Curriculum Vitae

Gwangsun Kim

67 Cheongam-Ro
R4 #4411
Nam-Gu, Pohang, Gyeongbuk, Republic of Korea 37673

Email: g.kim@postech.ac.kr
URL: https://psal.postech.ac.kr/gkim
Phone: +82-54-279-2260

RESEARCH INTERESTS

Computer architecture, Near-data processing, Systems for ML, GPU computing, Memory system, Large-scale systems, Programming Model, Neural Processing Units, Storage-class memory, Interconnection networks.

PROFESSIONAL EXPERIENCE

Nov. 2018 – Present	**Assistant Professor.** Department of Computer Science and Engineering. Pohang University of Science and Technology (POSTECH).
Mar. 2018 – Oct. 2018	**Senior Research Engineer.** Architecture Research Group. Arm Inc., Austin, Texas, USA.
Sep. 2016 – Mar. 2018	**Senior Performance Engineer.** Future System Design Team, Architecture and Technology Group. Arm Inc., Austin, Texas, USA.
Jun. 2015 – Sep. 2015	**Research Intern.** Architecture Research Group. NVIDIA Research, Austin, Texas, USA.
Jul. 2014 – Sep. 2014	**Research Intern.** Samsung Electronics, Hwasung, Republic of Korea.

EDUCATION

Feb. 2012 – Aug. 2016	**Ph.D. in Computer Science.** Korea Advanced Institute of Science and Technology (KAIST). Thesis: *High-Throughput System Design with Memory Networks.* Advisor: Prof. John Kim.
Feb. 2010 – Feb. 2012	**M.S. in Computer Science.** Korea Advanced Institute of Science and Technology (KAIST). Thesis: *LIBRA: Multi-mode On-Chip Network Arbitration for Locality-Oblivious Task Placement.* Advisor: Prof. John Kim.
Mar. 2005 – Feb. 2010	**B.S. in Electronic and Electrical Engineering.** **Double major: Computer Science and Engineering.** Pohang University of Science and Technology (POSTECH). **Magna cum laude.**

AWARDS AND HONORS

Feb. 2023	**Exemplary Practices in Learner-Centered Teaching and Learning Utilizing EduTech.** Innovation Center for Education, POSTECH.
Apr. 2021	**CSE Award.** In recognition of excellence in teaching and service. Department of Computer Science and Engineering, POSTECH.
Aug. 2020	**Excellence Award at Online Teaching Strategy Contest.** Innovation Center for Education, POSTECH.
Mar. 2014	**Best Invention Award for Industry-Academia Cooperation Projects.** Patent: Memory-centric System Interconnect Structure, Korea Patent 1020572460000. Inventors: Gwangsun Kim, John Kim, and Jung Ho Ahn. SK hynix.
Sep. 2013	**Best Paper Award.** Authors: Gwangsun Kim, John Kim, Jung Ho Ahn, and Jaeha Kim. The 22nd International Conference on Parallel Architectures and Compilation Techniques (PACT).
Dec. 2008	**Best Design Project Award.** Recipients: Gwangsun Kim and Sungjae Ha. Department of Electronic and Electrical Engineering, POSTECH.
2005 - 2009	**National Science and Technology Scholarship.** Korea Research Foundation.

REFEREED INTERNATIONAL CONFERENCE PUBLICATIONS

(*: Co-first authors)

1. Hyungkyu Ham*, Jeongmin Hong*, Geonwoo Park, Yunseon Shin, Okkyun Woo, Wonhyuk Yang, Jinhoon Bae, Eunhyeok Park, Hyojin Sung, Euicheol Lim, and Gwangsun Kim, "Low-overhead General-purpose Near-Data Processing in CXL Memory Expanders." *In Proceedings of the 57th IEEE/ACM International Symposium on Microarchitecture (MICRO), November 2024.* Acceptance rate: 22.7%
2. Shinnung Jeong, Sungjun Cho, Yongwoo Lee, Hyunjun Park, Seonyeong Heo, Gwangsun Kim, Youngsok Kim, Hanjun Kim, "CR2: Community-aware Compressed Regular Representation for Graph Processing on a GPU." *In Proceedings of the 53rd International Conference on Parallel Processing (ICPP),* August 2024.
3. Guseul Heo, Sangyeop Lee, Jaehong Cho, Hyunmin Choi, Sanghyeon Lee, Hyungkyu Ham, Gwangsun Kim, Divya Mahajan, and Jongse Park, "NeuPIMs: NPU-PIM Heterogeneous Acceleration for Batched LLM Inferencing." *In Proceedings of the 29th International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, May 2024.
4. Jeongmin Hong, Sungjun Cho, Geonwoo Park, Wonhyuk Yang, Young-Ho Gong, and Gwangsun Kim, "Bandwidth-Effective DRAM Cache for GPUs with Storage-Class Memory." *In Proceedings of the 30th International Symposium on High-Performance Computer Architecture (HPCA)*, March 2024. Acceptance rate: 18.3%
5. Hans Kasan, Gwangsun Kim, and John Kim, "Dynamic global adaptive routing in high-radix networks." *In Proceedings of the 49th International Symposium on Computer Architecture (ISCA)*, June 2022. Acceptance rate: 16%.
6. Gwangsun Kim, Hayoung Choi, and John Kim, "TCEP: Traffic Consolidation for Energy-Proportional High-Radix Networks." *In Proceedings of the 45th International Symposium on Computer Architecture (ISCA)*, June 2018. Acceptance rate: 16.9%.

7. Gwangsun Kim, Niladrish Chatterjee, Mike O'Connor, and Kevin Hsieh, "Toward Standardized Near-Data Processing with Unrestricted Data Placement for GPUs." *In Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (SC)*, November 2017. Acceptance rate: 18.7%.

8. Wonjun Song, Gwangsun Kim, Hyungjoon Jung, Jongwook Chung, Jung Ho Ahn, Jae W. Lee, and John Kim, "History-Based Arbitration for Fairness in Processor-Interconnect of NUMA Servers." *In Proceedings of the 22nd ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, April 2017. Acceptance rate: 17.4%.

9. Gwangsun Kim, Changhyun Kim, Jiyun Jeong, Mike Parker, and John Kim, "Contention-based Congestion Management in Large-Scale Networks." *In Proceedings of the 49th IEEE/ACM International Symposium on Microarchitecture (MICRO)*, October 2016. Acceptance rate: 21.6%.

10. Gwangsun Kim, Jiyun Jeong, John Kim, and Mark Stephenson, "Automatically Exploiting Implicit Pipeline Parallelism from Multiple Dependent Kernels for GPUs." *In Proceedings of the 25th International Conference on Parallel Architectures and Compilation Techniques (PACT)*, September 2016. Acceptance rate: 26.1%.

11. Byungchul Hong, Gwangsun Kim, Jung Ho Ahn, Yongkee Kwon, Hongsik Kim, and John Kim, "Accelerating Linked-list Traversal through Near-Data Processing." *In Proceedings of the 25th International Conference on Parallel Architectures and Compilation Techniques (PACT)*, September 2016. Acceptance rate: 26.1%.
    Nominated for the Best Paper Award.

12. Kevin Hsieh, Eiman Ebrahimi, Gwangsun Kim, Niladrish Chatterjee, Mike O'Connor, Nandita Vijaykumar, Onur Mutlu, and Stephen W. Keckler, "Transparent Offloading and Mapping (TOM): Enabling Programmer-Transparent Near-Data Processing in GPU Systems." *In Proceedings of the 43rd International Symposium on Computer Architecture (ISCA)*, June 2016. Acceptance rate: 19.6%.

13. Minseok Lee, Gwangsun Kim, John Kim, Woong Seo, Yeongon Cho, and Soojung Ryu, "iPAWS: Instruction-Issue Pattern-based Adaptive Warp Scheduling for GPGPUs." *In Proceedings of the 22nd International Symposium on High Performance Computer Architecture (HPCA)*, March 2016. Acceptance rate: 22.1%.

14. Jongmin Won, Gwangsun Kim, John Kim, Ted Jiang, Mike Parker, and Steve Scott, "Overcoming Far-end Congestion in Large-Scale Networks." *In Proceedings of the 21st International Symposium on High Performance Computer Architecture (HPCA)*, February 2015. Acceptance rate: 22.1%.

15. Gwangsun Kim, Minseok Lee, Jiyun Jeong, and John Kim, "Multi-GPU System Design with Memory Networks." *In Proceedings of the 47th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO)*, December 2014. Acceptance rate: 19.4%.

16. Hanjoon Kim, Gwangsun Kim, Hwasoo Yeo, Seungryoul Maeng, and John Kim, "Transportation-Network Inspired Network-on-Chip." *In Proceedings of the 20th International Symposium on High Performance Computer Architecture (HPCA)*, February 2014. Acceptance rate: 25.6%.

17. Gwangsun Kim, John Kim, Jung Ho Ahn, and Jaeha Kim, "Memory-centric System Interconnect Design with Hybrid Memory Cubes." *In Proceedings of the 22nd International Conference on Parallel Architectures and Compilation Techniques (PACT)*, September 2013. Acceptance rate:17.3%.
    Best Paper Award.

18. Hanjoon Kim, Gwangsun Kim, and John Kim, "Scalable On-chip Network in Power Constrained Manycore Processors." *In Proceedings of the 3rd IEEE International Green Computing Conference (IGCC)*, June 2012.

19. Gwangsun Kim, John Kim, and Sungjoo Yoo, "FlexiBuffer: Reducing Leakage Power in On-Chip Network Routers." *In Proceedings of the 48th ACM/EDAC/IEEE Design Automation Conference (DAC)*, June 2011. Acceptance rate: 23%.

REFEREED INTERNATIONAL JOURNAL PUBLICATIONS
(*: Co-first authors)

1. Byeori Kim, Changhun Lee, Gwangsun Kim, Eunhyeok Park, "Cost-effective Extension of DRAM-PIM for Group-wise LLM Quantization," *IEEE Computer Architecture Letters (accepted).*

2. Hyungkyu Ham*, Wonhyuk Yang*, Yunseon Shin, Okkyun Woo, Guseul Heo, Sangyeop Lee, Jongse Park, and Gwangsun Kim, “ONNXim: A Fast, Cycle-level Multi-core NPU Simulator,” *IEEE Computer Architecture Letters,* vol. 23, no. 2, pp. 219-222, July-Dec. 2024.
3. Hyungkyu Ham*, Hyunuk Cho*, Minjae Kim, Jueon Park, Jeongmin Hong, Hyojin Sung, Eunhyeok Park, Euicheol Lim, Gwangsun Kim, "Non-Invasive, Memory Access-Triggered Near-Data Processing for DNN Training Acceleration on GPUs." *IEEE Access*, vol. 12, pp. 142651-142667, 2024.
4. Jeongmin Hong*, Sungjun Cho*, and Gwangsun Kim, “Overcoming Memory Capacity Wall of GPUs With Heterogeneous Memory Stack.” *IEEE Computer Architecture Letters*, vol. 21, no. 2, July-December 2022.
5. Hyungkyu Ham*, Hyunuk Cho*, Minjae Kim, Jueon Park, Jeongmin Hong, Hyojin Sung, Eunhyeok Park, Euicheol Lim, and Gwangsun Kim, “Near-Data Processing in Memory Expander for DNN Acceleration on GPUs.” *IEEE Computer Architecture Letters*, vol. 20, no. 2, July-December 2021.
6. Hanjoon Kim, Gwangsun Kim, Hwasoo Yeo, John Kim, and Seungryoul Maeng, “Design and Analysis of Hybrid Flow Control for Hierarchical Ring Network-on-Chip.” *IEEE Transactions on Computers*, vol. 65, no. 2, pp. 480-494, February 2016.
7. Gwangsun Kim, Michael M. Lee, John Kim, Jae W. Lee, Dennis Abts, and Michael Marty, “Low-overhead Network-on-Chip Support for Location-oblivious Task Placement.” *IEEE Transactions on Computers*, vol. 63, no. 6, pp. 1487-1500, June 2014.

REFEREED DOMESTIC CONFERENCE PUBLICATIONS

1. Hyungkyu Ham, Wonhyuk Yang, Yunseon Shin, Okkyun Woo, Gwangsun Kim, “Compiler-integrated, High-speed Simulation Methodology for Designing AI Accelerators.” *In Proceedings of the Korea Computer Congress (KCC)*, June 2024.
2. Junho Lee and Gwangsun Kim, “Efficiency Analysis of Lossless Compression for Deep Neural Network Tensor.” *In Proceedings of the Korea Computer Congress (KCC)*, June 2022.
3. Junkyeong Choi and Gwangsun Kim, “Reducing DRAM Energy for Efficient Deep Learning Training on GPUs.” *In Proceedings of the Korea Software Congress (KSC)*, December 2020.

REFEREED WORKSHOP PUBLICATION

1. Gwangsun Kim, John Kim, Jung Ho Ahn, and Yongkee Kwon, “Memory Network: Enabling Technology for Scalable Near-Data Computing.” *In Proceedings of the 2nd Workshop on Near-Data Processing (WoNDP)*, December 2014.

PATENTS FILED / GRANTED

1. Controller For Memory Expansion, Memory Expander, And Data Processing Method Therefor.
   Inventors: Gwangsun Kim, Hyungkyu Ham, Jeongmin Hong.
   Korea Patent Application 10-2024-0028438. Filed February 2024.
   PCT Application PCT/KR2024/013322. Filed September 2024.
2. Graph Neural Network Machine Learning Apparatus and Method.
   Inventors: Gwangsun Kim, Sungjun Cho.
   Korea Patent Application 10-2023-0031819. Filed March 2023.
3. Host Device Performing Near Data Processing Function and Accelerator System Including the Same.
   Inventors: Hyungkyu Ham, Hyunuk Cho, Hyojin Sung, Eunhyeok Park, and Gwangsun Kim.
   US Patent Application US20230195651A1. Filed December 2022.
4. Memory Expansion Device Performing Near Data Processing Function and Accelerator System Including the Same.
   Inventors: Hyungkyu Ham, Hyunuk Cho, Hyojin Sung, Eunhyeok Park, and Gwangsun Kim.
   US Patent Application US20230195660A1. Filed December 2022.
5. Hybrid Memory Device and Managing Method Therefor.
   Inventors: Gwangsun Kim, Sungjun Cho, Jeongmin Hong.
   Korea Patent 1027462890000. **Granted** December 2024.

6. Host Device Performing Near Data Processing Function and Accelerator System Including the Same.
   Inventors: Hyungkyu Ham, Hyunuk Cho, Hyojin Sung, Eunhyeok Park, and Gwangsun Kim.
   Korea Patent Application 10-2022-0071298. Filed June 2022.
7. Memory Expander Performing Near Data Processing Function and Accelerator System Including the Same.
   Inventors: Hyungkyu Ham, Hyunuk Cho, Hyojin Sung, Eunhyeok Park, and Gwangsun Kim.
   Korea Patent Application 10-2022-0071298. Filed October 2022.
8. System, Device and/or Process for Hashing.
   Inventors: Gwangsun Kim and Dam Sunwoo.
   US Patent: US11640381B2. **Granted** May 2023.
9. Memory Network and System Including the same.
   Inventors: Gwangsun Kim, John Kim, and Yongkee Kwon.
   US Patent: US10447584B2. **Granted** October 2019.
   Korea Patent 10-2338266 B1. **Granted** December 2021.
10. Memory-centric System Interconnect Structure.
    Inventors: Gwangsun Kim, John Kim, and Jung Ho Ahn.
    Korea Patent 1020572460000. **Granted** December 2019.
11. Method of Arbitrating for Inter-Connecting within a Multi-Processor System and Apparatuses Performing the same.
    Inventors: John Kim, Wonjun Song, and Gwangsun Kim.
    Korea Patent 1019341950000. **Granted** December 2018.
12. Method and Multi-mode Arbiter with On-chip Network Arbitration for Location-Oblivious.
    Inventors: John Kim and Gwangsun Kim.
    Korea Patent 10-1428878. **Granted** July 2014.
13. Routing System, Flow Control Method, Buffer Management System and Buffer Management Method for Reducing Leakage Power in On-chip Network.
    Inventors: John Kim, Sungjoo Yoo, and Gwangsun Kim.
    Korea Patent 10-2011-0128821. **Granted** April 2013.

ACADEMIC SERVICES

- Program committee member, IEEE International Symposium on High-Performance Computer Architecture (HPCA) – 2024, 2025.
- Program committee member, IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS) – 2025.
- Program committee member, ACM International Conference on Supercomputing (ICS) – 2024.
- Program committee member, Architecture track, IEEE International Parallel & Distributed Processing Symposium (IPDPS) – 2023.
- External review committee, Annual International Symposium on Computer Architecture (ISCA) – 2022, 2024.
- Program committee member, IEEE/ACM International Symposium on Networks-on-Chip (NOCS) – 2020, 2021, 2022, 2023.
- Program committee member, IEEE International Conference on High Performance Computing, Data, and Analytics (HiPC) – 2020.
- Program committee member, IEEE International Conference on Computer Design (ICCD) – 2017, 2019.
- Program committee member, Arm Research Summit – 2018.
- Program committee member, International Symposium on Benchmarking, Measuring and Optimizing (Bench) – 2019, 2020, 2021, 2022.
- Reviewer, IEEE Transactions on Very Large Scale Integration Systems (TVLSI).
- Reviewer, IEEE Transactions on Computers (TC).
- Reviewer, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD).
- Reviewer, IEEE Transactions on Parallel and Distributed Systems (TPDS).
- Reviewer, ACM Transactions on Parallel Computing (TOPC).
- Reviewer, IEEE Computer Architecture Letters (CAL).
- Reviewer, Elsevier Journal of Parallel and Distributed Computing (JPDC).
- Reviewer, IET Computers & Digital Techniques.
- Board member, Korean Institute of Information Scientists and Engineers (KIISE), 2024.
- Vice Chair, ACM SIGARCH Korea Chapter, August 2023 – Present.

TEACHING EXPERIENCES

- CSED199: Freshman Research Participation, POSTECH, Fall 2021/2024.
- CSED311: Computer Architecture, POSTECH, Spring 2019/2020/2021/2022/2023/2024.
- CSED405: GPU and Accelerated Computing, POSTECH, Fall 2024.
- CSED490V: Parallel Architecture and Programming, POSTECH, Fall 2019/2022.
- CSED499: Research Project, POSTECH, Spring 2019.
- CSED503: Advanced Computer Architecture, POSTECH, Fall 2020/2021/2023.
- CSED800B: Computer Science Colloquium, Fall 2022.

INVITED TALKS / TUTORIALS

January 2025	Breaking the Memory Wall: Near-Data Processing for Hyperscale Applications, Invited talk at College of Computing, Yonsei University
December 2024	Breaking the Memory Wall: Near-Data Processing for Hyperscale Applications, Invited talk at KAIST School of Computing Colloquium
August 2024	Low-overhead General-purpose Near-Data Processing in CXL Memory Expanders for Hyperscale AI, Invited talk at KAIST AI PIM Research Center
July 2024	Low-overhead General-purpose Near-Data Processing in CXL Memory Expanders for Hyperscale AI, Invited talk at Rebellions Inc.
May 2024	Low-overhead General-purpose Near-Data Processing in CXL Memory Expanders, Invited talk at Open Compute Project CMS-Computational Programming workstream
May 2024	Cost-effective, General-purpose CXL-NDP for Hyperscale AI, Invited talk at MetisX, Korea
May 2024	Advances and Impacts of SmartNICs in Modern Datacenters, Panel Session at IEEE/IFIP Network Operations and Management Symposium (NOMS)
Nov. 2023	Cost-effective, General-purpose CXL-PNM for Hyperscale AI, Invited talk at Samsung Advanced Institute of Technology (SAIT), Korea
Apr. 2023	Disrupting Processor Architecture with Non-invasive Near-Data Processing: DNN Training Case Study, Invited talk at KAIST, Korea
Apr. 2023	Memory Access-Triggered Near-Data Processing for Accelerating DNN Training, Invited talk at SAPHEON, Korea
Sep.-Nov. 2022	Computer Architecture Course, DS University in Samsung Electronics, Korea
June 2022	Overcoming GPU Memory Wall with Heterogeneous Memory Stack and Near-Data Processing, Invited Talk at ETRI Workshop on Supercomputing
May-June 2022	Computer Architecture Tutorial, DS University in Samsung Electronics, Korea
Jul. 2021	Overcoming Memory Wall in DNN Training through NDP, SK Telecom, Korea
Oct. 2019	Reducing Data Movement Cost Through Near-Data Processing and Interconnect Power Management, Invited talk at Kyungpook National University, Korea
May 2019	Memory-Centric System Architecture for Data-Driven Computing, Invited talk at Ajou University, Korea
Jan. 2019	Memory-centric System Architecture for Data-Driven Computing. Invited talk at KIISE Computer System Society Winter Workshop, Korea.
Dec. 2018	Memory-centric System Architecture for Data-Driven Computing. Invited talk at Electronics and Telecommunications Research Institute (ETRI).
Oct. 2013	Memory-centric System Interconnect Design with Hybrid Memory Cubes. Invited talk at Seoul National University.
Oct. 2013	Network-on-Chip Simulation Methodology. Tutorial at Seoul National University.