Hyeongjun Cho

✉ cho0624ck@Gmail.com

linkedin.com/in/hjleonardocho

Research Interests

❍ Heterogeneous Computing Systems for AI/ML Acceleration

- Near-Data Processing (NDP) Architecture & Systems
  - PIM (Processing-In-Memory) architecture using DRAM/SRAM
  - Workload optimization for Large Language Models (LLMs)
  - Inter-device data movement optimization

- System-level Integration & Programmability
  - Workload-aware scheduling for heterogeneous devices (GPU, NPU, PIM)
  - Unified software stack and programming models

Education

Sungkyunkwan University, Korea

M.S. in Semiconductor Convergence Engineering — Feb, 2026

B.S. in Electronic and Computer Engineering — Feb, 2023

Professional Work Experience

HW Researcher — Jan 2023 – Jan2024

Vieworks, Machine Vision Camera department (Anyang, Korea)

- Developed power-efficient circuit designs to reduce signal noise and improve image sensor stability in machine vision cameras.

Publication

1st Author

- LibraPIM: Dynamic Load Rebalancing to Maximize Utilization in PIM-Assisted LLM Inference Systems
  Hyeongjun Cho*, Yoonho Jang, Hyungi Kim, Seongwook Kim, Keewon Kwon, Gwngsun Kim and Seokin Hong*
  The 34th International Conference on Parallel Architectures and Compilation Techniques (**PACT'25**) (Accepted)

2nd Author

- PIMPAL: Accelerating LLM Inference on Edge Devices via In-DRAM Arithmetic Lookup
  Yoonho Jang, ***Hyeongjun Cho****, Yesin Ryu, Jungrae Kim and Seokin Hong*
  Design Automation Conference (**DAC2025**)

- Redefining PIM Architecture with Compact and Power-Efficient Microscaling
  Yoonho Jang, ***Hyeongjun Cho*** *and Seokin Hong*
  International Conference on Electronics, Information, and Communications

Presentation & Poster

1st Author

- Bank-Split PIM: Enabling Concurrent PIM and Memory Operations for LLM Inference in Heterogeneous Systems
  Hyeongjun Cho*, Yoonho Jang and Seokin Hong*
  Design Automation Conference (**DAC2025**) (Poster)

Skills

- **Programming Languages**
  - C/C++ (Proficient, for simulator development)
  - Python (Proficient, for data analysis & modeling)
  - Verilog (Intermediate)
- **Architecture & Simulation Tools**
  - **Memory Simulators:** Ramulator, DRAMSim3
  - **Full-System Simulators:** gem5, Firesim
  - **Performance Profiling:** NVIDIA Nsight, TensorBoard
- **Machine Learning Frameworks**
  - PyTorch (Intermediate, for model implementation and analysis)

Teaching Assistant

Spring 2025: Computer Architecture Design

Fall 2025: Data Structure and Algorithm

Phone: (+82) 10-5190-7555

12-14, Giheung-ro 30beon-gil, Giheung-gu, Yongin-si, Gyeonggi-do