An Open Source Framework for Generating Modular DNN Accelerators supporting Flexible Dataflow

Tushar Krishna
Assistant Professor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Email: tushar@ece.gatech.edu

Efficiently tiling and mapping high-dimensional convolutions onto limited execution and buffering resources is a challenge faced by all deep learning accelerators today. We term each unique approach as dataflow. The dataflow determines overall throughput (utilization of the compute units) and energy-efficiency (reads, writes, and reuse of model parameters and partial sums across the accelerator’s memory hierarchy). The research community today lacks a simulation infrastructure to evaluate DNN dataflows and architectures systematically and reason about performance, power, and area implications of various design choices.

In this work, we first present a framework called MAESTRO to formally describe and analyze DNN dataflows, and predict roofline performance and energy-efficiency when running neural network layers, and report the hardware resources (size of buffers across the memory hierarchy, and network-on-chip (NoC) bandwidth) to support this dataflow. Using this, we demonstrate that adapting dataflow at runtime can provide significant benefits.

Adaptive dataflows require flexibility from the interconnect within the DNN accelerator. To address this need, next, we present MAERI, which is a programmable DNN accelerator built with a set of modular and configurable building blocks that can easily support myriad DNN partitions and mappings by appropriately configuring the interconnect within the accelerator. MAERI can run various DNN layers (CNNs, LSTMs, max pooling, fully-connected) and support myriad loop orderings, tiling strategies, and optimizations (such as weight-pruning), while providing near 100% utilization of the compute resources.

This work was in collaboration with Hyoukjun Kwon (PhD student, Georgia Tech) and Dr. Michael Pellauer (Research Scientist, NVIDIA).

References:
MAERI: http://synergy.ece.gatech.edu/tools/maeri/