BITFLEX: A Dynamic Runtime Library for Bit-Level Precision Manipulation and Approximate Computing

Ryan Barton
Tokyo Institute of Technology
AIST Real World Big-Data Computation Open Innovation Laboratory (RWBC-OIL)
Tokyo, Japan
barton.r.aa@m.titech.ac.jp

Mohamed Wahib
AIST Real World Big-Data Computation Open Innovation Laboratory (RWBC-OIL)
Tokyo, Japan
mohamed.attia@aist.go.jp

Artur Podobas
RIKEN Center for Computational Science
Kobe, Japan
artur.podobas@riken.jp

Satoshi Matsuoka
RIKEN Center for Computational Science
Kobe, Japan
matsu@acm.org

1 INTRODUCTION

Varying precision in floating-point arithmetic operations is an attractive means of boosting the performance of High-Performance Computing (HPC) applications. However, current approaches needlessly burden the developer with both the implementation of mixed precision, conventionally done by manually modifying the application, and ensuring numerical stability. In this work, we propose a method that transparently allows free control over numerical precision at runtime.

We build a custom framework called BITFLEX, based on the open-source MCXX compiler from Barcelona Supercomputing Center and the OpenMP parallel programming model. Fundamentally it enables the user to intercept and modify floating-point operations and their format completely at will. We further generalize the framework to support nondeterministic applications by designing a new OpenMP construct, denoted as **#pragma omp nondeter <domain type> <parameter(s)>**.

The motivation for this approach is threefold. Firstly, implementing BITFLEX is as simple as denoting a code block to intercept, rather than overhauling source code. Secondly, it gives HPC developers the ability to observe error propagation at a granularity of their choosing. Lastly, it offers insight into how HPC applications can utilize specialized hardware of the future.

2 METHODOLOGY AND RESULTS

We initially demonstrate the impact and implications of manipulated precision using Intel's PIN instrumentation tool. Our original pre-processing directive on a demo application reveals inflection points from truncation error to round-off error as precision increases. This is an intuition corroborated by floating-point compression work at Lawrence Livermore National Lab [1].

BITFLEX is also designed with a variety of approximate computing domain parameters. These include the following outputs: accuracy (error tolerance), repeat (iterations), adaptive precision (varying by a bit mask function), range (upper and lower bounds), and confidence (statistical confidence value between 0.0 and 1.0). Furthermore, we extend functionality of ADAPT [2], an error propagation analysis tool that employs an automatic differentiation (AD) based greedy algorithm to offer mixed precision suggestions in a target application. Other application possibilities include PDE solvers, quantum simulators, and DNNs. We will analyze error propagation in our candidate applications and reason around the expected increase in performance that our method indirectly yields.

3 CONCLUSION

BITFLEX serves as a contribution to the error tracking and approximate computing domains in HPC. As supercomputers move toward exascale, computation will become more inexact, and data more inaccurate and unpredictable. Our work keeps in mind specialized chipsets, low-power memory, and other emerging technologies, while also realizing performance gains. To that end, we demonstrate the flexibility of our library and propose it as a potential addition to a future release of the OpenMP standard.

REFERENCES

[1] James Diffenderfer, Alyson Fox, Jeffrey Hittinger, Geoffrey Sanders, and Peter Lindstrom. 2019. Error Analysis of ZFP Compression for Floating-Point Data. *SIAM Journal on Scientific Computing* 41 (02 2019), A1867–A1898. https://doi.org/10.1137/18M1168832

[2] Harshitha Menon, Michael O. Lam, Daniel Osei-Kuffuor, Markus Schordan, Scott Lloyd, Kathryn Mohror, and Jeffrey Hittinger. 2018. ADAPT: Algorithmic Differentiation Applied to Floating-point Precision Tuning. In *SC18*. IEEE Press, Piscataway, NJ, USA, Article 48, 13 pages. https://doi.org/10.1109/SC.2018.00051