YUQI XUE

Email: yuqixue2@illinois.edu ◇ Phone: +1-217-721-4958
xzman.github.io/yuqixue

RESEARCH INTERESTS

Computer Architecture and Systems. AI Infrastructure. Machine Learning Accelerators.

EDUCATION

University of Illinois at Urbana-Champaign *Aug 2021 – Present*
Ph.D. in Electrical and Computer Engineering
Advised by Prof. Jian Huang

University of Illinois at Urbana-Champaign *Aug 2017 – May 2021*
Bachelor of Science in Computer Engineering with High Honors
Bachelor of Science in Mathematics with Highest Distinction

PUBLICATION

G10: Enabling An Efficient Unified GPU Memory and Storage Architecture with Smart Tensor Migrations. Haoyang Zhang*, Yirui Eric Zhou*, Yuqi Xue, Yiqi Liu, and Jian Huang. In *Proceedings of the 56th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO'23)*. 2023.
*Co-primary authors.

RackBlox: A Software-Defined Rack-Scale Storage System with Network-Storage Co-Design. Benjamin Reidys, Yuqi Xue, Daixuan Li, Bharat Sukhwanim, Wen-mei Hwu, Deming Chen, Sameh Asaad, and Jian Huang. In *29th ACM Symposium on Operating Systems Principles (SOSP'23)*. 2023.

System Virtualization for Neural Processing Units. Yuqi Xue, Yiqi Liu, and Jian Huang. In *Proceedings of the 19th Workshop on Hot Topics in Operating Systems (HotOS'23)*. 2023.

V10: Hardware-Assisted NPU Multi-tenancy for Improved Resource Utilization and Fairness. Yuqi Xue, Yiqi Liu, Lifeng Nai, and Jian Huang. In *Proceedings of the 50th International Symposium on Computer Architecture (ISCA'23)*. 2023.

IceClave: A Trusted Execution Environment for In-Storage Computing. Luyi Kang*, Yuqi Xue*, Weiwei Jia*, Xiaohao Wang, Jongryool Kim, Changhwan Youn, Myeong Joon Kang, Hyung Jin Lim, Bruce Jacob, and Jian Huang. In *Proceedings of the 54th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO'21)*. 2021.
*Co-primary authors.

RESEARCH PROJECTS

Hardware-Assisted Multi-tenant Neural Processing Unit (NPU) *2022*
Paper accepted to ISCA'23 *UIUC*

- Conducts a thorough study of the resource utilization of modern NPUs with real hardware devices.
- Proposes a multi-tenant NPU design for improving resource utilization and fairness.
- Designs a tensor operator scheduler that improves the fairness and throughput for collocated ML workloads.
- Develops a clustering scheme to identify pairs of collocatable DNN workloads.

Software-Defined Rack-Scale Storage System with Network-Storage Co-Design *2022*
Paper accepted to SOSP'23 *UIUC*

· Proposes a rack-scale storage system that co-designs software-defined network and software-defined storage.
· Develops a new I/O scheduling mechanism across network and storage for predictable performance.
· Develops a coordinated GC mechanism for a rack of SSDs to minimize the impact on the storage performance.
· Develops a rack-scale wear leveling mechanism for ensuring the uniform lifetime of a rack of SSDs.

Trusted Execution Environment for In-Storage Computing *2021*
Paper accepted to MICRO'21 *UIUC*

· Develops *IceClave*, a trusted execution environment for in-storage computing in modern solid-state drives.
· Enables security isolation between in-storage apps and the flash translation layer with ARM TrustZone.
· Proposes efficient memory encryption and verification method to protect user data against physical attacks.

WORK EXPERIENCE

Systems Platform Research Group, UIUC *May 2021 - Present*
Graduate Research Assistant *Urbana, IL*

Google LLC *Fall 2023*
Student Researcher *Remote*

Google LLC *Summer 2023*
Research Intern *Sunnyvale, CA*

TEACHING

ECE 522: Emerging Memory and Storage Systems, UIUC, *Graduate TA* *Spring 2023*

ECE 511: Computer Architecture, UIUC, *Graduate TA* *Fall 2022*

ECE 411: Computer Organization and Design, UIUC, *Volunteer TA* *Spring 2022, Fall 2023*

ECE 310: Digital Signal Processing, UIUC, *Undergraduate Grader* *Jan 2020 - Jul 2020*

SERVICES

Artifact Evaluation Committee

· IEEE/ACM International Symposium on Microarchitecture (MICRO), 2022
· USENIX Symposium on Operating Systems Design and Implementation (OSDI), 2022, 2023
· USENIX Annual Technical Conference (ATC), 2022, 2023
· IEEE International Symposium on High-Performance Computer Architecture (HPCA), 2023

Reviewer

· IEEE Computer Architecture Letters, 2022, 2023

Organizing Committee

· The 1st Workshop on Hot Topics in System Infrastructure (HotInfra'23), Web Chair, 2023