Hyunwon Chung

INTERESTS Efficient AI, LLM Inference Approximation,
Reconfigurable architecture, VLIW Processor, Dataflow Architecture, High-performance SoC
Ultra Low-Power Neural Engine

EDUCATION

University of Michigan, Ann Arbor
Ph.D. in Electrical and Computer Engineering 2024 – May 2028 (expected)
M.S. in Electrical and Computer Engineering 2022 – 2024
Research Advisors: Prof. David Blaauw and Prof. Hun-Seok Kim

Korea University, Seoul
BS. in Electrical Engineering (Include 2-year military service) Mar 2015 – Feb 2022

RESEARCH

Activation-Aware Approximation Algorithm and Hardware Co-Design for Billion-Parameter Foundation Models
Co-advisors: Prof. David Blaauw and Prof. Hun-Seok Kim

- Developed activation-aware approximation algorithms for efficient inference in large-scale foundation models, including LLMs, vision-language models, and diffusion models by dynamically approximating input activations
- Proposed on-the-fly low-rank activation approximation using alternating least-squares (ALS) and a hybrid weight–activation approximation scheme, showing improved accuracy under matched FLOPs budgets compared to weight-only low-rank approximation
- Designed approximation-aware accelerator architectures and performance/energy models to analyze FLOPs reduction, memory traffic, compute intensity, data movement, memory hierarchy, and dataflow trade-offs for efficient foundation-model inference

Ultra-Low-Power Neural Audio Compression Engine
Co-advisors: Prof. David Blaauw and Prof. Hun-Seok Kim

- Developed and quantized neural network models for real-time audio compression, co-designing algorithms with hardware constraints to target low-latency, low-power inference.
- Built performance and power models to quantify energy and latency trade-offs across loop ordering, data reuse, memory access patterns, and on-chip buffer/SRAM sizing for convolution-dominated workloads.
- Explored systolic-array architectures by evaluating array dimensions, processing-element (PE) utilization, dataflow strategies, and microarchitectural variants for compressed neural network models.
- Tapeout submitted using TSMC 22 nm technology.

A Streaming Processor with Hardware-Level Scheduling for Adavanced Spectrum Sensing
Co-advisors: Prof. David Blaauw and Prof. Hun-Seok Kim

- Developed a 28 nm CMOS streaming processor with a hardware kernel launcher, enabling 10 ns runtime reconfiguration while reducing CPU scheduling overhead for data-dependent RF workloads
- Designed a dynamic hardware scheduling unit that autonomously manages graph-based kernel dispatch, resource availability, and multi-channel execution without CPU intervention
- Implemented concurrent execution and speculative preloading to improve PE utilization by up to 5× and achieve up to 8.3× end-to-end performance improvement
- Published as first author at the 2026 IEEE Symposium on VLSI Circuits, achieving 248.5 GFLOPS/mm² compute density and 1.67 TFLOPS/W energy efficiency in 28 nm CMOS

COCHON (Configurable Optical Communications via Heterogeneous-Processing Optimized Node)
Co-advisors: Prof. David Blaauw and Prof. Hun-Seok Kim

- Contributed to a 224.72 mm² GF 12 nm heterogeneous coherent optical baseband SoC integrating five accelerators—RX baseband, two FEC decoders, and RX/TX bit-domain processors—under a Cortex-M4 host

- Led top-level SoC integration and physical design across multiple accelerator teams, driving chip-level floorplanning, multi-clock timing closure, power delivery, and tapeout execution
- Designed multi-clock SoC integration infrastructure, including robust CDC architecture, asynchronous streaming interfaces, and timing-safe data transfer across independently clocked accelerator domains
- Taped out end-to-end silicon supporting 10 baseband kernels, 8 FEC code types, 2 bit-framing modes, and complete coherent optical RX processing for OpenZR+ and SDA OCT configurations

DASH-SoC (Domain-Focused Advanced Software Reconfiguration Heterogeneous)
Co-advisors: Prof. David Blaauw and Prof. Hun-Seok Kim

- Contributed to the architecture and implementation of two reconfigurable accelerators for a 159.6 mm^2 DARPA-funded DASH SoC, enabling domain-adaptive processing across diverse communication workloads
- Redesigned the Domain Adaptive Processor (DAP) compute architecture from a 16-bit to 32-bit datapath, improving arithmetic capability, kernel flexibility, and workload coverage
- Implemented multi-kernel execution support in DAP and optimized physical design to reduce routing congestion, improve timing closure, and enhance silicon implementation efficiency
- Designed a fully reconfigurable FEC accelerator supporting LTE/5G, Wi-Fi, and SDA OCT standards by integrating six FEC modes within a unified processor architecture
- Co-authored related work published in 2026 IEEE Transactions on Circuits and Systems I (TCAS-I)

Low Power Accelerators for ML & Communication Systems
Advisor: Prof. Jongsun Park Korea University

- Designed a DCT-based JPEG encoder/decoder architecture and synthesized the RTL in 65nm TSMC Technology
- Developed a low-power hardware implementation of a Viterbi decoder architecture
- Proposed an 8-bit fixed-point low-power accelerator for convolutional neural networks
- Designed a folded architecture for fully-connected layers to improve compute efficiency

PUBLICATIONS

[10] *IEEE International Solid-State Circuits Conference (ISSCC)*, 2027. **Under review.**

[9] Daewon Chae, **Hyunwon Chung**, Changwoo Lee, and Hun-Seok Kim. "DyRA: Dynamic Residual Approximation for Efficient Matrix Multiplication in DNNs." *Conference on Neural Information Processing Systems (NeurIPS)*, 2026. **Under review.**

[8] "A 165μW Long-Range Crystal-less IoT Localization Tag with an OFDMA Baseband Processor." *IEEE Journal of Solid-State Circuits*, 2026. **Under review.**

[7] Anish Vipperla, **Hyunwon Chung**, David Blaauw, Hun-Seok Kim, Ali Akoglu, and Chaitali Chakrabarti. "TOCSA: Throughput-Optimized Compilation Framework for Spatial Architectures." *ACM Transactions on Reconfigurable Technology and Systems*, 2026. **Under review.**

[6] Xiangdong Wei, **Hyunwon Chung**, Yufan Yue, Huanshihong Deng, Chieh-Shen Chen, Yejoong Kim, Seungkyu Choi, Thang Pham, Owen Ma, Alex Chiriyath, Ilya Kogan, Jacob Holtom, Tutu Ajayi, Long Nguyen, Jimmy Sa, Tuan Nguyen, Daniel Bliss, David Blaauw, and Hun-Seok Kim. "COCHON: A Configurable Coherent Heterogeneous Baseband SoC for High-Speed Optical Communication Networks." **Under review**, 2026.

[5] Hyunwon Chung, Parin Senta, Jason Yu, Yukun Fang, Pierre Abillama, Kuan-Yu Chen, Ronald Dreslinski, David Blaauw, and Hun-Seok Kim. "A 248.5 GFLOPS/mm^2, 1.67 TFLOPS/W Streaming Processor with Hardware-Level Scheduling for Advanced Spectrum Sensing." *IEEE Symposium on VLSI Technology and Circuits*, 2026. **Best Student Paper Nomination.**

[4] Sanghyuck Moon, Ashfakh Huluvallay, **Hyunwon Chung**, Myeongsu Ko, Seokhyeon Jeong, Jungho Lee, Mohammad Khreishah, Jeongtaek Chang, Hung Do, Caitlyn Sutherland, Jason Kapit, David Nicholson, William Reinhardt, Dennis Sylvester, Mark Miskin, and David Blaauw. "A 95.0 dB Dynamic Range Zero-Bias PV Light-to-Digital Converter for Seawater Monitoring with Single Point Calibration." *IEEE Symposium on VLSI Technology and Circuits*, 2026.

[3] Yufan Yue, Kuan-Yu Chen, Xiangdong Wei, Tutu Ajayi, **Hyunwon Chung**, Ronald Dreslinski, David Blaauw, and Hun-Seok Kim. "QFEC: A 9.97 Gb/s Fully Configurable Quad-Mode Decoder for LDPC, Polar, Turbo, and Convolutional Codes." *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 73, no. 6, pp. 4385–4397, 2026.

[2] Jiahao Lin, H. Umut Suluhan, **Hyunwon Chung**, Arindam Dutta, Anish Vipperla, Gerard Gubash, Jacob Holtom, Bernd-Peter Paris, Chaitali Chakrabarti, Daniel W. Bliss, David Blaauw, Hun-Seok Kim, Ali Akoglu, and Umit Y. Ogras. "An Overview of Challenges and Requirements for Real-Time Spectrum Sensing in Modern RF Autonomy Systems." *IEEE Design & Test*, vol. 42, no. 6, pp. 111–126, 2025.

[1] Xiangyu Zhao, Ryan Aridi, Jacob Hume, Swetha Subbiah, Xingqi Wu, **Hyunwon Chung**, Yutao Qin, and Yogesh B. Gianchandani. "Automatic Peak Detection Algorithm Based on Continuous Wavelet Transform for Complex Chromatograms from Multi-Detector Micro-Scale Gas Chromatographs." *Journal of Chromatography A*, vol. 1714, article 464582, 2024.

SELECTED PROJECTS

Parameter-Efficient Fine-Tuning of Vision Transformers

- *EECS 553 course project at the University of Michigan*
- Applied parameter-efficient fine-tuning methods, including LoRA, LoHa, and LoKr, to Vision Transformer image classification across six benchmark datasets: CIFAR-10, CIFAR-100, Caltech101, Caltech256, DTD, and SUN397.
- Built a flexible ViT fine-tuning framework supporting selective adaptation of attention layers, MLP layers, and bias parameters, enabling controlled comparison across different training configurations.
- Managed and analyzed approximately 100 fine-tuning experiments over 100+ GPU hours, showing that LoKr can outperform LoRA with fewer trainable parameters and that attention-layer adaptation is often more effective than MLP-layer adaptation.

PIM-Based Secure Memory Acceleration for Confidential Computing

- *EECS 598 course project at the University of Michigan*
- Investigated a processing-in-memory architecture for secure memory systems, focusing on moving AES encryption and decryption into 3D-stacked memory to reduce off-chip data exposure and improve confidential computing performance.
- Analyzed AES dataflow for in-memory execution, including SubBytes, MixColumns, and AddRoundKey operations, and estimated a baseline execution cost of 361 cycles per AES block before memory read/write overhead.
- Evaluated PIM simulation infrastructure using PIMSim, DRAMSim2, MARSSx86, and gem5, and conducted trace-based simulation showing approximately 441 cycles per block and 1.64 s latency for 1 GB AES encryption on an HMC-style PIM system.

A Dynamic Quantized CNN Processor with Analog Computing

- *EECS 627 course project at the University of Michigan*
- Designed analog MAC unit which includes digital to pulse converter, analog datapath, and flash ADC
- Proposed dynamic quantization method which allows each layer have different scaling
- Post-quantization training and quantization-aware training are both used for 5-bit quantization, to get proper accuracy on MNIST and Cifar-10 dataset

16 bit Microprocessor with In-memory computing technology based on 6T SRAM architecture

- *EECS 427 course project at the University of Michigan*
- Developed 16-bit Microprocessor based on RISC-V ISA, which includes RF, ALU, Controller, Program Counter, Shifter, and IMEM/DMEM
- Designed customized 6T-SRAM based in-memory computing (IMC) module into 16-bit microprocessor

The World Embedded Software Contest - AI Humanoid Division

- Developed a program for real-time image processing and humanoid robot control using python
- Won 3rd place and Awarded Embedded SW & System Industry Association Chairman's award

SKILLS

Programming & ML Frameworks: C, C++, Python, PyTorch, TensorFlow
GPU & Parallel Programming: CUDA, Triton, OpenMP
Hardware Simulation: SystemVerilog, Verilog, VCS, ModelSim, HSPICE
Circuit Design Tools: Cadence Virtuoso, Cadence Innovus, Calibre

MILITARY SERVICE

Korean Augmentation To the United States Army,
Squad Leader, Sergeant, Headquarter, 23rd CBERN, 2ID, Camp Humphreys Sep 2016 – Jun 2018

AWARDS & SCHOLARSHIPS

Best Student Paper Nomination, IEEE Symposium on VLSI Technology and Circuits	2026
IEEE Solid-State Circuits Society (SSCS) Travel Grant	2026
IEEE Symposium on VLSI Circuits Travel Grant	2026
University of Michigan Rackham Conference Travel Grant	2026
Electrical and Computer Engineering Department Graduate Fellowship, University of Michigan	2025
Dean’s list at School of Electrical Engineering, Korea University	2018 - 2021
Third Place and Embedded SW & System Industry Association Chairman’s Award World Embedded Software Contest, AI Humanoid Division	2019
Miraero Scholarship at School of Electrical Engineering, Korea University	2016
National Scholarship (Admission with highest distinction), Korea University	2015