Kai Wu

+1 517-763-1599 | kaiwu0126@gmail.com | kaikylewu.com

Summary of Qualifications

- 5+ years of research and work experience in designing and developing high-performance and scalable systems
- Research interests: Memory/Storage System, High Performance Computing (HPC), Heterogeneous Computing and Persistent Memory (Non-Volatile Memory)
- Programming skills: C/C++, Java, Python, Fortran, SQL, Shell scripting and JavaScript

Education

University of California, Merced — Merced, CA
Ph.D., in Electrical Engineering and Computer Science — *June 2016 – Dec 2020*

Michigan State University — East Lansing, MI
M.S., in Computer Science and Engineering — *August 2014 – May 2016*

Harbin Normal University — Harbin, China
B.S., in Digital Media Technology — *September 2010 – June 2014*

Experience

Researcher/Engineer — Feb 2021 – Present
Bytedance/Tiktok (System Infrastructure Lab), Mountain View, CA

- Collaboratively designed and developed a new multi-tier cache system (based on DRAM, persistent memory and SSD) to support multiple data-intensive services at ByteDance/Tiktok
- Led the design and implementation of a persistent memory-based object store storage system for a new distributed graph analytic platform to support trillion edge-level graph processing and graph mining
- Responsible for exploring and prototyping to apply new memory technologies (e.g., persistent memory, CPU/GPU unified memory, disaggregated memory and software-defined memory) to existing infrastructure systems

Graduate Student Researcher — June 2016 – Dec 2020
University of California Merced, Merced, CA — *Advisor: Prof.Dong Li*

- Ph.D. Dissertation: Runtime Data Management on Non-Volatile Memory-Based High Performance Systems, invited talk in SC'20 Doctoral Showcase
- Designed and implemented data/page placement and migration mechanisms on non-volatile memory-based heterogeneous memory systems for HPC and deep learning applications [SC'17, SC'18, HPCA'21]
- Designed and implemented high-performance system consistency and fault tolerance mechanisms for non-volatile memory-based systems [FAST'21, PACT'20, CLUSTER'20, MCHPC'18]
- Characterized and optimized HPC applications on non-volatile memory devices (e.g, Intel Optane DC Persistent Memory and Optane SSD) [IPDPS'20, NAS'17]

Research Intern — May 2020 – November 2020
Bytedance/Tiktok (System Infrastructure Lab), Mountain View, CA

- Worked on hybrid transactional/analytical processing (HTAP) system; designed an adaptive data compaction/compression algorithm; lead the performance benchmarking and optimization of HTAP store; coordinated with other team members to complete prototype development of a new large-scale HTAP system

Research Intern — May 2018 – August 2018
Lawrence Livermore National Laboratory, Livermore, CA — *Mentor: Dr.Maya Gokhale*

- Worked on memory caching mechanism; explored pre-fetch and eviction optimizations using the user faulted approach for efficient access to persistent memory for data-intensive applications such as out-of-core sorting, graph application (i.e., process Graph500) and asteroid detection application (i.e., process imaging data sets of the sky generated by the optical cameras like the Dark Energy camera in Chile); achieved up to 2.5x speedups [MCHPC'19]

Research Intern — May 2017 – August 2017
Los Alamos National Laboratory, Los Alamos, NM — *Mentor: Dr.Nathan DeBardeleben, Dr.Qiang Guan*

- Worked on system resilience and reliability for large-scale parallel HPC applications; built an analytical model to predict the fault injection result of the application running in large-scale based on fault injection results of the application running in small-scale and serial [ICPP'18, SC'17 poster]

Refereed Conference Publications

[**FAST'21**] **Kai Wu**, Jie Ren, Ivy Peng and Dong Li. "ArchTM: Architecture-Aware, High Performance Transaction for Persistent Memory". In 19th USENIX Conference on File and Storage Technologies, 2021.
[**HPCA'21**] Jie Ren, Jiaolin Luo, **Kai Wu**, Minjia Zhang, Hyeran Jeon and Dong Li. "Efficient Tensor Migration and Allocation on Heterogeneous Memory Systems for Deep Learning". In The 27th IEEE International Symposium on High-Performance Computer Architecture, 2021.
[**ICS'21**] Jie Ren, Jiaolin Luo, Ivy Peng, **Kai Wu** and Dong Li. "Performance Analysis and Optimization of Electromagnetic Particle-In-Cell Method on Emerging Persistent Memory-based Platform". In 35th International Conference on Supercomputing, 2021.
[**PACT'20**] **Kai Wu**, Ivy B. Peng, Jie Ren and Dong Li. "Ribbon: High Performance Cache Line Flushing for Persistent Memory". In 29th International Conference on Parallel Architectures and Compilation Techniques, 2020.
[**IPDPS'20**] Ivy B. Peng, **Kai Wu**, Jie Ren, Dong Li and Maya Gokhale. "Demystifying the Performance of HPC Scientific Applications on NVM-based Memory Systems". In 34rd IEEE International Parallel and Distributed Processing Symposium, 2020.
[**CLUSTER'20**] Jie Ren, **Kai Wu** and Dong Li. "Exploring Non-Volatility of Non-Volatile Memory for High Performance Computing Under Failures". In 22th IEEE Cluster Conference, 2020.
[**SC'18**] **Kai Wu**, Jie Ren and Dong Li. "Runtime Data Management on Non-Volatile Memory-based Heterogeneous Memory for Task-Parallel Programs". In 30th ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis, 2018.
[**ICPP'18**] **Kai Wu**, Wenqian Dong, Qiang Guan, Nathan Debardeleben and Dong Li. "Modeling Application Resilience in Large Scale Parallel Execution". In 47th International Conference on Parallel Processing, 2018.
[**SC'17**] **Kai Wu**, Yingchao Huang and Dong Li. "Unimem: Runtime Data Management in Non-Volatile Memory-based Heterogeneous Main Memory". In 29th ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis, 2017.
[**CLUSTER'17**] Shuo Yang, **Kai Wu**, Yifan Qiao, Dong Li and Jidong Zhai. "Algorithm-Directed Crash Consistence in Non-Volatile Memory for HPC". In 19th IEEE Cluster Conference, 2017.
[**NAS'17**] Wei Liu, **Kai Wu**, Jialin Liu, Feng Chen and Dong Li. "Performance Evaluation and Modeling of HPC I/O on Non-Volatile Memory". In 12th International Conference on Networking, Architecture, and Storage, 2017.

Patents

[**In application**] Jianjun Chen, Yonghua Ding, Ye Liu, Fangshi Li, Li Zhang, Mingyi Zhang, Kui Wei, Wei Ding, **Kai Wu** and Jason Yang Sun. "Storage Engine for Hybrid Data Processing".

SERVICES

Technical Program Committee: Eurosys'21 (shaow PC), HPCC'20 & 2021
Reviewer: DOE SBIR/STTR, IEEE Transactions on Parallel and Distributed Systems, Future Generation Computer Systems, IEEE Access, IEEE IPDPS'21, NPC'20, IEEE CLUSTER'20, NPC'19, ICPP'19, ACM/IEEE SC'18, IEEE IPDPS'17, IEEE CLUSTER'17, IEEE HPCC'17, IEEE NAS'17
Student Volunteer: SC'20, SC'19, SC'18, SC'16
Graduate Student Representative, UC Merced EECS, January 2020 - December 2020

AWARDS

Conference Student Grant: OSDI'20, NVMW'20, SC'19, NVMW'19, SC'18, OSDI'18, ASPLOS'18, NVMW'18, CLUSTER'17, NVMW'17, SC'16
UC Merced Graduate Travel Fellowship: 2018 & 2020
UC Merced Bobcat Graduate Research Fellowship: 2017
China National Scholarship: 2013
First Class Scholarship of Harbin Normal University: 2011 & 2012 & 2013 & 2014