Shinnung Jeong

Curriculum Vitae

Contact Information

College of Computing, Georgia Institute of Technology
266 Ferst Drive, KACB 2206, Atlanta, GA 30332-0765

sjeong306@gatech.edu
https://shin0403.github.io/

Education

Yonsei University, Seoul, Republic of Korea
Integrated M.S./Ph.D. student, March 2019 – February 2025
Advisor: Prof. Hanjun Kim

Pohang University of Science and Technology (POSTECH), Pohang, Republic of Korea
Bachelor of Science in Creative IT Engineering, March 2015 – February 2019

Experience

Postdoctoral Researcher, May 2025 – Present
High Performance Architecture Lab (HPArch), *Georgia Institute of Technology*, Atlanta, Georgia, USA

- Supervisors: Prof. Haesun Park and Prof. Hyesoon Kim
- Design and implement a compiler and microarchitecture for the open-source RISC-V GPU, Vortex GPU [4, 16].
- Develop GPU clustering optimizations using non-negative matrix factorization. [2]
- Projects related to GPU microarchitecture and compiler optimizations [3, 1].

Research Assistant, March 2019 – February 2025
Compiler Research Laboratory (Corelab), *Yonsei University*, Seoul, Republic of Korea

- Proposed a graph-processing interface to enlarge the design space for GPU [10].
- Proposed a locality-aware graph storage format for GPUs [7].
- Proposed ML compiler optimizations including memory optimization and real-time system support [8, 9, 11].
- Developed a thread-aware, area-optimized high-level synthesis framework for IoT devices [13, 20, 21].
- Proposed compiler optimization for Homomorphic Encryption [12].

Visiting Scholar, March 2023 – March 2024
High Performance Architecture Lab (HPArch), *Georgia Institute of Technology*, Atlanta, Georgia, USA

- Advisor: Prof. Hyesoon Kim
- Proposed a microarchitecture that converts sparse operations into dense operations for graph processing [5].
- Developed the software stack for the open-source Vortex GPU [14, 17, 18].

Undergraduate Research Assistant, December 2017 – June 2018
Compiler Research Laboratory (Corelab), *POSTECH*, Pohang, Republic of Korea

Undergraduate Research Assistant, June – August 2016, January – February 2017
POSTECH Database and Data Mining Lab (Big data lab), *POSTECH*, Pohang, Republic of Korea

Undergraduate Student Intern, September – December 2016
Excem, Pohang, Republic of Korea

Exchange Student, September 2015 – December 2015
University of California, Berkeley, United States of America

Activities

International Conference Committee

- Program Committee, IEEE/ACM International Symposium on Microarchitecture (MICRO), 2026
- Program Committee, International Symposium on Computer Architecture (ISCA), 2026
- Program Committee, IEEE International Symposium on High-Performance Computer Architecture (HPCA), 2026, 2027
- Program Committee, IEEE International Symposium on Workload Characterization (IISWC), 2026
- Program Committee, IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS), 2026
- Student Research Competition Committee, 58th IEEE/ACM International Symposium on Microarchitecture (MICRO), 2025

Workshop/Tutorial Committee

- Vortex Workshop and Tutorial (co-located with IEEE/ACM International Symposium on Microarchitecture), 2023, 2025, 2026

International Journal Reviewer

- Peer Reviewer, ACM Transactions on Architecture and Code Optimization (TACO), 2023, 2025, 2026
- Peer Reviewer, IEEE MICRO, 2025

Invited Talks

- "Decoupling Schedule and Storage Format for Balanced Graph Processing on a GPU" presented at KYUNG HEE UNIVERSITY, June 23, 2025.

Teaching

- OMSCS 8903: GPU ML Optimization Project
  Teaching Assistant, Fall 2025
- EEE3313-01: Basic Digital Experiments, Yonsei University
  Teaching Assistant, Spring 2021, Spring 2020, Spring 2019
  Taught weekly lab sessions on writing RTL Verilog and designing FPGA-based accelerators (FPGA-ARM).
- EEE3540-01: Microprocessor, Yonsei University
  Teaching Assistant, Fall 2019
  Taught weekly lab sessions on microprocessor programming

Publications

Refereed Conference Publications

[1] Jae Hyung Ju, Euijun Chung, Hritvik Taneja, Anish Saxena, Shinnung Jeong, Hyesoon Kim, Moinuddin K Qureshi, "TileLens: Efficiently Using Large-Granularity Memory Systems with Transparent Two-Dimensional Memory Layout," in *HPCA*, 2027. (To Appear).

[2] Benjamin Cobb*, Shinnung Jeong*, Hyesoon Kim, Richard Vuduc, and Haesun Park, "Accelerating Large-Scale Clustering with Hierarchical NMF-Guided Centroid Initialization," in *IEEE International Conference on Big Data*, 2026. (To Appear, *Co-first author)

[3] Jisheng Zhao*, Huanzhi Pu*, Shinnung Jeong, Chihyo Ahn, and Hyesoon Kim, "CuLifter: Lifting GPU Binaries to Typed IR," in *Proceedings of the 59th IEEE/ACM International Symposium on Microarchitecture (MICRO)*, 2026. (*Co-first author)

[4] Shinnung Jeong, Chihyo Ahn, Huanzhi Pu, Jisheng Zhao, Hyesoon Kim, and Blaise Tine, "Inside VOLT: Designing an Open-Source GPU Compiler," in *Proceedings of the 35th ACM SIGPLAN International Conference on Compiler Construction (CC)*, January 2026.

[5] Shinnung Jeong, Liam Paul Cooper, Ju Min Lee, Heelim Choi, Nicholas Parnenzini, Chihyo Ahn, Yongwoo Lee, Hanjun Kim, and Hyesoon Kim, "SparseWeaver: Converting Sparse Operations as Dense Operations on GPUs for Graph Workloads," in *2025 IEEE International Symposium on High-Performance Computer Architecture (HPCA)*, March 2025.

[6] Shinnung Jeong and Hanjun Kim, "Matrix Pattern-aware Partitioning and Auto-tuning Space Pruning for Graph Processing on GPU," in *The 9th International Conference On Consumer Electronics Asia (ICCE-ASIA)*, November 2024.

[7] Shinnung Jeong, Sungjun Cho, Yongwoo Lee, Hyunjun Park, Seonyeong Heo, Gwangsun Kim, Youngsok Kim, and 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.

[8] Seungbin Song, Ju Min Lee, Haeeun Jeong, Hyunho Kwon, Shinnung Jeong, Jaeho Lee, and Hanjun Kim, "TeMCO: Tensor Memory Compiler Optimization across Tensor Decompositions in Deep Learning Inference," in *Proceedings of the 53rd International Conference on Parallel Processing (ICPP)*, August 2024.

[9] Jaeho Lee, Shinnung Jeong, Seungbin Song, Kunwoo Kim, Heelim Choi, Youngsok Kim, and Hanjun Kim, "Occamy: Memory-efficient GPU Compiler for DNN Inference," in *Proceedings of the 60th Annual Design Automation Conference 2023 (DAC)*, July 2023.

[10] Shinnung Jeong, Yongwoo Lee, Jaeho Lee, Heelim Choi, Seungbin Song, Jinho Lee, Youngsok Kim, and Hanjun Kim, "Decoupling Schedule, Topology Layout, and Algorithm to Easily Enlarge the Tuning Space of GPU Graph Processing," in *Proceedings of the 31st International Conference on Parallel Architectures and Compilation Techniques (PACT)*, October 2022.

[11] Seonyeong Heo, Shinnung Jeong, and Hanjun Kim, "RTScale: Sensitivity-Aware Adaptive Image Scaling for Real-Time Object Detection," in *34th Euromicro Conference on Real-Time Systems (ECRTS)*, July 2022.

[12] Yongwoo Lee, Seonyeong Heo, Seonyoung Cheon, Shinnung Jeong, Changsu Kim, Eunkyung Kim, Dongyoon Lee, and Hanjun Kim, "HECATE: Performance-Aware Scale Optimization for Homomorphic Encryption Compiler," in *Proceedings of the 2022 International Symposium on Code Generation and Optimization (CGO)*, April 2022.

[13] Changsu Kim, Shinnung Jeong, Sungjun Cho, Yongwoo Lee, William Song, Youngsok Kim, and Hanjun Kim, "Thread-Aware Area-Efficient High-Level Synthesis Compiler for Embedded Devices," in *Proceedings of the 2021 International Symposium on Code Generation and Optimization (CGO)*, March 2021.

Refereed Journal Publications

[14] Chihyo Ahn, Shinnung Jeong, Liam Paul Cooper, Ruobing Han, Huanzhi Pu, Nicholas Parnenzini, Jisheng Zhao, Blaise Tine, and Hyesoon Kim, "FastTrackGPU: Static-Analysis–Guided Analytical Modeling for Softcore GPUs," in *IEEE Computer Architecture Letters (CAL)*, 2026.

[15] Bongjun Kim, Seonyeong Heo, Jaeho Lee, Shinnung Jeong, Yongwoo Lee, and Hanjun Kim, "Compiler-assisted Semantic-aware Encryption for Efficient and Secure Serverless Computing," in *IEEE Internet of Things Journal*, April 2021.

Refereed Workshop Publications

[16] Shinnung Jeong, Chihyo Ahn, Sai Hemanth Reddy Bheemreddy, Liam Cooper, Krishil Gandhi, Chulhyung Park, Vincent Pham, Huanzhi Pu, Saurabh Singh, Mitul Tandon, Jisheng Zhao, and Hyesoon Kim, "Toward End-to-End ML Support on Vortex," in *Open-Source Computer Architecture Research (OSCAR)*, 2026.

[17] Huanzhi Pu, Rishabh Ravi, Shinnung Jeong, Udit Subramanya, Euijun Chung, Jisheng Zhao, Chihyo Ahn, and Hyesoon Kim, "Hardware vs. Software Implementation of Warp-Level Features in Vortex RISC-V GPU ," in *Open Source Solutions for Massively Parallel Integrated Circuits (OSSMPIC)*, April 2025.

[18] Chihyo Ahn, Shinnung Jeong, Liam Paul Cooper, Nicholas Parnenzini, and Hyesoon Kim, "Comparative Analysis of Executing GPU Applications on FPGA: HLS vs. Soft GPU Approaches ," in *The Third International Workshop on Coarse-Grained Reconfigurable Architectures for High-Performance Computing (CGRA)*, May 2024.

Refereed Poster Publications

[19] Chan Lee, Shinnung Jeong, Heelim Choi, Jaeho Lee, Haeeun Jeong, Hoyun Youm, Ju Min Lee, and Hanjun Kim, "Approximation-based Inter-PE Communication-free Image Filtering for Commodity PIM," in *Proceedings of the 62th Annual Design Automation Conference 2025 - (Poster) (DAC)*, June 2025.

[20] Changsu Kim, Yongwoo Lee, Shinnung Jeong, and Hanjun Kim, "Logic Deduplication with Decentralized Pointer Analysis in HLS for Post-Quantum Cryptography Algorithms," in *Proceedings of the 57th Annual Design Automation Conference 2020 - (Poster) (DAC)*, July 2020.

[21] Changsu Kim, Yongwoo Lee, Shinnung Jeong, Wen Wang, Jakub Szefer, and Hanjun Kim, "Pipeline-aware Logic Deduplication in High-Level Synthesis for Post-Quantum Cryptography Algorithms," in *Proceedings of the 2020 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays (FPGA)*, February 2020.

Open-access Preprint Publication

[22] Liam Cooper and Shinnung Jeong and Hyeran Jeon and Jeffrey Young and Hyesoon Kim "Accelerating the Mitigation of LLM Inference Nondeterminism Across GPU Architectures," in Arxiv, 2026

Ph.D thesis

[23] Shinnung Jeong, "Decoupling Scheduling and Storage Formats for Balanced Graph Processing on a GPU," Ph.D. Dissertation, Yonsei University, February 2025.

Patents

[24] Hanjun Kim, Chan Lee, Heelim Choi, Haeeun Jeong, Shinnung Jeong, Ju Min Lee, Jaeho Lee, and Hoyun Youm, "Method of Filtering Image," KR Patent App. 10-2025-0181300, November 2025.

[25] Hanjun Kim, Bongjun Kim, Jaeho Lee, Seonyeong Heo, Shinnung Jeong, and Yongwoo Lee, "IoT Service Providing Method Based on Adaptive Encryption and IoT Apparatus," KR Patent Number 10-2508448-0000, March 2024.

[26] Hanjun Kim, Youngsok Kim, Changsu Kim, Shinnung Jeong, Yongwoo Lee, Sungjun Cho, and William Song, "High-Level Synthesis Method and Apparatus for Hardware Optimization," KR Patent App. 10-2022-0023420, February 2022.