Data Integration Tasks on Heterogenous Systems Using OpenCL

Clayton J. Faber
Anthony M. Cabrera
Orondé Booker
Gabe Maayan
Roger D. Chamberlain

Clayton J. Faber, Anthony M. Cabrera, Orondé Booker, Gabe Maayan, and Roger D. Chamberlain, “Data Integration Tasks on Heterogeneous Systems Using OpenCL,” in *Proc. of 7th International Workshop on OpenCL* (IWOCL), May 2019. DOI: 10.1145/3318170.3318187

Washington University in St. Louis
Rensselaer Polytechnic Institute

Data Integration Tasks on Heterogeneous Systems Using OpenCL

Clayton J. Faber
Washington University in St. Louis
St. Louis, MO, USA
cfaber@wustl.edu

Anthony M. Cabrera
Washington University in St. Louis
St. Louis, MO, USA
acabrera@wustl.edu

Orondé Booker
Washington University in St. Louis
St. Louis, MO, USA
booker.oronde@wustl.edu

Gabe Maayan
Rensselaer Polytechnic Institute
Troy, NY, USA
maayag@rpi.edu

Roger D. Chamberlain
Washington University in St. Louis
St. Louis, MO, USA
roger@wustl.edu

ABSTRACT

In the era of big data, many new algorithms are developed to try and find the most efficient way to perform computations with massive amounts of data. However, what is often overlooked is the preprocessing step for many of these applications. The Data Integration Benchmark Suite (DIBS) [1] was designed to understand the characteristics of dataset transformations in a hardware agnostic way. While on the surface these applications have a high amount of data parallelism, there are caveats in their specification that can potentially affect this characteristic. Even still, OpenCL can be an effective deployment environment for these applications.

In this work we take a subset of the data transformations from each category presented in DIBS and implement them in OpenCL to evaluate their performance for heterogeneous systems. For targeting heterogeneous systems, we take a common application and attempt to deploy it to three platforms targetable by OpenCL (CPU, GPU, and FPGA). The applications are evaluated by their average transformation data rate (see Figure 1). We illustrate the advantages of each compute device in the data integration space along with different communications schemes allowed for host/device communication in the OpenCL platform.

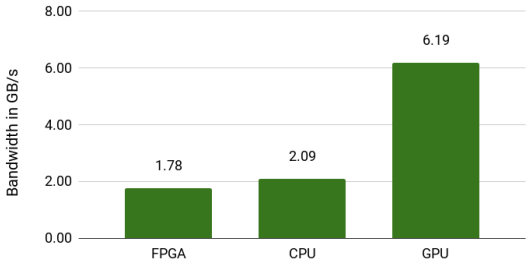

Figure 1: Performance results for IDX→TIFF application.

Figure 1 shows the performance for one of the DIBS applications. In this case, the GPU performance is clearly superior to both the CPU and the FPGA. The CPU execution is on a single core, exploiting vector instructions for parallelism. The FPGA is an unrolled `for`-loop pipelined implementation. The FPGA performance represents a speedup of 134× over the baseline sequential implementation from [1].

The primary distinguishing factor among the applications we consider is the following: whether or not there is an apparent sequential dependency in the specification of the data integration task to be performed. Several of the applications have no such dependency (i.e., they are embarrassingly parallel at the level of individual data elements), and subsequently perform quite well on each of the target platforms. The more interesting cases are those for which there is a sequential dependency (e.g., parsing comma-separated fields), and considerably more effort must be expended to enable these applications to perform well.

The IDX→TIFF application has such a sequential dependency, which has a negative impact on the FPGA performance specifically. The remaining benchmarks illustrate a range of circumstances in this regard.

CCS CONCEPTS

• **Computer systems organization → Reconfigurable computing**; **Heterogeneous (hybrid) systems**.

ACM Reference Format:
Clayton J. Faber, Anthony M. Cabrera, Orondé Booker, Gabe Maayan, and Roger D. Chamberlain. 2019. Data Integration Tasks on Heterogeneous Systems Using OpenCL. In *International Workshop on OpenCL (IWOCL'19), May 13–15, 2019, Boston, MA, USA.* ACM, New York, NY, USA, 1 page. https://doi.org/10.1145/3318170.3318187

ACKNOWLEDGMENTS

Supported by NSF grants CNS-1205721, CNS-1527510, CCF-1527692, and CNS-1763503. Thanks to Intel for access to the CPU+FPGA system through the Hardware Accelerator Research Program.

REFERENCES

[1] Anthony M Cabrera, Clayton J Faber, Kyle Cepeda, Robert Derber, Cooper Epstein, Jason Zheng, Ron K Cytron, and Roger D Chamberlain. 2018. DIBS: A Data Integration Benchmark Suite. In *Proc. of ACM/SPEC International Conference on Performance Engineering Companion.* ACM, 25–28.

IWOCL'19, May 13–15, 2019, Boston, MA, USA

ACM ISBN 978-1-4503-6230-6/19/05.
https://doi.org/10.1145/3318170.3318187