

OOKAMI PROJECT APPLICATION

Date: Sept 21, 2021

Project Title: Deep Learning Compilers To Enhance ML Performance on Ookami

Usage:

- Testbed

Principal Investigator: Tony Curtis

- University/Company/Institute: Stony Brook University
- Mailing address including country:

Institute for Advanced Computational Science,
Stony Brook University,
Stony Brook,
NY 11794-5250,
USA

- Phone number: (631) 632-4629
- Email: anthony.curtis@stonybrook.edu

Names & Email of initial project users:

- Gaurav Verma <gaurav.verma@stonybrook.edu>
- Smeet Chheda <schheda@cs.stonybrook.edu>
- Abid M. Malik <amalik@bnl.gov>

Usage Description:

We will be performing an in-depth study of DL Compiler architectures developed to support Machine Learning. Initially, we will explore TF-XLA and Apache TVM to analyze existing compiler passes and propose new possible optimizations, improving inference. Further, it will include profiling of various optimizations, lowering paradigms and a comparative analysis of the results (e.g. power consumption, memory usage, inference time, etc.).

At the moment, we have limited knowledge about how Ookami can efficiently support the popular (deep) machine learning compilers such as TVM and XLA. We expect to learn how these compilers can collaborate with the A64fx architecture, as various models typically include many highly parallel operations. In this work, we will also optimize the different operation mappings between the A64fx architecture and DL Compilers to leverage SVE.

Computational Resources:

- Total node hours per year: estimate 1000
- Size (nodes) and duration (hours) for a typical batch job: Runs can range from using a single GPU to all available GPUs on a single node. Runs often range from a few minutes to a few hours (e.g. running DL benchmarks, building DL compilers, XLA, TVM).
- Disk space (home, project, scratch): 40GB, 4TB, 4TB

Personnel Resources (assistance in porting/tuning, or training for your users):

None anticipated

Required software:

Probably none extra.

If your research is supported by US federal agencies:

- Agency: Department of Energy
 - Grant number(s): CREDIT <http://credit.pvamu.edu/index.html>
-