Contribution ID: **190** Type: **Poster**

Evaluating Portable Parallelization Strategies for Heterogeneous Architectures

Wednesday, 26 October 2022 11:00 (30 minutes)

High-energy physics (HEP) experiments have developed millions of lines of code over decades that are optimized to run on traditional x86 CPU systems. However we are seeing a rapidly increasing fraction of floating point computing power in leadership-class computing facilities and traditional data centers coming from new accelerator architectures, such as GPUs. HEP experiments are now faced with the untenable prospect of rewriting millions of lines of x86 CPU code, for the increasingly dominant architectures found in these computational accelerators. This task is made more challenging by the architecture specific languages and APIs promoted by manufacturers such as NVIDIA, Intel and AMD. Producing multiple, architecture specific implementations is not a viable scenario, given the available person power and code maintenance issues.

The Portable Parallelization Strategies team of the HEP Center for Computational Excellence is investigating the use of Kokkos, SYCL, OpenMP, std::execution::parallel and Alpaka as potential portability solutions that promise to execute on multiple architectures from the same source code, using an assortment of representative use cases from DUNE, LHC ATLAS and CMS experiments. Central to the project is to develop a list of metrics that evaluate the suitability of each portability layer for the various testbeds. This list includes both subjective ratings, such as the ease of learning the language, and objective criteria such as performance.

We report on the status of these projects, the development and evaluation of the metrics, as well as the current benchmarks and evaluations of the portability layers for the testbeds under study and recommendations for HEP experiments seeking forward looking portability solutions.

Significance

Porting code originally written for CPUs to diverse heterogeneous architectures is currently an unsolved problem in the HEP community. While some experiments have ported some code bases to a single or a small number of platforms as they have already purchased their selected hardware backends, there has not been a systematic study of problem addressing all currently available heterogeneous architectures.

The HEP-CCE/PPS effort is the only cross experiment investigation that is tackling the issue of software portability on heterogeneous architectures with a very broad selection of portability solutions. We are addressing the needs of both large and small experiments with representative testbeds taken from a broad variety of sources. We are working in close proximity with the various experiments, with core developers from the experiments being part of the CCE/PPS team, which facilitates the cross pollination of knowledge and experiences, and feedback cycles with the experiments.

References

- Childers, Taylor, et al. "Porting CMS Heterogeneous Pixel Reconstruction to Kokkos."vCHEP 2021. arXiv:2104.06573v1.
- Dong, Zhihua, et al. "Porting HEP Parameterized Calorimeter Simulation Code to GPUs."Frontiers in Big Data. arXiv:2103.14737v2.
- Kortelainen, Matti J., et al. "Performance of CUDA Unified Memory in CMS Heterogeneous Pixel Reconstruction."vCHEP 2021.

- Pascuzzi, Vincent R., Goli, Mehdi. "Achieving Near Native Runtime Performance and Cross-Platform Performance Portability for Random Number Generation Through SYCL Interoperability."arXiv:2109.01329
- Yu, Haiwang, et al. "Evaluation of Portable Acceleration Solutions for LArTPC Simulation Using Wire-Cell Toolkit."vCHEP 2021. arXiv:2104.08265v1.

Experiment context, if any

We have taken representative testbeds to evaluate the various portability layers from a number of current HEP experiments, namely Patatrack (pixel tracking) from CMS, p2r (central "propagate to R" tracker from CMS), Wirecell (Liquid Argon Time Projection Chamber toolkit) from DUNE, FastCaloSim (parametrized Liquid Argon Calorimeter Simulation) from ATLAS, and ACTS (detector agnostic tracking toolkit) with developers from ATLAS, LHCb, and sPHENIX. All these experiments are investigating software solutions for heterogeneous architectures for current and future runs. While some have made interim decisions for their current runs that are focussed on NVIDIA hardware, all are facing the prospect of diversifying heterogeneous architectures for future experimental phases, and are seeking portable solutions to address them.

Primary authors: STRELCHENKO, Alexei (Fermi National Accelerator Lab. (US)); YEO, Beomki; VIREN, Brett (Brookhaven National Laboratory); LEGGETT, Charles (Lawrence Berkeley National Lab (US)); YU, Haiwang; KWOK, Ka Hei Martin (Fermi National Accelerator Lab. (US)); KNOEPFEL, Kyle; DEWING, Mark; KORTELAINEN, Matti (Fermi National Accelerator Lab. (US)); BATTACHARYA, Meghna; LIN, Meifeng (Brookhaven National Laboratory (US)); ATIF, Mohhamad (BNL); GUTSCHE, Oliver (Fermi National Accelerator Lab. (US)); CALAFIURA, Paolo (Lawrence Berkeley National Lab. (US)); HABIB, Salman (Argonne National Laboratory); CHILDERS, Taylor (Argonne National Laboratory (US)); WANG, Tianle (Brookhaven National Laboratory); TSULAIA, Vakho (Lawrence Berkeley National Lab. (US)); PASCUZZI, Vincent (Brookhaven National Laboratory); DONG, Zhihua

Presenter: LEGGETT, Charles (Lawrence Berkeley National Lab (US))

Session Classification: Poster session with coffee break

Track Classification: Track 1: Computing Technology for Physics Research