Hyoukjun Kwon

Assistant Professor @ UCI (EECS dept.)
2200 Engineering Hall, CA 92697
(+1) 404-539-4457 | hyoukjun.kwon@uci.edu | www.hyoukjunkwon.com | hyoukjun-kwon

Professional Experience

University of California, Irvine, Irvine, CA
Assistant Professor, EECS — *Jan. 2023 - present*
Tenure-track position

Meta (Facebook), Sunnyvale, CA
Research Scientist at Meta (Facebook) Reality Labs — *Oct. 2020 - Dec. 2022*
Manager: Dr. Liangzhen Lai

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 — *May. 2018 - Aug. 2018*
Manager: Dr. Steve Keckler, Mentor: Dr. Michael Pellauer

NVIDIA, Westford, MA
Research Intern at Architecture Research Group — *May. 2017 - Aug. 2017*
Manager: Dr. Steve Keckler, Mentor: Dr. Michael Pellauer

Georgia Institute of Technology, Atlanta, GA
Graduate Research Assistant — *Aug. 2015 - Jul. 2020*
Advisor: Prof. Tushar Krishna

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 SIGACH/IEEE CS TCCA Outstanding dissertation award, **(Selected as one of the three best PhD dissertations in the computer architecture area in 2020)**

Seoul National University (SNU)
BS in CSE (Computer Science and Engineering)
BS in EMS (Environmental Material Science) — *Mar. 2007 - Feb. 2015*

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

Honors & Awards

2025 **Best Paper Award**, IEEE ISPASS 2025
2022 **Best Paper Award**, MLBench 2022 Workshop at MLSys 2022
2021 **Honorable Mention**, ACM SIGARCH/IEEE CS TCCA Outstanding Dissertation Award 2021
2020 **Best Paper Award**, IEEE 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) workloads
Mapping and dataflow optimization on accelerators
Cross-stack optimization of AI systems
Network-on-Chips (NoCs)
Efficient AI Models

Book

[B1] Tushar Krishna, **Hyoukjun Kwon**, Angshuman Parashar, Michael Pellauer, and Ananda Samajdar ***(alphabetical order)***, *"Synthesis lecture on computer architecture: Data Orchestration in Deep Learning Accelerators (Link)"*, *Morgan & Claypool Publishers* , *August 2020*

Peer-reviewed Publications

[C26] Yongfan Liu, **Hyoukjun Kwon**, *"Efficient Depth Estimation for Unstable Stereo Camera Systems on AR Glasses"*, *In Proc. of the Conference on Computer Vision and Pattern Recognition (**CVPR**)* , *2025*

[C25] Rachid Karami, Sheng-Chun Kao, **Hyoukjun Kwon**, *"Understanding the Performance Horizon of the Latest ML Workloads with NonGEMM Workloads"*, *In Proc. of the IEEE International Symposium on Performance Analysis of Systems and Software (**ISPASS**)* , *2025* ***Received the best paper award***

[C24] Jamin Seo, Jianming Tong, Tushar Krishna, **Hyoukjun Kwon**, *"Constrained Dataflow Accelerator for Real-Time Multi-Task Multi-Model Machine Learning Workloads"*, *In Proc. of the IEEE International Symposium on Performance Analysis of Systems and Software (**ISPASS**)* , *2025*

[C23] Mohanad Odema, Luke Chen, **Hyoukjun Kwon**, Mohammad Al Faruque, *"Performance Implications of Multi-Chiplet Neural Processing Units on Autonomous Driving Perception"*, *In Proc. of the Design, Automation and Test in Europe Conference (**DATE**)* , *2025*

[C22] Mohanad Odema, Luke Chen, **Hyoukjun Kwon**, Mohammad Al Faruque, *"SCAR: Scheduling Multi-Model AI Workloads on Heterogeneous Multi-Chiplet Module Accelerators"*, *In Proc. of the IEEE/ACM International Symposium on Microarchitecture (**MICRO**)*, *2024*

[C21] Chakshu Moar, Faraz Tahmasebi, Michael Pellauer, **Hyoukjun Kwon**, *"Characterizing the Accuracy – Efficiency Trade-off of Low-rank Decomposition in Language Models"*, *In Proc. of the 2024 IEEE International Symposium on Workload Characterization (**IISWC**)*, *2024*

[C20] Seah Kim, **Hyoukjun Kwon**, Jinook Song, Jihyuck Jo, Yu-Hsin Chen, Liangzhen Lai, Vikas Chandra, *"DREAM: A Dynamic Scheduler for Dynamic Real-time Multi-model ML Workloads"*, *In Proc. of the 29rd ACM International Conference on Architectural Support for Programming Languages and Operating Systems (**ASPLOS**)*, *2023*

[C19] **Hyoukjun Kwon**, Krishnakumar Nair, Jamin Seo, Jason Yik, Debabrata Mohapatra, Dongyuan Zhan, Jinook Song, Peter Capak, Peizhao Zhang, Peter Vajda, Colby Banbury, Mark Mazumder, Liangzhen Lai, Ashish Sirasao, Tushar Krishna, Harshit Khaitan, Vikas Chandra, Vijay Janapa Reddi, *"XRBench: An Extended Reality (XR) Machine Learning Benchmark Suite for the Metaverse"*, *Sixth Conference on Machine Learning and Systems (**MLSys**)*, *2023*

[W2] **Hyoukjun Kwon**, Krishnakumar Nair, Jinook Song, Colby Banbury, Mark Mazumder, Peter Capak, Yu-Hsin Chen, Liangzhen Lai, Tushar Krishna, Harshit Khaitan, Vikas Chandra, Vijay Janapa Reddi, *"MetaBench: Real-Time Multi-Modal Benchmark for Metaverse"*, *Third Workshop on Benchmarking Machine Learning Workloads on Emerging Hardware @ MLSys2022 (**MLBench**)*, *2022,* ***Received the best paper award***

[C18] Sheng-Chun Kao, **Hyoukjun Kwon**, Michael Pellauer, Angshuman Parashar, Tushar Krishna, *"A Formalism of DNN Accelerator Flexibility"*, *ACM SIGMETRICS/IFIP PERFORMANCE Joint International Conference on Measurement and Modeling of Computer Systems (**SIGMETRICS**)*, *2022*

[C17] Jiaqi Gu, **Hyoukjun Kwon**, Dilin Wang, Wei Ye, Meng Li, Yu-Hsin Chen, Liangzhen Lai, Vikas Chandra, and David Z. Pan, *"Multi-Scale High-Resolution Vision Transformer for Semantic Segmentation"*, *IEEE/CVF Conference on Computer Vision and Pattern Recognition (**CVPR**)*, *2022*

[J6] 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*

[J5] 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*

[C16] 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*

[C15] **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*

[J4] **Hyoukjun Kwon**, Michael Pellauer, Angshuman Parashar, Tushar Krishna, *"Flexion: A Quantitative Metric for Flexibility in DNN Accelerators"*, *IEEE Computer Architecture Letters (**CAL**)*, *2021*

[C14] **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*

[J3] 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*

[C13] 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*

[J2] **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*

[C12] 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

[C11] 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*

[C10] **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

[C9] 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*

[C8] 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*

[J1] **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*

[C7] 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*

[C6] **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

[C5] **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

[W1] 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*

[C4] **Hyoukjun Kwon**, Ananda Samajdar, Tushar Krishna, *"Rethinking NoCs for Spatial Neural Network Accelerators"*, *The 11th International Symposium on Networks-on-Chips (**NOCS**), 2017*

[C3] 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*

[C2] **Hyoukjun Kwon**, William Harris, Hadi Esmaeilzadeh, *"Proving Flow Security of Sequential Logic via Automatically Synthesized Relational Invariants"*, *The 34th Computer Security Foundations (**CSF**), 2017*

[C1] **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*

Invited Talks

ML Workloads in AR/VR and Their Implication to ML System Design
EMC2 workshop at HPCA 2025 — *Mar. 2025*

FlexiBit: Fully Flexible Precision Bit-parallel Accelerator Architecture for Arbitrary Mixed Precision AI
AMD Tech Talk — *Feb. 2025*

ML Workloads in AR/VR and Their Implication to ML System Design
Rutgers EFficient AI (REFAI) Seminar — *Oct. 2024*

ML Workloads in AR/VR and Their Implication to ML System Design
The The Industry-Academia Partnership (IAP) — *May. 2024*

ML Workloads in AR/VR and Their Implication to ML System Design
The Fourth Workshop on Benchmarking Machine Learning Workloads on Emerging Hardware (MLBench'23) — *Jun. 2023*

XRBench: An Extended Reality (XR) Machine Learning Benchmark Suite for the Metaverse
Industry-Academia Partnership Workshop- AI and Cloud Workshop @ UCSD — *Apr. 2023*

XRBench: An Extended Reality Machine Learning Benchmark Suite for the Metaverse
ML Performance Benchmarking Deep Learning Systems Workshop (MLPerf-Bench'23) — *Feb. 2023*

Accelerator System Design Challenges from Real-time and Multi-DNN Workloads
IEEE International Conference on Artificial Intelligence Circuits and Systems - Tutorial — *Jun. 2022*

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*

Professional Activities

Organizing Committee: Student Travel Grant Chair
International Symposium on Computer Architecture (ISCA) — *2025)*

Organizing Committee: Publication Co-chair
IEEE International Symposium on High-Performance Computer Architecture (HPCA) — *2025*

Organizing Committee: Artifact Evaluation Co-chair
The ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS) — *2025*

Workshop co-organizer (main PoC)
Architecture, Compiler, and System Support for Multi-model DNN Workloads Workshop (2021 at MICRO link) , (2022 at ISCA link) — *2021 (MICRO), 2022 (ISCA)*

Program Committee (PC)
Conference on Machine Learning and System (MLSys) — *2024, 2025*

Program Committee (PC)
The ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS) — *2023 (Fall), 2024*

Technical Program Committee (TPC)
Design Automation Conference (DAC) — *2023, 2024*

Technical Program Committee (TPC)
IEEE International Parallel & Distributed Processing Symposium (IPDPS) *2024*

Technical Program Committee (TPC)
Design, Automation and Test in Europe Conference | The European Event for Electronic System Design & Test (DATE) *2023, 2024, 2025*

Program Committee (PC)
IEEE/ACM International Symposium on Computer Architecture (ISCA) *2023, 2025*

Program Committee (PC)
IEEE/ACM International Symposium on Microarchitecture (MICRO) *2025*

Technical Program Committee (TPC)
IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS) *2023, 2024, 2025*

Technical Program Committee (TPC)
IEEE/ACM International Conference on Computer-Aided Design (ICCAD) *2023, 2024, 2025*

Program Committee (PC)
IACM International Conference on Supercomputing (ICS) *2025*

External Review Committee (ERC)
International Conference on Computer Vision (ICCV) *2023, 2025*

External Review Committee (ERC)
IEEE International Symposium on High-Performance Computer Architecture (HPCA). *2023*

External Review Committee (ERC)
The ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS) *2023 (Spring and Summer)*

External Review Committee (ERC)
The European Conference on Computer Vision (ECCV) *2024*

External Review Committee (ERC)
International Conference on Computer Vision (ICCV) *2023*

External Review Committee (ERC)
Conference on Computer Vision and Pattern Recognition (CVPR) *2023, 2024, 2025*

Technical Program Committee (TPC)
The International Conference for High Performance Computing, Networking, Storage, and Analysis (SC) — *2022, 2024*

Program Committee (PC)
IEEE International Symposium on Workload Characterization (IISWC) — *2022*

External Review Committee (ERC)
IEEE/ACM International Symposium on Computer Architecture (ISCA) — *2021, 2022, 2024*

External Review Committee (ERC)
IEEE/ACM International Symposium on Microarchitecture (MICRO) — *2021, 2022*

Journal Reviewer
IEEE Computer Architecture Letters — *2020, 2021*

Journal Reviewer
IEEE MICRO — *2019, 2022, 2023*

Journal Reviewer
IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems — *2021*

Journal Reviewer
ACM Transactions on Architecture and Code Optimization (TACO) — *2019, 2020, 2021, 2022, 2023, 2024*

Journal Reviewer
IEEE Transactions on Emerging Topics in Computing — *2022*

Journal Reviewer
IEEE Transactions on Computers (TC) — *2019, 2020, 2021, 2022, 2023, 2024*

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*