<Professor Minsoo Rhu's Lab>

Vertically Integrated Architecture (VIA) Research Group

■ Contact information

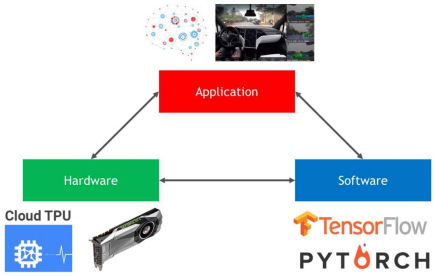
Professor : Bldg. N1, #809 TEL : 042-350-7547
Lab. : Bldg. N1, #818
Website : https://sites.google.com/view/kaist-via

■ Current state of the Lab. (in 2025 Spring Semester)

Postdoctoral Fellows : 0 PhD Students: 13 Master's Student: 7

■ Research Areas

Vertically Integrated Architecture (VIA) research group conducts research in the domain of computer architecture with a vertically integrated approach. By co-optimizing VLSI circuit technology, computer system architecture, and application & algorithms (with an emphasis on machine learning and computer vision), our research mission is to build high-performance computing platform for future "intelligent" systems that are programmable, robust, reliable, secure, and energy-efficient.



■ Recommended courses & Career after graduation

- Courses: computer architecture, data structures, system programming, digital logic design, compilers, operating systems, computer networks
- Careers: During your graduate studies, we strongly encourage you to take internships in the industry (preferably in bleeding-edge IT companies like Google, Facebook, NVIDIA, Samsung, Microsoft, and Intel) so that you get practical, hands-on experience within the electrical and computer engineering discipline.

■ Introduction to other activities besides research

Professor Rhu is a huge sports fan and encourages students to engage in extra-curricular activities as means to pursue a (mentally & physically) healthy graduate school life. We also encourage lab members to get together outside of the laboratory so that they maintain good social relationships with each other. There are frequent (un)official get-togethers and we plan on having regular team-building events during summer & winter breaks.

■ Introduction to the Lab.

Professor Minsoo Rhu has spent three years working at NVIDIA Research as a Senior Research Scientist. He worked in several domains within the computer system stack, including ASIC designs, computer system architecture, runtime systems, and application & workload characterization with an emphasis on machine learning (ML) and computer vision (CV). As such, our research mission is to train students to become computer system architects that understands *both* the hardware and software system, enabling you to optimize any target application (e.g., ML or CV) for the underlying computing stack. Our group is currently funded by several *research-oriented* projects, for instance, the ERC-AI (by National Research Foundation), Neural Processor Research Center (by Samsung Research), Samsung Future Research Funding and Incubation Center for Future Technology, and others.

■ Recent research achievements ('24~'25)

[1] Yunjae Lee*, Juntaek Lim*, Jehyeon Bang, Eunyeong Cho, Huijong Jeong, Taesu Kim, Hyungjun Kim, Joonhyung Lee, Jinseop Im, Ranggi Hwang, Se Jung Kwon, Dongsoo Lee, and Minsoo Rhu, "Debunking the CUDA Myth Towards GPU-based AI Systems", The 52nd IEEE/ACM International Symposium on Computer Architecture (**ISCA-52**), Tokyo, Japan, Jun. 2025

[2] Dongjae Lee, Bongjoon Hyun, Taehun Kim, and Minsoo Rhu, "PIM-MMU: A Memory Management Unit for Accelerating Data Transfers in Commercial PIM Systems," The 57th IEEE/ACM International Symposium on Microarchitecture (**MICRO-57**), Austin, TX, Nov. 2024

[3] Jehyeon Bang, Yujeong Choi, Myeongwoo Kim, Yongdeok Kim, and Minsoo Rhu, "vTrain: A Simulation Framework for Evaluating Cost-effective and Compute-optimal Large Language Model Training," The 57th IEEE/ACM International Symposium on Microarchitecture (**MICRO-57**), Austin, TX, Nov. 2024

[4] Yunjae Lee*, Hyeseong Kim*, and Minsoo Rhu, "PreSto: An In-Storage Data Preprocessing System for Training Recommendation Models," The 51st IEEE/ACM International Symposium on Computer Architecture (**ISCA-51**), Buenos Aires, Argentina, Jun. 2024

[5] Yujeong Choi, Jiin Kim, and Minsoo Rhu, "ElasticRec: A Microservice-based Model Serving Architecture Enabling Elastic Resource Scaling for Recommendation Models," The 51st IEEE/ACM International Symposium on Computer Architecture (**ISCA-51**), Buenos Aires, Argentina, Jun. 2024

[6] Juntaek Lim, Youngeun Kwon, Ranggi Hwang, Kiwan Maeng, Edward Suh, and Minsoo Rhu, "LazyDP: Co-Designing Algorithm-Software for Scalable Training of Differentially Private Recommendation Models," The 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems (**ASPLOS-29**), San Diego, CA, Apr. 2024

[7] Bongjoon Hyun, Taehun Kim, Dongjae Lee, and Minsoo Rhu, "Pathfinding Future PIM Architectures by Demystifying a Commercial PIM Technology," The 30th IEEE International Symposium on High-Performance Computer Architecture (**HPCA-30**), Edinburgh, UK, Feb. 2024