Evaluation of Performant Portable Solutions for Computing on Heterogeneous Architectures Using ATLAS Fast Calorimeter Simulation

A Snowmass 2021 Letter of Interest in Experimental Algorithm Parallelization (CompF1)

C. Leggett for the HEP-CCE/PPS[1] working group

Contact: Charles Leggett *cgleggett@lbl.gov*

Modeling physics interactions in the ATLAS detector is an essential aspect of the analysis workflow, required for proper understanding of the detector and background subtraction. Doing this simulation currently requires about 40% of ATLAS's CPU resources, and this amount is expected to grow as we enter the HL-LHC era. When using Geant4 to do the full simulation, 90% of the time is spent inside the Liquid Argon Calorimeter. Instead of doing a full simulation, a parametrized simulation of the calorimeter can produce similar physics results with a greatly reduced computational load. A version of this code, called FastCaloSim, has been ported to run on NVidia GPUs using CUDA, which significantly improves performance.

Current and next generation HPCs, as well as commercial cloud systems, are becoming increasingly heterogeneous, with ubiquitous use of GPUs. Though NVidia is the currently dominant manufacturer of GPUs, AMD and Intel are well represented in the design of several exascale HPCs that are under construction, and CUDA code is unable to run on these platforms. In order to make use of all current and future facilities, code bases must be either ported to hardware specific languages (which would be prohibitively time consuming), or to a common, hardware agnostic portable implementation. As part of the HEP-CCE/PPS project we are using FastCaloSim as a testbed to evaluate several portability solutions, beginning with Kokkos and SYCL, and continuing with OpenACC and Alpaka. If other promising portability solutions present themselves, such as standardized extensions to the C++ language, we will incorporate them into the process.

[1] https://hepcce.org

Software development does not exist in a vacuum. The FastCaloSim testbed was extracted from the ATLAS code repository into a slim, standalone package to ease code development and deployment on HPCs. This has required regular input from domain experts in ATLAS as well as software professionals. Porting the code to various portability layers is only the first part of the process - we must also validate the results, run performance tests on various hardware backends, and evaluate the entire process based on a set of metrics that we have developed.

Once the entire process is complete, recommendations will be made back to ATLAS as to how to make best use of available and future heterogeneous resources based on our experiences. The code must also be re-integrated into the ATLAS ecosystem, in such a way as to not disrupt existing activities, and also to be used as a template for developing portable solutions for utilizing heterogeneous architectures.

This LOI serves as an introduction to a whitepaper that will be used to document the process and the lessons learned, expand on the evaluation of the metrics, and detail our recommendations to not only ATLAS, but the rest of the HEP community, as well as all necessary steps that are required for re-integration of the ported code back into the ATLAS repository.