Exploiting HPC for HEP towards Exascale

Wahid Bhimji[1], Taylor Childers[2], Lisa Gerhardt[1], Jeff Porter[1]
[1]Lawrence Berkeley National Laboratory, [2]Argonne National Laboratory

Significant investment is being made in High Performance Computing on the 2020 timeframe. This includes the Exascale Computing Initiative in the US [1], [2] as well as similar initiatives in China, Japan, and Europe, which plan to offer an increase in computing power for scientific applications of at least 50x compared to today's 20 Petaflop systems. HPC machines offer increasingly more resources than existing and planned HEP-specific resources. Effectively exploiting HPC resources offers a way to fill expected gaps in resource requirements for HL-LHC and other HEP experiments. In addition, addressing the challenges of using supercomputers, e.g. increased parallelism, decreased memory capacity, and increased I/O performance, will lead to improved application performance on dedicated HEP resources.

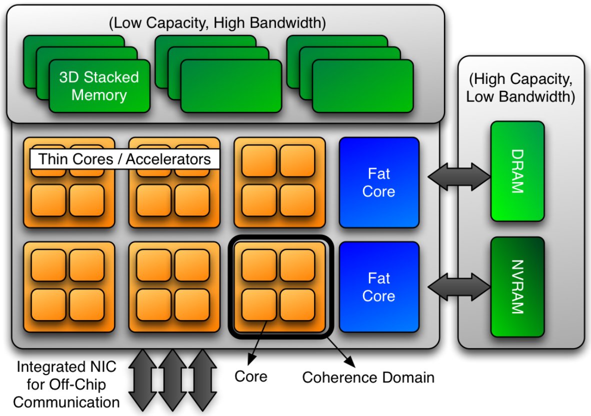
There is already considerable work ongoing to enable existing experimental-HEP workflows on HPC (e.g. [3]), and existing HPC facilities have welcomed HEP's data-intensive workflows. However current approaches mostly attempt to fit current HEP applications and grid workflows onto these systems and restrict themselves to certain workloads. Tackling issues to *maximally exploit* future HPC architectures for HL-LHC, will require a redesign to match HEP workflows to HPC resources together with shifts in the software, workflows and computing models.

We briefly highlight some features of likely HPC architectures in this timeframe, as well as some challenges arising from this architecture and the nature of HPC facilities, before proposing the activity required to meet these challenges and exploit this architecture.

Emerging architecture of HPC in 2020s

Salient features of future HPC include:

- Thin cores/ increased parallelism
- Heterogeneous compute resources
- Reduced memory capacity, high bandwidth memory.
- Integrated NIC (e.g. Intel Omnipath)
- Access from accelerators or thin cores to DRAM (e.g. via NVLINK)
- Storage performance hierarchy / Burst Buffers



From Ang et al.[4]

Challenges and opportunities in exploiting HPCs for HEP

Workflow/Policy challenges include:

- Philosophy / culture of deploying large interconnected systems using first-generation chips, as well as increasing node (or component) counts, leads to reliability and availability constraints.
- Storage hierarchy means that user has to effectively manage diverse storage resources. Data must be migrated from more-performant but impermanent layers in order to be retained. Future file systems will be focused on high bandwidth and IOPS but provide few tools for users to search the file system or migrate data, making data management challenging at the 10s of PB scale.
- Very dense systems limits architecture compared to what HEP resources may be elsewhere particularly in terms of external networking, on-node storage, etc.
- Lack of a strong federated identity makes access and data movement across multiple centers problematic.

Application challenges include:

- Lightweight cores provide the bulk of system performance, requiring increased parallelism and use of different instructions to obtain performance.
- Integrated NIC allows for improved across node performance, though alongside this data locality is increasingly important.
- Memory hierarchy with low capacity on high-bandwidth tier further increases importance of locality of compute to data.
- Heterogenous hardware provides coding and portability challenges.

For the workflow piece, DOE has begun a recent initiative around ‘Superfacilities’ [5] to connect experimental science into one or more HPC facility using widely applicable and high-performance tools and APIs. There is an opportunity therefore to build on this activity towards a common HL-LHC Superfacility that tackles many of the issues described above.

For the application piece, HPC centres have also recently extended application porting initiatives to target experimental data use-cases [6], [7]. These can be leveraged to enable applications to run well on future architectures.

Conclusions: Implications for HEP computing model / software

The use of HPC resources is crucial to meeting future HEP computing requirements. To do that effectively requires activity for both workflows and applications, on both the HPC and HEP sides and targeting both short-term evolution and a longer-term effort.

For workflow/policy challenges, other projects and documents have proposed a variety of activities to address some of these and enable better exploitation ([2],[3],[8]). These include working with facilities to add services, such as an edge service interface and a data transfer interface, containerization, high-speed data transfer and workflow tools. A longer-term effort to properly exploit exascale machines for HL-LHC should take a longer view of addressing these areas of workflows and data management. For example, workflows could be adapted to allow completion an entire MC campaign or reprocessing in a single job submission. Long-term workflow activity should link these edge services with the ‘Superfacility’ concept mentioned above, to build ground-up from the best of current LHC approaches and involve close collaboration between experiments and facilities, but aim to provide *common, scalable* approaches, filling gaps with development effort where needed.

For applications, in addition to current porting efforts, HEP workflows must evolve to match HPC architectures. For example they could exploit parallelism in multiple dimensions (including breaking per-event parallelism) to allow optimisation of on-node and off-node communication, memory use and data locality. Evolution of data formats and access protocols would allow effective use of storage hierarchy and available storage resources at diverse sites. There are opportunities for benefiting from expertise via application-performance programmes at HPC centres (mentioned above), but again these should be aimed at producing products that are broadly applicable as well as high performance.

Taken together these activities will allow moving beyond considering HPC machines solely as opportunistic backfill sites to embrace them as a truly high-performance, capability resources for HEP.

[1] https://exascaleproject.org/
[2] “ASCR/HEP Exascale Requirements Review Report” https://arxiv.org/abs/1603.09303
NP Exascale Requirements Review Report [In Press]
[3] http://hepcce.org/
[4] “Abstract Machine Models and Proxy Architectures for Exascale Computing” Ang et. al. DOI: 10.1109/Co-HPC.2014.4
[5] https://science.energy.gov/~/media/ascr/ascac/pdf/meetings/201609/Yelick_Superfacility-ASCAC_2016.pdf
[6] https://www.alcf.anl.gov/alcf-data-science-program
[7] http://www.nersc.gov/news-publications/nersc-news/nersc-center-news/2016/nersc-issues-nesap-for-data-call-for-proposals/
[8]“Management, Analysis and Visualization of Experimental and Observational Data” workshop https://science.energy.gov/~/media/ascr/pdf/programdocuments/docs/ascr-eod-workshop-2015-report_160524.pdf