Seoyoon Chae

Suwon, South Korea | sychaeee@g.skku.edu | chae-sy.github.io |

Google Scholar | LinkedIn | GitHub

Profile

Senior undergraduate specializing in AI hardware acceleration, with hands-on experience in Verilog-based ASIC/FPGA design, full digital tape-out flow, GPU CUDA kernel optimization, and neural network quantization. Passionate about co-designing ML models and custom accelerator architectures for energy-efficient edge AI.

Education

SungKyunKwan University (SKKU), B.S. in Electronic and Electrical Engineering Feb. 2022 – Aug. 2026

- **GPA:** 3.89 / 4.5 (Equivalent to 3.65 / 4.0)
- **Coursework:** AI Integrated Circuit Design (A), Digital Systems Design (A+), Electronic & Electrical Programming Lab (A+), Electronic Circuits I (A), Signals and Systems (A), Introduction to Machine Learning (A)
- **Teaching Experience:** Thermodynamics [ISS3317] (Teaching Assistant, Summer 2025)

Publications and Patents

Conference

C2 **Seoyoon Chae**, and Taewook Kang. "LLM on FPGA: Squeezing Language Models by Quantization and Multi-Query Attention and Its Efficient Hardware Architecture," in *2025 IEEE 22nd International SoC Design Conference (ISOCC)*. (acceptance rate : ∼40%) **[paper] [github] [link]**

C1 Daehee Lee, Sewon Kim, Jungmin Yoo, Yewon Jeong, **Seoyoon Chae**, Chaegyu Lee, Hyunseung Choo, and Jongpil Jeong. "Rotor Fault Diagnosis Method Based on 1D-CNN LSTM," in *2023 O2O Integrated Conference of the Institute of Internet, Broadcasting and Communication*, pp. 4–6. (in Korean) **[link]**

Patent

P1–P2 Jongpil Jeong, Daehee Lee, Jungmin Yoo, **Seoyoon Chae**, Sewon Kim, Chaegyu Lee, and Yewon Jeong. "Method for Rotor Failure Detection Based on Artificial Intelligence," Korean Patents 10-2024-0008103 and 10-2024-0008072, filed Jan. 2024 (pending).

Research Experience

Efficient and Intelligent Computing (EIC) Lab, SKKU - *Undergraduate Researcher* Jul. 2025 – Present
Advisor: Prof. Yulhwa Kim

- `[Dec.2025-Present]` Developed adaptive KV-cache quantization and tile-level precision control for LLM inference, co-designing quantization policies with memory bandwidth and hardware dataflow constraints to reduce KV memory traffic while preserving model quality.
- `[Jul.2025-Dec.2025]` Built optimized CUDA kernels for LLM inference leveraging block tiling, vectorized memory access, shared-memory reuse, warp-level tiling, and Tensor Core WMMA instructions. Reached near–PyTorch training/inference performance.

Kang Research Group, SKKU - *Undergraduate Researcher* Jul. 2024 – Jul. 2025
Advisor: Prof. Taewook Kang

- `[Jul.2024-Dec.2024]` Fine-tuned CNN keyword-spotting (KWS) models for hardware-friendly deployment using framewise incremental computation scheme; designed Verilog RTL blocks for efficient CNN operations. Participated in the Multi-Project Wafer (MPW) program for undergraduate and graduate students organized by IDEC and contributed to the **tape-out** of KWS ASIC tape-out. **[link]**
- `[Jan.2025-Jun.2025]` Quantized BC-ResNet for keyword spotting(KWS) and speaker-verification(SV) embeddings; implemented PTQ/QAT (CLE, bias correction, AWQ, SmoothQuant, PACT). Participated in the Multi-Project Wafer (MPW) program organized by IDEC and contributed to KWS and SV ASIC **tape-out** . **[link]**

- [Jun.2025-Jul.2025] Implemented Multi-Query Attention and W4A4 for FPGA LLM deployment; designed RTL in SystemVerilog for Transformer blocks; Conducted synthesis and bitstream generation for FPGA implementation in Xilinx Vivado. Resulted in publication [C2] **[link]**

Project Experience

Senior Capstone Design – Project Leader **[github] [link]** Mar. 2025 – Jun. 2025

- Led a team of four undergraduates to implement and train a CNN-based subpixel rendering algorithm, optimizing network parameters using Incremental Network Quantization scheme.
- Designed its hardware accelerator in Verilog: developed a pipelined structure and validated its functionality.
- Executed the full digital ASIC tape-out flow, including RTL synthesis, timing-driven place-and-route, power and timing sign-off, and final GDSII generation. Awarded Engineering Innovation Prize.

Smart Factory Capstone Design – Undergraduate Researcher **[link]** Aug. 2023 – Dec. 2023

- Developed a 1D-CNN LSTM–based fault diagnosis system for rotating machinery using vibration and acoustic signals, achieving up to 97.4% classification accuracy across multiple fault conditions.
- Utilized time-series feature extraction and sequence learning techniques to detect mechanical anomalies such as mechanical looseness, improving predictive maintenance capability. Resulted in [C1] and [P1-2]

Skills

Programming & ML: CUDA (kernel optimization), C/C++, Python (PyTorch), LLVM

Hardware & EDA: (System)Verilog, Synopsys IC Compiler II, Design Compiler, Xilinx Vivado, Cadence Virtuoso

Simulation & Verification: gem5, NVIDIA Nsight Compute, MATLAB

Workshops & Training

- **IDEC Workshop – Synopsys IC Compiler II for Block-Level Auto Place & Route** *Sep. 17–19, 2025*
  Hands-on training in ASIC backend design flow including floorplanning, placement, clock tree synthesis (CTS), routing, timing closure, and post-layout verification using Synopsys IC Compiler II.

Awards and Scholarships

Senior Capstone Design Contest - Engineering Innovation Prize Jun. 2025

Extra-Curricular Challenge Scholarship (Top 100 Volunteers in SKKU) Mar. 2025

Academic Excellence Scholarship (Top 12% GPA recipients) Fall 2022

Extra-Curricular Experience

Teach for Korea Feb. 2023 – Dec. 2024

- Provided high-school mathematics instruction to underserved students, offering emotional support and career guidance; counseled students weekly on university applications and study habits.
- Supported organizational operations: scheduled sessions and coordinated volunteers

Google Developer Students' Club Sep. 2022 – Aug. 2023

- Built and maintained a React front end (using Hooks and Context API) for a sentiment-analysis diary app; integrated AI sentiment API and published the app on Google Play Store.

SungKyunKwan University Amateur Orchestra Mar. 2022 – Dec. 2023

- Performed as second violinist in concerts; coordinated rehearsal schedules and managed stage logistics.
- Designed fundraising materials—posters, merchandise, pamphlets—for club events.

Proficiency

- TOEFL (obtained Mar. 2025): 103 (Reading: 28, Listening: 28, Speaking: 23, Writing: 24)
- GRE (obtained Nov. 2025): Verbal (150, 39%), Quantitative (168, 81%), Writing (3.5, 43%)