

Hyoukjun Kwon

RESEARCH SCIENTIST @ META (FACEBOOK)
1 Hacker Way, Menlo Park, CA 94025

☎ (+1) 404-539-4457 | ✉ hyoukjun.kwon.p@gmail.com | 🏠 www.hyoukjunkwon.com | 🌐 hyoukjun-kwon

Professional Experience

Meta (Facebook), Menlo Park, CA

Research Scientist at Meta (Facebook) Reality Labs
Manager: Dr. Liangzhen Lai

Oct. 2020 - Present

Facebook, Menlo Park, CA

Research Intern at AR/VR AI Research
Manager: Dr. Vikas Chandra, Mentor: Dr. Liangzhen Lai

May. 2019 - July. 2019

NVIDIA, Westford, MA

Research Intern at Architecture Research Group
Manager: Dr. Steve Keckler, Mentor: Dr. Michael Pellauer

May. 2018 - Aug. 2018

NVIDIA, Westford, MA

Research Intern at Architecture Research Group
Manager: Dr. Steve Keckler, Mentor: Dr. Michael Pellauer

May. 2017 - Aug. 2017

Georgia Institute of Technology, Atlanta, GA

Graduate Research Assistant

Aug. 2015 - Jul. 2020

Education

Georgia Institute of Technology

PhD in Computer Science

Aug. 2015 - Jul. 2020

- Advisor: Prof. Tushar Krishna and Dr. Michael Pellauer (co-advisor)
- Committee: Prof. Vivek Sarkar, Prof. Hyesoon Kim, and Prof. Alexey Tumanov
- Thesis Title: Data- and Communication-centric Approaches to Model and Design Flexible Deep Neural Network Accelerators
- Honor: Honorable mention, ACM SIGARCH/IEEE CS TCCA Outstanding dissertation award, (Selected as one of three best dissertations in computer architecture in 2020)

Seoul National University (SNU)

BS in CSE (Computer Science and Engineering)
BS in EMS (Environmental Material Science)

Mar. 2007 - Feb. 2015

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

Honors & Awards

- 2021 **Honorable Mention**, ACM SIGARCH/IEEE CS TCCA Outstanding Dissertation Award 2021
- 2020 **Best Paper Award**, HPCA 2020
- 2020 **Top Pick**, IEEE MICRO Top Picks from 2019 Computer Architecture Conferences
- 2019 **Finalist**, Qualcomm innovation fellowship
- 2019 **Honorable Mention, Top Pick**, IEEE MICRO Top Picks from 2018 Computer Architecture Conferences
- 2018 **Finalist**, ACM Student research competition (SRC) at MICRO 2018

Research Interests

Computer architecture

HW accelerator for deep learning (DL)

Mapping and dataflow optimization on accelerators

Cross-stack optimization of AI systems

Network-on-Chips (NoCs)

Machine learning

Book

Tushar Krishna, **Hyoukjun Kwon**, Angshuman Parashar, Michael Pellauer, and Ananda Samajdar (*alphabetical order*), “*Synthesis lecture on computer architecture: Data Orchestration in Deep Learning Accelerators*”, Morgan & Claypool Publishers, August 2020

Peer-reviewed Publications

Prasanth Chatarasi, **Hyoukjun Kwon**, Angshuman Parashar, Michael Pellauer, Tushar Krishna and Vivek Sarkar, “*Marvel: A Data-Centric Approach for Mapping Deep Learning Operators on Spatial Accelerators*”, *ACM Transactions on Architecture and Code Optimization (TACO)*, 2021

Gordon E Moon, **Hyoukjun Kwon**, Geonhwa Jeong, Prasanth Chatarasi, Sivasankaran Rajamanickam, Tushar Krishna, “*Evaluating Spatial Accelerator Architectures with Tiled Matrix-Matrix Multiplication*”, *IEEE Transactions on Parallel and Distributed Systems (TPDS)*, 2021

Eric Qin, Geonhwa Jeong, William Won, Sheng-Chun Kao, **Hyoukjun Kwon**, Sudarshan Srinivasan, Dipankar Das, Gordon E. Moon, Sivasankaran Rajamanickam, Tushar Krishna, “*Extending Sparse Tensor Accelerators to Support Multiple Compression Formats*”, *The 35th IEEE International Parallel & Distributed Processing Symposium (IPDPS)*, 2021

Hyoukjun Kwon, Liangzhen Lai, Michael Pellauer, Tushar Krishna, Yu-Hsin Chen, Vikas Chandra, “*Heterogeneous Dataflow Accelerators for Multi-DNN Workloads*”, *The 27th IEEE International Symposium on High-Performance Computer Architecture (HPCA)*, 2021

Hyoukjun Kwon, Michael Pellauer, Angshuman Parashar, Tushar Krishna, “*Flexion: A Quantitative Metric for Flexibility in DNN Accelerators*”, *IEEE Computer Architecture Letters (CAL)*, 2021

Robert Guirado and Hyoukjun Kwon (equal contribution), Sergi Abadal, Eduard Alarcon, Tushar Krishna, “*Dataflow-Architecture Co-Design for 2.5D DNN Accelerators using Wireless Network-on-Package*”, *The 26th Asia and South Pacific Design Automation Conference (ASP-DAC)*, 2021

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, Sung Kyu Lim, “*Architecture, Chip, and Package Co-design Flow for 2.5D Integration of Reusable IP Chiplets*”, *IEEE Transactions on Very Large Scale Integration (VLSI) Systems (VLSI)*, 2020

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

Hyoukjun Kwon, Prasanth Chatarasi, Michael Pellauer, Angshuman Parashar, Vivek Sarkar, Tushar Krishna, “MAESTRO: A Data-Centric Approach to Understand Reuse, Performance, and Hardware Cost of DNN Dataflows”, *IEEE MICRO: Top-Picks in Computer Architecture Conferences in 2019 (Top-Picks)*, 2020

Eric Qin, Ananda Samajdar, **Hyoukjun Kwon**, Vineet Nadella, Sudarshan Srinivasan, Dipankar Das, Bharat Kaul, Tushar Krishna, “SIGMA: A Sparse and Irregular GEMM Accelerator with Flexible Interconnects for DNN Training”, *The 26th IEEE International Symposium on High-Performance Computer Architecture (HPCA)*, 2020
Received the best paper award

Robert Guirado, **Hyoukjun Kwon**, Sergi Abadal, Eduard Alarcon, Tushar Krishna, “Understanding the Impact of On-Chip Communication on DNN Accelerator Performance”, *The 26th IEEE International Conference on Electronics Circuits and Systems (ICECS)*, 2019

Hyoukjun Kwon, Prasanth Chatarasi, Michael Pellauer, Angshuman Parashar, Vivek Sarkar, Tushar Krishna, “Understanding Reuse, Performance, and Hardware Cost of DNN Dataflows: A Data-Centric Approach”, *The 52nd IEEE/ACM International Symposium on Microarchitecture (MICRO)*, 2019
Selected as Top Picks in Computer Architecture Conferences in 2019

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, Sung Kyu Lim, “Architecture, Chip, and Package Co-design Flow for 2.5D Integration of Reusable IP Chiplets”, *The 56th Design Automation Conference (DAC)*, 2019

Zhongyuan Zhao, **Hyoukjun Kwon**, Sachit Kuhar, Weiguang Sheng, Zhigang Mao, Tushar Krishna, “mRNA: Enabling Efficient Mapping Space Exploration on a Reconfigurable Neural Accelerator”, *The 20th IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS)*, 2019

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

Brian Lebiednik, Sergi Abadal, **Hyoukjun Kwon**, Tushar Krishna, “Architecting a Secure Wireless Network-on-Chip”, *The 12th IEEE/ACM International Symposium on Networks-on-Chip (NOCS)*, 2018

Hyoukjun Kwon, Ananda Samajdar, Tushar Krishna, “MAERI: Enabling Flexible Dataflow Mapping over DNN Accelerators via Reconfigurable Interconnects”, *The 23rd ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, 2018
Honorable mention for Top Picks in Computer Architecture Conferences in 2018

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

Brian Lebiednik, Sergi Abadal, **Hyoukjun Kwon**, Tushar Krishna, “Spoofing Prevention via RF Power Profiling in Wireless Network-on-Chip”, *The 3rd International Workshop on Advanced Interconnect Solutions and Technologies for Emerging Computing Systems (AISTECS)*, 2018

Hyoukjun Kwon, Ananda Samajdar, Tushar Krishna, “Rethinking NoCs for Spatial Neural Network Accelerators”, *The 11th International Symposium on Networks-on-Chips (NOCS)*, 2017

Janardhan Rao Doppa, Ryan Gary Kim, Mihailo Isakov, Michel A. Kinsy, **Hyoukjun Kwon**, Tushar Krishna, “Adaptive Manycore Architectures for Big Data Computing”, *The 11th International Symposium on Networks-on-Chips (NOCS)*, 2017

Hyoukjun Kwon, William Harris, Hadi Esmaeilzadeh, “Proving Flow Security of Sequential Logic via Automatically Synthesized Relational Invariants”, *The 34th Computer Security Foundations (CSF)*, 2017

Hyoukjun Kwon, Tushar Krishna, “OpenSMART: Single-Cycle Multi-hop NoC Generator in BSV and Chisel”, *The 18th IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS)*, 2017

Hyoukjun Kwon, Dohyun Kim, Jisung Park, 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

Preprint

Jiaqi Gu, **Hyoukjun Kwon**, Dilin Wang, Wei Ye, Meng Li, Yu-hsin Chen, Liangzhen Lai, Vikas Chandra, David Z. Pan, “HRViT: Multi-Scale High-Resolution Vision Transformer”, *Arxiv Preprint: arXiv:2111.01236*, 2021

Invited Talks

Heterogeneous Dataflow Accelerators for AR/VR Workload

ACM SigArch Korea Workshop

Aug. 2021

Understanding hardware-mapping-model co-design space for efficient deep learning inference

Seoul National University: AI Summer School 2021

Aug. 2021

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

Pohang University of Science and Technology (Postech); Online Invited Talk – AI Seminar Series

Aug. 2020

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

IBM Research, Yorktown Heights, New York

Nov. 2018

Services

Architecture, Compiler, and System Support for Multi-model DNN Workloads Workshop (at MICRO 2021)

Workshop co-organizer, PoC

2021

<https://research.fb.com/arch-comp-sys-support-for-multi-model-dnn-workshop/>

External Review Committee (ERC)

IEEE/ACM International Symposium on Computer Architecture (ISCA)

2021, 2022

External Review Committee (ERC)

IEEE/ACM International Symposium on Microarchitecture (MICRO)

2021

Journal Reviewer

IEEE Computer Architecture Letters

2020, 2021

Journal Reviewer

IEEE MICRO

2019

Journal Reviewer

ACM Transactions on Architecture and Code Optimization (TACO)

2019, 2020, 2021

Journal Reviewer

IEEE Transactions on Computers (TOC)

2019, 2020, 2021

Journal Reviewer

IEEE Journal on Emerging and Selected Topics in Circuits and Systems (JETCAS)

2020

Journal Reviewer

IEEE Transactions on Neural Networks and Learning Systems (TNNLS)

2020

Journal Reviewer

IEEE Open Journal of Circuits and Systems (CAS)

2020

Talks in Industry/National Labs

Understanding Dataflow in DNN Accelerators and Modeling Them with MAESTRO

Sandia National Lab and Pacific Northwest National Lab (Online Talk)

Dec. 2019

Understanding Dataflows in DNN Accelerators

Facebook, Menlo Park, California

Jun. 2019

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

Heterogeneous Dataflow Accelerators for Multi-DNN Workloads

HPCA 2021, Online

Mar. 2021

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

Skills and Experiences

Software languages

C/C++, Matlab, and Python

Typesetting System

Latex

Machine Learning Frameworks

PyTorch

Formal verification tools

Coq Proof Assistant

Hardware languages

Verilog and Bluespec System Verilog

ASIC CAD 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)