HONGSEUNG YU

hongseung@snu.ac.kr hongseung-yu github.com/HongseungYu hongseungyu.com

INTERESTS

AI Systems and Hardware, Hardware-Software Co-design, Edge and Mobile Environments, Operating Systems

EDUCATION

Seoul National University Mar. 2025 - Feb. 2027 (Expected)
M.S. in Electrical and Computer Engineering
Overall GPA : 4.3/4.3
Advised by Prof. Kyunghan Lee

Seoul National University Mar. 2019 - Feb. 2025
B.S. in Electrical and Computer Engineering, summa cum laude
Overall GPA: 4.04/4.3 (Senior Year: 4.3/4.3)
2-year absence due to military service (Apr. 2021 - Jan. 2023)

PUBLICATIONS

mzCache: Mobile LLM Memory Management for Responsive Inference under Multitasking
Hongseung Yu, Minsung Kim, Jongseok Park, Kyunghan Lee
Under Review at ACM MobiCom 2026

You Only Profile Once: Roofline-Guided DVFS for Energy-Efficient LLM Inference on Mobile Devices
Yonghwa Cho, Minsung Kim, Wonjun Bang, Hyunah Kim, **Hongseung Yu**, Seyeon Kim, Kyunghan Lee
Under Review at ACM MobiCom 2026

BVSR: BitVector-Sum Representation for Distribution-Aware LLM Weight Quantization
Wonjun Bang, Jongseok Park, **Hongseung Yu**, Sanghyun Han, Kyungmin Bin, Kyunghan Lee
Available at https://arxiv.org/abs/2509.18172

HONORS & AWARDS

Grand Prize (Minister's Award, 1st Place) Jan. 2025
2024 AI Semiconductor Idea & Design Competition (AI-ON), Ministry of Science and ICT, Republic of Korea.
Prize includes a 10-day Silicon Valley research training program.

Scholarship to Academic Excellence 2019, 2023, and 2024
Tuition Scholarship, Seoul National University

Recognition of Superior Performance, Ulchi Freedom Shield 22-2 Sep. 2022
Presented by Lt. Gen. Scott L. Pleus (U.S. Air Force) and Lt. Gen. Ha Sik Park (ROK Air Force).
Selected as one of only two soldiers recognized for exceptional performance during the exercise.

Service Award Dec. 2020
Recognized for outstanding contributions to the student council in the ECE Department.

RESEARCH & ENGINEERING EXPERIENCE

M.S. Research Project Mar. 2025 - Present
Advisor: Prof. Kyunghan Lee *Seoul National University*

- **mzCache: On-device LLM inference system adaptive to multitasking memory pressure**
  - Maintains responsive LLM inference on smartphones under multitasking-induced memory pressure
  - Restructures weight and KV cache allocation in llama.cpp to enable fine-grained eviction and rapid restoration
  - Built and deployed as an Android application using OpenCL kernels optimized for mobile Adreno GPUs
- **ARM NEON CPU Kernel for BVSR Weight Quantization**

- Implemented the ARM NEON CPU kernel for BVSR weight quantization; details TBD (see Publications).

RADACS: Resolution-Aware Diffusion Accelerator Nov. 2024 - Jan. 2025
AI Semiconductor Idea & Design Competition *Ministry of Science and ICT*

- *Grand Prize, 1st of 73 teams.* First accelerator to bake *computational staleness* into hardware.
- Designed HMMU/LMMU matmul units in DU-NE/QU-NE engines connected by a staleness-aware feature cache.
- Vitis HLS on Xilinx Zynq-7000: 1.17×/1.47× speedups over a single-MMU baseline. **[press (KR)]** **[pdf]**

Mixed-Format MAC Unit for MXFP8 / MXFP6 / MXFP4 Jun. 2024 - Dec. 2024
Advisor: Prof. Jae-Joon Kim *Undergraduate Thesis*

- Designed a Verilog MAC unit handling arbitrary mixes of MXFP8/6/4 sub-formats via per-element alignment.
- 36-bit adder tree + 67-bit accumulator with a barrel shifter for single-cycle normalization.
- Synthesized on Samsung 28nm CMOS: 46% lower area, 42% lower power vs. FP32 baseline. **[poster]**

High-Speed SHA-256 Hardware Accelerator Apr. 2024 - Aug. 2025
Advisor: Prof. Deog-Kyoon Jeong (Emeritus) *Undergraduate Internship*

- Designed the full RTL (datapath and control logic) of a SHA-256 accelerator in Verilog.
- Optimized via loop unrolling, quasi-pipelining, and logic/gate-level refinements for low latency.
- Synthesis and preliminary P&R with Synopsys DC/ICC and Cadence Virtuoso. **[page]** **[github]**

DON'T LXXK UP — 24/7 Unmanned Photo Booth Kiosk Jan. 2023 - Present
Co-founder & Software Lead *Mapo-gu, Seoul, Republic of Korea*

- Co-founded Korea's first unmanned high-angle photo booth, defining a new self-photo category.
- Built a C#/WinForms kiosk app coordinating DSLR, photo printer, payment terminal, and touchscreen UI.
- Designed for 24/7 unattended operation with crash tolerance; running for 3+ years since 2023. **[page]**

LEADERSHIP & MENTORING

STEM (SNU Tomorrow's Engineers Membership) Sep. 2024 - Mar. 2026
Honor Society of College of Engineering *Seoul National University*

- **Vision Mentor — High-School Outreach**
  - Talks on study methods, CSAT prep, and GPA management for ˜40 high-school students.
  - Held 1:1 career mentoring on engineering majors and university life.
- **MasterpiECE DAY Mentor — Department-Level Career Mentoring**
  - Mentor for Electrical and Computer Engineering first- and second-year students.
  - Open Q&A on undergrad experience, careers, grad school, and electives.

Student Council Mar. 2019 - Feb. 2021
Department of Electrical and Computer Engineering *Seoul National University*

- **Team Leader — eSports Tournament Team**
  - Led the eSports tournament team for the annual ECE department-wide event.
  - Organized cross-cohort competitions (League of Legends, StarCraft) between students and faculty.
- **Freshman Mentor**
  - Mentored 10 first-year students through course registration under COVID-19.
  - Organized peer meetups so freshmen could connect during the remote semester.

SKILLS

- **Software Programming:** C/C++/C#, CUDA, OpenCL, ARM NEON, AVX2, OpenMP, Python, Java
- **Hardware Design:** Verilog, Xilinx Vivado, Vitis HLS, FPGA Development, Cadence Virtuoso, Synopsys Design Compiler