Zhuangzhuang Zhou

471B Rhodes Hall – Ithaca, NY, 14850
+1 (607) 379 8536 • zz586@cornell.edu • zzhou612.com

Research Interests

I am broadly interested in improving the *performance* and *efficiency* of computer systems. My recent research focuses on improving the performance of cloud computing systems with QoS-and-uncertainty-aware resource management, efficient serverless computing architecture, and characterization and optimization of datacenter tax. **I am actively looking for full-time job opportunities in the industry**.

Education

Cornell University
Ph.D., Computer Engineering *Aug 2019 – Aug 2024 (Expected)*
Computer Systems Laboratory, Advisor: Prof. Christina Delimitrou.

Shanghai Jiao Tong University
B.S., Electrical and Computer Engineering *Aug 2014 – Aug 2019*
Graduation deferred due to a car accident in 2017. Advisor: Prof. Weikang Qian.

Professional Experience

Industry

Google **Sunnyvale, CA**
Research Intern *May 2023 – Nov 2023*
Characterized a warehouse-scale memory allocator, TCMalloc. Based on the performance insights gained, redesigned each tier in the TCMalloc cache hierarchy for warehouse-scale environments, resulting in 1.4% and 3.4% improvement in fleet-wide throughput and memory usage respectively.

Intel **Hillsboro, OR**
Research Intern *May 2021 – Aug 2021*
Extended DeathStarBench, a multi-tier microservice benchmark, with machine learning and image compression components. Added supports for avx2, avx512 accelerations for ML components and evaluated their performance impacts.

Cisco **Shanghai, China**
Software Engineer Intern *Jan 2018 – May 2018*
Developed the Event-driven Test Automation Dashboard for IoT Systems and Software Group as a full-stack developer.

Academia

Computer Systems Laboratory, Cornell University **Ithaca, NY**
Research assistant *Aug 2019 – Now*
Focused on improving the performance and efficiency of cloud computing systems.

- **Resource management for serverless workflows**: Proposed Aquatope, a QoS-and-uncertainty-aware resource scheduler for serverless workflows, which uses scalable Bayesian models to pre-warm function containers ahead of invocations, and allocates appropriate resources at function granularity to meet the QoS target while minimizing required resources.
- **Efficient serverless workflow engine**: Designed Meteion, a fast and efficient serverless workflow engine that decouples the control plane from the workflow execution, and leverages lightweight per-function engines to enable decentralized workflow orchestration and direct inter-function communication.
- **Resource management for microservices**: Worked on ML-driven and analytical-model-based resource management frameworks for multi-tier microservices.

Emerging Computing Technology Laboratory, Shanghai Jiao Tong University **Shanghai, China**
Research assistant *Aug 2017 – May 2019*

- Developed a delay-driven approximate synthesis framework to synthesize approximate circuits with optimal delays.

Publications

[1] **Zhuangzhuang Zhou**, Vaibhav Gogte, Nilay Vaish, Chris Kennelly, Patrick Xia, Svilen Kanev, Tipp Moseley, Christina Delimitrou, and Parthasarathy Ranganathan. Characterizing a warehouse-scale memory allocator. In *Submission.*, 2024.

[2] **Zhuangzhuang Zhou**, Yanqi Zhang, and Christina Delimitrou. Meteion: Fast and efficient serverless workflows for latency-critical interactive applications. In *Submission.*, 2024.

[3] Yanqi Zhang, **Zhuangzhuang Zhou**, Sameh Elnikety, and Christina Delimitrou. Ursa: Lightweight resource management for cloud-native microservices. In *The 30th IEEE HPCA International Conference on High Performance Computer Architecture (HPCA)*, 2024.

[4] **Zhuangzhuang Zhou**, Yanqi Zhang, and Christina Delimitrou. Aquatope: Qos-and-uncertainty-aware resource management for multi-stage serverless workflows. In *Proceedings of the 28th ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, 2023.

[5] Yanqi Zhang, Weizhe Hua, **Zhuangzhuang Zhou**, G. Edward Suh, and Christina Delimitrou. Sinan: Ml-based and qos-aware resource management for cloud microservices. In *Proceedings of the 26th ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, 2021.

[6] Chang Meng, **Zhuangzhuang Zhou**, Yue Yao, Shuyang Huang, Yuhang Chen, and Weikang Qian. Hedals: Highly efficient delay-driven approximate logic synthesis. In *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)*, 2023.

[7] Zuodong Zhang, Runsheng Wang, Zhe Zhang, Ru Huang, Chang Meng, Weikang Qian, and **Zhuangzhuang Zhou**. Reliability-enhanced circuit design flow based on approximate logic synthesis. In *Proceedings of the 2020 on Great Lakes Symposium on VLSI (GLSVLSI)*, 2020.

[8] **Zhuangzhuang Zhou**, Yue Yao, Shuyang Huang, Sanbao Su, Chang Meng, and Weikang Qian. Dals: Delay-driven approximate logic synthesis. In *2018 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 2018.

Teaching Experience

Cornell University:

- *Head teaching assistant* of ECE4750/CS4420 Computer Architecture (Fall 2020, Fall 2021).

Shanghai Jiao Tong University:

- *Teaching assistant* of VG281 Data Structures and Algorithms (Fall 2017).
- *Teaching assistant* of VG280 Programming and Elementary Data Structure (Summer 2018).

Awards

- Cornell Graduate School Fellowship (Jacobs Scholarship), 2019
- First Place, Undergraduate Category, ACM Student Research Competition Grand Finals 2019.
- Gold Medal, Undergraduate Category, ACM Student Research Competition at ICCAD 2018.
- The Yu Liming Scholarship 2018 (top 1%).
- The Mong Man Wai International Exchange Fund 2018.
- The Fung Scholarship 2017.
- First Prize, National Olympiad in Informatics in Provinces (NOIP) 2011.