

Weijian Zheng

Phone: (317) 370-4736
Homepage: <http://www.weijianzheng.net/>

Email: wzheng@anl.gov / zwj3652006@gmail.com

Research Interests:

Distributed Systems, AI4Science, Parallel Algorithms Design,
Parallel Machine Learning, High-Performance Computing, LLMs

Professional Experiences

Argonne National Laboratory Assistant Computer Scientist, Dec. 2025 - current	Lemont, IL, USA
Argonne National Laboratory Postdoc Appointee, Jan. 2023 - Dec. 2025	Lemont, IL, USA
Oak Ridge National Laboratory Research Assistant, Aug. 2018 - Dec. 2022	Oak Ridge, TN, USA
Oak Ridge National Laboratory Research Intern, May 2018 - Aug. 2018	Oak Ridge, TN, USA
Indiana U Purdue U Indianapolis Graduate Research Assistant, Aug. 2014 - Dec. 2022	Indianapolis, IN, USA
Ball State University Teaching Assistant, June 2013 - July 2013	Muncie, IN, USA

Education

Ph.D. in Computer Science, Purdue University, Dec. 2022	West Lafayette, IN, USA
B.S. in Computer Science (<i>Summa cum Laude</i>), Ball State University, July 2014	Muncie, IN, USA

Publications

- Sumera Anjum*, **Zheng, Weijian**, Rajkumar Kettimuthu, Heng Fan, and Yunhe Feng. ProMCP: Profiling Token Flows and Latency Costs in Model Context Protocol-Based LLM Agents. In *Findings of the Association for Computational Linguistics: ACL 2026*, pages 39476–39487, 2026. *Student advised
- Dali Wang, Chen Wang, Qinglei Cao, Jayesh Krishna, **Zheng, Weijian**, Peter Schwartz, Fengming Yuan, Daniel Ricciuto, Peter Thornton, and Shih-Chieh Kao. Enabling kilometer-scale E3SM land model simulation over North America: A new integrated framework solution. *The International Journal of High Performance Computing Applications*, page 10943420261446003, 2026
- Ziyu Hu*, Jiamin Wang, Zhiqing Zhong, **Zheng, Weijian**, Hemant Sharma, Jun-Sang Park, Peter Kenesei, Antonino Miceli, Zhaorui Zhang, Rajkumar Kettimuthu, et al. FastREI: Fast Rare Event Identification on X-ray Data with Cross-Stage Optimizations. In *2025 IEEE International Conference on Big Data (BigData)*, pages 2169–2176. IEEE, 2025a. *Student advised
- Austin Yunker*, **Weijian Zheng***, and Rajkumar Kettimuthu. InferCT: An Efficient and Generalizable Framework to Enable 3D Machine Learning for Computed Tomography. In *Proceedings of the SC'25 Workshops of the International Conference for High Performance Computing, Networking, Storage and Analysis*, pages 71–77, 2025. *Equal contribution
- Flavio Castro, **Zheng, Weijian**, Joaquin Chung, Ian Foster, and Raj Kettimuthu. To Stream or Not to Stream: Towards A Quantitative Model for Remote HPC Processing Decisions. In *Proceedings of the SC'25 Workshops of the International Conference for High Performance Computing, Networking, Storage and Analysis (best paper award)*, pages 834–840, 2025

6. Ziyu Hu*, Zhiqing Zhong, **Zheng, Weijian**, Zhijing Ye, Xuwei Tan, Xueru Zhang, Zheng Xie, Rajkumar Kettimuthu, and Xiaodong Yu. DABench-LLM: Standardized and In-Depth Benchmarking of Post-Moore Dataflow AI Accelerators for LLMs. In *2025 IEEE International Symposium on Workload Characterization (IISWC)*, pages 127–141. IEEE, 2025b.
*Student advised
7. Dali Wang, Chen Wang, Qinglei Cao, Jayesh Krishna, Danqing Wu, **Zheng, Weijian**, Peter Schwartz, Fengming Yuan, Kathryn Mohror, and Peter Thornton. Scaling Ultrahigh-Resolution E3SM Land Model for Leadership-Class Supercomputers. In *2025 IEEE 25th International Symposium on Cluster, Cloud and Internet Computing Workshops (CC-GridW)*, pages 1–4. IEEE, 2025
8. Hanzhi Zhang, Sumera Anjum, Heng Fan, **Zheng, Weijian**, Yan Huang, and Yunhe Feng. Poly-FEVER: A Multilingual Fact Verification Benchmark for Hallucination Detection in Large Language Models. *arXiv preprint arXiv:2503.16541*, 2025
9. Chengming Zhang, Xinheng Ding, Baixi Sun, Xiaodong Yu, **Zheng, Weijian**, Zhen Xie, and Dingwen Tao. GFormer: Accelerating Large Language Models with Optimized Transformers on Gaudi Processors. *arXiv preprint arXiv:2412.19829*, 2024
10. **Zheng, Weijian**, Hemant Sharma, Ryan Chard, Peter Kenesei, Jun-Sang Park, Nicholas Schwarz, Antonino Miceli, Ian T Foster, and Rajkumar Kettimuthu. Model and Data Management for Machine Learning (M2ML): Integrating Instruments, Edge and HPC for Accelerated Machine Learning. In *2024 IEEE International Conference on Big Data (BigData)*, pages 4275–4282. IEEE, 2024c
11. **Zheng, Weijian**, Jack Kordas, Tyler J Skluzacek, Raj Kettimuthu, and Ian Foster. Globus service enhancements for exascale applications and facilities. *The International Journal of High Performance Computing Applications*, 38(6): 658–670, 2024a
12. **Zheng, Weijian**, J-S Park, Peter Kenesei, Ahsan Ali, Zhengchun Liu, Ian Foster, Nicholas Schwarz, Rajkumar Kettimuthu, Antonino Miceli, and Hemant Sharma. Rapid detection of rare events from in situ X-ray diffraction data using machine learning. *Journal of Applied Crystallography*, 57(4), 2024b
13. Kareem Shaik, Dali Wang, **Zheng, Weijian**, Qinglei Cao, Heng Fan, Peter Schwartz, and Yunhe Feng. S3LLM: Large Scale Scientific Software Understanding with LLMs Using Source, Metadata, and Document. In *International Conference on Computational Science*, pages 222–230. Springer, 2024
14. Shihui Song, Yafan Huang, Peng Jiang, Xiaodong Yu, **Zheng, Weijian**, Sheng Di, Qinglei Cao, Yunhe Feng, Zhen Xie, and Franck Cappello. Ceresz: Enabling and scaling error-bounded lossy compression on cerebras cs-2. In *Proceedings of the 33rd International Symposium on High-Performance Parallel and Distributed Computing*, pages 309–321, 2024
15. Jim Pruyne, Valerie Hayot-Sasson, **Zheng, Weijian**, Ryan Chard, Justin M Wozniak, Tekin Bicer, Kyle Chard, and Ian T Foster. Steering a fleet: Adaptation for large-scale, workflow-based experiments. *arXiv preprint arXiv:2403.06077*, 2024
16. Chengming Zhang, Baixi Sun, Xiaodong Yu, Zhen Xie, **Zheng, Weijian**, Kamil A Iskra, Pete Beckman, and Dingwen Tao. Benchmarking and in-depth performance study of large language models on habana gaudi processors. In *Proceedings of the SC'23 Workshops of the International Conference on High Performance Computing, Network, Storage, and Analysis*, pages 1759–1766, 2023
17. Aniket Tekawade, Viktor Nikitin, Yashas Satapathy, Zhengchun Liu, Xuan Zhang, Peter Kenesei, **Zheng, Weijian**, Francesco De Carlo, Ian Foster, and Rajkumar Kettimuthu. Tomo2Mesh: Fast porosity mapping and visualization for synchrotron tomography. In *2023 IEEE 19th International Conference on e-Science (e-Science)*, pages 1–10. IEEE, 2023

18. Yunhe Feng, Sreecharan Vanam, Manasa Cherukupally, **Zheng, Weijian**, Meikang Qiu, and Haihua Chen. Investigating Code Generation Performance of Chat-GPT with Crowdsourcing Social Data. In *Proceedings of the 47th IEEE Computer Software and Applications Conference*, pages 1–10, 2023
19. **Zheng, Weijian**, Dali Wang, and Fengguang Song. A Distributed-GPU Deep Reinforcement Learning System for Solving Large Graph Optimization Problems. *ACM Transactions on Parallel Computing*, 2023
20. Dali Wang, Peter Schwartz, Fengming Yuan, Peter Thornton, and **Zheng, Weijian**. Towards Ultra-high-resolution E3SM Land Modeling on Exascale Computers. *Computing in Science & Engineering*, 2022
21. Jian Zhou, **Zheng, Weijian**, Dali Wang, and David W Coit. A resilient network recovery framework against cascading failures with deep graph learning. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, page 1748006X221128869, 2022
22. **Zheng, Weijian**, Dali Wang, and Fengguang Song. Designing a parallel Feel-the-Way clustering algorithm on HPC systems. *The International Journal of High Performance Computing Applications (IJHPCA)*, 35(2):154–169, 2021.
23. **Zheng, Weijian**, Dali Wang, and Fengguang Song. OpenGraphGym: A parallel reinforcement learning framework for graph optimization problems. In *International Conference on Computational Science (ICCS)*, pages 439–452. Springer, 2020.
24. **Zheng, Weijian**, Dali Wang, and Fengguang Song. FQL: An extensible feature query language and toolkit on searching software characteristics for HPC applications. In *Tools and Techniques for High Performance Computing (SE-HER)*, pages 129–142. Springer, 2019b.
25. **Zheng, Weijian**, Dali Wang, and Fengguang Song. XScan: An Integrated Tool for Understanding Open Source Community-Based Scientific Code. In *International Conference on Computational Science (ICCS)*, pages 226–237. Springer, 2019a.
26. **Zheng, Weijian**, Fengguang Song, Lan Lin, and Zizhong Chen. Scaling Up Parallel Computation of Tiled QR Factorizations by a Distributed Scheduling Runtime System and Analytical Modeling. *Parallel Processing Letters (PPL)*, 28(01): 1850004, 2018.
27. Dali Wang, **Zheng, Weijian**, and Fengguang Song. Application Software Analytics Toolkit for Facilitating the Understanding, Componentization, and Refactoring of Large-Scale Scientific Models. Technical report, Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States), 2018.
28. **Zheng, Weijian**, Fengguang Song, and Lan Lin. Designing a Synchronization-reducing Clustering Method on Many-cores: Some Issues and Improvements. In *Proceedings of the Machine Learning on HPC Environments (MLHPC)*, page 9. ACM, 2017.
29. **Zheng, Weijian**, Fengguang Song, Lan Lin, and Zizhong Chen. suCAQR: A Simplified Communication-Avoiding QR Factorization Solver Using the TBLAS Framework. In *2016 IEEE 22nd International Conference on Parallel and Distributed Systems (ICPADS)*, pages 1092–1099. IEEE, 2016.

Grants and Awards

Awards

2025 Argonne Postdoctoral Performance Awards – 12 recipients in 2025

Grant Awards

“Multimodal AI Foundation Modeling of Turbulent, Multiphase, and Reacting Flows for Propulsion and Power Applications.”

- Argonne Laboratory Directed Research & Development (LDRD) Prime Funding, 2026,
- \$400,000, Role: Co-PI

“Use of Generative AI and LLMs for Accelerator Design and Optimization.”

- Argonne Laboratory Directed Research & Development (LDRD) AI for Science Project Funding, 2025,
- \$100,000, Role: Investigator

“Towards Fast Quantitative Regression Analysis with Vision Transformers and Bayesian Neural Networks.”

- Argonne Laboratory Directed Research & Development (LDRD) Innovate (Seed) Funding, 2024,
- \$25,000, Role: Sole PI

Computing Resource Awards

“Accelerating 3D Tomography Analysis with High-Performance Computing.”

- ALCF Director’s Discretionary (DD) Allocation, Argonne National Laboratory, 2025,
- 3,000 node-hours on Aurora Supercomputer, Role: PI

“Co-optimize AI Algorithm and Computer System for Accelerating High-resolution Image Generation.”:

- ALCF Director’s Discretionary (DD) Allocation, Argonne National Laboratory, 2024,
- 3,000 node-hours on Polaris and Sophia Supercomputer, Role: PI

“Advanced Applications and Performance Benchmarking of the Cerebras CS-2 in High-Volume Image Processing and Computational Analysis.”

- NSF-funded ACCESS Allocation at Pittsburgh Supercomputing Center (PSC) 2024,
- 400,000 ACCESS credits on Neocortex, Role: PI

“Evaluation of NAS for X-ray applications on next generation AI hardware.”

- ALCF Director’s Discretionary (DD) Allocation, Argonne National Laboratory, 2024,
- AI-testbed (Cerebras, Graphcore, and SambaNova), Role: PI

Professional Activities

Conferences Program Committee Members and Reviewers

- TPC for IEEE Cloud Summit, 2025, 26
- TPC for IEEE Intl. Conf. on Distributed Computing Systems (ICDCS), 2025, 26
- PC Member for Intl. Conf. on Dependable Systems and Networks (DSN), 2025
- Reviewer for Intl. Conf. for High Performance Computing, Networking, Storage, and Analysis (SC), 2025
- Reviewer for ACM Intl. Conf. on Knowledge Discovery and Data Mining (KDD), 2025, 26
- Reviewer for IEEE Intl. Conf. on Cloud Computing (CLOUD), 2024, 2025
- Reviewer for ACM Intl. Conf. on Supercomputing (ICS), 2025
- Reviewer for IEEE Intl. Conf. on Advanced Networks and Telecommunications Systems (ANT), 2024
- Reviewer for Intl. Conf. on Parallel Processing (ICPP), 2024
- Reviewer for European Conf. on Parallel Processing (Euro-Par), 2024, 25
- Reviewer for IEEE Intl. Symposium on Cluster, Cloud, and Internet Computing (CCGRID), 2023

Journal Reviewers

- Reviewer for IEEE Transactions on Parallel and Distributed Systems (TPDS)
- Reviewer for Information Sciences
- Reviewer for IEEE Transactions on Systems, Man, and Cybernetics: Systems

Reviewer for IEEE Transactions on Big Data

Reviewer for IEEE Network Magazine

Reviewer for IEEE Internet of Things Journal

Reviewer for Journal of Signal Processing Systems

Reviewer for Digital Communications and Networks

Last updated: July 17, 2026