Sowoong Kim

Ph.D. Candidate, Computer Science and Engineering, UNIST
bioloid@unist.ac.kr · bioloid.github.io · Google Scholar
Room 607-2, Bldg. 106, UNIST, Ulsan, Republic of Korea

Research Interests

Systems and software for high-performance and efficient generative AI, including on-device inference; computer security, including microarchitectural side-channel attacks.

Education

Ph.D. in Computer Science and Engineering 2021 – Present
Ulsan National Institute of Science and Technology (UNIST), Ulsan, Korea
Advisor: Prof. Woongki Baek

B.S. in Computer Science and Engineering 2017 – 2021
Ulsan National Institute of Science and Technology (UNIST), Ulsan, Korea
Minor in Electrical Engineering

Publications

Conference Papers

[C5] **Sowoong Kim** and Woongki Baek, "Deco: High-Throughput MoE Inference through Decoupled Batching and Cooperative Execution," in *Proceedings of the 44th IEEE International Conference on Computer Design (ICCD)*, Nov. 2026. (To appear)

[C4] **Sowoong Kim**, Youngsam Shin, YeonGon Cho, and Woongki Baek, "Harmonia: QoS-Aware and High-Throughput Generative Inference with a Single GPU," in *Proceedings of the 32nd International European Conference on Parallel and Distributed Computing (Euro-Par)*, Aug. 2026. doi:10.1007/978-3-032-35251-4_25

[C3] **Sowoong Kim**, Eunyeong Sim, Youngsam Shin, YeonGon Cho, and Woongki Baek, "Activation Sequence Caching: High-Throughput and Memory-Efficient Generative Inference with a Single GPU," in *Proceedings of the 33rd International Conference on Parallel Architectures and Compilation Techniques (PACT)*, Oct. 2024. doi:10.1145/3656019.3676945

[C2] **Sowoong Kim**, Myeonggyun Han, and Woongki Baek, "MARF: A Memory-Aware CLFLUSH-Based Intra- and Inter-CPU Side-Channel Attack," in *Proceedings of the 28th European Symposium on Research in Computer Security (ESORICS)*, Sep. 2023. doi:10.1007/978-3-031-51479-1_7

[C1] **Sowoong Kim**, Myeonggyun Han, and Woongki Baek, "DPrime+DAbort: A High-Precision and Timer-Free Directory-Based Side-Channel Attack in Non-Inclusive Cache Hierarchies using Intel TSX," in *Proceedings of the 28th IEEE International Symposium on High-Performance Computer Architecture (HPCA)*, Apr. 2022. doi:10.1109/HPCA53966.2022.00014

Patents

[P5] Woongki Baek, **Sowoong Kim**, Youngsam Shin, and YeonGon Cho, "Method and Apparatus for Processing Generative Model Using Multi-Processing Group," Korean Patent Application No. 10-2025-0077369, filed Jun. 2025; U.S. Patent Application No. 19/335,591, filed Sep. 2025; also filed in Europe and China. (Pending)
Related publication: [C4]

[P4] Woongki Baek, **Sowoong Kim**, Eunyeong Sim, and Youngsam Shin, "Method and Apparatus for Inference Using Generative Model," Korean Patent Application No. 10-2024-0046554, filed Apr. 2024; U.S. Patent Application No. 19/011,709, filed Jan. 2025; also filed in Europe, China, and Taiwan. (Pending)
Related publication: [C3]

[P3] Woongki Baek, **Sowoong Kim**, and Myeonggyun Han, "Apparatus for Performing Cache Side-Channel Attacks on Multiple CPUs and Attack Method Thereof," Korean Patent No. 10-2816368, registered May 2025.
Related publication: [C2]

[P2] Woongki Baek, **Sowoong Kim**, and Myeonggyun Han, "Method for Defending Against Side-Channel Attacks Based on Transactional Memory," Korean Patent No. 10-2724715, registered Oct. 2024.
Related publication: [C1]

[P1] Woongki Baek, **Sowoong Kim**, and Myeonggyun Han, "Side-Channel Attack Method for CPUs with Non-Inclusive Cache Hierarchies," Korean Patent No. 10-2724686, registered Oct. 2024.
Related publication: [C1]

Honors and Awards

NAVER Ph.D. Fellowship Dec. 2022
NAVER Corporation (KRW 5 million)

Last updated: November 28, 2026