Jaehyeong Sim

+82-2-3277-6547 | jh.sim@ewha.ac.kr | https://acpl.ewha.ac.kr/professor/

LinkedIn | Google Scholar | Updated 2025-08-28

Seoul, 03760, Republic of Korea

Summary

I am an accomplished researcher and academic with extensive expertise in hardware-software co-design, neural processing units (NPUs), digital systems, and efficient deep learning. I currently serve as an Assistant Professor at the Department of Computer Science and Engineering, Ewha Womans University, where he leads the AI Computing Platform Laboratory, supervising 7 graduate students. My professional experience includes practical work at Samsung Advanced Institute of Technology, contributing to industrial NPU hardware architecture and performance analysis. I have spearheaded diverse projects, such as on-device learning for embedded systems and automated FPGA-based AI semiconductor design frameworks. I hold numerous patents, have published prolifically in premier journals and conferences, and received the Best Paper Award at IEEE ICCD. With my interdisciplinary expertise and a track record of bridging hardware and software for AI systems, I believe that I am well-positioned to excel in roles driving innovation at the intersection of deep learning and hardware design.

Experience

- **Ewha Womans University** *April 2021 - Present*
  Assistant Professor (Tenure Track) Seoul, Republic of Korea
  - Department of Computer Science and Engineering
  - Principal investigator at AI Computing Platform Laboratory
  - Department of Artificial Intelligence (Former Adjuct)
  - HOKMA College of General Education (Former Adjuct)
- **Samsung Advanced Institute of Technology (SAIT)** *March 2020 - March 2021*
  Staff Researcher Suwon, Republic of Korea
  - Developed datacenter NPU hardware architecture.
  - Developed variable-precision processing elements for NPU.
  - Conducted performance and power analysis for NPU and its PEs.
- **Korea Advanced Institute of Science and Technology** *March 2019 - February 2020*
  Postdoctoral Researcher Daejeon, Republic of Korea
  - Actively participated in a hardware-software co-design framework to accelerate CNN models for a super-resolution task.
  - Actively participated in developing customized 4T embedded DRAM cell array architecture for accelerating binary neural networks.
  - Actively participated in developing processing in-memory architecture for accelerating LSTM based on SOT-MRAM.

Education

- **Korea Advanced Institute of Science and Technology (KAIST)** *March 2014 - February 2019*
  PhD in Electrical Engineering Daejeon, Republic of Korea
  - Advisor: Prof. Lee-Sup Kim, *IEEE Fellow*
  - Thesis: Energy Efficient Processors and In-DRAM Processing Framework for Deep Convolutional Neural Network
  - Have experience of 6 silicon tapeouts.
- **Korea Advanced Institute of Science and Technology (KAIST)** *February 2012 - February 2014*
  Master in Electrical Engineering Daejeon, Republic of Korea
  - Advisor: Prof. Lee-Sup Kim, *IEEE Fellow*
  - Thesis: Timing Error Masking by Exploiting Operand Value Locality in SIMD Architecture
- **Korea Advanced Institute of Science and Technology (KAIST)** *February 2007 - February 2012*
  Bachelor in Electrical Engineering Daejeon, Republic of Korea
  - Cum Laude

Research Projects

- **Basic Research Laboratory, National Research Foundation of Korea (NRF)** *June 2025 - May 2028*
  Energy-Efficient, General-Purpose Multi-Modal AI with Heterogeneous Computing Accelerators Principle Investigator
  ◦ Developing an NPU architecture for multi-modal AI with multi-tenancy.
- **LG Electronics** *March 2025 - December 2025*
  Transformer-based On-Device Segmentation for Next Generation Robotic Vacuum Cleaner Principle Investigator
  ◦ Improving on-device efficiency of SAM2 model on NVIDIA Jetson Orin Nano platform.
- **LG Electronics** *March 2024 - December 2024*
  On-Device Learning of an Object Detection Model for Robotic Vacuum Cleaners Principle Investigator
  ◦ Improved detection accuracy of YOLO-X model in a personalized in-door environment using knowledge distillation, few-shot learning, and domain adaptation.
  ◦ Converted the proposed model into TFLite, and implemented C code to execute it on the commercial NPU developed by LG Electronics.
  ◦ Implemented a few-shot learning and domain adaptation methods in C to execute them on the SoC embedded in the commercial robotic vacuum cleaner.
- **First Research in Life Time, National Research Foundation of Korea (NRF)** *March 2023 - February 2026*
  End-to-End Design Automation Framework for FPGA-Based AI Semiconductor Principle Investigator
  ◦ Developed a hardware-aware neural architecture search (NAS) framework for deep learning acceleration in a FPGA.
  ◦ Developed an automatic architecture search framework for a FPGA-based deep learning accelerator.
- **Artificial Intelligence Innovation Hub, IITP** *July 2021 - December 2025*
  Self-evolving Multi-task Intelligence Co-Investigator
  ◦ Developed a hardware-aware automatic operation search framework for a squash function for capsule networks.
  ◦ Developed a novel token merging method to accelerate vision transformer models.
- **AI Semiconductor Processing SW Research Center, IITP** *July 2022 - December 2029*
  Hardware-Aware Neural Architecture Search and Accelerator Design Co-Investigator
  ◦ Developed a hardware-aware neural architecture search framework for an NPU.
  ◦ Developed an energy-efficient accelerator for YOLOX inference.
  ◦ Developed a bit-serial architecture exploiting weight bit sparsity for efficient deep learning acceleration.

Publications

International Conference

[C.23] Soeun Choi, **Jaehyeong Sim**. **LoRA-PIM: In-Memory Delta-Weight Injection for Multi-Adapter LLM Serving**. In *22st International SoC Design Conference (ISOCC)*, 2025.

[C.22] Eunjin Lee, Eunseo Kim, Eunjoung Yoo, **Jaehyeong Sim**. **GATHER: A Gated-Attention Accelerator for Efficient LLM Inference**. In *22st International SoC Design Conference (ISOCC)*, 2025.

[C.21] Eunjoung Yoo, **Jaehyeong Sim**. **ViT-Slim: Genetic Alorighm-based NAS Framework for Efficient Vision Transformer Design**. In *IEEE International Conference on Artificial Intelligence (CAI)*, 2025.

[C.20] Chaeyun Kim, Eunseo Kim, Yeonhee Kim, **Jaehyeong Sim**, Jonkil Kim. **Enhancing Gender Prediction Model Performance through Automatic Individual Entity Extraction and Class Balance**. In *IEEE International Conference on Big Data and Smart Computing (BigComp)*, 2025.

[C.19] Jieui Kang, Subean Lee, Eunseo Kim, Soeun Choi, **Jaehyeong Sim**. **AutoCaps-Zero: Searching for Hardware-Efficient Squash Function in Capsule Networks**. In *International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI)*, 2024.

[C.18] Jieui Kang, Subean Lee, Eunseo Kim, Soeun Choi, **Jaehyeong Sim**. **OCW: Enhancing Few-Shot Learning with Optimized Class-Weighting Methods**. In *International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI)*, 2024.

[C.17] Kyungmi Kim, Soeun Choi, Eunkyeol Hong, Yoonseo Jang, **Jaehyeong Sim**. **An Energy-Efficient Hardware Accelerator for On-Device Inference of YOLOX**. In *21st International SoC Design Conference (ISOCC)*, 2024.

[C.16] Eunseo Kim, Subean Lee, Chaeyun Kim, HaYoung Lim, Jimin Nam, **Jaehyeong Sim**. **BS2: Bit-Serial Architecture Exploiting Weight Bit Sparsity for Efficient Deep Learning Acceleration**. In *21st International SoC Design Conference (ISOCC)*, 2024.

[C.15] Jiho Lee, Jieui Kang, Eunjin Lee, Yejin Lee, **Jaehyeong Sim**. **AlphaAccelerator: An Automatic Neural FPGA Accelerator Design Framework Based on GNNs**. In *21st International SoC Design Conference (ISOCC)*, 2024.

[C.14] Jieui Kang, **Jaehyeong Sim**, Hyokyung Bahn. **Optimization of the Modified Gaussian Filter for Mobile GPU Usage in Game Workloads**. In *International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI)*, 2023.

[C.13] HaYoung Lim, Yeseo Jang, Juyeon Kim, **Jaehyeong Sim**. **TD-NAAS: Template-Based Differentiable Neural Architecture Accelerator Search**. In *20th International SoC Design Conference (ISOCC)*, 2023.

[C.12] Seungkyu Choi, **Jaehyeong Sim**, Myeonggu Kang, Yeongjae Choi, Hyeonuk Kim, Lee-Sup Kim. **A 47.4 uJ/epoch Trainable Deep Convolutional Neural Network Accelerator for In-Situ Personalization on Smart Devices**. In *IEEE Asian Solid-State Circuits Conference (A-SSCC)*, 2019.

[C.11] Youngbeom Jung, Yeongjae Choi, **Jaehyeong Sim**, Lee-Sup Kim. **eSRCNN: A Framework for Optimizing Super-Resolution Tasks on Diverse Embedded CNN Accelerators**. In *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 2019.

[C.10] Hyein Shin, **Jaehyeong Sim**, Daewoong Lee, Lee-Sup Kim. **A PVT-Robust Customized 4T Embedded DRAM Cell Array for Accelerating Binary Neural Networks**. In *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 2019.

[C.9] Kyeonghan Kim, Hyein Shin, **Jaehyeong Sim**, Myeonggu Kang, Lee-Sup Kim. **An Energy-Efficient Processing-in-Memory Architecture for Long Short Term Memory in Spin Orbit Torque MRAM**. In *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 2019.

[C.8] Hyeonuk Kim, **Jaehyeong Sim**, Yeongjae Choi, Lee-Sup Kim. **NAND-Net: Minimizing Computational Complexity of In-Memory Processing for Binary Neural Networks**. In *IEEE International Symposium on High Performance Computer Architecture (HPCA)*, 2019.

[C.7] **Jaehyeong Sim**, Hoseok Seol, Lee-Sup Kim. **NID: Processing Binary Convolutional Neural Network in Commodity DRAM**. In *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 2018.

[C.6] Seungkyu Choi, **Jaehyeong Sim**, Myeonggu Kang, Lee-Sup Kim. **TrainWare: A Memory Optimized Weight Update Architecture for On-Device Convolutional Neural Network Training**. In *ACM/IEEE International Symposium on Low Power Electronics and Design (ISLPED)*, 2018.

[C.5] Hyeonuk Kim, **Jaehyeong Sim**, Yeongjae Choi, Lee-Sup Kim. **A Kernel Decomposition Architecture for Binary-Weight Convolutional Neural Networks**. In *IEEE/ACM 54th Annual Design Automation Conference (DAC)*, 2017.

[C.4] Myung-Hoon Choi, Seungkyu Choi, **Jaehyeong Sim**, Lee-Sup Kim. **SENIN: An Energy-Efficient Sparse Neuromorphic System with On-Chip Learning**. In *ACM/IEEE International Symposium on Low Power Electronics and Design (ISLPED)*, 2017.

[C.3] **Jaehyeong Sim**, Jun-Seok Park, Minhye Kim, Dongmyung Bae, Yeongjae Choi, Lee-Sup Kim. **A 1.42 TOPS/W Deep Convolutional Neural Network Recognition Processor for Intelligent IoE Systems**. In *IEEE International Solid-State Circuits Conference (ISSCC)*, 2016.

[C.2] **Jaehyeong Sim**, Jun-Seok Park, Seungwook Paek, Lee-Sup Kim. **Timing Error Masking by Exploiting Operand Value Locality in SIMD Architecture**. In *IEEE 32nd International Conference on Computer Design (ICCD)*, 2014. **Best Paper Award recipient.**

[C.1] Seungwook Paek, Seok-Hwan Moon, Wongyu Shin, **Jaehyeong Sim**, Lee-Sup Kim. **PowerField: A Transient Temperature-to-Power Technique Based on Markov Random Field Theory**. In *IEEE/ACM 49th Annual Design Automation Conference (DAC)*, 2012.

International Journal

[J.10] Jieui Kang, Hyungon Ryu, **Jaehyeong Sim**. **PRISM-Med: Parameter-efficient Robust Interdomain Specialty Model for Medical Language Tasks**. *IEEE Access*, Vol. 13, pp. 4957-4965, 2025.

[J.9] Jieui Kang, Soeun Choi, Eunjin Lee, **Jaehyeong Sim**. **SpDRAM: Efficient In-DRAM Acceleration of Sparse Matrix-Vector Multiplication**. *IEEE Access*, Vol. 12, pp. 176009-176021, 2024.

[J.8] Jieui Kang, Jihye Park, Soeun Choi, **Jaehyeong Sim** **Q-LAtte: An Efficient and Versatile LSTM Model for Quantized Attention-Based Time Series Forecasting in Building Energy Applications**. *IEEE Access*, Vol. 12, pp. 69325-69341, 2024.

[J.7] Myeonggu Kang, Hyeonuk Kim, Hyein Shin, Jaehyeong Sim, Kyeonghan Kim, Lee-Sup Kim **S-FLASH: A NAND Flash-Based Deep Neural Network Accelerator Exploiting Bit-Level Sparsity**. *IEEE Transactions on Computers*, Vol. 71, Issue 6, pp. 1291-1304, 2021.

[J.6] Yeongjae Choi, **Jaehyeong Sim**, Lee-Sup Kim **CREMON: Cryptography Embedded on the Convolutional Neural Network Accelerator**. *IEEE Transactions on Circuits and Systems II: Express Briefs*, Vol. 67, Issue 12, pp. 3337-3341, 2020.

[J.5] Seungkyu Choi, **Jaehyeong Sim**, Myeonggu Kang, Yeongjae Choi, Hyeonuk Kim, Lee-Sup Kim. **An Energy-Efficient Deep Convolutional Neural Network Training Accelerator for In Situ Personalization on Smart Devices**. *IEEE Journal of Solid-State Circuits*, Vol. 55, Issue 10, pp. 2691-2702, 2020.

[J.4] **Jaehyeong Sim**, Somin Lee, Lee-Sup Kim. **An Energy-Efficient Deep Convolutional Neural Network Inference Processor with Enhanced Output Stationary Dataflow in 65-nm CMOS**. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 28, Issue 1, pp. 87-100, 2019.

[J.3] Yeongjae Choi, Dongmyung Bae, **Jaehyeong Sim**, Seungkyu Choi, Minhye Kim, Lee-Sup Kim **Energy-Efficient Design of Processing Element for Convolutional Neural Network**. *IEEE Transactions on Circuits and Systems II: Express Briefs*, Vol. 64, Issue 11, pp. 1332-1336, 2017.

[J.2] Taeho Lee, Yong-Hun Kim, **Jaehyeong Sim**, Jun-Seok Park, Lee-Sup Kim. **A 5-Gb/s 2.67-mW/Gb/s Digital Clock and Data Recovery with Hybrid Dithering Using a Time-Dithered DdeltaSigma Modulator**. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 24, Issue 4, pp. 1450-1459, 2015.

[J.1] Seungwook Paek, Wongyu Shin, **Jaehyeong Sim**, Lee-Sup Kim. **PowerField: A Probabilistic Approach for Temperature-to-Power Conversion Based on Markov Random Field Theory**. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 32, Issue 10, pp. 1509-1519, 2013.

PATENTS

Granted Patents

[P.5] **Apparatus for Processing Image Based on Gaussian Plus Filter and Method Thereof**. Korean Intellectual Property Office (KIPO), Patent No. KR10-2820700, 2025.

[P.4] **Computing device and method for allocating resources using cost matrix**. USPTO, Patent No. US12175299, 2024.

[P.3] **Accelerator, method of operating an accelerator, and electronic device including an accelerator**. USPTO, Patent No. US12130756, 2024.

[P.2] **Method and apparatus for performing convolution operation in neural network**. Korean Intellectual Property Office (KIPO), Patent No. KR10-2452951, 2022.

[P.1] **Neural network method and apparatus**. USPTO, Patent No. US10699160, 2020.

Patent Applications

[A.21] **Apparatus and Method of Processing Images Based on Deep Learning**. Korean Intellectual Property Office (KIPO), Application No. KR10-2025-0105299, 2025.

[A.20] **Sentence-Based Knowledge Distillation Apparatus and Operation Method Using Clustering-Based Sentence Pruning**. Korean Intellectual Property Office (KIPO), Application No. KR10-2025-0099179, 2025.

[A.19] **Task-Aware Knowledge Distillation Apparatus and Method for Performing Task-Aware Knowledge Distillation Based on Dynamic Token Selection and Dynamic Token Integration**. Korean Intellectual Property Office (KIPO), Application No. KR10-2025-0069129, 2025.

[A.18] **Apparatus and Method of Knowledge Distillation for Language Model Based on Structured Hierarchical Attention Rank Projection**. Korean Intellectual Property Office (KIPO), Application No. KR10-2025-0052980, 2025.

[A.17] **Vision Transformer Apparatus and Vision Transformer Method using Token Merging**. PCT, Application No. PCT/KR2024/018690, 2024.

[A.16] **Apparatus and Method of Processing Memory Computing by Using Weight Matrix**. Korean Intellectual Property Office (KIPO), Application No. KR10-2024-0114013, 2024.

[A.15] **Apparatus and Method of Processing Bit Serial Operations**. Korean Intellectual Property Office (KIPO), Application No. KR10-2024-0108146, 2024.

[A.14] **Hardware Architecture Design Apparatus and Method for Accelerating Hardware Architecture Using Graph Neural Network**. Korean Intellectual Property Office (KIPO), Application No. KR10-2024-0108145, 2024.

[A.13] **Domain-Adaptive Language Model Processing Apparatus and Method**. Korean Intellectual Property Office (KIPO), Application No. KR10-2024-0094108, 2024.

[A.12] **Vision Transformer Apparatus and Vision Transformer Method using Token Merging**. Korean Intellectual Property Office (KIPO), Application No. KR10-2024-0065166, 2024.

[A.11] **Apparatus and Method of Processing Quantization Artificial Intelligence Learning Using Loss Function and Distribution Information and a Method Thereof**. PCT, Application No. PCT/KR2024/006433, 2024.

[A.10] **Artificial intelligence-based smart window control system and control method**. Korean Intellectual Property Office (KIPO), Application No. KR10-2024-0057005, 2024.

[A.9] **Apparatus for Searching Neural Architecture Based on Template and Method Thereof**. PCT, Application No. PCT/KR2024/005651, 2024.

[A.8] **Apparatus and Method for Searching Squash Function in Capsule Network**. PCT, Application No. PCT/KR2024/003896, 2024.

[A.7] **Artificial intelligence-based smart window control system and control method**. Korean Intellectual Property Office (KIPO), Application No. KR10-2024-0057005, 2024.

[A.6] **Apparatus and Method of Processing Artificial Intelligence Learning by Using Optimized Class Weight**. Korean Intellectual Property Office (KIPO), Application No. KR10-2024-0031351, 2024.

[A.5] **Apparatus and Method of Processing Quantization Artificial Intelligence Learning Using Loss Function and Distribution Information and a Method Thereof**. Korean Intellectual Property Office (KIPO), Application No. KR10-2023-0194206, 2023.

[A.4] **Apparatus for Searching Neural Architecture Based on Template and Method Thereof**. Korean Intellectual Property Office (KIPO), Application No. KR10-2023-0178909, 2023.

[A.3] **Apparatus and Method for Searching Squash Function in Capsule Network**. Korean Intellectual Property Office (KIPO), Application No. KR10-2023-0121855, 2023.

[A.2] **Computing device and method**. USPTO, Application No. US20220083390, 2021.

[A.1] **Method and apparatus with deep learning operations**. USPTO, Application No. US20220164164, 2021.

Skills

- **Programming Languages:** C, C++, Python, Java, Lua, Node.js, CUDA, Verilog HDL, System Verilog, SystemC, MATLAB, MySQL
- **Machine Learning Framework:** Caffe2, TensorFlow, Keras, PyTorch
- **Hardware Design Tools:** NC-Verilog, Synopsys Design Compiler, Synopsys Fusion Compiler, Synopsys PrimeTime-PX, Synopsys IC Compiler, Synopsys Astro, Simvision, Verdi, Cadence Virtuoso, Cadence Allegro, OrCAD, Xilinx Vivado, Mentor Catapult HLS
- **Hardware Simulation Tools:** GPGPU-Sim, SCALE-Sim, MAESTRO, CACTI, DRAMSim2, HotSpot
- **Cloud Technologies:** Amazon Web Service (EC2, Lambda, S3, DynamoDB, RDS, SageMaker)
- **Specialized Area:** Digital System Design, NPU, Parallel Computing, Performance and Power Modeling, Deep Learning, Hardware-Software Co-Design, AutoML, Processing-in-Memory

Honors and Awards

- **Teaching Excellence Award** *2025*
  EWha Womans University
- **Rising Star** *2023*
  Electronic & Information Research Information Center (EIRIC) [🌐]
- **Encouragement Award at Undergraduate Student Research Paper Competition** *2023*
  Autumn Annual Conference of IEIE
- **Outstanding Paper Poster Award** *2019*
  Samsung-KAIST Cooperation Meeting
- **IDEC Congress Chip Design Contest 2019 Best Design Award** *2019*
  IC Design Education Center (IDEC)
- **IDEC Congress Chip Design Contest 2019 Best Poster Award** *2019*
  IC Design Education Center (IDEC)
- **Best Paper Award** *2014*
  IEEE 32nd International Conference on Computer Design (ICCD)

Academic Services

- **IEEE Transactions on Circuit and Systems I: Regular Papers** *2022, 2024, 2025*
  External Reviewer
- **IEEE Transactions on Circuit and Systems II: Express Briefs** *2019 - 2025*
  External Reviewer
- **IEEE Transactions on Very Large Scale Integration (VLSI) Systems** *2020, 2023, 2025*
  External Reviewer
- **IEEE Transactions on Computers** *2024 - 2025*
  External Reviewer

- **IEEE Access** *2023 - 2025*
  External Reviewer
- **IEEE Transactions on Knowledge and Data Engineering** *2025*
  External Reviewer
- **IEEE Transactions on Artificial Intelligence** *2025*
  External Reviewer
- **IEEE Transactions on Neural Networks and Learning Systems** *2025*
  External Reviewer
- **IEEE Transactions on Reliability** *2024*
  External Reviewer
- **IEEE Journal on Emerging and Selected Topics in Circuits and Systems** *2023*
  External Reviewer
- **IEEE Open Journal of the Solid-State Circuit Society** *2022*
  External Reviewer
- **IEEE Transactions on Signal Processing** *2021*
  External Reviewer
- **IEEE International Symposium on Circuits and Systems (ISCAS)** *2021*
  External Reviewer
- **Scientific Reports (Springer)** *2024 - 2025*
  External Reviewer
- **Journal of Real-Time Image Processing (Springer)** *2024 - 2025*
  External Reviewer
- **Signal, Image and Video Processing (Springer)** *2024 - 2025*
  External Reviewer
- **Analog Integrated Circuits and Signal Processing (Springer)** *2025*
  External Reviewer
- **Journal of Nondestructive Evaluation (Springer)** *2024*
  External Reviewer
- **The Journal of Supercomputing (Springer)** *2024*
  External Reviewer
- **Nature (Springer)** *2018*
  External Reviewer

Additional Information

Languages: Korean (Native), English (Intermediate High)