Xiaoyang Lu

917-755-1369 | xlu40@hawk.iit.edu | 819 Pomeroon Street, Naperville, IL

Education

Illinois Institute of Technology — Chicago, IL
Ph.D. in Computer Science, Department of Computer Science — Aug 2017 – Present

New York University — New York, NY
M.S. in Computer Engineering, Department of Electrical and Computer Engineering — Aug 2015 – May 2017

Zhejiang University — Hangzhou, China
B.E. in Electronic Science and Technology — Aug 2011 – July 2015

Research Experience

Research Assistant — Jan 2020 – Present
Illinois Institute of Technology — Chicago, IL

- Conduct comprehensive research in memory-centric computer architectures, scalable memory systems, and high-performance computing domains
- Specialize in memory performance optimization, developing advanced memory performance models, and innovating ML-assisted memory architectures
- Developed and critically analyzed a cutting-edge performance model for modern hierarchical memory systems, enhancing system efficiency
- Developed intelligent, streamlined frameworks to significantly enhance cache performance, prioritizing efficiency and lightweight design
- Developed advanced machine learning models to drive optimizations in computer architecture and system performance
- Developed specialized accelerators for machine learning workloads, optimizing computational speed and efficiency
- Provided mentorship and training to graduate students on research projects, fostering academic growth and practical skills

Research Aide — May 2020 – Aug 2020
Argonne National Laboratory — Lemont, IL

- Conducted comprehensive performance testing on disaggregated memory systems, identifying key areas for improvement
- Developed and refined performance models for disaggregated memory systems, enhancing predictive accuracy and system efficiency
- Quantified and mitigated interference in disaggregated memory systems, ensuring optimal operation and reliability

Publications

- [**ASPLOS 2024**] ACES: Accelerating Sparse Matrix Multiplication with Adaptive Execution Flow and Concurrency-Aware Cache Optimizations

  Xiaoyang Lu[*], Boyu Long[*], Xiaoming Chen, Yinhe Han, Xian-He Sun

  In the Proceedings of the International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS), 2024

- [**HPCA 2024**] CHROME: Concurrency-Aware Holistic Cache Management Framework with Online Reinforcement Learning

  Xiaoyang Lu, Hamed Najafi, Jason Liu, Xian-He Sun

  In the Proceedings of the International Symposium on High-Performance Computer Architecture (HPCA), 2024

- [**HPCA 2023**] CARE: A Concurrency-Aware Enhanced Lightweight Cache Management Framework

  Xiaoyang Lu, Rujia Wang, Xian-He Sun

  In the Proceedings of the International Symposium on High-Performance Computer Architecture (HPCA), 2023

- [**JCST 2023**] The Memory-Bounded Speedup Model and its Impacts in Computing
  Xian-He Sun, **Xiaoyang Lu**
  Journal of Computer Science and Technology, 2023, 38(1): 64-79
- [**WSC 2022**] A Generalized Model For Modern Hierarchical Memory System
  Hamed Najafi, **Xiaoyang Lu**, Jason Liu, Xian-He Sun
  In the Proceedings of the Winter Simulation Conference (WSC), 2022
- [**ICCD 2021**] Premier: A Concurrency-Aware Pseudo-Partitioning Framework for Shared Last-Level Cache
  Xiaoyang Lu, Rujia Wang, Xian-He Sun
  In the Proceedings of the 39th International Conference on Computer Design (ICCD), 2021
- [**ISLPED 2021**] CoPIM: A Concurrency-Aware PIM Workload Offloading Architecture for Graph Applications
  Liang Yan, Mingzhe Zhang, Rujia Wang, Xiaoming Chen, Xingqi Zou, **Xiaoyang Lu**, Yinhe Han, Xian-He Sun
  In the Proceedings of the International Symposium on Low Power Electronics and Design (ISLPED), 2021
- [**ICCD 2020**] APAC: An Accurate and Adaptive Prefetch Framework with Concurrent Memory Access Analysis
  Xiaoyang Lu, Rujia Wang, Xian-He Sun
  In the Proceedings of the 38th International Conference on Computer Design (ICCD), 2020

Leadership Experience

Teaching Assistant Aug 2017 – May 2022
Illinois Institute of Technology Chicago, IL

- Assisted in teaching five graduate courses, each with 9-60 students, covering topics such as Java Programming (CS 401), Software Engineering (CS 487), Parallel and Distributed Processing (CS 546), Advanced Operating Systems (CS 550), and Advanced Computer Architecture (CS 570)
- Developed and prepared comprehensive course materials, including laboratory experiments, lectures, exams, homework, and practice problems
- Led weekly lab sessions and problem-solving discussions for groups of up to 30 students, enhancing their understanding and application of course materials
- Supervised and guided students in final projects, provided detailed feedback, and graded exams and weekly homework assignments

Academic Honors and Awards

- 2023 HPCA Student Travel Award
- 2015 New York University Scholarship
- 2015 Zhejiang University Excellent Bachelor Thesis Award

Services

Invited Reviewer for Journals & Transactions:

- IEEE Transactions on Parallel and Distributed Systems
- Future Generation Computer Systems
- Simulation: Transactions of the Society for Modeling and Simulation International