

Hyoukjun Kwon

PH.D. CANDIDATE

Klaus Advanced Computing Building 3305, 266 Ferst Dr. NW, Atlanta, GA 30332

☎ (+1) 404-539-4457 | ✉ hyoukjun@gatech.edu | 🏠 www.hyoukjunkwon.com | 📺 hyoukjun-kwon

Professional Experience

Facebook, Menlo Park, CA

Research Scientist at AR/VR AI Research (Accepted Offer)

Sep. 2020 -

Manager: Dr. Vikas Chandra

Facebook, Menlo Park, CA

Research Intern at AR/VR AI Research

May. 2019 - July. 2019

Manager: Dr. Vikas Chandra, Mentor: Dr. Liangzhen Lai

NVIDIA, Westford, MA

Research Intern at Architecture Research Group (ARG)

May. 2018 - Aug. 2018

Manager: Dr. Steve Keckler, Mentor: Dr. Michael Pellauer

NVIDIA, Westford, MA

Research Intern at Architecture Research Group (ARG)

May. 2017 - Aug. 2017

Manager: Dr. Steve Keckler, Mentor: Dr. Michael Pellauer

Georgia Institute of Technology, Atlanta, GA

Graduate Research Assistant

Aug. 2015 - present

Education

Georgia Institute of Technology

Ph.D. student in Computer Science

Aug. 2015 - Current

Advisor: Dr. Tushar Krishna

Seoul National University (SNU)

B.S. in CSE (Computer Science and Engineering)

Mar. 2007 - Feb. 2015

B.S. in EMS (Environmental Material Science)

Advisor: Dr. Jihong Kim (CSE) and Dr. Junjae Lee (EMS)

Research Interest

Deep neural network (DNN) accelerators

DNN dataflow optimization

Interconnection network (NoCs)

Reconfigurable architecture

Computer architecture

Machine learning

Design automation

Hardware security

Services

Journal Reviewer

IEEE Computer Architecture Letters

2020

Journal Reviewer

IEEE MICRO

2019

Journal Reviewer

ACM Transactions on Architecture and Code Optimization (TACO)

2019, 2020

Journal Reviewer

IEEE Transactions on Computers

2019, 2020

Talks in Industry

Understanding Dataflow in DNN Accelerators and Modeling Them with MAESTRO

Sandia National Lab (Online Talk)

Dec. 2019

Understanding Dataflows in DNN Accelerators

Facebook, Menlo Park, California

Jun. 2019

An Open Source Framework for Exploring Dataflow and Generating DNN Accelerators Supporting Flexible Dataflow

IBM Research, Yorktown Heights, New York

Nov. 2018

Communication-driven Approach to Design DNN Accelerators

Western Digital, Online Talk

Sep. 2018

Analyzing Dataflows in Accelerators

NVIDIA, Westford, Massachusetts

Aug. 2018

Optimizing Networks-On-Chip for Deep Learning Accelerators using Micro-switches

NVIDIA, Westford, Massachusetts

Aug. 2017

Automatic generation of low-latency networks-on-chip

Bluespec Inc., Framingham, Massachusetts

Jul. 2017

Talks in Academia

Understanding Reuse, Performance, and Hardware Cost of DNN Accelerator Dataflows

Seoul National University, Seoul, Korea

Jan. 2020

Modeling and Analyzing Dataflows in DNN Accelerators

Tokyo City University, 2018, Tokyo, Japan

Dec. 2018

Enabling Rapid Design Space Exploration and Prototyping of DNN Accelerators

Massachusetts Institute of Technology (MIT), 2018, Cambridge, Massachusetts

Jul. 2018

A Communication-driven Approach to Designing Flexible DNN Accelerators

CMU, Pittsburgh, Pennsylvania

May. 2018

[Teaching] Designing CNN Accelerators using Bluespec System Verilog

Seoul National University (SNU), Seoul, Korea

A three-day lecture for undergraduate students

(<https://github.com/hyoukjun/DesignCNNAccelerators>)

Dec. 2017

Light-weight and High-performance NoC for DNN accelerators

Konkuk University, Seoul, Korea

Oct. 2017

Talks at Conferences/Workshops

Understanding the Impact of On-Chip Communication on DNN Accelerator Performance

MICRO 2019, Columbus, Ohio

Oct. 2019

[Tutorial] Enabling Rapid Design Space Exploration and Prototyping of DNN Accelerators

HPCA 2019, Washington D.C.

<http://synergy.ece.gatech.edu/tools/maeri/maeri-tutorial-hpca-2019/>

Feb. 2019

[Tutorial] MAERI: Enabling Rapid Design Space Exploration and Prototyping of DNN Accelerators

ISCA 2018, Los Angeles, California

http://synergy.ece.gatech.edu/tools/maeri/maeri_tutorial_isca2018/

Mar. 2018

MAERI: Enabling Flexible Dataflow Mapping over DNN Accelerators via Reconfigurable Interconnects

ASPLOS 2018, Williamsburg, Virginia

Mar. 2018

[Demo] MAESTRO: An Open-source Infrastructure for Modeling Dataflows within Deep Learning Accelerators

Corgarch 2018 (colocated with ASPLOS 2018), Williamsburg, Virginia

Mar. 2018

[Demo] OpenSMART: An Opensource Single-cycle Multi-hop NoC Generator

SC17 OpenSuCo 2017, Denver, Colorado

Nov. 2017

Rethinking NoCs for Spatial Neural Network Accelerators

NOCS 2017, Seoul, Korea

Oct. 2017

Adaptive Manycore Architectures for Big Data Computing

NOCS 2017, Seoul, Korea

Oct. 2017

Proving Flow Security of Sequential Logic via Automatically-Synthesized Relational Invariants

CSF 2017, Santa Barbara, California

Aug. 2017

OpenSMART: Single-cycle Multi-hop NoC Generator in BSV and Chisel

ISPASS 2017, Santa Rosa, California

May. 2017

Peer-reviewed Publications

Lei Yang, Zheyu Yan, Meng Li, **Hyoukjun Kwon**, Liangzhen Lai, Tushar Krishna, Vikas Chandra, Weiwen Jiang, Yiyu Shi, “Co-Exploration of Neural Architectures and Heterogeneous ASIC Accelerator Designs Targeting Multiple Tasks”, *The 57th Annual Design Automation Conference (DAC)*, 2020 (to appear)

Eric Qin, Ananda Samajdar, **Hyoukjun Kwon**, Vineet Nadella, Sudarshan Srinivasan, Dipankar Das, Bharat Kaul, and Tushar Krishna, “SIGMA: A Sparse and Irregular GEMM Accelerator with Flexible Interconnects for DNN Training”, *International Symposium on High-Performance Computer Architecture (HPCA)*, 2020

Received the best paper award

Robert Guirado, **Hyoukjun Kwon**, Sergi Abadal, Eduard Alarcon, and Tushar Krishna, “Understanding the Impact of On-Chip Communication on DNN Accelerator Performance”, *International Conference on Electronics Circuits and System (ICECS)*, 2019

Hyoukjun Kwon, Prasanth Chatarasi, Michael Pellauer, Angshuman Parashar, Vivek Sarkar, and Tushar Krishna, “Understanding Reuse, Performance, and Hardware Cost of DNN Dataflows: A Data-Centric Approach”, *International Symposium on Microarchitecture (MICRO)*, 2019

Selected as One of the IEEE Micro Top Picks (2020 issue)

Jinwoo Kim, Gauthaman Murali, Heechun Park, Eric Qin, **Hyoukjun Kwon**, Venkata Chaitanya Krishna, Nihar Dasari, Arvind Singh, Minah Lee, Hakki Torun, Kallol Roy, Madhavan Swaminathan, Saibal Mukhopadhyay, Tushar Krishna, and Sung Kyu Lim, “mRNA: Enabling Efficient Mapping Space Exploration on a Reconfigurable Neural Accelerator”, *Design Automation Conference (DAC)*, 2019

Zhongyuan Zhao, **Hyoukjun Kwon**, Sachit Kuhar, Weiguang Sheng, Zhigang Mao, and Tushar Krishna, “mRNA: Enabling Efficient Mapping Space Exploration on a Reconfigurable Neural Accelerator”, *International Symposium on Performance Analysis of Systems and Software (ISPASS)*, 2019

Hyoukjun Kwon, Ananda Samajdar, and Tushar Krishna, “A Communication-driven Approach for Designing Flexible DNN Accelerators”, *IEEE Micro Special Issue on Hardware Acceleration (IEEE Micro)*, 2018

Brian Lebednik, Sergi Abadal, **Hyoukjun Kwon** and Tushar Krishna, “Architecting a Secure Wireless Network-on-Chip”, *International Symposium on Network-on-Chips (NOCS)*, 2018

Hyoukjun Kwon, Ananda Samajdar, and Tushar Krishna, “MAERI: Enabling Flexible Dataflow Mapping over DNN Accelerators via Reconfigurable Interconnects”, *International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, 2018
Honorable mention in IEEE Micro Top Picks (2019 issue)

Hyoukjun Kwon, Ananda Samajdar, and Tushar Krishna, “MAERI: Enabling Flexible Dataflow Mapping over DNN Accelerators via Reconfigurable Interconnects”, *The Inaugural Sysml Conference (Sysml)*, not archived, 2018

Brian Lebednik, Sergi Abadal, **Hyoukjun Kwon**, and Tushar Krishna, “Spoofing Prevention via RF Power Profiling in Wireless Network-on-Chip”, *International Workshop on Advanced Interconnect Solutions and Technologies for Emerging Computing Systems (AISTECS)*, 2018

Hyoukjun Kwon, Ananda Samajdar, and Tushar Krishna, “Rethinking NoCs for Spatial Neural Network Accelerators”, *International Symposium on Networks-on-Chips (NOCS)*, 2017

Janardhan Rao Doppa, Ryan Gary Kim, Mihailo Isakov, Michel A. Kinsy, **Hyoukjun Kwon**, and Tushar Krishna, “Adaptive Manycore Architectures for Big Data Computing”, *International Symposium on Networks-on-Chips (NOCS)*, 2017

Hyoukjun Kwon, William Harris, and Hadi Esmaeilzadeh, “Proving Flow Security of Sequential Logic via Automatically Synthesized Relational Invariants”, *Computer Security Foundations (CSF)*, 2017

Hyoukjun Kwon and Tushar Krishna, “OpenSMART: Single-Cycle Multi-hop NoC Generator in BSV and Chisel”, *International Symposium on Performance Analysis of Systems and Software (ISPASS)*, 2017

Hyoukjun Kwon, Dohyun Kim, Jisung Park, and Jihong Kim, “Improving the Lifetime of NAND Flash-based Storages Using MADE (Minhash-Assisted Delta-compression Engine)”, *Korean Institute of Information Science and Engineers Annual Conference (KIISE)*, 2014

Book Chapter

Tushar Krishna, **Hyoukjun Kwon**, Angshuman Parashar, Michael Pellauer, and Ananda Samajdar (lexicographical order of last names), “Synthesis lecture on computer architecture: Data Orchestration in DNN Accelerators”, *Morgan & Claypool*, Planned to be published in 2020

Preprint

Hyoukjun Kwon, Liangzhen Lai, Tushar Krishnam and Vikas Chandra, “HERALD: Optimizing Heterogeneous DNN Accelerators for Edge Devices”, *Arxiv preprint*, 2019

Research Projects

Reserch on artificial intelligence-focused architectures and algorithms

Georgia Institute of Technology and Sandia National Lab

Dec. 2019 - PRESENT

- Tech lead role from Georgia Tech side
- <https://www.sandia.gov/news/publications/labnews/articles/2019/11-08/ARIAA.html>

Exploiting data and compute reuse in DNN accelerators

Georgia Institute of Technology

Oct. 2017 - PRESENT

- Developing a DNN accelerator-framework co-design method to exploit both of data and compute reuse in DNN accelerators

Heterogeneous DNN Accelerator Design

Facebook and Georgia Institute of Technology

May. 2019 - PRESENT

- Explored costs and benefits of heterogeneous DNN accelerators
- Developed a heterogeneous DNN accelerator optimization framework
- Submitted a paper as the first a author

Modeling performance and energy cost/benefits of neural network mapping (dataflows) on accelerators (MAESTRO)

Georgia Institute of Technology and NVIDIA

Oct. 2017 - Sep. 2019

- Developed an analytic model (MAESTRO) for CNN problem mapping on accelerator architectures
- Published a paper as the first a author in MICRO 2019

Fine-grained programmability support in DNN accelerators

Georgia Institute of Technology

Oct. 2017 - PRESENT

- Developing a fine-grained ISA (Instruction Set Architecture for DNN accelerators to support dataflow programming
- Submitted a paper as the second author

DARPA - CHIPS project (Communication architecture support for highly modular chip design)

Georgia Institute of Technology

Sep. 2017 - Aug. 2019

- Worked on architecture level exploration and NoC designs

Enabling flexible dataflow in DNN accelerators

NVIDIA and Georgia Institute of Technology

May. 2017 - PRESENT

- Developed a light-weight and reconfigurable NoC architecture to support flexible dataflow in DNN acclerators
- Developed a dataflow cost analysis framework whose dataflow description can be plugged into DNN accelerator RTL as configuration

A communication-aware flexible dataflow DNN accelerator (MAERI)

Georgia Institute of Technology

Jan. 2017 - Jan. 2018

- Developed a new DNN accelerator microarchitecture for CNN and RNN that offloads computation into network side and maximize the utilization of computation units.
- Implemented RTL using Bluespec System Verilog.
- Published a paper as the first author in ASPLOS 2018 and Sysml.

Secure wireless NoC system

Georgia Institute of Technology

Aug. 2016 - Jan. 2018

- Worked on a method that prevents hardware Trojan that monopolies wireless network channel bandwidth and slows down other cores.
- Published a paper as a co-author in AISTECS 2018

Scalable NoC design for DNN accelerators

Georgia Institute of Technology

Aug. 2016 - May. 2017

- Developed a light-weight and low-latency interconnection network for CNN accelerators.
- Implemented distributed architecture of network fabric that maximizes PE utilization.
- Implemented RTL using Bluespec System Verilog.
- Published a paper as the first author in NOCS 2017.

Advanced low-latency network-on-chip generator

Georgia Institute of Technology

Nov. 2015 - Oct. 2016

- Developed a network-on-chip RTL generator that provides state-of-the-art mesh and single-cycle-multi-hop (SMART) network.
- Verified the design comparing RTL simulation results with those of Garnet under synthetic workloads and synthesized the design for ASIC and FPGA
- Provided the source code as opensource; now it is used by Bluespec Inc. IIT Madras for RISC-V Shakti Project, and Intrinsix Corp.
- Published a paper as the first author in ISPASS 2017.

Information flow-secure FPGA accelerators

Georgia Institute of Technology

Oct. 2015 - Aug. 2016

- Developed an information flow policy language and a policy checker using a state-of-the-art formal verification method.
- Successfully identified insecure open-source hardware designs within several minutes on a laptop.
- Published a paper as the first author in CSF 2017.

Endurance-aware flash transition layer design

Seoul National University

Dec. 2014 - Sept. 2015

- Developed an FTL (flash-transition-layer) that delta-compresses write data toward flash memory and efficiently manages the mapping information of delta-compression
- Reduced significant write traffic toward flash memory, which enhances the lifetime of flash memories.
- Published a paper as the first author in KIISE 2014.

Teaching

Teaching Assistant in Hardware for Machine Learning Course (ECE8893B)

Georgia Institute of Technology

Sep. 2018 - Dec. 2018

- Taught some of classes; dataflow modeling, LSTM, and a case study
- Developed two lab assignments; design and mapping space exploration
- Q&A via Piazza forum

Teaching Assistant in Advanced Computer Architecture Course (CS6290/ECE6100)

Georgia Institute of Technology

Sep. 2018 - Dec. 2018

- Helped students in recitation session and office hour
- Q&A via Piazza forum

Special Lecture: Designing a CNN accelerator Using Bluespec System Verilog

Seoul National University

Dec. 2017

- Taught a 3-day course about CNN accelerator design for CS and ECE undergraduate students.
- Developed lecture materials and lab code, which lets students implement a full CNN accelerator.

Teaching Assistant in Processor Design Course

Georgia Institute of Technology

Jan. 2017 - May 2017

- Helped students with processor design assignments on Altera FPGA boards.
- Graded exams and assignments

Teaching Assistant in Computer Architecture Course

Seoul National University

Sep. 2014 - Dec. 2014, Mar.
2015 - Jun. 2015

- Developed lab assignments that implements a Y86 processor in Bluespec System Verilog.
- Verified the Y86 processor design on an FPGA.
- Helped developing course material.

Skills and Experiences

Software languages

C/C++, JAVA, VBA, OCaml, Scheme, Matlab, and Python

Typesetting System

Latex

Machine Learning Frameworks

PyTorch

Formal verification tools

Coq Proof Assistant and ABC

Hardware languages

Verilog, Chisel, System C, and Bluespec System Verilog

ASIC synthesis tools

Synopsys Design Compiler and Cadence Encounter(Innovus)

FPGA synthesis tools

Xilinx Vivado and Altera Quartus

Parallel programming

OpenMP, OpenCL, and MPI

Image processing

OpenCV and Matlab

Languages

Korean (Native), English (Fluent), and Japanese (Proficient)