Hongsun Jang

Accelerated Intelligent Systems Lab (AISys) | Seoul National University
hongsun.jang@snu.ac.kr | github.com/hongsunjang | Google Scholar

Research Interests

Large LMs (Training/Deployment), Near-Data Processing (Storage, DRAM), Distributed Training, FPGA-based Acceleration

Education

Seoul National University — Seoul, South Korea
Integrated M.S./Ph.D. in Electrical & Computer Engineering (Advisor: Prof. Jinho Lee) — *Mar. 2023 – Present*
Selected Coursework:
- *Systems & Arch.:* Advanced OS, Advanced Comp. Arch., Memory Centric Systems, Advanced Digital IC
- *AI & Applications:* Embedded ML, NLP, Robot Learning

Yonsei University — Seoul, South Korea
B.S. in Computer Science — *Mar. 2017 – Aug. 2022*
Relevant Coursework: Multi-core & GPU Prog., Linux Sys. Prog., Comp. Arch., OS, ML, Computer Vision

Publications

A Cost-Effective Near-Storage Processing Solution for Offline Inference of Long-Context LLMs
Hongsun Jang, Jaeyong Song, Changmin Shin, Siung Noh, Jaewon Jung, Jisung Park, Jinho Lee
ASPLOS 2026 (To appear)

FALA: Locality-Aware PIM-Host Cooperation for Graph Processing with Fine-Grained Column Access
Changmin Shin, Jaeyong Song, Seongmin Na, Jun Sung, **Hongsun Jang**, Jinho Lee
MICRO 2025 [Link]

INF^2: High-Throughput Generative Inference of Large Language Models using Near-Storage Processing
Hongsun Jang, Siung Noh, Changmin Shin, Jaewon Jung, Jaeyong Song, Jinho Lee
arXiv preprint [Link]

Piccolo: Large-Scale Graph Processing with Fine-Grained In-Memory Scatter-Gather
Changmin Shin, Jaeyong Song, **Hongsun Jang**, Dogeun Kim, Jun Sung, Taehee Kwon, Jae Hyung Ju, Frank Liu, Yeonkyu Choi, Jinho Lee
HPCA 2025 [Link]

GraNNDis: Efficient Unified Distributed Training Framework for Deep GNNs on Large Clusters
Jaeyong Song, **Hongsun Jang**, Jaewon Jung, Youngsok Kim, Jinho Lee
PACT 2024 [Link]

PeerAiD: Improving Adversarial Distillation from a Specialized Peer Tutor
Jaewon Jung, **Hongsun Jang**, Jaeyong Song, Jinho Lee
CVPR 2024 [Link]

Smart-Infinity: Fast Large Language Model Training using Near-Storage Processing on a Real System
Hongsun Jang, Jaeyong Song, Jaewon Jung, Jaeyoung Park, Youngsok Kim, Jinho Lee
HPCA 2024 – Acceptance Rate: 18% **(Best Paper Honorable Mention)** [Link]

Pipette: Automatic Fine-Grained Large Language Model Training Configurator for Real-World Clusters
Jinkyu Yim[1], Jaeyong Song[1], Yerim Choi, Jaebeen Lee, Jaewon Jung, **Hongsun Jang**, Jinho Lee
DATE 2024 [Link]

Fast Adversarial Training with Dynamic Batch-level Attack Control
Jaewon Jung, Jaeyong Song, **Hongsun Jang**, Hyeyoon Lee, Kanghyun Choi, Noseong Park, Jinho Lee
DAC 2023 [Link]

Pipe-BD: Pipelined Parallel Blockwise Distillation
Hongsun Jang, Jaewon Jung, Jaeyong Song, Joonsang Yu, Youngsok Kim, Jinho Lee
DATE 2023 – Acceptance Rate: 25% [Link]

Optimus-CC: Efficient Large NLP Model Training with 3D Parallelism Aware Communication Compression
Jaeyong Song[1], Jinkyu Yim[1], Jaewon Jung, **Hongsun Jang**, Hyung-jin Kim, Youngsok Kim, Jinho Lee
ASPLOS 2023 [Link]

Note: 1 indicates equal contribution.

Honors and Awards

Best Paper Award – Honorable Mention HPCA 2024
The 30th IEEE Int'l Symp. on High-Performance Computer Architecture (Top 3/410) *Edinburgh, UK*
1st Graduate School Presidential Science Scholarship 2023 – Present
Government of the Republic of Korea (Top 120/2,980)
Samsung Humantech Paper Award Feb. 2024
Samsung Electronics (Top 115/1,189)
Academic Excellence Award, Seoul National University Spring 2023
Academic Excellence Award, Yonsei University Fall 2022

Skills

Languages: C/C++, Python, OpenCL, CUDA, Verilog
Frameworks & Tools: PyTorch, Vitis, Xilinx Runtime (XRT), Linux
Spoken Languages: Korean (Native), English (Fluent)

Projects

Co-design Framework for DNN and SSD/NPU Systems 2024
Industry-Academic Project with Samsung Electronics *[Code]*
Fast Distributed Graph Neural Network Training Framework 2024
Unified Distributed Training Framework for Deep GNNs *[Code]*
Distributed NLP Training Acceleration with Gradient Compression 2023
Optimized Communication for Large-scale NLP Training *[Code]*
Parallel Algorithm for NAS with Blockwise Knowledge Distillation 2022
Undergraduate Research *[Code] [Video]*
Korean Sentence Relationship Classification Competition (NLI) 2022
Dacon Competition (Ranked 33rd of 1,353) – Finetuning NLP models *[Site]*
Network-aware Resource Scheduling in Edge Computing using Kubernetes 2021
Undergraduate Project *[Video]*

Industrial Experience

Samsung Electronics, Memory Solutions Team Hwaseong, South Korea
Engineering Intern (Mentor: Gitae Na) *Mar. 2021 – Jun. 2021*
Eugene Investment & Futures, IT Team Seoul, South Korea
Field Practice Intern *Jan. 2021 – Feb. 2021*

Academic Service & Teaching

- **Reviewer:** MLSys 2026 (ERC), CVPR 2026, IEEE TETC 2025 (Invited), ACM TACO 2024 (Invited), PACT SRC 2024
- **Teaching Assistant (Seoul National University):**
  - Design Project for Electrical Devices & Systems (Fall 2024)
  - Embedded System Design (Fall 2024)
  - Digital System Design and Practices (Fall 2023)
  - Programming Methodology (Spring 2023)

Military Service

Republic of Korea Army South Korea
Sergeant (Honorably Discharged) *Aug. 2018 – Mar. 2020*