Seung Hun Chung

sh.c@live.ca
shchung.com

Education

2018 - 2021 **Master of Applied Science (MASc)**, *University of Toronto*
Supervisor: Tarek S. Abdelrahman, Computer Engineering (CGPA: 4.0/4.0).
Research: Optimizing compiler-generated OpenCL kernels for FPGA acceleration of CNNs
2019-2020 QE II Graduate Student Scholarships in Science and Technology (QEII-GSST) Recipient

S. Chung and T. S. Abdelrahman, "**Optimization of Compiler-Generated OpenCL CNN Kernels and Runtime for FPGAs**", 2022 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), 2022.

S. Chung and T. S. Abdelrahman, "**A Compilation Flow for the Generation of CNN Inference Accelerators on FPGAs**", arXiv:2203.04015 [cs.DC].

2013 - 2018 **Bachelor of Applied Science (BASc)**, *University of Toronto*
Graduated with honours distinction in Computer Engineering (CGPA: 3.4/4.0)

Experience

Aug 2021 - Present **Senior Software Engineer, Compilers**, *Untether AI*
- Creating an optimizing compiler on the MLIR compiler stack for runAI200: a spatial, in-memory architecture for deep learning inference acceleration
- Design kernel level dialect, conversion passes from high-level tensor operations to kernel programs, and integrating a constraint programming solver for performance-aware kernel and parameter selection
- Worked with kernel developers to expand compiler coverage and flexibly support a variety of input models, including classification, detection, and segmentation networks
- Resulting front-end compiler infrastructure improved automatic compilation capabilities, reduced turnaround time on internal/client demos and customer model compilation while simultaneously improving model throughput
- Optimized average compile time across 8 benchmark models by 10×

Sep 2018 - Apr 2020 **Teaching Assistant**, *University of Toronto*
- Head teaching assistant for a C++ programming fundamentals course, responsible for operations and grading
- Supported students with C programming assignments for operating systems and a SW project course

May 2016 - Aug 2017 **Software Engineering Intern**, *AMD Radeon Technologies Group*
- 16-month internship for Windows AMD GPU driver development (PPLib)
- Awarded 1st place in AMD Markham Innovation Showcase (2016) for Advanced Application Profiling, a feature to dynamically change chip voltage and clock on a per-application basis, saving up to 20% power
- Built a fully-automated research infrastructure for GPU power-saving initiatives
- Implemented features related to voltage, clock, and fan control
- Debugged driver issues related to device thermals, kernel-mode crashes, and gaming performance
- Replaced driver component testing with Jenkins, migrated existing tests and expanded test coverage

Summer 2015 **Operation Lead**, *ARB Labs*
- Startup consisted of 5 engineers building prototypes for casino gaming security
- Created a card detection device for blackjack tables, wrote software for card gaming analytics including image capture, classification APIs, databases

Summer 2014 **IT Intern**, *HDI-Gerling Industrial Insurance*
- Troubleshoot IT-related issues and provided technical support to office as only member of on-site IT
- Wrote internal documentation for proprietary data entry software

Projects

Reconfigurable Hardware Implementations of Neural Networks, *Capstone Project*
Implemented a matrix-multiply engine on Xilinx Zynq-7000 SoC/FPGA with Xilinx Vivado HDL C++ and Caffe for the purpose of accelerating inference with small CNNs on FPGA