Johnny Rhe

SungKyunKwan University: 2066, Seobu-ro, Jangan-gu, Suwon-si, Gyeonggi-do, Korea

📱 (+82)10-5236-5648 | ✉ djwhsdj@skku.edu | ⌨ djwhsdj.github.io | in linkedin.com/in/djwhsdj

Research Interests

I am currently pursuing a combined M.S./Ph.D. degree in the Department of Electrical and Computer Engineering at Sungkyunkwan University. My research focuses on mapping Deep/Convolutional Neural Networks (D/CNNs) onto In-Memory Computing (IMC) architectures to enable **energy-efficient and high-speed D/CNN inference.** Recently, I have expanded my research interests to **Large Language Models (LLMs)**, aiming to optimize their performance and inference efficiency in IMC systems. To achieve this, I am exploring various hardware-aware optimization techniques, including mapping, dataflow, and model compression.

Education

Combined M.S-Ph.D. Electrical and Computer Engineering   *Sept. 2019 - Present*

SungKyunKwan University, Suwon, Korea

- Intelligent & Resource-Efficient Image Processing & Systems Design (IRIS) Lab
- **Thesis Advisor: Prof. Jong Hwan Ko**

B.S. School of Electronic and Electrical Engineering   *Mar. 2011 - Feb. 2019*

SungKyunKwan University, Suwon, Korea

Research Experience

Graduate Research Assistant in SungKyunKwan University   *Sept. 2019 - Present*

1. Mapping Method for Energy-efficient IMC-based D/CNN Inference

- Developed an optimal weight mapping method to minimize computing cycles and enhance array utilization
- Proposed a mapping precision controller and adder tree for flexible DNN inference in the digital IMC system, improving computational efficiency

2. Model Compression Techniques for Efficient Computation in IMC

- Designed array/row/cell-level optimization methods using channel-, pattern-, and column-wise pruning to maximize array utilization while minimizing accuracy loss

3. Hardware-Based Simulation of Weight Mapping Methods

- Conducted visualization and NeuroSim-based evaluation of the weight mapping methods to assess their impact on power efficiency and computational performance

Projects

A Study on Collaborative Mapping and Training Techniques for Energy-Efficient In-Memory CNN Inference   *Sept 2024 - Present*

National Research Foundation of Korea (NRF)

- Leading research as the principal investigator

HW-SW co-design technique for low-power real-time in-memory deep learning operation   *Mar 2023 - Present*

Samsung Advanced Institute of Technology (SAIT)

- Study about energy-efficient weight mapping for the IMC systems

Student Activites

[S6] Poster Presentation at PhD Forum   *Mar, 2025*
  Design, Automation and Test in Europe Conference (DATE) 2025

[S5] Poster Presentation at PhD Forum   *May, 2024*
  IEEE International Symposium on Circuits and Systems (ISCAS) 2024

[S4] Poster Presentation at Student Research Forum   *Jan, 2024*
  29th Asia South Pacific Design Automation Conference (ASP-DAC) 2024

[S3] IEEE CASS Non-Flagship Student Travel Grant Selection - ISLPED 2023   *Aug, 2023*
  IEEE Circuits and System Society (CAS)

[S2] SKKU Innovation Research Fellowship Scholarship Selection   *Oct, 2022*
  BK Graduate School Innovation Support Project

[S1] Introduction and Demonstration of RC Car Remote Control Using 3D Hand Motion Recognition   *Apr, 2022*
Information Technology Research Center (ITRC), Korea

Awards

[A8] Outstanding Research Award   *Nov, 2024*
   Korea Collaborative & High-tech Initiative for Prospective Semiconductor Research (K-CHIPS)

[A7] Outstanding Research Award   *Oct, 2024*
   Korea Institute of Energy Technology Evaluation and Planning (KETEP)

[A6] Outstanding Paper Award   *Aug, 2024*
   Samsung Electronics

[A5] The Grand Prize at the Exynos AI Challenger   *July, 2024*
   Samsung Electronics' S.LSI

[A4] 3rd Place Winner in the System Design Contest - GPU Track   *June, 2024*
   Design Automation Conference (DAC)

[A3] Most Popular Poster Award at the Student Research Forum   *Jan, 2024*
   29th Asia South Pacific Design Automation Conference (ASP-DAC) 2024

[A2] Best Paper Award   *July, 2023*
   Korean Artificial Intelligence Association

[A1] 2nd Place Winner of Artificial Intelligence Grand Challenge   *July, 2020*
   Institute of Information & Communications Technology Planning & Evaluation, Korea

Publications

Journal Articles

[J5] **Johnny Rhe**, Kang Eun Jeon and Jong Hwan Ko, "Genetic Algorithm-Aided Row-Skipping for SDK-Based Convolutional Weight Mapping," *Journal of Systems Architecture (JSA)*, In Review (JCR Q1).

[J4] Juhong Park, **Johnny Rhe**, and Jong Hwan Ko, "Input/Mapping Precision Controllable Digital PIM with Adaptive Adder Tree Architecture for Flexible DNN Inference," *Journal of Systems Architecture (JSA)*, Feb. 2025 (JCR Q1).

[J3] **Johnny Rhe**, Kang Eun Jeon, Joo Chan Lee, Seong Moon Jeong, and Jong Hwan Ko, "KERNTROL: Kernel Shape Control Toward Ultimate Memory Utilization for In-Memory Convolutional Weight Mapping," *IEEE Transactions on Circuits and Systems I: Regular Papers (TCAS-I) vol. 71, no. 12, pp. 6138-6151*, Nov. 2024.

[J2] **Johnny Rhe**, Sungmin Moon, and Jong Hwan Ko, "VWC-SDK: Convolutional Weight Mapping Using Shifted and Duplicated Kernel with Variable Windows and Channels," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems (JETCAS), vol. 12, no. 2, pp. 408-421*, May. 2022 (impact factor: 5.877).

[J1] Eunyoung Lee, Taeyoung Han, Donguk Seo, Gicheol Shin, Jaerok Kim, Seonho Kim, Soyoun Jeong, **Johnny Rhe**, Jaeyung Park, Jong Hwan Ko, Yoonmyung Lee, "A Charge-Domain Scalable-Weight In-Memory Computing Macro with Dual-SRAM Architecture for Precision-Scalable DNN Accelerators," in *IEEE Transactions Circuits and System I (TCAS-I) vol. 68, no. 8, pp. 3305-3316*, May. 2021.

Conference Proceedings

[C11] Kang Eun Jeon, **Johnny Rhe**, and Jong Hwan Ko, "Low-Rank Compression for IMC Arrays," *Design, Automation & Test in Europe Conference (DATE)*, Mar. 2025.

[C10] **Johnny Rhe**, and Jong Hwan Ko, "Row-Efficient Pruning for In-Memory Convolutional Weight Mapping," *International SoC Design Conference (ISOCC)*, Aug. 2024.

[C9] Kang Eun Jeon[1], Wooram Seo[1], **Johnny Rhe**, and Jong Hwan Ko, "ConvMapSim: Modeling and Simulating Convolutional Network Mapping on PIM Arrays," *IEEE Artificial Intelligence Circuits and Systems (AICAS)*, Apr. 2024.

[C8] Chanwook Hwang, Jaehyeon So, **Johnny Rhe**, Jiyoon Kim, Juhong Park, Kang Eun Jeon, and Jong Hwan Ko, "An Efficient Ventricular Arrhythmias Detection on Microcontrollers with Optimized 1D CNN," *IEEE Artificial Intelligence Circuits and Systems (AICAS)*, Apr. 2024.

[C7] Juhong Park, **Johnny Rhe**, and Jong Hwan Ko, "KARS: Kernel-Grouping Aided Row-Skipping for SDK-based Weight Compression in PIM Arrays," *IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2024.

[C6] Hyeonsu Bang, Kang Eun Jeon, **Johnny Rhe**, and Jong Hwan Ko, "DCR: Decomposition-Aware Column Re-Mapping for Stuck-At-Fault Tolerance in ReRAM Arrays," *IEEE International Conference on Computer Design (ICCD)*, Nov. 2023.

[C5] **Johnny Rhe**, Kang Eun Jeon, Joo Chan Lee, Seong Moon Jeong, and Jong Hwan Ko, "Kernel Shape Control for Row-Efficient Convolution on Processing-In-Memory Arrays," *ACM/IEEE International Conference on Computer-Aided Design (ICCAD)*, Oct. 2023.

[C4] **Johnny Rhe**, Kang Eun Jeon, and Jong Hwan Ko, "PAIRS: Pruning-AIded Row-Skipping for SDK-Based Convolutional Weight Mapping in Processing-In-Memory Architectures," *ACM/IEEE International Symposium on Low Power Electronics and Design (ISLPED 2023)*, Aug. 2023.

[C3] Kang Eun Jeon, **Johnny Rhe**, and Jong Hwan Ko, "Weight-Aware Activation Mapping for Energy-Efficient Convolution on PIM Arrays," *ACM/IEEE International Symposium on Low Power Electronics and Design (ISLPED)*, Aug. 2023.

[C2] **Johnny Rhe**, Sungmin Moon, and Jong Hwan Ko, "VW-SDK: Efficient Convolutional Weight Mapping Using Variable Windows for Processing-In-Memory Architectures," *Design, Automation, and Test in Europe (DATE)*, Mar. 2022 (Oral, acceptance rate: 25%).

[C1] Gicheol Shin, Donguk Seo, Jaerok Kim, **Johnny Rhe**, Eunyoung Lee, Seonho Kim, Soyoun Jeong, Jong Hwan Ko, Yoonmyung Lee, "A charge-Domain Computation-in-Memory Macro with Versatile All-Around-Wire-Capacitor for Variable-Precision Computation and Array-Embedded DA/AD Conversions," *European Conference on Solid-State Circuits (ESSCIRC)*, Sept. 2021.

Patents

[P9] **Johnny Rhe**, Kang Eun Jeon, and Jong Hwan Ko, US Patent Application No.19069410, METHOD AND ELECTRONIC DEVICE WITH WEIGHT PRUNING CROSS-REFERENCE TO RELATED APPLICATION, Mar. 2025.

[P8] **Johnny Rhe**, Kang Eun Jeon, Joo Chan Lee, Seongmoon Jeong, and Jong Hwan Ko, PCT Patent Application PCT/KR2024/016234, Convolution operation method and apparatus utilization kernel shape control, Oct. 2024.

[P7] **Johnny Rhe**, Kang Eun Jeon, and Jong Hwan Ko, KR Patent Application No.10-2024-0058360, Pattern-based weight pruning method and electronic devices, May. 2024.

[P6] **Johnny Rhe** and Jong Hwan Ko, US Patent Application No.18398389, Apparatus and method for controlling processing-in-memory by accelerating convolution operation based on arranging pattern of weight in kernel, and storage medium storing instructions to perform method for controlling processing-in-memory, Dec. 2023.

[P5] Juhong Park, **Johnny Rhe**, Jong Hwan Ko, KR Patent Application No. 10-2023-0157705, Adaptive adder tree and digital in-memory architecture supporting multi-bit partitioning and improved column-wise mapping for depth-wise convolution layers, Nov. 2023.

[P4] **Johnny Rhe**, Kang Eun Jeon, Joo Chan Lee, Seongmoon Jeong, and Jong Hwan Ko, KR Patent Application No.10-2023-0171519, Convolution operation method and apparatus utilization kernel shape control, Nov. 2023.

[P3] **Johnny Rhe**, Sungmin Moon, and Jong Hwan Ko, US Patent Application No.18090628, Memory device for optimizing computation of convolutional layer, method for controlling memory device, and recording medium storing instructions to perform method for controlling memory device, Dec. 2022.

[P2] **Johnny Rhe** and Jong Hwan Ko, KR Patent Application No.10-2022-0187676, Apparatus, method, computer-readable storage medium and computer program for controlling processing-in-memory by accelerating convolution operation based on arranging pattern of weight in kernel, Dec. 2022.

[P1] **Johnny Rhe**, Sungmin Moon, and Jong Hwan Ko, KR Patent No.10-2021-0191289, Memory device for optimizing operation of convolutional layer and method for controlling the same, Dec. 2021.

Reference

Dr. Jong Hwan Ko
Professor, College of ICE, SungKyunKwan University
jhko@skku.edu