

CONTACT INFORMATION	KAIST N1 818 291 Daehak-Ro, Yuseong-Gu Daejeon, South Korea	E-mail: jinha.chung@kaist.ac.kr Website: https://jinhachung.github.io
RESEARCH INTERESTS	My research interests are in computer architecture and systems for agentic applications, with a particular focus on serving AI agents on edge devices.	
EDUCATION	Korea Advanced Institute of Science and Technology (KAIST) , Daejeon, South Korea Integrated M.S. & Ph.D., School of Electrical Engineering, March 2021 – present Advisor: Minsoo Rhu B.S., School of Electrical Engineering, March 2016 – February 2021 Double major in Mathematical Sciences <i>Cum Laude</i>	
PUBLICATIONS	- Jehyeon Bang, Eunyeong Cho, Ranggi Hwang, Jinha Chung, and Minsoo Rhu, “Spec-MoE: A Fast and Efficient Mixture-of-Experts Inference via Self-Assisted Speculative Decoding,” <i>The 63rd ACM/ESDA/IEEE Design Automation Conference (DAC-63)</i>, Long Beach, CA, July 2026 (acceptance rate: 22%, 543/2,435) Jinha Chung, Byeongjun Shin, Jiin Kim, and Minsoo Rhu, “Agent-X: Full Pipeline Acceleration of On-device AI Agents,” <i>The 24th International Conference on Mobile Systems, Applications, and Services (MobiSys-24)</i>, Cambridge, UK, June 2026 (acceptance rate: 22%, 68/304) - Jiin Kim, Byeongjun Shin, Jinha Chung, and Minsoo Rhu, “The Cost of Dynamic Reasoning: Demystifying AI Agents and Test-Time Scaling from an AI Infrastructure Perspective,” <i>The 32nd IEEE International Symposium on High-Performance Computer Architecture (HPCA-32)</i>, Sydney, Australia, February 2026 (acceptance rate: 20%, 119/602) - Yunjae Lee, Jinha Chung, and Minsoo Rhu, “SmartSAGE: Training Large-scale Graph Neural Networks Using In-storage Processing Architectures,” <i>The 49th International Symposium on Computer Architecture (ISCA-49)</i>, New York, NY, June 2022 (acceptance rate: 16%, 67/400)	
PROFESSIONAL SERVICES	Artifact Evaluation Committee Symposium on Microarchitecture (MICRO-59) Symposium on High-Performance Computer Architecture (HPCA-32)	2026 2026
	Journal Reviewer IEEE Computer Architecture Letters (CAL)	2026
EXPERIENCE	Samsung Advanced Institute of Technology (SAIT) , Suwon, South Korea <i>Research Engineer Intern</i> Exploring various use-cases of CXL-PNM Mentors: Yong Min Tai and Jinsu Park Manager: Yeongon Cho	

June 2023 – September 2023

VIA Research Group, KAIST, Daejeon, South Korea
Undergraduate Research Intern
March 2019 – February 2021

Building a TPU cycle-level simulator (March 2019 – August 2019)

Building recommendation system model accelerators (January 2020 – February 2021)

Mentors: Youngeun Kwon and Yujeong Choi Advisor: Minsoo Rhu

TEACHING**EE209: Programming Structure for Electrical Engineering, KAIST**

Spring 2022 — Head TA, Best TA Award

Fall 2023

EE312: Introduction to Computer Architecture, KAIST

Fall 2021

Fall 2022

Spring 2024

RECOGNITION**KAIST**Best TA Award (EE209 Programming Structure for Electrical Engineering) **Spring 2022**2nd Place, KAIST URP Program **2018****RELEVANT****COURSEWORK****Computer Systems**

Digital System Design, Computer Architecture, Operating System

Parallel Computing Systems for Deep Learning

Hardware Acceleration for Machine Learning

Hardware Security Techniques, Datacenter Security, Secure Computer Architecture

Flash Storage and File Systems

Mathematics and Applied Mathematics

Machine Learning, Advanced Intelligence

Probability Theory, Mathematical Statistics, Engineering Random Processes

Analysis I & II, Fourier Analysis and Applications

Linear Algebra, Modern Algebra I & II, Cryptography

SKILLS**Languages**

C, C++, Python, Bash (Unix Shell)

Frameworks

PyTorch, Hugging Face Transformers, MLX