Last updated in April 2026

Sunwoo Kim

PhD Student — Compiler-aided reasoning for chip design

Ithaca, NY | sk3463@cornell.edu | (607) 262-0961 | sunwookim028.github.io | github.com/sunwookim028

Education

Cornell University, PhD in Computer Engineering — Aug 2025 – present

- Advisor: Prof. Zhiru Zhang
- Selected coursework: Advanced Compilers; High-Level Digital Design Automation

Seoul National University, BS in Electrical & Computer Engineering (cum laude) — Mar 2019 – Feb 2025

Experience

Student Researcher, Computer Systems Lab, Cornell University – Ithaca, NY — Aug 2025 – present

- Research: Chip design infrastructure using MLIR-based abstractions and LLM agents (see Projects)
- Mentoring: 14 undergraduate researchers across two semesters; several since interned at AMD and Micron
- Service: Artifact Evaluation Committee, MLSys 2026

Teaching Assistant, School of ECE, Cornell University – Ithaca, NY — Jan 2026 – present

- Developing hardware design and programming labs for a new ECE course on AI hardware, taught by Prof. Zhang

Research Intern, Accelerated Intelligent Systems Lab, SNU – Seoul, Korea — Jan 2024 – June 2025

- Co-led kernel design and indexing optimization for G^3SA, a GPU-accelerated sequence alignment library
- Co-first author on ACM ICS 2025 publication

Projects

Compiler-Aided Reasoning for Agentic RTL Design *(link)* — Sept 2025 – present

Compiler infrastructure and composability as grounding for LLM-based hardware design.

- Implemented a parametrized TPU in Allo; extended Allo with scratchpad modeling and kernel-launch programming to support ASIC beyond FPGA backends (~15× fewer lines of code than our Verilog baseline)
- Compiler-based composition of HLS and IP modules for verifiable assembly, targeting a CGRA design
- MLIR-based microarchitecture abstractions above HLS/RTL for automated ISA extraction and compiler backend generation, targeting the TPU design
- Task decomposition and position-independent context caching for large-scale design tasks
- Harness design and post-training for EDA-specific tool use and reasoning

Agentic PyTorch Backend for Heterogeneous Acceleration *(link)* — Oct 2025 – present

Agent-driven backend adaptation for evolving heterogeneous hardware (e.g., edge NPUs, chiplet-based SiPs) and kernel libraries.

- Program partitioning (e.g., PD disagg.) and compilation for heterogeneous systems (NPU+GPU, CGRA+PIM)
- Demonstrated on AMD Strix Halo with IRON and ROCm libraries
- Automated kernel-library registration and partitioning-logic updates via LLM agents

Hardware-Accelerated Scalable Tooling *(link)* — Jan 2024 – present

Multi-GPU software for genomics; GPU-accelerated simulation for chip design.

- Multi-GPU acceleration for high-throughput NGS pipelines (kernels for all seed, chain, extend alignment stages)
- Triton migration for long-term maintainability
- GPU-accelerated architecture simulation for chip design

Selected Honors

- Fulbright Award for Graduate Studies in STEM (graciously declined), $40k (2024)
- Korea Presidential Science Scholarship, full tuition and stipend (2019–2024)
- SNU Semiconductor Specialization Scholarship, $11k (2023)

Publications

G^3SA: A GPU-Accelerated Gold Standard Genomics Library for End-to-End Sequence Alignment *(link)* ICS 2025

Yeejoo Han*, **Sunwoo Kim***, Seongyeon Park, Jinho Lee *(* equal contribution)*

From Pragmas to Partners: A Symbiotic Evolution of Agentic High-Level Synthesis LATTE 2026 (Workshop)
(link)

Niansong Zhang, **Sunwoo Kim**, Shreesha Srinath, Zhiru Zhang

Skills

Programming Languages: C, C++, CUDA, Python, Bash, Verilog, SystemVerilog

Frameworks & Tools: MLIR, PyTorch, Allo, Vitis HLS, cocotb, Vivado, SPICE, Make, gdb, git