

HYUNGYO KIM

◇ E-mail: hyungyo2@illinois.edu

EDUCATION

University of Illinois at Urbana-Champaign Ph.D candidate in Electrical and Computer Engineering	<i>August 2021 - Current</i>
Seoul National University B.S. in Electrical and Computer Engineering (<i>summa cum laude</i>)	<i>March 2014 - February 2020</i>
Gyeonggi Science High School Specialized high school for students talented in math and science	<i>March 2011 - February 2014</i>

EXPERIENCE

University of Illinois at Urbana-Champaign Graduate Research Assistant (Advisor: Prof. Nam Sung Kim)	<i>August 2023 - Current</i> Urbana, IL
IBM Research Research Intern (Supervisor: Dr. Apoorve Mohan)	<i>May 2025 - August 2025</i> Yorktown Heights, NY
Intel Corporation Research Intern (Supervisor: Dr. Frank Hady)	<i>May 2024 - August 2024</i> Champaign, IL
Samsung Electronics Research Intern (Supervisor: Jinin So)	<i>July 2023 - August 2023</i> Hwaseong, Korea
University of Illinois at Urbana-Champaign Graduate Research Assistant (Advisor: Prof. Naresh R. Shanbhag)	<i>August 2021 - June 2023</i> Urbana, IL
Seoul National University Research Intern (Advisor: Prof. Jong-Ho Lee)	<i>June 2020 - July 2021</i> Seoul, Korea
ROK Marine Corps Force Recon Reconnaissance Marine	<i>October 2015 - July 2017</i> Baengnyeong-Island, Korea

HONORS & AWARDS

ML and Systems Rising Star ML Commons	<i>2026</i>
IEEE CAL Best Paper Award IEEE Computer Society Publications Board	<i>2025</i>
IIDAI Student Best Poster Award IBM-Illinois Discovery Accelerator Institute	<i>2025</i>
List of Teachers Ranked as Excellent University of Illinois at Urbana-Champaign	<i>2023</i>
PhD Fellowship Samsung Electronics	<i>2023</i>
Korean Government Scholarship for Study Overseas Korean Government	<i>2021</i>
Graduated with Highest Honors (<i>Summa Cum Laude</i>) Seoul National University	<i>2020</i>
Eminence Scholarship Seoul National University	<i>2019</i>
National Scholarship for Science and Engineering Korea Student Aid Foundation	<i>2018</i>
Merit-Based Scholarship Seoul National University	<i>2014</i>

PROFESSIONAL SERVICES AND ACTIVITIES

Reviewer

IEEE Computer Architecture Letters (CAL)

2025, 2026

IEEE Transactions on Parallel and Distributed Systems (TPDS)

2025

PUBLICATIONS

(† : equal contribution)

12. **Hyungyo Kim**[†], Qirong Xia[†], Jinghan Huang, Nachuan Wang, Jung Ho Ahn, Younjoo Lee, Wajdi K Feghali, Ren Wang, and Nam Sung Kim, “LILo: Harnessing the On-chip Accelerators in Intel CPUs for Compressed LLM Inference Acceleration,” *IEEE International Symposium on High-Performance Computer Architecture (HPCA)*, January 2026.
11. Shuo Li, Chihun Song, **Hyungyo Kim**, Nam Sung Kim, and Naresh Shanbhag, “Characterizing the Intrinsic Bank-level Accuracy vs. Energy Trade-off of SRAM-Based In-Memory Computing Architectures in 28nm CMOS,” *IEEE Transactions on Circuits and Systems I: Regular Papers (TCAS-I)*, October 2025.
10. Nachuan Wang, **Hyungyo Kim**, and Nam Sung Kim, “LPC: Lossless Parameter Compression for Deploying LLM Inference on Edge Systems,” *IEEE Embedded Systems Letters (ESL)*, October 2025.
9. Sungmin Yun, Seonyong Park, Hwayong Nam, Younjoo Lee, Gunjun Lee, Kwanhee Kyung, Sangpyo Kim, Nam Sung Kim, Jongmin Kim, **Hyungyo Kim**, Juhwan Cho, Seungmin Baek, and Jung Ho Ahn, “The New LLM Bottleneck: A Systems Perspective on Latent Attention and Mixture-of-Experts,” *arXiv preprint*, July 2025.
8. Jinghan Huang, **Hyungyo Kim**, Nachuan Wang, Jaeyoung Kang, Hrishi Shah, Eun Kyung Lee, Minjia Zhang, Fan Lai, and Nam Sung Kim, “NetZIP: Algorithm/Hardware Co-design of In-network Lossless Compression for Distributed Large Model Training,” *58th IEEE/ACM International Symposium on Microarchitecture (MICRO)*, October 2025.
7. Yue Pan, Zihan Xia, Po-Kai Hsu, Lanxiang Hu, **Hyungyo Kim**, Janak Sharda, Minxuan Zhou, Nam Sung Kim, Shimeng Yu, Tajana Rosing, and Mingu Kang, “Stratum: System-Hardware Co-design with Tiered Monolithic 3D-DRAM for Efficient MoE Serving,” *58th IEEE/ACM International Symposium on Microarchitecture (MICRO)*, October 2025.
6. **Hyungyo Kim**, Nachuan Wang, Qirong Xia, Jinghan Huang, Amir Yazdanbakhsh, and Nam Sung Kim, “LIA: A Single-GPU LLM Inference Acceleration with Cooperative AMX-Enabled CPU-GPU Computation and CXL Offloading,” *2025 IEEE International Symposium on Computer Architecture (ISCA)*, June 2025.
5. **Hyungyo Kim**, Gaohan Ye, Nachuan Wang, Amir Yazdanbakhsh, and Nam Sung Kim, “Exploiting Intel[®] Advanced Matrix Extensions (AMX) for Large Language Model Inference,” *IEEE Computer Architecture Letters (CAL)*, May 2024.
4. **Hyungyo Kim** and Naresh Shanbhag, “Enhancing the Accuracy of 6T SRAM-based In-Memory Architecture via Maximum Likelihood Detection,” *IEEE Transactions on Signal Processing (TSP)*, April 2024.

3. Sang-Soo Park, KyungSoo Kim, Jinin So, Jin Jung, Jonggeon Lee, Kyoungwan Woo, Nayeon Kim, Younghyun Lee, **Hyungyo Kim**, Yongsuk Kwon, Jinhyun Kim, Jieun Lee, YeonGon Cho, Yongmin Tai, Jeonghyeon Cho, Hoyoung Song, Jung Ho Ahn, and Nam Sung Kim, “An LPDDR-based CXL-PNM platform for TCO-efficient inference of Transformer-based large language models,” *IEEE International Symposium on High-Performance Computer Architecture (HPCA)*, March 2024.
2. **Hyungyo Kim** and Naresh Shanbhag, “Boosting The Accuracy of SRAM-based In-memory Architectures via Maximum Likelihood-based Error Compensation Methods,” *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, June 2023.
1. **Hyungyo Kim**[†], Joon Hwang[†], Dongseok Kwon, Jangsaeng Kim, Jiseong Im, Byung-Gook Park, and Jong-Ho Lee, “Direct Gradient Calculation: Simple and Variation-Tolerant On-Chip Training Method for Neural Networks,” *Advanced Intelligent Systems*, August 2021.

PATENTS

1. **Hyungyo Kim**, Alper Buyuktosunoglu, Apoorve Mohan, and Cedric Lichtenau, “Split Execution of Artificial Intelligence Inference between Disaggregated Compute Elements,” U.S. Patent, filed February 2026.