

Zhen Jin

✉ jin_zhen@zju.edu.com

☎ +8618980467627

📍 38 Zheda Road, Xihu District, Hangzhou, Zhejiang, China

🏠 <https://wrestle189.github.io/>



About Me

I am currently a fourth-year Ph.D. student at ARClab, Zhejiang University, China, under the supervision of Professor Wenzhi Chen. My Ph.D. program follows a five-year track, and I expect to graduate in 2026. At the same time, I am also working as a research intern at Alibaba Cloud, under the supervision of Senior Engineer Yiquan Chen. Before completing my Ph.D., I hope to join a leading research group as a visiting student. Through this opportunity, I aim to broaden my academic horizons and enhance my research capabilities. I also hope the experience helps me better understand whether to pursue an academic career in the future.

Education

- 2021 – Present 📖 **Ph.D., Zhejiang University** — College of Computer Science and Technology
Research focus: *Storage Systems for AI, NVMe Storage Virtualization*
- 2017 – 2021 📖 **B.Eng., Northwestern Polytechnical University** — College of Computer Science and Technology
GPA Ranking: 13/235 (Top 5.53%)
Relevant coursework includes: *Operating Systems, Computer Organization and Architecture, Assembly Language and Interface Programming*

Internship

- 2021 – Present 📖 **Academic Intern, Alibaba Cloud** — Alibaba Infrastructure Service (AIS)
Research focus: *NVMe storage virtualization, object storage client offloading, and remote inference*

Research Publications

First-Author Publications

- 1 Y. Chen*, **Z. Jin***, Y. Wang, Y. Chen, J. Xu, H. Yu, J. Chen, W. Lin, K. Fang, K. Zhang, C. Wei, Q. Liu, Y. Xie, and W. Chen, “NVMePass: A Lightweight, High-performance and Scalable NVMe Virtualization Architecture with I/O Queues Passthrough”, in *2025 IEEE International Symposium on High Performance Computer Architecture (HPCA)*, Los Alamitos, CA, USA: IEEE Computer Society, Mar. 2025, pp. 1395–1407. 📄 DOI: 10.1109/HPCA61900.2025.00105, **Category: NVMe storage**.
- 2 **Z. Jin**, Y. Chen, M. Liang, Y. Wang, G. Fang, A. Zhou, K. Zhang, J. Xu, W. Lin, Y. Lin, S. Zhao, W. Shi, Z. He, S. Cai, and W. Chen, “OS2G: A high-performance DPU offloading architecture for GPU-based deep learning with object storage”, in *Proceedings of the 30th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2*, ser. ASPLOS ’25, Rotterdam, Netherlands: Association for Computing Machinery, 2025, pp. 750–765, ISBN: 9798400710797. 📄 DOI: 10.1145/3676641.3716265, **Category: Offloading**.
- 3 **Z. Jin**, Y. Chen, Y. Wang, Y. Du, K. Zhang, J. Xu, W. Lin, J. Qin, K. Fang, and W. Chen, “rInfer: A generic and high-performance framework for remote inference with heterogeneous Accelerators”, in *Proceedings of the IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing (CCGrid)*, 2025, **Category: Remote inference**.

Co-authored Publications

- 1 Y. Chen, Y. Xie, Y. Wang, J. Xu, **Z. Jin**, A. Li, X. Fu, Q. Liu, and W. Chen, "Optimizing NVMe storage for large-Scale deployment: Key technologies and strategies in Alibaba Cloud", *IEEE Micro*, vol. 44, no. 5, pp. 47–56, 2024. [DOI: 10.1109/MM.2024.3426514](#), **Category: NVMe storage**.
- 2 J. Xu, Y. Chen, Y. Wang, W. Shi, G. Fang, Y. Chen, H. Liao, Y. Wang, H. Lin, **Z. Jin**, Q. Liu, and W. Chen, "LightPool: A NVMe-oF-based high-performance and lightweight storage pool architecture for cloud-Native distributed Database", in *2024 IEEE International Symposium on High-Performance Computer Architecture (HPCA)*, 2024, pp. 983–995. [DOI: 10.1109/HPCA57654.2024.00079](#), **Category: NVMe storage**.
- 3 Y. Chen, J. Chen, Y. Wang, Y. Chen, **Z. Jin**, J. Xu, G. Fang, W. Lin, C. Wei, and W. Chen, "HyQ: Hybrid I/O queue architecture for NVMe over fabrics to enable high- performance hardware offloading", in *2023 IEEE/ACM 23rd International Symposium on Cluster, Cloud and Internet Computing (CCGrid)*, 2023, pp. 13–24. [DOI: 10.1109/CCGrid57682.2023.00012](#), **Category: NVMe storage**.

Current Research

My current research focuses on multi-tier memory management in training scenarios with constrained GPU memory. I aim to leverage host memory and SSDs as an extension to GPU memory, designing a system that delivers high performance and efficient resource utilization.

Skills

Research Attitude		Always passionate about new directions and solving real-world problems
Coding		C/C++, Python
Tools		Linux, Docker, Git

Awards

2022		Award of Honor for Graduate , Zhejiang University.
2021		Outstanding Undergraduate Graduate , Northwestern Polytechnical University
2018		National Scholarship , Ministry of Education of the People's Republic of China

Volunteer

2019.03 – 2019.07	Volunteer Companion , Tian'ai Rehabilitation Center, Xi'an, China Volunteered weekly with children with intellectual disabilities, offering companionship and emotional support through play-based activities.
2018.07	Volunteer Math Teacher , Rural Primary School, Weinan, Shaanxi, China Designed and delivered math lessons to primary school students in a remote village. Contributed to promoting education equity and rural development through hands-on teaching.
2017.09 – 2019.09	Library Volunteer , Northwestern Polytechnical University Library, Xi'an, Shaanxi, China Assisted in book organization, archiving, and supported daily operations as part of the campus volunteer team.

Hobbies

Sports (Badminton, Cycling, Swimming, Table Tennis), Photography, Travel