

JIACHENG MA

2100 Logic Dr
San Jose, CA 95124, USA

i@jcma.me
<https://jcma.me>

RESEARCH INTERESTS

My research is at the intersection of hardware and software. I am interested in improving the programmability, reliability, debuggability, and deployability of heterogeneous systems by building systems support such as hypervisors, compilers, debuggers, and runtimes. I am also interested in system virtualization and software-hardware co-design.

EDUCATION

University of Michigan	<i>Ann Arbor, MI, USA</i>
<i>Ph.D. in Computer Science and Engineering</i>	Sept. 2018 – Jan. 2024
Thesis: Systems and Debugging Supports for Hardware Designs	
Advisor: Prof. Baris Kasikci	
University of Michigan	<i>Ann Arbor, MI, USA</i>
<i>M.S.E. in Computer Science and Engineering</i>	Sept. 2018 – 2021
Shanghai Jiao Tong University	<i>Shanghai, China</i>
<i>B.E. in Software Engineering</i>	Sept. 2014 – June 2018
Thesis: Efficient GPU Live Migration Optimized by Software Dirty Page for Full Virtualization	
Advisor: Prof. Zhengwei Qi	

EMPLOYMENT

Advanced Micro Devices	<i>San Jose, California</i>
<i>Sr. Software Engineer</i>	Jan 2024 – Present
Supervisor: Dr. Alejandro Rico	
Alibaba Group (U.S.)	<i>Sunnyvale, California</i>
<i>Research Intern</i>	May 2022 – Aug. 2022
Supervisor: Dr. Dimin Niu	
VMware Research	<i>Remote</i>
<i>Research Intern</i>	May 2021 – Aug. 2021
Supervisors: Dr. Marcos K. Aguilera, Dr. Irina Calciu	
Intel Lab	<i>Remote</i>
<i>Graduate Technical Intern</i>	June 2020 – Aug. 2020
Supervisor: Dr. Sanjay Kumar	
Intel Asia-Pacific Research & Development Ltd	<i>Shanghai, China</i>
<i>Software Developer Intern</i>	July 2016 – June 2018
Supervisor: Dr. Eddie (Yaozu) Dong	

PEER-REVIEWED CONFERENCE PUBLICATIONS

- [1] Vidi: Record Replay for Reconfigurable Hardware. Gefei Zuo, Jiacheng Ma, Andrew Quinn, and Baris Kasikci. *Proceedings of the 28th International Conference on Architectural Support for Programming Languages and Operating Systems*, 2023.

- [2] Debugging in the Brave New World of Reconfigurable Hardware. Jiacheng Ma, Gefei Zuo, Kevin Loughlin, Haoyang Zhang, Andrew Quinn, and Baris Kasikci. *Proceedings of the 27th International Conference on Architectural Support for Programming Languages and Operating Systems*, 2022.
- [3] MEGATRON: Software-Managed Device TLB for Shared-Memory FPGA Virtualization. Yanqiang Liu, Jiacheng Ma, Zhengjun Zhang, Linsheng Li, Zhengwei Qi, and Haibing Guan. *The 58th Design Automation Conference*, 2021.
- [4] DOLMA: Securing Speculation with the Principle of Transient Non-Observability. Kevin Loughlin, Ian Neal, Jiacheng Ma, Elisa Tsai, Ofir Weisse, Satish Narayanasamy, and Baris Kasikci. *30th USENIX Security Symposium (USENIX Security 21)*, 2021.
- [5] Execution reconstruction: Harnessing failure reoccurrences for failure reproduction. Gefei Zuo, Jiacheng Ma, Andrew Quinn, Pramod Bhatotia, Pedro Fonseca, and Baris Kasikci. *Proceedings of the 42nd ACM SIGPLAN International Conference on Programming Language Design and Implementation*, pages 1155–1170, 2021.
- [6] A Hypervisor for Shared-Memory FPGA Platforms. Jiacheng Ma, Gefei Zuo, Kevin Loughlin, Xiaohe Cheng, Yanqiang Liu, Abel Mulugeta Eneyew, Zhengwei Qi, and Baris Kasikci. *Proceedings of the 25th International Conference on Architectural Support for Programming Languages and Operating Systems*, 2020.
- [7] gRemote: API-Forwarding Powered Cloud Rendering. Dongjie Tang, Yun Wang, Linsheng Li, Jiacheng Ma, Xue Liu, Zhengwei Qi, and Haibing Guan. *Proceedings of the 29th International Symposium on High-Performance Parallel and Distributed Computing*, pages 197–201, 2020.
- [8] gMig: Efficient GPU Live Migration Optimized by Software Dirty Page for Full Virtualization. Jiacheng Ma, Xiao Zheng, Yaozu Dong, Wentai Li, Zhengwei Qi, Bingsheng He, and Haibing Guan. *Proceedings of the 14th ACM SIGPLAN/SIGOPS International Conference on Virtual Execution Environments*, pages 31–44, 2018.
- [9] gScale: Scaling up GPU Virtualization with Dynamic Sharing of Graphics Memory Space. Mochi Xue, Kun Tian, Yaozu Dong, Jiacheng Ma, Jiajun Wang, Zhengwei Qi, Bingsheng He, and Haibing Guan. *Proceedings of the 2016 USENIX Conference on Usenix Annual Technical Conference*, pages 579–590, 2016.

PEER-REVIEWED JOURNAL PUBLICATIONS

- [10] gRemote: Cloud rendering on GPU resource pool based on API-forwarding. Dongjie Tang, Linsheng Li, Jiacheng Ma, Xue Liu, Zhengwei Qi, and Haibing Guan. *Journal of Systems Architecture*, 116:102055, 2021.
- [11] gMig: Efficient vGPU Live Migration with Overlapped Software-based Dirty Page Verification. Qiumin Lu, Xiao Zheng, Jiacheng Ma, Yaozu Dong, Zhengwei Qi, Jianguo Yao, Bingsheng He, and Haibing Guan. *IEEE Transactions on Parallel and Distributed Systems*, 2019.
- [12] Scalable GPU Virtualization with Dynamic Sharing of Graphics Memory Space. Mochi Xue, Jiacheng Ma, Wentai Li, Kun Tian, Yaozu Dong, Jinyu Wu, Zhengwei Qi, Bingsheng He, and Haibing Guan. *IEEE Transactions on Parallel and Distributed Systems*, 29(8):1823–1836, 2018.

U.S. PATENT

- [13] Jiacheng Ma, Haibing Guan, Zhengwei Qi, and Yongbiao Chen. System, apparatus, and method for optimizing a scalable GPU virtualization, October 1 2019. US Patent 10,430,991.

SELECTED PROJECTS

Hardware Defect Detection

Sept. 2022 – Jan. 2024

- This on-going work explores techniques to construct software test cases that detects aging-related silent data corruptions inside a chip. It would enable data center operators to conduct frequent chip testing and detect hardware failures in a cost-efficient way.

FPGA Record/Replay [1]

Nov. 2021 – July 2023

- This on-going work explores record/replay techniques on FPGAs. In this work, we build the framework that can not only record and replay a hardware execution on an FPGA, but also enable other developers to build more sophisticated development tools.

FPGA Bug Study and Debugging Techniques [2]

Sept. 2020 – Nov. 2021

- This work explores debugging techniques for FPGAs. In this work, we perform a comprehensive study on real-world bugs in FPGA projects, classify these bugs based on their root causes and symptoms, and propose techniques to help bug localization.

Secure Speculative Execution [4]

Jan. 2020 – Apr. 2020

- This work presents the first defense scheme to provide automatic comprehensive protection against all known transient execution attacks.

Hardware-Assisted Bug Reproduction [5]

Aug. 2019 – Apr. 2020

- This work combines online recording and offline symbolic execution to recover failing program executions with low overhead and high accuracy. In the online phase, the control-flow of a program execution is recorded; the control-flow is then used in the offline phase to guide symbolic execution.

FPGA Virtualization [6]

Apr. 2018 – Aug. 2019

- In this project, we create the first scalable hypervisor for shared-memory FPGA platforms. The hypervisor, called OPTIMUS, supports both spatial multiplexing and temporal multiplexing, and scales linearly until the memory bandwidth is saturated. OPTIMUS helps data center operators to deploy multiple shared-memory accelerators on the same FPGA, thus maximizing the resource utilization.

Remote Rendering [7, 10]

Mar. 2018 – May 2018

- This work enables API-forwarding based cloud rendering for OpenGL applications on a resource pool.

Optane as Memory

Aug. 2017 – June 2018

- This work explores using NVM (Intel Optane) as main memory for KVM guests.

vGPU Live Migration [11, 8]

Dec. 2016 – July 2017

- Intel GVT-g is an open-source KVM/Xen based full GPU virtualization solution; however, it lacks the feature of live migration. This work enables live migrating vGPUs for cloud applications such as virtual desktops, cloud gaming farms, cloud transcoding services, etc.

Scaling up GPU Virtualization [12, 9, 13]

Jan. 2016 – Mar. 2017

- This project scales up the maximum number of vGPUs in Intel GVT-g (i.e., Intel's GPU virtualization technology) for 5× while minimizing the performance overhead. It makes GVT-g more consolidated, since more VMs with vGPU can be depolyed on one physical machine.

TEACHING

Programming and Data Structure (SE 117)

Shanghai, China

Teaching Assistant

Feb. 2016 – June 2016

Advanced Operating Systems (EECS 582)

Ann Arbor, MI

Graduate Student Instructor

Sept. 2021 – Dec. 2021

ACADEMIC MENTORING

Xiaohe Cheng (HKUST BSc → Google)

2019

Abel Mulugeta Eneyew (Addis Ababa Institute of Technology)

2019

Haoyang Zhang (UMich BSc → UIUC PhD)

2021 – 2022

Wentao Zhang (SJTU MSc → UIUC PhD)	2022
Yin Yuan (UMich BSc/MSc)	2022

PROFESSIONAL SERVICES

Reviewer for IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems	2023-2024
Artifact Evaluation Committee for ACM SIGPLAN Conference on Programming Language Design and Implementation	2023
External Reviewer for International Conference on Architectural Support for Programming Languages and Operating Systems	2022
Artifact Evaluation Committee for Journal of Systems Research	2021
Artifact Evaluation Committee for Symposium on Operating Systems Principles	2021
Reviewer for IEEE Transactions on Parallel and Distributed Systems	2018

TECHNICAL SKILLS

Programming Language: C, C++, Verilog, System Verilog
OS & Virtualization: Linux Kernel Development, KVM, QEMU, Mediated Pass-Through
Program Analysis: LLVM, Klee, Yosys, Pyverilog