machine Learning technologies applied to online event selection - Maurizio

Close to a signal processing problem - there is synergy here.
Perhaps this is really needed for HL-LHC? Even use Run 3 as a testbed.
Technique could be used to inform detector design itself for optimal application.
Is this too vague? Could be a moonshot. Clearly a lot of excitement in this area, is it the right solution though? Develop understanding. Formulate better.

Machine Learning for Fast Simulation - Sofia

How generalisable is the approach? Curse of current fast simulation is that it's very specific. A. This is the R&D that needs to be done, but strong indications that the approach can be made flexible enough to do this and be generic.
Instead of integrating into the current simulation frameworks would it not be far better to have this kind of approach engineered to produce the specific simulation directly for a particular analysis?
- This is actually a huge research problem - would be at the cutting edge of ML techniques and requires real specialists. Dealing with very dense sets of information.
- Need to train people as well - these would be new techniques that are radically different from those that were learned by people 10, 20, 30 years ago.

Resources for input data (full sim) plus training - might not that outweigh the gains from the fast simulation itself. A. Training sets so far are relatively small and training time is reasonably quick. Then generation of showers is extremely quick. So we do expect really substantial gains in the end.
Has quite a lot of coverage already (does it need extra investment?). Would it be EP strategic area?
Per analysis - how would you validate? Particularly with no clear physics tuning available.

Automation of Data quality and certification with Deep Learning - Giovanni

Is it possible to use the previous knowledge of a trained network to optimise the evolution of the network for new situation? I.e., not have to do a full retrain. A. Partly R&D issue, but also can use wide shallow networks to train fast off the captured knowledge in a DNN.
Should look at current use in LHCb and ALICE. We do need an estimator and this is a statistical problem. Relevant.

Software for Detector Optimisation Studies in EP-LCD - Marko

This is a requirement of the R&D projects themselves. Provides for very quick feedback between R&D developers and users of the software. Makes it much more likely that success is obtained. Engineering problem, but a critical one.
See next talk...

Turnkey software solutions - Benedikt

Would this be a full framework? Yes, but with simple/dummy plugins for some components. Lack of this provoked ACTS to develop a ~100 line OpenMP mini-framework. Would very much like to drop that.
Does Gaudi offer the same concurrent functionality as CMSSW? Yes, it's able to cover the same space, albeit with some slightly different concepts.
Does FairROOT offer the same functionality?
(Didn't discuss link to Grid and production systems.)
Moving in this direction on a few fronts. Not so much R&D per-se. How much *extra* investment is really needed? What are the expected benefits.

High throughput data analysis on future heterogeneous platforms - Danilo

OpenCL? OpenCL hasn't really delivered performance and is maybe not suited to our problems (analysis). Do share a backend (LLVM), but this approach is more runtime using a JIT.
Validation is important, but arithmetic reproducibility on modern hardware is a lot better than in the LEP days (FORTRAN + x87). But do we also need to adapt algorithms to hardware, but that will lead to issues in the validation.
See next talk...

Heterogeneous computing - Felice

Does this look like the ALICE/FAIR O2, with message passing? Maybe some common ideas, but think there scheduling is still static.
Consider the cost/benefit of optimizing software vs. optimizing hardware (example: TPUs)
This is real challenge (see also VecFlow, same area). Need to protect physics coders and have good backends and frameworks to tackle this problem. Discussion with WLCG about links to infrastructure (Naples ws). Need more details on expectations.

Browser as a platform / compute device - Jakob

What is (fundamentally) different about the browser being open with the device being 'closed' if the browser could run any code? Isn't this like a shell? Running in a sandbox, but in some sense all apps do.
How closed will the future ecosystem be? Not clear.
Browser as a gateway to other computing resources - looks/feels local but is running remote.
Is this a better idea than a VM or a container?
What is the model for physics in the future? Will everything be 'remote' (cloud-like)?
Ergonomics of the browser is very convenient (e.g., SWAN).

Thematic point - are we missing a focus on the end user? A lot of data centre ideas.
HPCs are missing from our considerations (heterogeneities).