

## RESEARCH INTERESTS

---

- **Optimized Resource Management Per Application Characterizations, Systems for ML, Distributed Systems, AI Workloads Serving Platforms, Cloud Computing, Operating Systems**

My research is centered on enhancing computational resource efficiency throughout the software stack, with a specific focus on AI workloads. This entails developing strategies to mitigate memory pressure, optimize memory bandwidth, minimize data redundancy, and reduce communication overhead. These challenges demand tailored solutions depending on the application's nature, particularly in AI workloads, where the extensive use of accelerators introduces complex programming models and distinctive resource usage patterns. A comprehensive understanding of the interaction between host systems and accelerators, alongside synchronization and pipelining techniques, is vital for effective resource management. I am particularly interested in exploring the parallels and distinctions between optimizing traditional HPC and AI workloads, with a strong emphasis on refining the software stack. This includes focusing on optimizing memory bandwidth, computational efficiency, and communication layers.

## EDUCATION

---

- **Georgia Institute of Technology** Atlanta, GA  
*Ph.D. Student in Computer Science* Aug. 2019 – Present
  - **Advisor:** Ada Gavrilovska
- **Ewha Womans University** Seoul, South Korea  
*Bachelor of Economics; Magna Cum Laude* Mar. 2008 – Feb. 2014

## EXPERIENCE

---

- **Microsoft Research** Redmond, WA  
*Research Intern* May. 2024 – Aug. 2024
  - **AI Compiler Optimization:** Analyzed and addressed the limitations of current AI compilers, focusing on balancing the constraints of dynamic optimization with the rigidity of static approaches. Developed methodologies to harness workload information typically underutilized by state-of-the-art AI compilers and implemented a runtime compiler stack to dynamically combine compilation primitives, thereby enhancing versatility and performance in AI compilation.
- **Intel Corporation** Hillsboro, OR  
*Graduate Research Intern* May. 2022 – Aug. 2022
  - **HPC Operating Systems and AI Analytics:** Research on low-level characteristics of AI/DL applications running on mOS, concentrating on memory access patterns and effects of accelerators.
- **Intel Corporation** Hillsboro, OR  
*Graduate Research Intern* May. 2021 – Aug. 2021
  - **HPC Operating Systems and AI Analytics:** Research on improving the performance and functionality of AI/DL applications running on mOS, to match the requirements and characteristics of emerging AI/DL applications while scaling it across large supercomputers.
- **Georgia Institute of Technology** Atlanta, GA  
*Graduate Research Assistant* Aug. 2019 – Present
  - **Resource Management in Fixed-resource Computing Environment:** Research on edge computing infrastructure from aspects including how the software stack of edge infrastructure should look like, with regard to resource efficiency, scalability, performance while serving AI workloads.
- **Samsung Electronics—Semiconductor Division** Hwaseong, South Korea  
*System Software Engineer* Aug. 2015 – Jul. 2019
  - **Mobile GPU:** GPUs are becoming more and more important in recent mobile devices; They are required to meet not only conventional graphics functionality requirements but also new requirements such as neural computations while retaining the temperature and power consumption low. Worked on developing debuggers for the mobile GPUs.

- **USB Communication IC:** USB implementations have to comply with the low-level specification and high level protocols in developing software solutions relevant to universal interface. Worked on the development of USB communication chip firmware.
- **NAVER** Seongnam, South Korea  
*UX Designer/Content Editor* *Jan. 2014 – Dec. 2014*
  - **Hot Topics for 20s:** Worked on designing the composition of webpage and editing contents to deliver trendy content of interest to those in their 20s..
  - **Contents Search Design:** Contents search is one of the signatrues of NAVER search service; it delivers the search results accurately desired by users in a more organized manner. Worked on organizing and designing the information.
- **Korea Trade Insurance Corporation** Seoul, South Korea  
*Intern* *Jan. 2013 – May. 2013*
  - **Domestic Claims & Recoveries Department:** Worked on examining and compensating unpaid receivables from international trade deals.

---

## PUBLICATION

- **Pocket: ML Serving from the Edge.** Misun Park, Ketan Bhardwaj, Ada Gavrilovska, European Conference on Computer Systems (EuroSys'23), May. 2023.
- **Toward Lighter Containers for the Edge.** Misun Park, Ketan Bhardwaj, Ada Gavrilovska, USENIX Workshop on Hot Topics in Edge Computing (HotEdge'20), Jun. 2020.

---

## PRESENTATION

- **Fine-grained Memory Management For Data-intensive Workloads at the Edge.** SRC TechCon 2022, Misun Park, Ketan Bhardwaj, Ada Gavrilovska, 2022.
- **Lightweight Containers for the Edge.** SRC TechCon 2020, Misun Park, Ketan Bhardwaj, Ada Gavrilovska, 2020.
- **Toward Lighter Containers for the Edge.** 3rd USENIX Workshop on Hot Topics in Edge Computing (HotEdge 20), Misun Park, Ketan Bhardwaj, Ada Gavrilovska, 2020.
- **Lightweight Containers for the Edge.** ADA Liaison Meeting, Misun Park, 2020.

---

## HONORS AND AWARDS

- **1st Prize** from Samsung Convergence Software Academy (First Half of the Program), 2014
- **1st Prize** from Samsung Convergence Software Academy (Second Half of the Program), 2015

---

## TECHNICAL SKILLS

- **Languages:** C/C++ (C++11-19), Python, ARM Assembly, Java, Bash Scripting
- **Technologies/Tools/Knowledge:** Strong understanding in MPI, PyTorch, CUDA, vLLM, Linux Kernel, Docker, ARM Architecture, vTune, PyTorch, OpenGL ES, Performance Profiling Tools(BCC, Perf, Strace, ...), JTAG debugging