frontiers in Big Data

EDITORIAL
published: 01 July 2021
doi: 10.3389/fdata.2021.652881

Editorial: Heterogeneous Computing for AI and Big Data in High Energy Physics

Daniele D'Agostino[1]* *and Daniele Cesini*[2]

[1]IEIIT-National Research Council of Italy, Genoa, Italy, [2]CNAF-Italian Institute for Nuclear Physics, Bologna, Italy

Keywords: high performance computing, big data, artificial intelligence, high energy physics, heterogeneous computing

Editorial on the Research Topic

Heterogeneous Computing for AI and Big Data in High Energy Physics

OPEN ACCESS

Edited by:
Jean-Roch Vlimant, California Institute of Technology, United States

Reviewed by:
Anushree Ghosh, University of Padua, Italy
Xiangyang Ju, Lawrence Berkeley National Laboratory, United States
Daniele Bonacorsi, University of Bologna, Italy

***Correspondence:**
Daniele D'Agostino
daniele.dagostino@ieiit.cnr.it

Specialty section:
This article was submitted to Big Data and AI in High Energy Physics, a section of the journal Frontiers in Big Data

Received: *21 May 2021*
Accepted: *21 June 2021*
Published: *01 July 2021*

Citation:
D'Agostino D and Cesini D (2021) Editorial: Heterogeneous Computing for AI and Big Data in High Energy Physics. Front. Big Data 4:652881. doi: 10.3389/fdata.2021.652881

Heterogeneous computing denotes a scenario where different computing platforms are exploited for specific applications (Danovaro et al., 2014). While the demand for computational resources continues to grow with increasing need for querying and analyzing the volumes and rates of Big Data, energy efficiency is limiting the traditional approach to improve the compute capabilities of a data center by adding thousands of state-of-the-art x86 machines to an existing infrastructure in favor of adopting energy savvy devices (Cesini et al., 2017; D'Agostino et al., 2019). The result is that the computing nodes in data centers have different execution models, ranging from the traditional x68 architecture to GPUs, FPGAs (Papadimitriou et al., 2020) and other processor types like ARMs or more specialized processors as TPUs (Albrecht et al., 2019; Cass, 2019). For example, GPUs are used in many scientific applications based on regular domains and are delivering performance that is orders of magnitude better than traditional cores. They are also widely used in deep learning, especially the machine learning training phase. The FPGAs, being integrated circuits that can be configured by the programmer to implement a certain function, tries to close the gap between hardware and software.

Within this context the research topic collects five papers showing very interesting experiences in adopting heterogeneous architectures, for AI and Big Data applications in High Energy Physics.

In GPU-accelerated machine learning inference as a service for computing in neutrino experiments (Wang et al.) the authors discuss the performance achieved by exploiting GPU resources as a service for the ProtoDUNE-SP reconstruction chain developed in the context of the Deep Underground Neutrino Experiment (DUNE). This contribution represents one of the first experiences in which machine learning is acceleratedis accelerated with GPUs in neutrino software frameworks. The most time-consuming task, the track and particle shower hit identification, has been accelerated by a factor of 17.

In Heterogeneous reconstruction of tracks and primary vertices with the CMS pixel tracker (Bocci et al.) the authors describe an heterogeneous implementation of pixel tracks and vertices reconstruction chain on GPUs able to achieve high performance speedup values. The resulting framework has been developed for being integrated in the CMS particle detector reconstruction software, CMSSW (http://cms-sw.github.io), which is employed to detect particle and phenomena resulting from high-energy collisions in the LHC by the CMS experiment.

In Distance-Weighted Graph Neural Networks on FPGAs for Real-Time Particle Reconstruction in High Energy Physics (Iiyama et al.) the authors present a novel method to export graph neural networks from complex modern machine learning packages to an efficient FPGA implementation.

The main characteristics is that this implementation is able to perform nontrivial physics tasks with a sub-microsecond latency, therefore it represents a suitable solution for low-latency applications, such as the Level-1 trigger stage of LHC.

In CLUE: A Fast Parallel Clustering Algorithm for High Granularity Calorimeters in High-Energy Physics (Rovere et al.) the authors propose a novel density-based clustering algorithm exploiting a new, highly parallelizable way of assigning points to clusters. Experiments have been performed on a synthetic dataset that resembles high occupancy events in high granularity calorimeters operated at HL-LHC with speedup values up to 112 on a GPU.

In Porting HEP Parameterized Calorimeter Simulation Code to GPUs (Leggett et al.) the authors describe an accelerated version of the ATLAS Calorimeter Fast-Simulation on GPUs. Several techniques have been incorporated the implementation to increase the GPU overall utilization, because the compute-to-I/O ratio is not very large, and the achieved results are interesting. A key topic of the work is represented by the introduction of performance portability through Kokkos, a C++ cross-device compatibility library. Authos discuss the different issues encountered during the porting process, the adopted solutions and present a comparison of the achievable performance with respect to previous approaches.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

REFERENCES

Albrecht, J., Alves, A. A., Amadio, G., Andronico, G., Anh-Ky, N., Aphecetche, L., et al. (2019). A Roadmap for HEP Software and Computing R&D for the 2020s. *Comput. Softw. big Sci.* 3 (1), 7. doi:10.1007/s41781-018-0018-8

Cass, S. (2019). Taking AI to the Edge: Google's TPU Now Comes in a Maker-Friendly Package. *IEEE Spectr.* 56 (5), 16–17. doi:10.1109/mspec.2019.8701189

Cesini, D., Corni, E., Falabella, A., Ferraro, A., Morganti, L., Calore, E., et al. (2017). *Power-efficient Computing: Experiences from the COSA Project*. Scientific Programming, 7206595.

D'Agostino, D., Quarati, A., Clematis, A., Morganti, L., Corni, E., Giansanti, V., et al. (2019). SoC-based Computing Infrastructures for Scientific Applications and Commercial Services: Performance and Economic Evaluations. *Future Generation Comp. Syst.* 96, 11–22. doi:10.1016/j.future.2019.01.024

Danovaro, E., Clematis, A., Galizia, A., Ripepi, G., Quarati, A., and D'Agostino, D. (2014). Heterogeneous Architectures for Computational Intensive Applications: A Cost-Effectiveness Analysis. *J. Comput. Appl. Math.* 270, 63–77. doi:10.1016/j.cam.2014.02.022

Papadimitriou, G., Chatzidimitriou, A., Gizopoulos, D., Reddi, V. J., Leng, J., Salami, B., et al. (2020). Exceeding Conservative Limits: A Consolidated Analysis on Modern Hardware Margins. *IEEE Trans. Device Mater. Relib.* 20 (2), 341–350. doi:10.1109/tdmr.2020.2989813

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.