Wonung **Kim**

COMPUTER ARCHITECT

291 Daehak-ro, Eoeun-dong, Yuseong-gu, E3-5 #410, Daejeon, 34141, South Korea

☎ (+82) 10-6326-5969 | ✉ waneon.kim@gmail.com | 🏠 www.waneon.me | ⓖ waneon | 🔗 wonung-kim-4585872b2

Education

KAIST   *Daejeon, Korea*
PH.D. IN GRADUATE SCHOOL OF SEMICONDUCTOR   *Sep. 2025 - Present*
- Advisor: Prof. Jongse Park

KAIST   *Daejeon, Korea*
M.S. IN SCHOOL OF COMPUTING   *Sep. 2023 - Aug. 2025*
- Advisor: Prof. Jongse Park

Korea University   *Seoul, Korea*
B.S. IN COMPUTER SCIENCE AND ENGINEERING   *Mar. 2020 - Aug. 2023*
- Cumulative GPA: 4.28/4.5   Major GPA: 4.38/4.5

Research Interests

- **Processing-in-memory architectures**
- **Emerging model architecture acceleration**
- **Heterogeneous system architecture for efficient AI**

Publications

Pimba: A Processing-in-Memory Acceleration for Post-Transformer Large Language Model Serving   *MICRO 2025*

W. KIM, Y. LEE, Y. KIM, J. HWANG, S. OH, J. JUNG, A. HUSEYNOV, W. G. PARK, C. H. PARK, D. MAHAJAN, AND J. PARK   *Oct. 2025*
- **Summary:** This work presents the first PIM accelerator designed to support both transformer and post-transformer LLMs. It conducts an in-depth analysis of post-transformer models and, based on these insights, proposes Pimba, which leverages access interleaving and an MX-based processing engine for efficient state-update acceleration.
- **Role:** Led the research project and proposed the core architecture of Pimba. Implemented the simulator and accuracy evaluation code to validate the design.

DaCapo: Accelerating Continuous Learning in Autonomous Systems for Video Analytics   *ISCA 2024*

Y. KIM, C. OH, J. HWANG, **W. KIM**, S. OH, Y. LEE, H. SHARMA, A. YAZDANBAKHSH, AND J. PARK   *July. 2024*
- **Summary:** This work presents DaCapo, an acceleration system for continuous learning. It incorporates a novel spatially partitionable and precision-flexible accelerator that enables parallel execution of training and inference kernels on sub-accelerators at their respective precisions.
- **Role:** Contributed to the implementation of precision-flexible GEMM kernels, organized experimental datasets, and contributed the DaCapo continuous learning emulation framework.

Honors & Awards

AWARDS

2025	**Gold Prize, 1st Author**, 31st Samsung Humantech Paper Award. Paper Name: Pimba	*Samsung*
2024	**Distinguished Artifact Award**, ISCA 2024. Paper Name: DaCapo	*ISCA*
2022	**Academic Excellence Award**, Korea University	*Korea University*

HONORS

2025	**Student Travel Grant**, MICRO 2025	*MICRO*
2023	**Student Travel Grant**, MICRO 2023	*MICRO*
2023	**Scholarship**, National Science & Technology Scholarship	*KOSAF*
2020	**Scholarship**, SK C&C IT Scholarship	*SK C&C*

Services

2025	**Artifact Evaluation Committee**, MICRO 2025	*MICRO*

Teaching Assistants

| 2024 | **CS230,** System Programming | *KAIST* |
| 2024 | **CS610,** Parallel Processing | *KAIST* |

Skills

Programming	C/C++, Python, Triton, CUDA, Rust, Go, LaTEX, Lua, Javascript
Frameworks	PyTorch, Akita
Hardware Langauges	Systemverilog, Verilog, Chisel
Languages	Korean, English
Development Stacks	Neovim, Nix