HUAICHENG LI

CONTACT

Phone Number: (+1) 872-333-1908
5000 Forbes Avenue, Pittsburgh, PA

Email: huaichel@andrew.cmu.edu
Website: https://huaicheng.github.io

RESEARCH INTERESTS

Areas: Operating Systems, Storage and Memory Systems, Systems Support for Emerging Hardware/Applications

Focus: Design and build novel computing systems for emerging storage/memory hardware to achieve

(1) Performance: Software/Hardware co-design for low and predictable end-to-end latencies and high throughput
(2) Efficiency: Offloaded and disaggregated system architecture designs for improved resource and cost efficiency
(3) Programmability: Systems support for emerging I/O and acceleration technologies to ease development efforts

RESEARCH IMPACT

My research has been published at top Systems venues, *e.g.*, SOSP (×2), ASPLOS, and FAST (×3). My work has received recognition for its novelty and generated real-world impact. The systems and tools I developed have seen broad adoption in the research community and industry:

(1) TTFLASH [C7] and FailSlow [C5] received Best Paper Nominations at USENIX FAST 2017 and 2018
(2) LeapIO [C3] approach has been adopted in Microsoft datacenters to serve production workloads
(3) MittOS [C6] findings have been partially merged to the Linux kernel OpenChannel-SSD subsystem
(4) FEMU [C4], a modern storage research platform, has been widely used by the research community

ACADEMIC POSITIONS

Carnegie Mellon University — Pittsburgh, PA
Postdoctoral Researcher, Parallel Data Lab (PDL) — 2020–Present
Supervisor: Gregory R. Ganger

EDUCATION

University of Chicago — Chicago, IL
Ph.D. in Computer Science (*M.S. conferred in 2018*) — 2015–2020
Advisor: Haryadi S. Gunawi
Thesis: Evolving Storage Stack for Predictability and Efficiency [D1]

Wuhan University — Wuhan, China
M.S. in Computer Science (*dropped out to attend the Ph.D. program*) — 2013–2015
B.S. in Computer Science and Technology — 2009–2013

CONFERENCE PUBLICATIONS

Bibliometrics on Google Scholar; In-Submission (InSub) drafts (title anonymized) available upon request.

InSub'22 [C1] **Huaicheng Li**, Daniel S. Berger, Stanko Novakovic, Lisa Hsu, Dan Ernst, Monish Shah, Ishwar Agarwal, Mark Hill, Marcus Fontoura, Ricardo Bianchini. **Title Intentionally Omitted Due to Potential Conflicts for Double-blind Reviewing**. *In (Re)Submission to Some Conference with a Word Limit. Good Spirit Arxived in 2022.*

SOSP'21 [C2] **Huaicheng Li**, Martin L. Putra, Ronald Shi, Xing Lin, Gregory R. Ganger, Haryadi S. Gunawi. **IODA: A Host/Device Co-Design for Strong Predictability Contract on Modern Flash Storage.** *In Proceedings of the 28th ACM Symposium on Operating Systems Principles (SOSP), 2021.* [54/348=15.5%]
Pre-accepted sans discussion at the PC meeting; Artifact Available, Functional and Results Reproduced

ASPLOS'20 [C3] **Huaicheng Li**, Mingzhe Hao, Stanko Novakovic, Vaibhav Gogte, Sriram Govindan, Dan R. K. Ports, Irene Zhang, Ricardo Bianchini, Haryadi S. Gunawi, Anirudh Badam. **LeapIO: Efficient and Portable Virtual NVMe Storage on ARM SoCs.** *In Proceedings of the 25th ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS), 2020.* [86/476=18.1%]
A LeapIO variant has been deployed in Microsoft datacenters

FAST'18 [C4] **Huaicheng Li**, Mingzhe Hao, Michael Hao Tong, Swaminatahan Sundararaman, Matias Bjørling, Haryadi S. Gunawi. **The CASE of FEMU: Cheap, Accurate, Scalable and Extensible Flash Emulator.** *In Proceedings of the 16th USENIX Conference on File and Storage Technologies (FAST), 2018.* [23/139=16.5%]
A popular storage research platform widely used by recent top venue papers at OSDI, SOSP, ASPLOS, FAST, etc.

FAST'18 [C5] Haryadi S. Gunawi, Riza Suminto, Russell Sears, Casey Golliher, Swaminatahan Sundararaman, Xing Lin, Tim Emami, Weiguang Sheng, Nematollah Bidokhti, Caitie McCaffrey, Gary Grider, Parks M. Fields, Kevin Harms, Robert B. Ross, Andree Jacobson, Robert Ricci, Kirk Webb, Peter Alvaro, H. Birali Runesh, Mingzhe Hao, **Huaicheng Li**. **Fail-Slow at Scale: Evidence of Hardware Performance Faults in Large Production Systems.** *In Proceedings of the 16th USENIX Conference on File and Storage Technologies (FAST), 2018.* [23/139=16.5%] **Best Paper Nominee**

Fast-Tracked to ACM Transactions on Storage; and Invited to USENIX ;login:
Top-2 most cited papers at FAST 2018
Featured on "the Morning Paper", "ZDNet", and "HackerNews"

SOSP'17 [C6] Mingzhe Hao, **Huaicheng Li***, Michael Hao Tong, Chrisma Pakha, Riza Suminto, Cesar A. Stuardo, Andrew A. Chien, Haryadi S. Gunawi. **MittOS: Supporting Millisecond Tail Tolerance with Fast Rejecting SLO-Aware OS Interface.** *In Proceedings of the 26th ACM Symposium on Operating Systems Principles (SOSP), 2017.* [39/232=16.8%]

***Almost equal-contribution as the first author**
My findings have been partially merged into the Linux OpenChannel-SSD subsystem

FAST'17 [C7] Shiqin Yan, **Huaicheng Li**, Mingzhe Hao, Michael Hao Tong, Swaminatahan Sundararaman, Andrew A. Chien, Haryadi S. Gunawi. **Tiny-Tail Flash: Near-Perfect Elimination of Garbage Collection Tail Latencies in NAND SSDs.** *In Proceedings of the 15th USENIX Conference on File and Storage Technologies (FAST), 2017.* [28/116=24.1%] **Best Paper Nominee**

Fast-Tracked to ACM Transactions on Storage
Discussed as an alternative for NVMe IOD interface in next-gen SSDs at industrial events (*e.g.*, SDC, FMS)
Top-2 most cited FAST papers in the last 5 years

Journal Publications

InSub'22 [J1] **Huaicheng Li**, Martin L. Putra, Ronald Shi, Fadhil I. Kurnia, Xing Lin, Jaeyoung Do, Achmad Imam Kistijantoro, Gregory R. Ganger, Haryadi S. Gunawi. **IODA: A Host/Device Co-Design for Strong Predictability Contract on Modern Flash Storage.** In *Submission to ACM Transactions on Storage (TOS)*, 2022. [Extended version of C2]

TOS'18 [J2] Haryadi S. Gunawi, Riza Suminto, Russell Sears, Casey Golliher, Swaminatahan Sundararaman, Xing Lin, Tim Emami, Weiguang Sheng, Nematollah Bidokhti, Caitie McCaffrey, Gary Grider, Parks M. Fields, Kevin Harms, Robert B. Ross, Andree Jacobson, Robert Ricci, Kirk Webb, Peter Alvaro, H. Birali Runesh, Mingzhe Hao, **Huaicheng Li**. **Fail-Slow at Scale: Evidence of Hardware Performance Faults in Large Production Systems.** *ACM Transactions on Storage (TOS), Volume 14, Issue 3, November 2018.* [Extended version of C5] **Fast-tracked**

Research highlights from FAST 2018, only 5 papers selected

TOS'17 [J3] Shiqin Yan, **Huaicheng Li**, Mingzhe Hao, Michael Hao Tong, Swaminatahan Sundararaman, Andrew A. Chien, Haryadi S. Gunawi. **Tiny-Tail Flash: Near-Perfect Elimination of Garbage Collection Tail Latencies in NAND SSDs.** *ACM Transactions on Storage (TOS), Volume 13, Issue 3, October 2017.* [Extended version of C7] **Fast-tracked**

Research highlights from FAST 2017, only 5 papers selected

Dissertation

[D1] Ph.D. Thesis, **Evolving Storage Stack for Predictability and Efficiency.** University of Chicago. 2020

[D2] M.S. Thesis, **FEMU: Fast, Accurate and Extensible Flash Emulator.** University of Chicago. 2018

Honors & Awards

Samsung UR Collaboration Award, $110,000, (w/ George, Greg, and James) 2021

Facebook Research Gift, (w/ George and Greg) 2021

SYSTOR'21 Distinguished Reviewer Award 2021

Nomination for the SIGOPS Dennis M. Ritchie Doctoral Dissertation Award (*1 per department*) 2020

University Unrestricted (UU) Fellowship, University of Chicago 2019

EuroSys'18 Best Reviewer Award (Shadow Program Committee) 2018

FAST'18 Best Paper Nominee: Fail-Slow at Scale [C5] 2018

FAST'17 Best Paper Nominee: Tiny-Tail Flash [C7] 2017

Travel Grants: OSDI'16, SOSP'17, FAST'18, FAST'20, NSDI'20, ASPLOS'20, ATC'20 2016–2020

Bachelor period: First-class Scholarship, National Endeavor Scholarship, Outstanding Graduate, etc. 2009–2013

Work Experience

Research Internships at Industrial Labs

Microsoft Research (Redmond), *Systems Research Group* Summer 2020
Research Intern working on resource disaggregation for datacenter deployment [InSub'22]
Collaborators: Ricardo Bianchini, Daniel Berger, Stanko Novakovic, Mark Hill, Lisa Hwu

Microsoft Research (Redmond), *Database Group* Summer 2019
Research Intern working on programmable storage
Collaborator: Jae Young Do

Microsoft Research (Redmond), *Systems Research Group* Summer 2018
Research Intern working on offloading cloud storage stack to ARM SoCs [ASPLOS'20]
Collaborators: Ricardo Bianchini, Anirudh Badam, Irene Zhang, Dan R. Ports

NetApp, *Advanced Technology Group (ATG)* Spring 2020
Research Intern working on new file system designs for emerging storage hardware
Collaborators: Xing Lin, Fan Ni, Art Harkin

Research Experience at Universities

Carnegie Mellon University, *Parallel Data Lab (PDL)* 2020–Present
Postdoctoral Researcher collaborating with Gregory R. Ganger, George Amvrosiadis,
David G. Andersen and CMU students on new storage and memory technologies [InSub'22]

University of Chicago, *Systems Group* 2015–2020
Graduate Student Researcher working on Operating and Storage Systems research
[SOSP'21, ASPLOS'20, FAST'18, SOSP'17, FAST'17]

Wuhan University, *Cloud Computing Lab* 2012–2015
Research Assistant working on I/O virtualization and cloud resource scheduling
Advisor: Yili Gong

Engineering Internship in Industry

Tencent (Shenzhen) Summer 2012
Undergraduate Intern working on cluster resource monitoring and kernel optimization

Service

Program Committee (PC)
APSys'21: The 12th ACM SIGOPS Asia-Pacific Workshop on Systems 2021
SYSTOR'21: The 14th ACM International Systems and Storage Conference 2021

External Review Committee (ERC)
ASPLOS'23: The 28th ACM Intl' Conf. on Architectural Support for PL and OS 2022

Shadow Program Committee
EuroSys'20: The 15th European Conference on Computer Systems 2020
EuroSys'18: The 13th European Conference on Computer Systems 2018

Journal Reviewer
TC: IEEE Transactions on Computers 2019, 2020
TOCS: ACM Transactions on Computer Systems 2019
TPDS: IEEE Transactions on Parallel and Distributed Systems 2019
JPDC: Journal of Parallel and Distributed Computing 2019, 2021
TOS: ACM Transactions on Storage 2018
CACM: Communications of the ACM 2018

External Reviewer

FAST'19: The 17th USENIX Conference on File and Storage Technologies 2019
ATC'18: The 2018 USENIX Annual Technical Conference 2018

Secondary/Sub-reviewer

FAST'20: The 18th USENIX Conference on File and Storage Technologies 2020
SOSP'19: The 27th ACM Symposium on Operating Systems Principles 2019
ASPLOS'19: The 24th ACM Intl' Conf. on Architectural Support for PL and OS 2019
FAST'18: The 16th USENIX Conference on File and Storage Technologies 2018

Artifact Evaluation Committee (AEC)

SOSP'21: The 28th ACM Symposium on Operating Systems Principles 2021

Departmental Service

Graduate Student Ministry - Minister for Faculty Hiring, CS Dept, University of Chicago 2019

Other Activities

SOSP'21 Mentoring Program 2021
Session Chair: SYSTOR'21 ("Storage Session") 2021
USENIX HotStorage'20 Program Committee Meeting Scribe 2020
Chameleon Cloud Testbed Student Ambassador 2020
USENIX ATC'18 Program Committee Meeting Scribe 2018

Talks

IODA: Host/Device Co-Design for Strong Predictability Contract on Modern Flash Storage

Conference Talk, SOSP'21, Online 2021
Parallel Data Lab (PDL) Seminar, Carnegie Mellon University, PA, USA 2021

Towards Hardware-based Memory Disaggregation

Invited Talk, Microsoft Research (Redmond), WA, USA 2020

NVMeFS: SmartNIC-centric File System Offloading

Invited Talk, NetApp, CA, USA 2020

Evolving Storage Stack for Predictability and Efficiency

Invited Talk, University of Illinois at Urbana-Champaign, IL, USA 2021
Ph.D. Thesis Defense, University of Chicago, IL, USA 2020
Invited Talk, Carnegie Mellon University, PA, USA 2020
Invited Talk, Microsoft Research - Cambridge, UK 2020
Invited Talk, University of California - Berkeley, CA, USA 2020
Invited Talk, University of Wisconsin - Madison, WI, USA 2020

LeapIO: Efficient and Portable Virtual NVMe Storage on ARM SoCs

Invited Talk, CAS ICT Young Scholar Forum, Beijing, China 2020
Conference Talk, ASPLOS'20, Lausanne, Switzerland 2020
Ph.D. Thesis Proposal, University of Chicago, IL, USA 2019
Invited Talk, Microsoft Research (Redmond), WA, USA 2018

The CASE of FEMU: Cheap, Accurate, Scalable and Extensible Flash Emulator

Master Thesis Defense, University of Chicago, IL, USA 2018
Conference Talk, FAST'18, Oakland, CA, USA 2018

Teaching Experience

Co-Instructor

18-746: Storage Systems (Fall 2021), Carnegie Mellon University 2021
Course website: https://course.ece.cmu.edu/ ece746/index.html
↪ Co-teaching with Gregory R. Ganger and George Amvrosiadis, ∼100 students (BS/MS/PhD)
↪ Designing and giving lectures on storage management, file systems, etc.
↪ End-to-end class administration: course website, weekly TA meetings, designing quizzes, grading, etc.

Guest Lecturer

18-746: Storage Systems (Fall 2020), Carnegie Mellon University 2020
↪ Topic: "Ins and Outs of Storage Offloading using ARM SoCs"

Teaching Assistant

CMSC 230: Operating Systems, University of Chicago, TA — 2015, 2018, 2019
↪ Hosting lab sessions about Pintos projects (tutorials/lectures, office hours, grading, etc.), Rating: 9.4/10

Computer Organization and Design, Wuhan University, TA — 2014

Student Mentoring

Students at Carnegie Mellon University I work with

Thomas Kim (CS Ph.D. student), on Zoned Storage (ZNS) — 2020–Present
Sara McAllister (CS Ph.D. student), on OS support for memory disaggregation — 2021–Present
Daiyaan Arfeen (CS Ph.D. student), on Computational Storage — 2021–Present
Nirjhar Mukherjee (CS Ph.D. student), on Computational Storage — 2021–Present
Sumanth Subramanya (MCSD master student), on Programmable Storage — 2022–Present
Jiuzhi Yu (MCSD master student), on Programmable Storage — 2022–Present
Zixu Chen (MCSD master student), on Programmable Storage — 2021–2022
Aditya Shetty (MCSD master student), on Programmable Storage — 2021–2022

Mentees at University of Chicago

As thesis research advisor:

Yesa Rahmad (BS/MS)), on FEMU scalability optimization — 2019–2020
Fadhil I. Kurnia (BS→PhD student at University of Massachusetts - Amherst) — 2018–2020
Martin L. Putra (BS→PhD student at University of Chicago, co-author of [C2]) — 2018–2020
Ronald Shi (BS/MS→Facebook, co-author of [C2]) — 2018–2019

As research project mentor:

UChicago: Kelvin Ho (MS), Yang Chen (MS), Jingchun Wang (MS), Jingjie Wan (MS) — 2017–2018
Remote: Sujin Park (BS→Ph.D. student at Georgia Institute of Technology) — 2019

Grant Writing

Co-PI, Microsoft Azure - Research Proposal, Memory Disaggregation — 2021
With Daniel S. Berger; *(ongoing)*

Co-PI, Samsung MSL - Research Proposal, Composable Storage Architecture — 2021
With George Amvrosiadis, Gregory R. Ganger, and James C. Hoe; (I'm the major proposal writer.)

Co-PI, Facebook Request for Proposals, Next-generation Data Infrastructure — 2021
With George Amvrosiadis and Gregory R. Ganger;

Software

IODA: https://github.com/huaicheng/IODA (3☆) — 2021
LeapIO: https://github.com/huaicheng/LeapIO (10☆) — 2020
↪ A LeapIO variant has been deployed in Microsoft datacenters.
FEMU: https://github.com/ucare-uchicago/FEMU (190☆) — 2018
↪ FEMU is a popular storage research platform widely used by top venue papers at ASPLOS, FAST, OSDI, and SOSP, etc.
MittSSD: https://github.com/ucare-uchicago/mittssd (6☆) — 2018
ttFlash: https://github.com/ucare-uchicago/tinyTailFlash (7☆) — 2017
Linux Kernel Contributor: Linux Open-Channel SSD Subsystem - pblk (120☆) — 2017

Skills

Linux Kernel: Hacking experience with KVM, MM, EXT4, BIO, MD, LightNVM, NVMe subsystems

Systems: Hacking experience with QEMU, SPDK, RocksDB, MongoDB, Spark, FEMU, SSDSim

Hardware Platforms: BlueField/StingRay SmartNIC, OpenChannel-SSD, OpenSSD, DFC, SmartSSD

Programming & Tools: C/C++, Python, Bash; Awk, Sed, Gnuplot, GDB, LaTeX, etc.

Media Coverage

Fail-Slow at Scale

The Morning Paper, https://blog.acolyer.org, search "fail slow at scale" — 2018
ZDNet, https://www.zdnet.com/article/how-clouds-fail-slow — 2018
Hacker News, https://news.ycombinator.com/item?id=16463714 — 2018

Recent Projects

[P1] **CXL-based Software/Hardware Co-Design for Memory Disaggregation** [InSub'22] 2020–2021

Memory disaggregation is a promising approach to achieve DRAM cost-efficiency for datacenters. Unfortunately, many existing disaggregation designs fail to meet requirements for cloud deployment at scale. We propose a full-stack Compute Express Link (CXL) based memory disaggregation design without jeopardizing QoS for practical deployment in large cloud platforms.

[P2] **IODA: Host/Device Co-Design for I/O Determinism** [SOSP'21] 2018–2021

Predictable latency on flash storage is a long-pursuit goal, yet, unpredictability stays due to disturbance from many SSD internal activities. IODA is an I/O deterministic flash array design built on top of small but powerful extensions to the NVMe I/O Determinism (IOD) interface for easy deployment. IODA advocates cross-device coordination and host/SSD co-design to guarantee predictable latencies.

[P3] **LeapIO: Efficient Virtual NVMe Storage on ARM SoCs** [ASPLOS'20] 2018–2020

Modern cloud storage stack is extremely resource hungry, burning 10-20% of data-center x86 cores, it is a major "storage tax" that cloud providers must pay. We design and build a next generation co-processor assisted cloud storage stack to achieve performance predictability and cost-efficiency while satisfying deployment requirements such as composability and extensibility.

[P4] **FEMU: Accurate, Scalable, Extensible NVMe SSD Emulator** [FAST'18] 2017–2018

Existing SSD research platforms have limited support for conducting research utilizing software defined flash and software/hardware co-design approaches. To bridge the gap, we build a full-system SSD emulator platform, FEMU, to foster future software/hardware storage research. FEMU is carefully designed and optimized to deliver high performance, accuracy, flexibility and extensibility.

[P5] **MittOS: OS Support for Tail Tolerance** [SOSP'17] 2016–2017

Current OS's best-effort policy in resource management would bring tail latencies and cause application level SLO violations. We argue that OS should be transparent of resource busyness and return `EBUSY` timely to allow applications to quickly failover to less-busy replicas. Thus, we design a novel fast-rejection SLO interface and demonstrate how they can help applications achieve better performance.

[P6] **ttFlash: Tiny-Tail Flash Storage Architecture** [FAST'17, TOS'17] 2015–2016

Traditional SSD controller designs fail to deliver low and stable latencies due to the unavoidable Garbage Collection (GC) process which interferes with user requests. We re-architect SSD architecture designs to tackle the problem via a fine-grained (chip-level) GC blocking mechanism, proactive regeneration of blocked-reads by exploiting parity redundancy and rotational GC scheduling policy.