

# Junsu Kim

145, Anam-ro, Seongbuk-gu, Seoul, Republic of Korea (Korea University)

☎ (+82) 10-8684-3631   ✉ j0807s@korea.ac.kr   🌐 github.com/junsu-kim0807   🏠 junsu-kim0807.github.io

## Research Interests

---

Computer Architecture, Systems for ML & ML for Systems

## Education

---

**Korea University, Seoul, Korea**

Sep. 2023 - Current

M.S. in Electrical Engineering (Advisor: Prof. Yunho Oh)

Cumulative GPA: 4.0/4.0

**Hanyang University, Seoul, Korea**

Mar. 2014 - Feb. 2021

B.S. in Electronic Engineering (Advisor: Prof. Ki-Seok Chung)

Cumulative GPA: 3.81/4.0 (Graduating with Honors - Summa Cum Laude)

## Publications

---

### Conference [C] and Journal [J] Papers

[J2] **Junsu Kim**, Jaebeom Jeon, Jaeyong Park, and Sangun Choi, Minseong Gil, Seokin Hong, Gunjae Koo, Myung Kuk Yoon, and Yunho Oh “Memory Oversubscription-Aware Tensor Migration Scheduling for GPU Unified Storage Architecture” The IEEE Computer Architecture Letters (CAL), 2025.

[C5] **Junsu Kim**, and Suhyun Kim, “Salient Frequency-aware Exemplar Compression for Resource-constrained Online Continual Learning”, The 39th Annual AAAI Conference on Artificial Intelligence (AAAI), 2025.

[C4] Seondeok Kim\*, Sangun Choi\*, Jaebeom Jeon, **Junsu Kim**, Minseong Gil, Jaehyeok Ryu, and Yunho Oh, “Kubism: Disassembling and Reassembling K-Means Clustering for Mobile Heterogeneous Platforms”, The 26th ACM SIGPLAN/SIGBED International Conference on Languages, Compilers, and Tools for Embedded Systems (LCTES), 2025.

[C3] Dongwon Yang, Jaebeom Jeon, Minseong Gil, **Junsu Kim**, Seondeok Kim, Gunjae Koo, Myung Kuk Yoon, and Yunho Oh, “SSFFT: Energy-Efficient Selective Scaling for Fast Fourier Transform in Embedded GPUs”, The 26th ACM SIGPLAN/SIGBED International Conference on Languages, Compilers, and Tools for Embedded Systems (LCTES), 2025.

[J1] Minseong Gil, Jaebeom Jeon, **Junsu Kim**, Sangun Choi, Gunjae Koo, Myung Kuk Yoon, and Yunho Oh, “TLP Balancer: Predictive Thread Allocation for Multi-Tenant Inference in Embedded GPUs”, The IEEE Embedded Systems Letters (ESL), 2024.

[C2] Jaebeom Jeon, Minsung Gil, **Junsu Kim**, Jaeyoung Park, Gunjae Koo, Myung-Kuk Yoon, and Yunho Oh. “VitBit: Enhancing Embedded GPU Performance for AI Workloads through Register Operand Packing”. The 53rd International Conference on Parallel Processing (ICPP), 2024

[C1] Kwangrae Kim, Jeonghyun Woo, **Junsu Kim**, and Ki-Seok Chung. “HammerFilter: Robust Protection and Low Hardware Overhead Method for RowHammer”. The 39th IEEE International Conference on Computer Design (ICCD), 2021

### Preprints (Project Names Only)

[P1] Co-author, “Mitigating FTL Overhead in Tiered Memory-based Accelerator for Training” *Under Review in MICRO*

[P2] Co-author, “Hardware Supports for Enabling Arbitrary Numeric Format on GPUs” *Under Review in MICRO*

[P3] Co-author, “On-Chip Memory Management System for Embedding Vector Operations” *Under Review in MICRO*

[P4] Co-author, “A Behavioral Analysis of CXL Memory Systems”

## Work Experience

---

**Korea University, Seoul, Korea**

Sep. 2023 - Current

Research Assistant at Computer Architecture and System Software Lab (ComSys)

Advisor: Prof. Yunho Oh

**Korea Institute of Science and Technology, Seoul, Korea**

May. 2022 - Aug. 2023

Research Assistant at Korea Data Science Team (KDST)

Supervisor: Dr. Suhyun Kim

**Hanyang University, Seoul, Korea**

Dec. 2019 - Mar. 2020, Aug. 2020 - Nov. 2020

Research Assistant at Embedded System on Chip Laboratory (ESOC Lab)

Advisor: Prof. Ki-Seok Chung

Research Assistant at Computer Architecture and System SW Lab (CASS Lab)

Advisor: Prof. Yongjun Park

## Teaching Experience

---

### Korea University, Seoul, Korea

Teaching Assistant for Computer Architecture

Spring 2024, Fall 2024

### School for the Blind, Chuncheon, Korea

Assistant Teacher (Alternative Military Service)

Mar. 2017 - Feb. 2019

## Honors and Awards

---

AAAI Student Travel Grant → ∼ \$1600 USD

Mar. 2025

Korea University Graduate School Scholarship → Half Tuition (∼ \$10000 CAD)

Spring 2024, Fall 2024

Korea University Graduate School Scholarship for Outstanding New Student → Half Tuition (∼ \$5000 CAD)

Fall 2023

Hanyang University Scholarship → Full Tuition (∼ \$20000 CAD)

2014 - 2019

## Research Projects

---

### Mitigating Software Overhead in Tiered Memory-based Accelerator for Training [Under Review]

Advisor: Prof. Yunho Oh, Korea University

Mar. 2024 - Nov. 2024

- ◇ Analyzed FTL overhead due to frequent promotion and demotion bottlenecked AI training as model sizes grow
- ◇ Proposed a unified address translation with dedicated IOMMU for each accelerator to reduce address translation overhead
- ◇ Proposed a migration scheduler that prefetches tensors at runtime, leveraging the predictability of AI workloads
- ◇ Contributions: co-author, motivation study, idea, implementation, paper write-up

### Hardware Supports for Enabling Arbitrary Numeric Format on GPUs [Under Review]

Advisor: Prof. Yunho Oh, Korea University

May. 2024 - Nov. 2024

- ◇ Observed GPU supports a limited set of numeric formats, wasting register files when processing arbitrary numeric formats
- ◇ Employed bitslice representation, which transposes the data elements, packing arbitrary numeric formats without register wastage.
- ◇ Proposed Bitslice Vector multiplier and adder, constructing a tree structure to replace a multiplication-adder tree in a Tensor core
- ◇ Contributions: co-author, motivation study, idea, implementation, paper write-up

### Memory Oversubscription-aware Tensor Migration Scheduling for GPU Unified Storage Architecture [CAL]

Advisor: Prof. Yunho Oh, Korea University

Feb. 2024 - Sep. 2024

- ◇ Analyzed page faults due to memory oversubscription stalled AI workloads when expanding GPU memory with SSD using UVM
- ◇ Proposed a tensor migration scheduling algorithm considering GPU memory oversubscription for GPU unified storage architecture
- ◇ Achieved the averaged speedup by 12.9% compared to G10, which was presented at MICRO 2023
- ◇ Contributions: 1st author, motivation study, idea, implementation, experiment, paper write-up

### A Behavioral Analysis of CXL Memory Systems [Under Review]

Advisor: Prof. Yunho Oh, Korea University

Sep. 2023 - Sep. 2024

Collaborator: SK hynix

- ◇ Observed the behavior of a real CXL-based system on datacenter and AI workloads in the CXL-based platform
- ◇ Analyzed how the different promotion and demotion methods for CXL devices affected the performance of the workloads
- ◇ Presented performance modeling for datacenter workloads using different system factors (e.g., memory bandwidth, memory latency)
- ◇ Contributions: co-author, experiment, analysis, paper write-up

### VitBit: Enhancing Embedded GPU Performance for AI Workloads through Register Operand Packing [ICPP'24]

Advisor: Prof. Yunho Oh, Korea University

Sep. 2023 - May. 2024

- ◇ Observed under-utilization of floating CUDA cores or Tensor cores when processing integer-quantized AI workloads
- ◇ Proposed a software technique for simultaneous computation on all heterogeneous cores on GPU to support arbitrary integer formats
- ◇ Proposed a software-based packing policy to support simultaneous processing of packed integers
- ◇ Contributions: co-author, motivation study, idea, implementation, paper write-up

### Salient Frequency-aware Exemplar Compression for Online Continual Learning [AAAI'25]

Supervisor: Dr. Suhyun Kim, Korea Institute of Science and Technology

Jan. 2023 - Nov. 2023

- ◇ Observed exemplar compression methods occupied limited GPU resources during online continual learning
- ◇ Proposed a computationally efficient compression algorithm using salient frequency
- ◇ Proposed a buffer management scheme to alleviate harmful effects from the compression artifacts remaining in the buffer
- ◇ Contributions: 1st author, motivation study, idea, implementation, paper write-up

## Skills

---

C/C++, Python, Tensorflow, Pytorch, Git, Verilog, Shell script