TAEHYUN KIM

E-mail : taehyunzzz@snu.ac.kr
Phone : (+81) 010-4507-7499

Research Interests

- **Large Transformer Model Applications**: Language, Vision, Multi-modal AI
- **Efficient AI Algorithms**: Parameter-Efficient Finetuning (PEFT), Speculative Decoding, Mixture-of-Experts (MoE)
- **Hardware Architecture**: Processing-in-Memory (PIM), Near-Data Processing (NDP), CXL Memory
- **System Design**: Accelerator Parallelism, Algorithm-System Co-design, Memory Efficiency

Work Experience

Computer Architecture and Parallel Processing Lab @ SNU — Mar 2019 – Present
Ph.D. Researcher — *Seoul, Republic of Korea*

Computer Architecture and Parallel Processing Lab @ SNU — Aug 2019 – Dec 2019
Undergraduate Researcher — *Seoul, Republic of Korea*

LG Electronics — Jan 2018 – Feb 2018
Undergraduate Intern — *Seoul, Republic of Korea*

Education

Seoul National University (SNU) — Mar 2019 – Present
Doctor of Philosophy in Electrical and Computer Engineering — *Seoul, Republic of Korea*
- Advisor : Prof. Hyuk-Jae Lee
- GPA: 3.94 / 4.3

Seoul National University (SNU) — Mar 2013 – Feb 2019
Bachelor of Science in Electrical and Computer Engineering — *Seoul, Republic of Korea*
- GPA: 3.61 / 4.3

Tokyo Korean School (TKS) — Apr 2010 – Feb 2013
High School Education — *Tokyo, Japan*

Publications (Featured)

MoNDE: Mixture of Near-Data Experts for Large-Scale Sparse Models
Taehyun Kim, Kwanseok Choi, Youngmock Cho, Jaehoon Cho, Hyuk-Jae Lee, Jaewoong Sim
61th ACM/IEEE Design Automation Conference (DAC), June 2024.

Virtual keyboards with real-time and robust deep learning-based gesture recognition
Tae-Ho Lee, Sunwoong Kim, **Taehyun Kim**, Jin-Sung Kim, Hyuk-Jae Lee
IEEE Transactions on Human-Machine Systems (THMS), Volume 52 Issue 4, April 2022.

Smart refrigerator inventory management using convolutional neural networks
Tae-Ho Lee, Shin-Woo Kang, **Taehyun Kim**, Jin-Sung Kim, Hyuk-Jae Lee
IEEE 3rd International Conference on Artificial Intelligence Circuits and Systems (AICAS), June 2021.

GradPIM: A Practical Processing-in-DRAM Architecture for Gradient Descent
Heesu Kim, Hanmin Park, **Taehyun Kim**, Kwanheum Cho, Eojin Lee, Soojung Ryu, Hyuk-Jae Lee, Kiyoung Choi, Jinho Lee
27th IEEE International Symposium on High-Performance Computer Architecture (HPCA), Jan. 2021.

Patents

METHOD FOR ROUTING TOKEN AND APPARATUS THEREFOR
Inventor: **Taehyun Kim**, Jaewoong Sim, Jaehun Cho, Hyuk-Jae Lee
Submitted: Dec 10, 2024. (*Under Review*)

NEAR-DATA PROCESSING SYSTEM FOR LARGE-SCALE MIXTURE-OF-EXPERTS AI MODEL
Inventor: **Taehyun Kim**, Hyuk-Jae Lee, Jaewoong Sim, Jaehun Cho, Kanseok Choi, Youngmock Cho
Submitted: Dec 29, 2023. (*Under Review*)

Research Projects

Intelligent in-memory error-correction device for high-reliability memory.

- March 2021 - Dec 2024
- Served as the lead manager of the project
- Developed and verified a hardware architecture of next-generation error-correcting code (ECC) memory
- Conducted research on near-data-processing (NDP) in emerging CXL memory that could work aside ECCs.
- 1 publication, 2 patent applications

Smart refrigerator stock management system

- June 2018 - Dec 2019
- Developed an automatic stock management system based on YOLOv3.
- Tracks the movement of stored items inside the refrigerator for stock management.
- Exhibited at CES 2020.

Skills & Tools

- Languages (proficiency): Korean (native), English (high), Japanese (high)
- Programming Languages: Python, C++, CUDA, Verilog
- DL Frameworks: PyTorch, HuggingFace, Deepspeed, TensorRT-LLM (FasterTransformer), vLLM
- Simulators: DRAMsim3, Ramulator
- Other tools: SystemC, Synopsys Design Compiler, ModelSim, NVIDIA NSight Systems, Xilinx FPGA, vim

Services

External Reviewer
42nd IEEE International Conference on Computer Design (ICCD) *2024*

Research Assistant
Graduation Project for Undergraduate Students *Spring 2019, Spring & Fall 2024*

Teaching Assistant
400.018 Creative Engineering Design *Fall 2019-2022*

Republic of Korea Army
Administrative staff @ AFMC *2015-2016*