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 September 2025

RESEARCH INTERESTS

Mainly in High-Performance Computing.
Inter-discipline of: Algorithmic Theory + Computer System Design + Computer Architecture.

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**: 3.9/4.0 **Ranking**: Top 3%

HONORS & AWARDS

- Michel and Kathy Doreau Graduate Fellowship (2024 - present).
- Lee-Stanziale Ohana Fellowship (2023 - 2024).
- Qualcomm PhD Fellowship Finalist (2024).
- Best Paper Runner-up, VLDB (2023) [5].
- Carnegie Mellon Institute of Technology Dean's Fellow (2021 - 2022).
- Honored Graduate, 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 (Research Scientist Intern), Redmond, WA, May 2025 – August 2025.

- Host: Chiao Liu, Ziyun Li, Barbara De Salvo.
- Research Topic: Efficient Multimodal Learning.

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

- Host: Chiao Liu, Ziyun Li, Sai Qian Zhang.
- Research Topic: Architecture & System Design for Edge Devices.

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 Accelerators.

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

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

Tsinghua University (Research Assistant), Beijing, China, January 2019 – May 2020.

- Advisor: Dan Pei.
- Research Topic: Fault Localization and Data Mining in Multi-dimensional Data.

PUBLICATIONS

In Preparation

Yiwei Zhao, Hongbo Kang, Ziyang Men, Yan Gu, Guy E. Blelloch, Laxman Dhulipala, Charles McGuffey, and Phillip B. Gibbons. 2025. [Under review. On efficient space-partitioning index on emerging hardware.]

Yiwei Zhao, Qiushi Lin, Hongbo Kang, Guy E. Blelloch, Laxman Dhulipala, Charles McGuffey, and Phillip B. Gibbons. 2025. [Under Review. On task-data orchestration for distributed systems, and distributed graph processing system.]

Hongbo Kang, Xiangyun Ding, Yingdi Shan, **Yiwei Zhao**, Guy E. Blelloch, Laxman Dhulipala, Yan Gu, Charles McGuffey, Mingxing Zhang, Yongwei Wu, and Phillip B. Gibbons. 2025. [Under review. On parallel learned indexes.]

Yiwei Zhao, Qiushi Lin, Hongbo Kang, Guy E. Blelloch, Laxman Dhulipala, and Phillip B. Gibbons. 2025. [In preparation. On architectural supports for emerging hardware.]

Yiwei Zhao, Yi Zheng, Jieyu Lin, Cijo Jose, Michael Ramamonjisoa, Patrick Labatut, Barbara De Salvo, Chiao Liu, Phillip B. Gibbons, Ziyun Li. 2025. [In preparation. On efficient inference for multimodal foundational models.]

Full Publications

[1] **Yiwei Zhao**, Hongbo Kang, Yan Gu, Guy E. Blelloch, Laxman Dhulipala, Charles McGuffey, and Phillip B. Gibbons. 2025. "**Optimal Batch-Dynamic kd-trees for Processing-In-Memory with Applications**". In Proceedings of the 37th ACM Symposium on Parallelism in Algorithms and Architectures (**SPAA** '25). Association for Computing Machinery, New York, NY, USA, 350–366. doi:10.1145/3694906.3743318.

[2] Hyoungjoo Kim, **Yiwei Zhao**, Andrew Pavlo, and Phillip B. Gibbons. 2025. "**No Cap, This Memory Slaps: Breaking Through the Memory Wall of Transactional Database Systems with Processing-in-Memory**". In Proceedings of the VLDB Endowment (**PVLDB**), 18(11): 4241-4254, July 2025. doi:10.14778/3749646.3749690.

[3] **Yiwei Zhao**, Jinhui Chen, Sai Qian Zhang, Syed Shakib Sarwar, Kleber Hugo Stangherlin, Jorge Tomas Gomez, Jae Sun Seo, Barbara De Salvo, Chiao Liu, Phillip B. Gibbons, 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). Association for Computing Machinery, New York, NY, USA, 1133–1141. doi:10.1145/3658617.3697627.

[4] 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.

[5] 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***.

[6] 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.

[7] 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.

Short Publications & Workshops

[8] **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 (Abstract)**". In Proceedings of the 3rd Highlights of Parallel Computing Workshop (**HOPC** '25), July 28, 2025, Portaland, OR, USA. doi:10.1145/3746238.3746241.

[9] **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. arXiv:2410.08326.

[10] 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. doi:10.1145/3597635.3598029.

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

- **37th ACM Symposium on Parallelism in Algorithms and Architectures** (**SPAA'25**): Junior Program Committee.
- **2025 International Conference on Parallel Architectures and Compilation Techniques** (**PACT'25**): External Review Committee.
- **Student Council for Departmental Faculty Hiring**: Chair, January 2024 – June 2025, CMU.
- **Student Council for Departmental Faculty Hiring**: Member, January 2023 – January 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.