Yiwei ZHAO | 赵奕炜

Curriculum Vitae

Carnegie Mellon University

Office: 41D7 Collaborative Innovation Center, 4720 Forbes Ave, Pittsburgh, PA, 15213

Email: yiweiz3@andrew.cmu.edu / yiweizhao@cmu.edu | Website: https://zhaoyw007.github.io/

Last modified in November 2024

RESEARCH INTERESTS

Inter-discipline of: Computer System/Architecture [1] + Algorithm Engineering [2]

[1] System/Architecture: Parallel & Distributed Computing; Heterogeneous-Memory Systems.

[2] Algorithm Engineering: Algorithms & System Engineering for Prevalent Workloads in Databases & MLSys.

EDUCATION

Carnegie Mellon University, Pittsburgh, Pennsylvania, 2021-present.

- Ph.D. candidate.
- Primary Advisor: Prof. Phillip B. Gibbons.

Tsinghua University, Beijing, China, 2017-2021.

- B.E. in Electronic Engineering. Graduated with highest honor.
- Double major in Economics & Finance.
- **Overall GPA**: 4.0/4.0   **Ranking**: Top 3%

HONORS & AWARDS

- Lee-Stanziale Ohana Fellowship (2023 - 2024).
- Qualcomm PhD Fellowship Finalist (2024).
- Best Paper Runner-up, VLDB (2023) [3].
- IBM PhD Fellowship Finalist (2023).
- Carnegie Mellon Institute of Technology Dean's Fellow (2021 - 2022).
- Excellence Honor for Undergraduate, Tsinghua University & Beijing (2021).
- Tsinghua University Fellowship for Undergrads (2017 - 2021).
- Shuping Fellowship for Undergrads (2017 - 2021).
- China National Fellowship for Undergrads (2018 - 2020).

RESEARCH EXPERIENCE

Carnegie Mellon University (Graduate Research Assistant), Pittsburgh, PA, August 2021 – present.

- Advisor: Phillip B. Gibbons.
- Research Topics: Processing-In-Memory; Database Index; Parallel Graph Processing; Parallel Algorithms and Data Structures; Computation Geometry; Vector Databases; Computer Architecture.

Meta Reality Lab (Research Scientist Intern), Redmond, WA, May 2023 – November 2023.

- Host: Chiao Liu, Ziyun Li.
- Research Topic: Architecture & System Design for Edge Devices; Compute-in-memory.

Massachusetts Institute of Technology (Research Assistant), Cambridge, MA, June 2020 – November 2020.

- Advisor: Julian Shun.
- Research Topic: Parallel Spatial Clustering Algorithms & Unsupervised Learning.

Tsinghua University (Undergraduate Research Assistant), Beijing, China, September 2020 – June 2021.
- Advisor: Yongpan Liu.
- Research Topic: Error Tolerant Designs for ReRAM based Compute-In-Memory Neural Network Accelerators.

Tsinghua University (Research Assistant), Beijing, China, January 2019 – Oct 2019.
- Advisor: Dan Pei.
- Research Topic: Fault Localization and Data Mining in Multi-dimensional Data.

Tsinghua University (SRT Research Assistant), Beijing, China, October 2018 – June 2019.
- Advisor: Yuantao Gu.
- Research Topics: Optimization, Statistical Learning, Graph Filtering, Sparse recovery, and NLP.

PUBLICATIONS

In Preparation

Yiwei Zhao, Jinhui Chen, Sai Qian Zhang, Syed Shakib Sarwar, Kleber Hugo Stangherlin, Jorge Tomas Gomez, Barbara De Salvo, Chiao Liu, Phillip B. Gibbons, Ziyun Li. 2025. [Paper title and conference name omitted to keep anonymity. On Neural Architecture Search (NAS) for transformers.]

Yiwei Zhao, Hongbo Kang, Guy E. Blelloch, Yan Gu, Laxman Dhulipala, Charles McGuffey, and Phillip B. Gibbons. 2025. [Paper title and conference name omitted to keep anonymity. On efficient space-partitioning index on emerging hardware.]

Yiwei Zhao, Hongbo Kang, Charles McGuffey, and Phillip B. Gibbons. 2025. [Paper title and conference name omitted to keep anonymity. On efficient task-data orchestration on distributed-memory systems.]

Full Publications

[1] **Yiwei Zhao**, Jinhui Chen, Sai Qian Zhang, Syed Shakib Sarwar, Kleber Hugo Stangherlin, Jorge Tomas Gomez, Jae Sun Seo, Phillip B. Gibbons, Barbara De Salvo, Chiao Liu, Ziyun Li. 2025. "**H4H: Hybrid Convolution-Transformer Architecture Search for NPU-CIM Heterogeneous Systems for AR/VR Applications**". In Proceedings of the 30th Asia and South Pacific Design Automation Conference (**ASPDAC '25**).

[2] Hongbo Kang, **Yiwei Zhao**, Guy E. Blelloch, Laxman Dhulipala, Yan Gu, Charles McGuffey, and Phillip B. Gibbons. 2023. "**PIM-trie: A Skew-Resistant Trie for Processing-in-Memory**". In Proceedings of the 35th ACM Symposium on Parallelism in Algorithms and Architectures (**SPAA '23**). Association for Computing Machinery, New York, NY, USA, pp. 1–14. doi:10.1145/3558481.3591070.

[3] Hongbo Kang, **Yiwei Zhao**, Guy E. Blelloch, Laxman Dhulipala, Yan Gu, Charles McGuffey, and Phillip B. Gibbons. 2022. "**PIM-tree: A Skew-resistant Index for Processing-in-Memory**". In Proceedings of the VLDB Endowment (**PVLDB**), 16(4): 946-958, December 2022. doi:10.14778/3574245.3574275. arXiv:2211.10516. *Best Research Paper Runner-up in VLDB 2023*.

[4] Zeyan Li, Junjie Chen, Yihao Chen, Chengyang Luo, **Yiwei Zhao**, Yongqian Sun, Kaixin Sui, Xiping Wang, Dapeng Liu, Xing Jin, Qi Wang, and Dan Pei. 2023. "**Generic and Robust Root Cause Localization for Multi-Dimensional Data in Online Service Systems**". In Journal of Systems and Software (**JSS**), Vol. 203, (2023), 111748. doi:10.1016/j.jss.2023.111748.

[5] Zeyan Li, Chengyang Luo, **Yiwei Zhao**, Yongqian Sun, Kaixin Sui, Xiping Wang, Dapeng Liu, Xing Jin, Qi Wang, and Dan Pei. 2019. "**Generic and Robust Localization of Multi-Dimensional Root Cause**". In the 30th International Symposium on Software Reliability Engineering (**ISSRE '19**). Oct. 28-31, 2019, Berlin. doi:10.1109/ISSRE.2019.00015.

<u>Short Publications</u>

[6] **Yiwei Zhao**, Ziyun Li, Win-San Khwa, Xiaoyu Sun, Sai Qian Zhang, Syed Shakib Sarwar, Kleber Hugo Stangherlin, Yi-Lun Lu, Jorge Tomas Gomez, Jae Sun Seo, Phillip B. Gibbons, Barbara De Salvo, Chiao Liu. 2024. "**Neural Architecture Search of Hybrid Models for NPU-CIM Heterogeneous AR/VR Devices**". In 61th ACM/IEEE Design Automation Conference (**DAC '24**), Poster Session, San Francisco, CA, USA, 2024. <u>arXiv:2410.08326</u>.

[7] Hongbo Kang, **Yiwei Zhao**, Guy E. Blelloch, Laxman Dhulipala, Yan Gu, Charles McGuffey, and Phillip B. Gibbons. 2023. "**PIM-tree: A Skew-resistant Index for Processing-in-Memory (Abstract)**". In Proceedings of the 2023 ACM Workshop on Highlights of Parallel Computing (**HOPC '23**), June 16, 2023, Orlando, FL, USA. <u>doi:10.1145/3597635.3598029</u>.

TEACHING EXPERIENCES

- **18-751 Applied Stochastic Processes, with Applications to AI/ML**: Teaching Assistant, Fall 2024, CMU.
- **18-742 Computer Architecture and Systems**: Teaching Assistant, Spring 2024, CMU.

SERVICES

- **Student Council for Departmental Faculty Hiring**: Chair, Jan 2024 – present, CMU.
- **Student Council for Departmental Faculty Hiring**: Member, Jan 2023 – Jan 2024, CMU.

SKILLS

- **Programming Languages**: C/C++, Java, Python, Shell Scripting, and Assembly.
- **Software Development Tools**: MATLAB, Git, SQL, and R.
- **Hardware Design Tools**: Verilog (HDL), Gem5, McPAT, ZSim, Multisim, and ADS.
- **Machine Learning Frameworks**: TensorFlow, PyTorch, and CUDA.
- **Other Software Tools**: Mathematica, Latex, AutoCAD, and Stata.