Jiho Kim

jiho.kim@gatech.edu | ❂ jihoray.github.io | Atlanta, USA

EDUCATION

Georgia Institute of Technology   Atlanta, USA
Ph.D. in Electrical and Computer Engineering; GPA: 3.8/4.0   *Aug 2023 − Present*

Korea Aerospace University   Gyeonggi, Korea
M.Sc. in Electrical and Information Engineering; GPA: 3.90/4.00   *Mar 2021 − Feb 2023*
Thesis: Design and implementation of artificial neural network inference system for time series data

Korea Aerospace University   Gyeonggi, Korea
*B.Sc. in Electrical and Information Engineering; GPA: 3.59/4.00 (**Cum Laude**)*   *Mar 2016 − Feb 2021*
Thesis: Optimizing the resource usage of the recurrent neural network inference accelerator (in Korean)

PUBLICATIONS

[Journal]

[J3] **Jiho Kim** and Tae-Hwan Kim. "ROSETTA: A resource and energy-efficient inference processor for recurrent neural networks based on programmable data formats and fine activation pruning," ***IEEE Trans. on Emerging Topics in Computing*** (**TETC**), vol.11, no.3, pp.650-663, Sept. 2023

[J2] **Jiho Kim**, Jihoon Shin, and Taehwan Kim. "Low-latency bearing fault diagnosis based on convolutional LSTM model," in ***Jrnl. of the Korean Institute of Electronics and Information Engineers*** (**IEIE**), 59(1): 124-130, 2022 (Invited to write a paper after [C1], in Korean)

[J1] Jinwon Kim, **Jiho Kim**, and Tae-Hwan Kim. "AERO: A 1.28 MOP/s/LUT reconfigurable inference processor for recurrent neural networks in a resource-limited FPGA," ***Electronics***, 10(11), May 2021

[Conference]

[C5] **Jiho Kim**, Cong (Callie) Hao. "RealProbe: An Automated and Lightweight Performance Profiler for In-FPGA of High-Level Synthesis Designs," submitted in ***The 33rd IEEE International Symposium on Field-Programmable Custom Computing Machines*** (**FCCM**), 2025 ★ Best Paper Nomination ★

[C4] Yuchen Xia, **Jiho Kim**, Yuhan Chen, Haojie Ye, Souvik Kundu, Cong (Callie) Hao, and Nishil Talati. "Understanding the Performance and Estimating the Cost of LLM Fine-Tuning," in ***IEEE International Symposium on Workload Characterization (IISWC)***, **2024**

[C3] Seongjae Hong, **Jiho Kim**, and Taehwan Kim. "An integrated verification system for an RNN inference processor," in ***Summer Conf. of the Institute of Electronics and Information Engineers*** (**Conf. of IEIE**), pages 1274-1277, Aug 2022 (in Korean)

[C2] **Jiho Kim**, Kwoanyoung Park, and Tae-Hwan Kim. "A reconfigurable inference processor for recurrent neural networks based on programmable data format in a resource-limited FPGA," In ***Asia and South Pacific Design Automation Conf.*** (**ASP-DAC**), pages 94-95, January 2022 (Video)

[C1] **Jiho Kim**, Jihoon Shin, and Taehwan Kim. "Low-latency bearing fault diagnosis based on convolutional long short-term memory," in ***Summer Conf. of the Institute of Electronics and Information Engineers*** (**Conf. of IEIE**), pages 1274-1277. **IEIE**, Aug 2021 (in Korean)

PATENTS

[P3] **Jiho Kim** and Taehwan Kim. 2022. Dynamic Pruning Apparatus and Method for Instruction-set Based Neural Network Inference Processor. Korean Patent, filed Jun 2022, pending

[P2] **Jiho Kim** and Taehwan Kim. 2021. Apparatus and Method for Hardware Implementation of Activation Function in Recurrent Neural Network. Korean Patent, filed Apr 2021, pending

[P1] **Jiho Kim** and Taehwan Kim. 2020. Apparatus and Method for Improving Inference Speed of Neural Network Model. Korean Patent No. 10-2020-0052268, filed Apr 2020, and issued Nov 2021

RESEARCH EXPERIENCE

Software/Hardware Co-Design Laboratory (Sharc Lab)   Atlanta, USA
Graduate Researcher (Advised by Prof. Cong "Callie" Hao)   *Aug 2023 – Present*
- **Project: Agile Hardware Development [C5]**
  - Developed a fully-automated, lightweight FPGA profiling tool that operates end-to-end from HLS source code to in-FPGA execution
- **Project: Accurate Interconnect Modeling for Chiplet-based 3D Architecture**
  - Developing a heterogeneous network interconnect simulator for chiplet-based 3D architectures that accurately models package and circuit technology advances
- **Project: Efficient GPU Architecture [C4]**
  - Characterizing the performance and estimating the cost of LLM Fine-Tuning

IBM Thomas J. Watson Research Center   Yorktown Heights, USA
Research Internship in System Hardware Architecture   *May 2024 – Aug 2024*
- **Project: Efficient Fine-Tuning algorithms and architecture co-design**
  - Beyond Recovery: Fine-Tuning for Increased Compression Rate and Accuracy Gains

Circuits and Systems Laboratory   Gyeonggi, Korea
Undergraduate & Graduate Researcher   *Feb 2020 – Feb 2023*
- **Project: Intelligent Semiconductor Working for the Multi-Band Smart Radar [J1], [J3], [C2], [P3]**
  - Designed and implemented recurrent neural network inference system for time series radar data
  - Developed a method of dynamic pruning for instruction-set based neural network inference processor
- **Project: 3D spatial data processing and applied technology research [C3], [P2], [P1]**
  - Developed a method for efficient implementation of activation function in hardware
  - Devised a method for improving inference speed in neural networks
- **Project: Image and sound fusion technology [J2], [C1]**
  - Developed a low-latency Convolutional-LSTM based fault diagnosis system for time-series data

Electronics and Telecommunication Research Institute (ETRI)   Daejeon, Korea
Leader of Undergraduate Researcher of Autonomous Driving System Research Group   *Dec 2018 – Feb 2019*
- **Project: Development of Driving Computing System Supporting Real-time Sensor Fusion Processing for Self-Driving Car**
  - Involved in preprocessing data, feeding it to neural network, and running it through jetson board
  - Supervised and assisted the undergraduate researchers for the entirety of the project

AWARDS & SCHOLARSHIPS & GRANTS

Qualcomm Innovation Fellowship - Finalist (Winner TBD)   Feb, 2025
- Title: "Network-on-X: Hierarchical Interconnect Modeling and DSE for 3D Heterogeneous Chiplet Architectures"

IISWC Travel Grant   Sep, 2024
Gem5 Bootcamp Travel Grant   July, 2024
DAC Young Fellow & Travel Grant   June, 2024
FCCM Travel Grant   May, 2024
University Merit Scholarship for Academic Excellence of M.Sc.   2021-2022, 4 semesters
Research Assistant & Teaching Assistant Scholarship   2021-2022, 4 semesters
University Merit Scholarship for Academic Excellence of B.Sc.   2017-2020, 7 semesters
Grand Prize ($3000), Korean Government-Sponsored Smart City IT Make-a-thon   2018
Volunteer Scholarship in Overseas Technical Fields & Education Service   2018

SKILLS

Programming (Advance): C, C++, Verilog HDL, Python, MATLAB, LaTeX

Tools (Advance): Xilinx Vitis HLS, Xilinx Vivado Design Suite, Intel Quartus Prime, Modelsim, Spyglass, VCS

Hardware (Advance): Altera Cyclone V FPGAs, Xilinx ZYNQ 7000 SoCs, Xilinx PYNQ-Z2

Machine Learning Frameworks: Pytorch, Pytorch Lightning, Tensorflow, Keras

Major Coursework

Microelectronics Systems Packaging; **Advanced Glass-Core Packaging**; **Reliable and secure computer architecture**; **HW-SW co-design for ML systems**; Microprocessor; Advanced Computer Architecture, Digital System Design; Advanced Digital System Design; Digital Logic Circuit; VLSI System; Parallel Computing; Advanced Programming Technique, Deep Learning; Recommendation System; AI Convergence Capstone Design; Electronic Hardware Design; Embedded System Design; Operating Systems;