

Xinwei (Mason) Fu

Email: foxinwe@amazon.com Web: <https://xinweif.github.io/>

RESEARCH INTEREST

- Machine Learning System
- Software Reliability
- Persistent Memory

EXPERIENCE

Amazon Web Services, Inc., Santa Clara, CA

Senior Applied Scientist

Dec 2024 - Present

Applied Scientist

Apr 2022 - Nov 2024

Applied Scientist Internship

May 2021 - Aug 2021

Virginia Tech, Blacksburg, VA

Jan 2016 - Dec 2021

Graduate Research/Teaching Assistant

EDUCATION

Virginia Tech, Blacksburg, VA

Jan 2016 - Feb 2022

Ph.D. in Computer Science

Sun Yat-Sen University (SYSU), Guangzhou, China

Sep 2008 - Jun 2015

M.S. in Microelectronics and Solid-state Electronics

B.E. in Microelectronics

PUBLICATIONS

[SOCC'24] **Xinwei Fu**, Zhen Zhang, Haozheng Fan, Guangtai Huang, Mohammad El-Shabani, Randy Huang, Rahul Solanki, Fei Wu, Ron Diamant, and Yida Wang, "Distributed Training of Large Language Models on AWS Trainium", 15th ACM Symposium on Cloud Computing (SoCC), Redmond, WA, November, 2024.

[BigData'24] Haozheng Fan, Hao Zhou, Guangtai Huang, Parameswaran Raman, **Xinwei Fu**, Gaurav Gupta, Dhananjay Ram, Yida Wang, and Jun Huan, "HLAT: High-quality Large Language Model Pre-trained on AWS Trainium", 2024 IEEE International Conference on Big Data (BigData), Washington DC, USA, December, 2024.

[SOSP'23] Zhuang Wang, Zhen Jia, Shuai Zheng, Zhen Zhang, **Xinwei Fu**, T. S. Eugene Ng, and Yida Wang, "Gemini: Fast Failure Recovery in Distributed Training with In-Memory Checkpoints", 29th ACM Symposium on Operating Systems Principles (SOSP), Koblenz, Germany, October 2023.

[OSDI'22] **Xinwei Fu**, Dongyoon Lee, and Changwoo Min, "Durinn: Adversarial Memory and Thread Interleaving for Detecting Durable Linearizability Bugs", 16th USENIX Symposium on Operating Systems Design and Implementation (OSDI), Carlsbad, CA, July 2022.

[NVMW'22] Jianping Zeng, Jongouk Choi, **Xinwei Fu**, Ajay P. Shreepathi, Dongyoon Lee, Changwoo Min, and Changhee Jung, "ReplayCache: Enabling Volatile Caches for Energy Harvesting Systems", 13th Non-Volatile Memories Workshop (NVMW), San Diego, CA, May 2022.

[SOSP'21] **Xinwei Fu**, Wook-Hee Kim, Ajay Paddayuru Shreepathi, Mohannad Ismail, Sunny Wadkar, Dongyoon Lee, and Changwoo Min, "Witcher: Systematic Crash Consistency Testing for Non-Volatile Memory Key-Value Stores", 28th ACM Symposium on Operating Systems Principles (SOSP), Virtual, October 2021.

[SOSP'21] Wook-Hee Kim, R. Madhava Krishnan, **Xinwei Fu**, Sanidhya Kashyap, and Changwoo Min, "PACTree: A High Performance Persistent Range Index Using PAC Guidelines", 28th ACM Symposium on Operating Systems Principles (SOSP), Virtual, October 2021.

[MICRO'21] Jianping Zeng, Jongouk Choi, **Xinwei Fu**, Ajay Paddayuru Shreepathi, Dongyoon Lee, Changwoo Min, and Changhee Jung, "ReplayCache: Enabling Volatile Caches for Energy Harvesting Systems", 54th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO), Virtual, October 2021.

[ATC'21] R. Madhava Krishnan, Wook-Hee Kim, **Xinwei Fu**, Sumit Kumar Monga, Hee Won Lee, Minsung Jang, Ajit Mathew, and Changwoo Min, "TIPS: Making Volatile Index Structures Persistent with DRAM-NVMM Tiering", USENIX Annual Technical Conference (ATC), Virtual, July 2021.

[NVMW'20] R.Madhava Krishnan, Jaeho Kim, Ajit Mathew, **Xinwei Fu**, Anthony Demeri, Changwoo Min and Sudarsun Kannan, "Durable Transactional Memory Can Scale with TimeStone", 11th Annual Non-Volatile Memories Workshop (NVMW), San Diego, CA, March 2020.

[ASPLOS'20] R.Madhava Krishnan, Jaeho Kim, Ajit Mathew, **Xinwei Fu**, Anthony Demeri, Changwoo Min and Sudarsun Kannan, "Durable Transactional Memory Can Scale with TimeStone", ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS), Lausanne, Switzerland, March, 2020.

[ATC'19] **Xinwei Fu**, Talha Ghaffar, James C. Davis and Dongyoon Lee, "A Better Stream Processing Engine for the Edge", USENIX Annual Technical Conference (ATC), Renton, WA, July 2019.

[CGO'18] **Xinwei Fu**, Dongyoon Lee, and Changhee Jung, "nAdroid: Statically Detecting Ordering Violations in Android Applications", IEEE/ACM International Symposium on Code Generation and Optimization (CGO), Vienna, Austria, February 2018.

SERVICES

Conference Committee Activities

- Program Committee, USENIX Annual Technical Conference (ATC). 2023, 2024, 2025
- Program Committee, Annual Conference on Machine Learning and Systems (MLSys). 2025
- Program Committee, IEEE International Parallel & Distributed Processing Symposium (IPDPS). 2025
- Program Committee, ACM SIGPLAN/SIGBED International Conference on Languages, Compilers, and Tools for Embedded Systems (LCTES). 2023
- Artifact Evaluation Program Committee, IEEE/ACM International Symposium on Code Generation and Optimization (CGO). 2019

Conference Organizing Activities

- Artifact Evaluation Co-Chair, ACM SIGPLAN/SIGBED International Conference on Languages, Compilers, and Tools for Embedded Systems (LCTES). 2023