SUNGJUN JUNG

+82 10-8289-3501 ◇ Engineering Building 303, 1, Gwanak-ro, Gwanak-gu, Seoul, Republic of Korea
miguel92@snu.ac.kr ◇ linkedin.com/in/migueljung/ ◇ arc.snu.ac.kr/people/jsj/

RESEARCH INTEREST

- CPU architecture
- Data prefetching
- Similarity search
- Graph processing
- HW/SW codesign
- Domain-specific hardware design
- ASIC/FPGA prototyping

EDUCATION

Seoul National University, Seoul, Korea **Mar 2018 - Present**
M.S./Ph.D candidate in Computer Science and Engineering
- Advisor: Professor Jae W. Lee
- Research areas: computer architecture, data prefetch, VLSI, similarity search, graph processing

Yonsei University, Seoul, Korea **Mar 2012 - Feb 2018**
B.S. in Electrical and Electronic Engineering

EXPERIENCE

Architecture and Code Optimization Lab (ARC Lab), Seoul, Korea **Mar 2018 - Present**
Ph.D. Candidate

- **Vector Indirect Memory PrefetchER for SpMM and SDDMM (VIPER)**
  Investigate the limitation of existing indirect memory prefetchers for supporting SpMM and SDDMM
  Proposed the prefetcher for SpMM and SDDMM, highly scalable to the number and dimension of vectors
  Modeled VIPER using gem5 simulator and evaluated performance

- **Algorithmic Optimization for Graph-based Approximate Nearest Neighbor Search (ADA-NNS)**
  Analyzed the performance bottleneck of the greedy search method using Intel VTune and Advisor
  Devised the angular distance-guided search method for graph-based ANNS to reduce the query latency
  Implemented ADA-NNS on various state-of-the-art graph-based algorithms and evaluated performance

- **RISC-V CPU with Hardware-based Demand Paging (HWDP)**
  Lead the ASIC prototyping of HWDP in TSMC 40nm GP
  Implemented I/O interface and emulated on Xilinx FPGA
  Managed both front-end and back-end processes of tape-out

- **Self-Attention Mechanism Accelerator (ELSA)**
  Estimated the power/area cost and performed PnR of ELSA
  Lead the ASIC prototyping of ELSA in TSMC 40nm GP
  Managed back-end process of tape-out

- **Inverted Index Search Accelerator (BOSS)**
  Lead ASIC prototyping of Inverted Index Search Accelerator in Samsung 28nm
  Designed a multiple modules and managed back-end process of tape-out

- **Specialized Hardware for Object Serialization (Cereal)**
  Designed the specialized hardware accelerator for memory object (de)serialization

Implemented the cycle-accurate simulator for Cereal

- **Attention Mechanism Accelerator (A^3)**
  Designed RTL model of accelerator for base and efficient candidate selection method of attention mechanism
  Estimated the power/area cost of the accelerator
  Lead the ASIC prototyping of A^3 in TSMC 40nm LP (Demo)
    * **First** ASIC prototype of ARC Lab

PUBLICATIONS

[WWW'25] Angular Distance-Guided Neighbor Selection for Graph-Based Approximate Nearest Neighbor Search
Sungjun Jung, Yongsang Park, Haeun Lee, Young H. Oh, Jae W. Lee
The Web Conference 2025, 2025 (To appear)
Acceptance rate: 19.8% (409/2062)

[ESSCIRC'22] A 40nm 5.6TOPS/W 239GOPS/mm^2 Self-Attention Processor with Sign Random Projection-based Approximation
Seong Hoon Seo, Soosung Kim, **Sung Jun Jung**, Sangwoo Kwon, Hyunseung Lee, Jae W. Lee
The 48th IEEE European Solid State Circuits Conference, 2022

[ISCA'21] BOSS: Bandwidth-Optimized Search Accelerator for Storage-Class Memory
Jun Heo, Seung Yul Lee, Sunhong Min, Yeonhong Park, **Sung Jun Jung**, Tae Jun Ham, Jae W. Lee
The 48th ACM/IEEE International Symposium on Computer Architecture, 2021
Acceptance rate: 18.7% (76/406)

[ISCA'21] ELSA: Hardware-Softaware Co-design for Efficient Lightweight Self-Attention Mechanism in Neural Networks
Tae Jun Ham, Yejin Lee, Seong Hoon Seo, Soosung Kim, Hyunji Choi, **Sung Jun Jung**, Jae W. Lee
The 48th ACM/IEEE International Symposium on Computer Architecture, 2021
Acceptance rate: 18.7% (76/406)

[ISCA'20] A Specialized Architecture for Object Serialization with Applications to Big Data Analytics
Jaeyoung Jang, **Sung Jun Jung**, Sunmin Jeong, Jun Heo, Hoon Shin, Tae Jun Ham, Jae W. Lee
The 47th ACM/IEEE International Symposium on Computer Architecture, 2020
Acceptance rate: 18.0% (77/428)

[HPCA'20] A^3: Accelerating Neural Network Attention Mechanism with Approximation
Tae Jun Ham, **Sung Jun Jung**, Seonghak Kim, Young H. Oh, Yoon Ho Song, Junghoon Park, Sanghee Lee, Kyoung Park, Jae W. Lee, Deog-Kyoon Jeong
The 26th International Symposium on High Performance Computer Architecture, 2020
Acceptance rate: 20.4% (48/235)

[MICRO'19] Charon: Specialized Near-Memory Processing Architecture for Clearing Dead Objects in Memory
Jaeyoung Jang, Jun Heo, Yejin Lee, Jaeyeon Won, Seonghak Kim, **Sung Jun Jung**, Hakbeom Jang, Tae Jun Ham, Jae W. Lee
The 52nd International Symposium on Microarchitecture, 2019
Acceptance rate: 23.0% (79/344)

[PACT'18] A portable automatic data quantizer for deep neural networks
Young H. Oh, Quan Quan, Daeyeon Kim, Seonghak Kim, Jun Heo, Jaeyoung Jang, **Sung Jun Jung**, Jae W. Lee
The 27th IEEE International Conference on Parallel Architectures and Compilation Techniques, 2018

PATENTS

Device for accelerating self-attention operation in neural networks (US Pending: 17/864.235) *with Jae W. Lee, Yejin Lee, Tae Jun Ham, Seong Hoon Seo, Soosung Kim, Hyunji Choi*

Device for accelerating self-attention operation in neural networks (KR Pending: 10-2021-0190300) *with Jae W. Lee, Yejin Lee, Tae Jun Ham, Seong Hoon Seo, Soosung Kim, Hyunji Choi*
Method for accelerating candidate selection based on similarity and accelerator for performing candidate selection (US Registration: US-20200311182-A1) *with Deog-Kyoon Jeong, Jae W. Lee, Tae Jun Ham, Seonghak Kim, Minsoo Lim*
Method for accelerating candidate selection based on similarity and accelerator for performing candidate selection (CN Registration: 111753251A) *with Deog-Kyoon Jeong, Jae W. Lee, Tae Jun Ham, Seonghak Kim, Minsoo Lim*
Method for accelerating candidate selection based on similarity and accelerator for performing candidate selection (KR Application: 1020200115102) *with Deog-Kyoon Jeong, Jae W. Lee, Tae Jun Ham, Seonghak Kim, Minsoo Lim*

POSTERS

The 19th International SoC Design Conference Chip Design Contest, Gangeung, Korea **Oct 2022**
A 28nm Inverted Index Search Accelerator Chip with Applications to High-Performance Full-Text Search

SNU Artificial Intelligence Institute Workshop, Seoul, Korea **Feb 2020**
A^3: Accelerating Neural Network Attention Mechanism with Approximation

SKILLS

Languages: C/C++, Bash, Verilog, Chisel3, LaTeX
Performance Analysis Tools: Gem5, ZSim, Intel Perf, VTune, Advisor
Applications & Frameworks: Xilinx Vivado, Synopsys VCS, Design Compiler, Formality, Primetime, IC Compiler, Cadence Virtuoso, Siemens Calibre-DRC, LVS

SERVICES

Reviewer, The Web Conference 2025
Student Volunteer, International Symposium on Code Generation and Optimization (CGO 2022)
Member, FireSim/Chipyard Developer Meeting

TEACHING EXPERIENCE

4190.206A Electrical and Electronic Circuit, *Teaching Assistant*, Seoul National Unversity **Fall 2018**
This course provides the basic theory of electrical and electronic ciruits. Covers from circuit analysis methods with resistive circuits to principles of BJT and MOSFET

4190.309A Hardware System Design, *Teaching Assistant*, Seoul National University **Spring 2018**
The course provides experience of designing digital hardware design in Verilog. Implemented DNN accelerator on Xilinx FPGA

REFERENCES

Jae W. Lee
Professor, Seoul National University
Department of Computer Science & Engineering
1 Gwanak-ro, Gwanak-gu, Seoul, Korea
Contact: jaewlee@snu.ac.kr / +82 2-880-1836