

# Zhiting Zhu, Ph.D.




✉ zhitingz@nvidia.com

in LinkedIn

G Google Scholar

ID Orcid

## Education







- 2014 - 2024     **Ph.D., Computer Science, The University of Texas at Austin**  
Thesis: *Exactly-once Semantics for Recoverable Data Processing Applications*  
Advisor: Emmett Witchel  
Research interests: *Database System, Operating Systems*
- 2010-2014     **B.S. Computer Engineering, University of Washington**
-  **B.S. Electrical Engineering, University of Washington**

## Research Publications

- 1    **ZhitingZhu**, Zhipeng Jia, Newton Ni, Dixin Tang, and Emmett Witchel. “Impeller: Stream Processing on Shared Logs”. In: *Proceedings of the Twentieth European Conference on Computer Systems*. EuroSys '25. 2025.
- 2    **ZhitingZhu**, Newton Ni, Yibo Huang, Yan Sun, Zhipeng Jia, Nam Sung Kim, and Emmett Witchel. “Lupin: Tolerating Partial Failures in a CXL Pod”. In: *Proceedings of the 2nd Workshop on Disruptive Memory Systems*. DIMES '24. 2024.
- 3    Henrique Fingler, **ZhitingZhu**, Esther Yoon, Zhipeng Jia, Emmett Witchel, and Christopher J. Rossbach. “Disaggregated GPU Acceleration for Serverless Applications”. In: *SIGOPS Oper. Syst. Rev.* 57.1 (June 2023). ISSN: 0163-5980.
- 4    Henrique Fingler, **ZhitingZhu**, Esther Yoon, Zhipeng Jia, Emmett Witchel, and Christopher J. Rossbach. “DGSF: Disaggregated GPUs for Serverless Functions”. In: *2022 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*. 2022.
- 5    Yige Hu, **ZhitingZhu**, Ian Neal, Youngjin Kwon, Tianyu Cheng, Vijay Chidambaram, and Emmett Witchel. “TxFS: Leveraging File-system Crash Consistency to Provide ACID Transactions”. In: *ACM Trans. Storage* 15.2 (May 2019). ISSN: 1553-3077.
- 6    Yige Hu, **ZhitingZhu**, Ian Neal, Youngjin Kwon, Tianyu Cheng, Vijay Chidambaram, and Emmett Witchel. “TxFS: Leveraging File-System Crash Consistency to Provide ACID Transactions”. In: *2018 USENIX Annual Technical Conference (USENIX ATC 18)*. July 2018.
- 7    Tyler Hunt, **ZhitingZhu**, Yuanzhong Xu, Simon Peter, and Emmett Witchel. “Ryoan: A Distributed Sandbox for Untrusted Computation on Secret Data”. In: *ACM Trans. Comput. Syst.* 35.4 (Dec. 2018). ISSN: 0734-2071.
- 8    **ZhitingZhu**, Sangman Kim, Yuri Rozhanski, Yige Hu, Emmett Witchel, and Mark Silberstein. “Understanding The Security of Discrete GPUs”. In: *Proceedings of the General Purpose GPUs*. GPGPU-10. 2017.
- 9    Tyler Hunt, **ZhitingZhu**, Yuanzhong Xu, Simon Peter, and Emmett Witchel. “Ryoan: A Distributed Sandbox for Untrusted Computation on Secret Data”. In: *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*. Nov. 2016.

## Experience

---

- Apr 2025 – ...     Senior system software engineer, NVIDIA. Santa Clara, CA  
Automated cluster network control-plane creation with Headscale and Kubernetes controllers.  
Improved accuracy and coverage of job failure analysis.  
Managed Ray clusters and Slurm metadata on Lepton; collected and analyzed cluster metrics.  
Anomaly detection for large-scale training clusters.
- Sept 2024 – ...     Software engineer, Lepton AI LLC. Cupertino, CA  
Built services for monitoring and diagnosing ML training cluster failures, identifying root causes in both software and hardware layers.  
Designed GPU memory isolation mechanisms using a custom device driver, enabling flexible GPU memory partitioning and allocation via Node Resource Interface (NRI).  
Developed an automated test suite to validate node health, GPU interconnect communication, Ethernet and InfiniBand connectivity, and storage bandwidth.
- May-Aug 2017     Software Engineer (Intern), Mozilla, San Francisco, CA  
Reimplemented DOM representation in SpiderMonkey as JS values for reduced overhead.  
Used Rust procedural macros to automatically generate cross-language accessors.  
Built static analysis passes for C/C++ header inspection in rust-bindgen, checking type traits and v-table properties.
- May-Aug 2016     Software Engineer (Intern), Google, Mountain View, CA  
Benchmarked memory allocation behavior for Android Watch.
- Jun-Aug 2015     Software Engineer (Intern), Google, San Francisco, CA  
Ported command-line utilities to Native Client.  
Added FreeBSD package manager support and implemented missing libc functions.
- Jun-Sept 2013     Software Engineer (Intern), Foursquare, San Francisco, CA  
Developed MapReduce analytics for mobile app metrics.

## TA experience

---

- |                                  |  |
|----------------------------------|--|
| 2018 – 2019, Aug 2020 – Dec 2020 |  CS 371M: Mobile Computing(Android)     |
| Jan 2017 – Aug 2017              |  CS 380L: Advanced Operating System     |
| Aug 2016 – Dec 2016              |  CS 371M: Mobile Computing(Android)     |
| Aug 2014 – Dec 2014              |  CS 439: Principles of Computer Systems |