Hyunwuk Lee

Samsung Electronics | hyunwuk.lee@yonsei.ac.kr,hyunwuklee0519@gmail.com
https://hyunwuklee0519@github.io | linkedin.com/in/hyunwuk-lee-56a7b4297/

Education

Yonsei University, Korea Mar. 2018 – Aug. 2024
Integrated Master and Ph.D. in School of Electrical and Electronic Engineering
- GPA: 4.25/4.5
- **Advisor:** Prof. Won Woo Ro
- **Dissertation:** Quantization and Accelerator Designs for Efficient Complex Valued Neural Networks

Yonsei University, Korea Mar. 2014 – Feb. 2018
B.S. in School of Electrical and Electronic Engineering
- GPA: 3.61/4.5
- **Advisor:** Prof. Won Woo Ro

Experience

System Engineer Sep. 2024 – Present
Samsung Electronics, Hwasung, Korea
- System Architect in SSD FTL Hardware Automation Solution
- Modeling, profiling, and designing hardware automation for FTL

Research Interest

Accelerating Neural Networks
- Accelerator Architecture and Scheduling
- Generative AI Acceleration (LLM, VLM, Diffusion Model, etc.)
- Neural Network Compression (Quantization, Pruning, etc.)
- Systems for Machine Learning

Graphic Processing Unit
- GPU Register and Memory System
- Multi-GPUs
- GPU Device Driver

Publications

COSMOS: An LLC Contention Slowdown Model for Heterogeneous Multi-core Systems Accepted (May 2025)
Yongju Lee, Jaewon Kwon, Cheolhwan Kim, Enhyeok Jang, Jiwon Lee, ***Hyunwuk Lee***, and Won Woo Ro
2025 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS 2025)

CVMAX: Accelerator Architecture with Polar Form Multiplication for Complex-Valued Neural Networks Accepted (Jun. 2025)
Hyunwuk Lee, Sungbin Kim, Sungwoo Kim, and Won Woo Ro
The 62nd Design Automation Conference (DAC 2025)

Ditto: Accelerating Diffusion Model via Temporal Value Similarity, Mar. 2025
Sungbin Kim*, ***Hyunwuk Lee***, Wonho Cho, Mincheol Park and Won Woo Ro
2025 IEEE International Symposium on High-Performance Computer Architecture (HPCA 2025)
* Co-First Author

REC: Enhancing fine-grained cache coherence protocol in multi-GPU systems 2025
Gun Ko, Jiwon Lee, Hongju Kal, ***Hyunwuk Lee***, and Won Woo Ro
Journal of Systems Architecture (2025)

AirGun: Adaptive Granularity Quantization for Accelerating Large Language Models Oct. 2024
Sungbin Kim, ***Hyunwuk Lee***, Sungwoo Kim, Cheolhwan Kim, and Won Woo Ro
The 42nd IEEE International Conference on Computer Design (ICCD 2024)

GUMSO: Gating Unnecessary On-Chip Memory Slices for Power Optimization on GPUs Aug. 2024
Seunghyun Jin, ***Hyunwuk Lee***, and Won Woo Ro
ACM/IEEE International Symposium on Low Power Electronics and Design (ISLPED 2024)

SHREG: Mitigating register redundancy in GPUs 2024
Seunghyun Jin, ***Hyunwuk Lee***, Jonghyun Lee, Junsung Kim, and Won Woo Ro
Journal of Systems Architecture (2024)

Exploiting Inherent Properties of Complex Numbers for Accelerating Complex Valued Neural Networks Oct. 2023
Hyunwuk Lee, Hyungjun Jang, Sungbin Kim, Sungwoo Kim, Wonho Cho, and Won Woo Ro
The 56th International Symposium on Microarchitecture (MICRO 2023)

Early-Adaptor: An Adaptive Framework for Proactive UVM Memory Management Apr. 2023
Seokjin Go, ***Hyunwuk Lee***, Junsung Kim, Jiwon Lee, Myung Kuk Yoon, and Won Woo Ro
2023 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS 2023)

Deep Learning with GPUs 2021
Won Jeon, Gun Ko, Jiwon Lee, ***Hyunwuk Lee***, Dongho Ha, and Won Woo Ro
Book Chapter

Patents

Memory device including a plurality of area having different refresh periods, memory controller controlling the same and memory system including the same
Hyunwuk Lee, Gun Ko, Ipoom Jeong, and Won Woo Ro
US Patent No. 11276452, Korea Application No. 10-2020-0045023, China Patent No. ZL 2020 1 1090146.1

A method for neural network processing including memory-optimization techniques
Hongju Kal, Cheonjun Park, **Hyunwook Lee**, Ipoom Jeong, Jiwon Lee, and Won Woo Ro
Korea Application No. 10-2022-0041848

Apparatus and method with register sharing
Seunghyun Jin, Jonghyun Lee, **Hyunwuk Lee**, and Won Woo Ro
US Application No. 18/ 315576, Korea Application No. 10-2022-0074653

Memory management unit and method of walking page table
Jiwon Lee, Ipoom Jeong, Hongju Kal, Gun Ko, **Hyunwuk Lee**, and Won Woo Ro
US Application No. 18/502058, Korea Application No. 10-2022-0175909, China Application No. 202311700848.0

Memory management apparatus and method for UVM
Seokjin Go, Junsung Kim, **Hyunwuk Lee**, Jiwon Lee, and Won Woo Ro

Korea Application No. 10-2023-0076606

Polar Form Aware Apparatus and Method for Complex Valued Neural Network and Rectangular Form Conversion Apparatus

Hyunwuk Lee, Hyungjun Jang, Sungbin Kim, Wonho Cho, Sungwoo Kim, and Won Woo Ro

Korea Application No. 10-2023-0158538

Projects

Memory-Centric Architecture using the Reconfigurable PIM Devices Jan. 2024 - Aug. 2024

Institute for Information & communication Technology Planning & evaluation (IITP), Korea

- Led the project team in eSCaL
- Developed virtualized multi-PIM API for DNN applications
- Designed data allocation schemes for multi-PIM architecture

Developing Data Processing Unit for AI Workloads July 2022 - Jan. 2024

Korea Evaluation Institute of Industrial Technology (Keit), Korea

- Led the project team in eSCaL
- Developed virtualized multi-PIM API for DNN applications
- Designed data allocation schemes for multi-PIM architecture

Analysis on High Performance GPU Workloads and Architecture Design May 2021 - Apr. 2022

Samsung Advanced Institute of Technology, Korea

- Led the project team in eSCaL
- Profiled GPU running neural network workloads
- Designed GPU register architecture sharing data inter warp and intra warp for general matrix multiplications
- Modeled the register architecture in GPU simulator

Development of High Performance Multi-GPU Memory System Mar. 2021 - Feb. 2022

National Research Foundation of Korea (NRF) , Korea

- Researched HW/SW memory management of modern multi-GPU system
- Modeled the memory management techniques in multi-GPU simulator

Architectural Exploration of Parallel Execution Processing Units for Supercomputer CPU July 2020 - Jan. 2024

National Research Foundation of Korea (NRF), Korea

- Led the project team in eSCaL
- Designed SIMD architecture for supercomputer CPU
- Modeled SIMD architecture for supercomputer based on RISC-V CPU SIMD extension version in CPU simulator
- Designed register architecture for SIMD unit in supercomputer CPU

Performance Analysis of Neural Network Workloads and Development of Energy Efficient Approximate Memory for Neural Networks July 2018 - June 201

SK Hynix, Korea

- Profiled GPU running neural network workloads
- Designed Approximate DRAM architecture for energy efficient neural networks
- Modeled DRAM architecture in DRAM simulator

Honor and Awards

SK Hynix Industry-Academic Research Project Outstanding Invention 2021

SK Hynix, Korea

Teaching Experiences

Teaching Assistant 2018

School of Electrical and Electronic Engineering, Yonsei University
Computer Architecture

Skills and Techniques

Programming Languages: C++, C, Python, CUDA

DNN Frameworks: PyTorch, Tensorflow

Simulator

- **DRAM simulator:** DRAMSim3, Ramulator
- **CPU simulator:** Gem5
- **GPU simulator:** GPGPU-sim, Accel-Sim, MGPUSim
- **NPU simulator:** SCALE-Sim
- Simpy

Languages: Korean, English

References

eSCaL, Electrical and Electronics Engineering, Yonsei University, Seoul

- **Advisor:** Professor Won Woo Ro, wro@yonsei.ac.kr
- **eSCaL home page:** escal.yonsei.ac.kr