Bongjun Kim

Curriculum Vitae

Contact Information

AI System Platform
Samsung Advanced Institute of Technology
130, Samsung-ro, Yeongtong-gu, Suwon-si
Gyeonggi-do, South Korea, 16678

+82-010-4110-9412
bkim.bongjun@gmail.com
bong90@postech.ac.kr
https://vongzun.github.io

Education

Pohang University of Science and Technology (POSTECH), Pohang-si, Republic of Korea
Ph.D. in Computer Science and Engineering, August 2021
Thesis: "Semantic-aware Automatic Program Partitioning and Privacy-preserving Selective Encryption for Fog Computing"
Advisor: Prof. Hanjun Kim and Prof. Jong Kim
March 2016 to August 2021

Pohang University of Science and Technology (POSTECH), Pohang-si, Republic of Korea
B.S. in Computer Science and Engineering, February 2016
March 2009 to February 2016

Experience

Staff Researcher, September 2021 to Present
AI System Platform, *Samsung Advanced Institute of Technology (SAIT)*, Suwon-si, Republic of Korea

- Leading the development of a Universal Deep Learning Compiler (UDLC), targeting heterogeneous AI accelerators including NPUs, GPUs, and memory-centric architectures with Processing-in-Memory (PIM).
- Presented the UDLC framework as a poster at the 4th Workshop on the Compilers for Machine Learning (C4ML), co-located with CGO 2023.
- Contributed to system-level enablement of HBM2-PIM architectures, including FPGA-based evaluation platforms (FCCM 2024).
- Collaborated with Codeplay (Intel) on heterogeneous programming model support for general-purpose PIM systems (ISC-HPC 2024).

Research Assistant, March 2016 to August 2021
Compiler Optimization Research Laboratory (CoreLab), *POSTECH / Yonsei University*, Republic of Korea

- Programming Models and Compilers for IoT: Developed high-level programming abstractions and compiler frameworks for IoT applications, reducing development effort (LCTES'17, IEEE IoT-J'21)
- Distributed Execution Frameworks for IoT: Developed compiler-runtime systems to automatically orchestrate computation across IoT/Edge devices, improving system-wide performance and reducing latency through intelligent partitioning and unified memory abstractions (IEEE Micro'16, CC'19)
- Resource-Efficient and Dependable IoT Services: Developed compiler-assisted techniques for enhancing security, efficiency, and reliability in IoT environments, addressing challenges in encryption, scheduling, and fault detection (IEEE IoT-J'21, RTAS'20, EMSOFT'21)

Research Intern, September 1, 2019 to January 31, 2020
Samsung Advanced Institute of Technology, Suwon-si, Republic of Korea

- Designed a compiler for server-scale many-core NPUs (Mentor: Jinseong Kim)

Recognition

- Silver Medal, Graduate Student Research Competition at International Symposium on Code Generation and Optimization (CGO), February 2020
- Best Oral Presentation Award, Student Workshop at Department of Computer Science and Engineering (POSTECH), September 2016

Projects

- MLIR-based Deep Learning Compiler for Heterogeneous Architectures
  - Develop a universal deep learning compiler targeting NPUs, GPUs, and PIM-equipped GPUs
  - Achieve 1.4x latency optimization for large language model inference compared to existing PIM AI Compiler
  - Samsung Electronics, September 2021 to Present
- Programming Model Extensions for Processing-In-Memory (PIM)
  - Designed and implemented a compiler to support extended OpenMP programming model for efficient general-purpose computing on GPU-PIM
  - Demonstrated 1.2x performance improvement on real-world HPC applications
  - Samsung Electronics, March 2023 to December 2023

Activities

International Journal Reviewer

- Journal Reviewer, The Journal of Supercomputing (3 papers), 2024
- Journal Reviewer, Cluster Computing, 2024
- Journal Reviewer, IEEE Transactions on Cloud Computing (TCC), 2021

International Conference Artifact Evaluation

- Artifact Evaluation Committee, International Symposium on Code Generation and Optimization (CGO), 2019

Teaching

- LLVM Compiler Tutorial on Program Analysis and Optimization
  Lecturer, LIG Nex1, 27 February 2018
  (Taught how to write LLVM passes for program analysis and optimization)
- CSED 423: Compiler Design
  Teaching Assistant, Spring 2017
  (Conducted weekly lab classes about compiler implementation and graded lab assignments)
- CSED 312: Operating System
  Teaching Assistant, Fall 2017
  (Conducted lab classes about operating system implementation and graded lab assignments)

Publications

Refereed Journal Publications

[1] Gyeongmin Lee, Bongjun Kim, Seungbin Song, Seonyeong Heo, and Hanjun Kim, "ComFlex: Composable and Flexible Resource Management for the IoT," in *IEEE Internet of Things Journal*, November 2021.
IF=9.936, Q1 (JCR 2019)

[2] 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.
IF=9.936, Q1 (JCR 2019)

[3] Bongjun Kim, Seonyeong Heo, Gyeongmin Lee, Soyeon Park, Hanjun Kim, and Jong Kim, "Heterogeneous Distributed Shared Memory for Lightweight Internet-of-Things Devices," in *IEEE Micro*, November 2016.
IF=1.933, Q2 (JCR 2016)

Refereed Conference Publications

[4] Jinwoo Choi, Yeonan Ha, Hanna Cha, Seil Lee, Sungchul Lee, Jounghoo Lee, Shin-haeng Kang, Bongjun Kim, Hanwoong Jung, Hanjun Kim, and Youngsok Kim, "MPC-Wrapper: Fully Harnessing the Potential of Samsung Aquabolt-XL HBM2-PIM on FPGAs," in *Proceedings of the 32nd IEEE International Symposium on Field-Programmable Custom Computing Machines (FCCM)*, May 2024.

[5] Hyesun Hong, Lukas Sommer, Bongjun Kim, Mikhail Kashkarov, Kumudha Narasimhan, Ilya Veselov, Mehdi Goli, Jaeyeon Kim, Ruyman Reyes Castro, and Hanwoong Jung, "Programming Model Extensions for General-Purpose Processing-In-Memory," in *Proceedings of the 39th International Conference on High Performance Computing, ISC High Performance 2024 (ISC-HPC)*, May 2024.

[6] Gyeongmin Lee, Bongjun Kim, Seungbin Song, Changsu Kim, Jong Kim, and Hanjun Kim, "Precise Correlation Extraction for IoT Fault Detection with Concurrent Activities," in *Proceedings of the International Conference on Embedded Software (EMSOFT)*, October 2021.

[7] Seonyeong Heo, Seungbin Song, Bongjun Kim, and Hanjun Kim, "Sharing-aware Data Acquisition Scheduling for Multiple Rules in the IoT," in *Proceedings of the IEEE Real-Time And Embedded Technology And Applications Symposium (RTAS)*, April 2020.

[8] Bongjun Kim, Seonyeong Heo, Gyeongmin Lee, Seungbin Song, Jong Kim, and Hanjun Kim, "Spinal Code: Automatic Code Extraction for Near-User Computation in Fogs," in *Proceedings of the 28th International Conference on Compiler Construction (CC)*, February 2019.

[9] Gyeongmin Lee, Seonyeong Heo, Bongjun Kim, Jong Kim, and Hanjun Kim, "Rapid prototyping of IoT applications with Esperanto compiler," in *Proceedings of the 28th International Symposium on Rapid System Prototyping (RSP)*, October 2017. Invited.

[10] Gyeongmin Lee, Seonyeong Heo, Bongjun Kim, Jong Kim, and Hanjun Kim, "Integrated IoT Programming with Selective Abstraction," in *Proceedings of the 18th ACM SIGPLAN/SIGBAD Conference on Languages, Compilers, Tools, and Theory for Embedded Systems (LCTES)*, June 2017.

Patents

[11] Gyeongmin Lee, Bongjun Kim, Seungwon Lee, and Hanwoong Jung, "Near-memory operator and method with accelerator performance improvement," US Patent No. US12524236B2, January 2026.

[12] Youngsok Kim, Sungchul Lee, Seil Lee, Suhyun Lee, Jounghoo Lee, Chaemin Lim, Hanna Cha, Jinwoo Choi, Yeonan Ha, Bongjun Kim, and Hanwoong Jung, "Electronic device and method with processing in memory operation," US Patent App. US20250335126A1, October 2025.

[13] Gyeongmin Lee, Bongjun Kim, and Hanwoong Jung, "Method and device with iterative compilation for deep learning," US Patent App. US20240411532A1, May 2024.

[14] Bongjun Kim, Gyeongmin Lee, and Hanwoong Jung, "Electric device and method with hardware-optimized compilation," US Patent App. US20240411531A1, May 2024.

[15] Youngsok Kim, Jounghoo Lee, Yeonan Ha, Sangsu Lee, and Bongjun Kim, "Electronic device and method with hardware acceleration," US Patent App. US20250045196A1, May 2024.

[16] 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 No. 10-2508448-0000, March 2023.

[17] Hanjun Kim, Seungbin Song, Bongjun Kim, and Seonyeong Heo, "Scheduling Apparatus and Method based on Data Sharing between Multiple Rules in IoT Environment," KR Patent No. 10-2382328-0000, March 2022.

[18] Hanjun Kim, Seonyeong Heo, Jong Kim, and Bongjun Kim, "Write Control Method and Disk Controller for Automated Backup and Recovery," KR Patent No. 10-21896070000, December 2020.

[19] Bongjun Kim, Jong Kim, Soyeon Park, Hanjun Kim, Seonyeong Heo, and Gyeongmin Lee, "Heterogeneous Distributed Shared Memory For IoT Devices," KR Patent No. 10-18579070000, May 2018.