

Shuang Chen

Email

chenshuang0804@gmail.com

chenshuang44@huawei.com

Website

<https://sc2682cornell.github.io>

Research Interests

Cloud Computing, Datacenter

Workload Characterization, Resource Management

QoS Optimization

Hardware-Software Co-Optimization

EDUCATION

Cornell University, USA.

May 2021

Ph.D. in Computer Engineering

- Committee: José F. Martínez (advisor), Christina Delimitrou, Robbert van Renesse
- Thesis: Improving Resource Efficiency for Colocation of Multiple Latency-Critical Applications in Datacenters

Shanghai Jiao Tong University, China

July 2015

B.S. in Computer Science, ACM Honored Class, Zhiyuan College

- Admitted in Aug 2010, waiving the college entrance exam

PROFESSIONAL EXPERIENCES

Researcher at Huawei Cloud, Hangzhou, China.

Aug 2021 ~ Present

Shuhai Lab, under Dr. Zhibin Yu

- Improving resource efficiency of cloud workloads through hardware-software co-optimization

Graduate Research Assistant at Cornell University, NY, USA.

Aug 2015 ~ May 2021

M³ Research Group, under Prof. José F. Martínez

- QoS-aware resource management of interactive services in datacenters

Research Intern at Facebook Inc., CA, USA.

Summer 2018

Capacity Engineering and Analysis team, under Dr. Banit Agrawal

- Resource monitoring and management of Facebook Containers
- GPU tracing and profiling of Facebook machine learning workloads

Software Engineering Intern at Cavium (now Marvell) Inc., CA, USA.

Summer 2016

Processor Architecture team, under Dr. Srilatha(Bobbie) Manne

- Improving performance of last-level cache in Cavium ThunderX2
- Performance analysis of interactive cloud applications in Cavium ThunderX2

Undergraduate Research Assistant at NTU, Singapore

Aug 2014 ~ Feb 2015

Xtra Computing Group, under Prof. Bingsheng He

- Optimizing sorting algorithms using approximate NVRAM

Undergraduate Research Assistant at Shanghai Jiao Tong University, China

July 2013 ~ June 2014

Advanced Computer Architecture Laboratory, under Prof. Xiaoyao Liang

- Optimization of GPGPU Register File

PROFESSIONAL SERVICES

- **Program Committee:** HPCA 2026, NSDI 2026, ChinaSys Spring 2024(PC Co-Chair), ASPLOS 2024/2023, YArch 2025/2024/2023
- **External Review Committee,** MICRO 2025, ISCA 2024, ASPLOS 2022, HPCA 2021

PUBLICATIONS

- Yajuan Peng*, **Shuang Chen***, Yi Zhao, Zhibin Yu, “UFO: The Ultimate QoS-Aware Core Management for Virtualized and Oversubscribed Public Clouds,” in *the 21th USENIX Symposium on Networked Systems Design and Implementation (NSDI)*, April 2024. (*co-first authors)
- **Shuang Chen**, Angela Jin, Christina Delimitrou, and José F. Martínez, “ReTail: Opting for Learning Simplicity to Enable QoS-Aware Power Management in the Cloud,” in *the 28th IEEE HPCA International Conference on High Performance Computer Architecture (HPCA)*, Feb 2022.
- **Shuang Chen**, Yi Jiang, Christina Delimitrou, and José F. Martínez, “PIMCloud: QoS-Aware Resource Management of Latency-Critical Applications in Clouds with Processing-in-Memory,” in *the 28th IEEE HPCA International Conference on High Performance Computer Architecture (HPCA)*, Feb 2022.
- **Shuang Chen**, Christina Delimitrou, and José F. Martínez, “PARTIES: QoS-Aware Resource Partitioning for Multiple Interactive Services,” in *the 24th ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, Apr 2019.
- **Shuang Chen**, Shay Galon, Christina Delimitrou, Srilatha Manne, and José F. Martínez, “Workload Characterization of Interactive Cloud Services on Big and Small Platforms,” in *the 2017 IEEE International Symposium on Workload Characterization (IISWC)*, Oct 2017.
- Xiaodong Wang, **Shuang Chen**, Jeff Setter and José F. Martínez, “SWAP: Effective Fine-grain Management of Shared Last-level Caches with Minimum Hardware Support,” in *the 23rd IEEE HPCA International Conference on High Performance Computer Architecture (HPCA)*, Feb 2017.
- **Shuang Chen**, Shunning Jiang, Bingsheng He and Xueyan Tang, “A Study of Sorting Algorithms Using Approximate Storage,” in *the 2016 ACM SIGMOD International Conference on Management of Data (SIGMOD)*, June 2016.
- Naifeng Jing, Shunning Jiang, **Shuang Chen**, Jinjie Zhang, Li Jiang, Chao Li and Xiaoyao Liang, “Bank Stealing for a Compact and Efficient Register File Architecture in GPGPU,” in *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 2016.
- Naifeng Jing, **Shuang Chen**, Shunning Jiang, Li Jiang and Xiaoyao Liang, “Bank Stealing for Conflict Mitigation in GPGPU Register File,” in *the 2015 ACM/IEEE International Symposium on Low Power Electronics and Design (ISLPED)*, July 2015.

TALKS

QoS-Aware Resource Management in the Cloud: From Academia to Industry

- Invited talk at *IIS of Tsinghua University*, hosted by Dr. Yi Wu, Feb 21, 2023
- Invited talk at *ICT, Chinese Academy of Sciences*, hosted by Dr. Chenxi Wang, Feb 22, 2023
- Invited talk at *Nankai University*, hosted by Dr. Yusen Li, Feb 27, 2023
- Invited talk at *Tianjin University*, hosted by Dr. Shanjiang Tang, Feb 28, 2023
- Invited talk at *Zhejiang University*, hosted by Dr. Zeke Wang, April 26, 2023

PIMCloud: QoS-Aware Resource Management of Interactive Services in Datacenters with Processing-in-Memory

- Poster session at *SRC TECHCON*, Austin, Sept 9, 2019
- Invited talk at *SRC JUMP e-Workshop*, Oct 15, 2019
- Poster session at *SRC JUMP Center for Research in Intelligent Storage (CRISP) Annual Review*, Charlottesville, USA, Nov 5, 2019
- Invited talk at *CCFSys*, online, Dec 15, 2021
- Conference talk at *the 28th IEEE HPCA International Conference on High Performance Computer Architecture (HPCA)*, online, Apr 6, 2022

ReTail: Opting for Learning Simplicity to Enable QoS-Aware Power Management in the Cloud

- Invited talk at *ChinaSys*, online, Dec 4, 2021
- Conference talk at *the 28th IEEE HPCA International Conference on High Performance Computer Architecture (HPCA)*, online, Apr 4, 2022

PARTIES: QoS-Aware Resource Partitioning for Multiple Interactive Services

- Invited talk at *Microsoft Research*, Redmond, Nov 29, 2017
- Conference talk at *the 24th ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, Providence, USA, Apr 15, 2019
- Research pitch competition and poster session at *Rising Stars in EECS Workshop*, University of Illinois at Urbana-Champaign, USA, Oct 30, 2019

Workload Characterization of Interactive Cloud Services on Big and Small Platforms

- Conference talk at *the 2017 IEEE International Symposium on Workload Characterization (IISWC)*, Seattle, USA, Oct 2, 2017

A Study of Sorting Algorithms Using Approximate Storage

- Conference talk at *the 2016 ACM SIGMOD International Conference on Management of Data (SIGMOD)*, San Francisco, USA, June 27, 2016

HONORS AND AWARDS

• Top Minds Program , Huawei, China	2021
• Selected for Rising Stars in EECS Workshop , UIUC, USA	2019
– Winner of the Research Pitch Competition (ECE Track)	
• ECE Woman of the Month , Cornell University, USA	Mar 2018
• Cornell Graduate Fellowship , Cornell University, USA	2015
• Outstanding Graduate (top 1%), Shanghai Jiao Tong University, China	2015
• Outstanding Graduate in Zhiyuan College (top 12), Shanghai Jiao Tong University, China	2015
• Kai Yuan Scholarship (top 20 in Zhiyuan College), Shanghai Jiao Tong University, China	2013
• ACM-ICPC (International Collegiate Programming Contest) Asia Regional	
– 4th place in Hsinchu Site	2011
– Gold medal in Beijing Site	2011
• Silver medal (ranked the 64th) in National Olympiad in Informatics, China	2010

TEACHING EXPERIENCES

Cornell University, USA

- **Teaching Assistant** of Advanced Computer Architecture (CS5420/ECE5750) Fall 2020, Fall 2019
- **Guest Lecturer** of Competition Programming and Problem Solving Seminar (CS5199) Sept 2019
- **Head Teaching Assistant** of Computer Systems Programming (ECE2400) Spring 2017
- **Teaching Assistant** of Computer Architecture (CS4420/ECE4750) Fall 2016

Shanghai Jiao Tong University, China

- **Teaching Assistant** of C++ Programming (CS110) Fall 2012