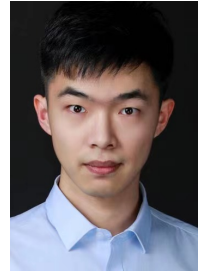


Tong Zhao, Ph.D.

✉ tongzhao.zhaot@gmail.com

☎ +86-19975519969



Academic Positions

2021 – now **Assistant Professor**, Computer Science Department, University of Chinese Academy of Sciences & High Performance Computer Research Center, Institute of Computing Technology, Chinese Academy of Sciences.

Education

2016 – 2021 **Ph.D. School of Mathematical Sciences, Fudan University**, Stochastic Control and Operation Research (supervisor: Shanjian Tang).
Thesis: *Interacting partical systems and deep learning*.

2012 – 2016 **B.Sc. School of Mathematical Sciences, Fudan University**, Mathematics (supervisor: Weiguo Gao).

Research Publications








- 1 G. Feng, H. Wang, Z. Guo, M. Li, T. Zhao, Z. Jin, W. Jia, G. Tan, and N. Sun, “Accelerating large-scale sparse lu factorization for rf circuit simulation,” in *European Conference on Parallel Processing*, Springer, 2024, pp. 182–195.
- 2 S. Hu, T. Zhao, Q. Sha, E. Li, X. Meng, L. Liu, L.-W. Wang, G. Tan, and W. Jia, “Training one deepmd model in minutes: A step towards online learning,” in *Proceedings of the 29th ACM SIGPLAN Annual Symposium on Principles and Practice of Parallel Programming*, 2024, pp. 257–269.
- 3 Y.-J. Yan, H.-B. Li, T. Zhao, L.-W. Wang, L. Shi, T. Liu, G.-M. Tan, W.-L. Jia, and N.-H. Sun, “10-million atoms simulation of first-principle package ls3df,” *Journal of Computer Science and Technology*, vol. 39, no. 1, pp. 45–62, 2024.
- 4 S. Hu, W. Zhang, Q. Sha, F. Pan, L.-W. Wang, W. Jia, G. Tan, and T. Zhao, “Rlekf: An optimizer for deep potential with ab initio accuracy,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, 2023, pp. 7910–7918.
- 5 J. Li, T. Zhao, Z. Guo, S. Shi, L. Liu, G. Tan, W. Jia, G. Yuan, and Z. Wang, “Enhance the strong scaling of lammps on fugaku,” in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, 2023, pp. 1–13.
- 6 T. Zhao, “Limits of one-dimensional interacting particle systems with two-scale interaction,” *Chinese Annals of Mathematics, Series B*, vol. 43, no. 2, pp. 195–208, 2022.

Preprint & Working Paper

- 1 Z. Chen, T. Zhao, and A. C. Zhou, *Interference-aware i/o scheduling*, working paper, 2024.
- 2 T. Zhao, *Momentum is generalization*, working paper, 2024.

- 3 T. Zhao, L. Jiacheng, G. Tan, and W. Jia, *Stack normalization*, working paper, 2024.
- 4 T. Zhao, X. Yang, Q. Sha, G. Tan, and W. Jia, *Kalman optimizer*, working paper, 2024.
- 5 T. Zhao, L. Jiacheng, Z. Yuanchang, G. Tan, and W. Jia, *Alto: Acceleration large-batch training optimizer*, Submitted, 2023.
- 6 T. Zhao, Y. Zhou, E. Shao, G. Tan, and W. Jia, *Bellman: An high throughput co-running scheduler for hpc and ai hybrid workloads*, preprint, 2023.
- 7 H. Li, G. Tan, and T. Zhao, *Backward error analysis of the lanczos bidiagonalization with reorthogonalization*, arXiv preprint arXiv:2210.10297, 2022.

Awards

- 2023  AAAI23 Oral.
- 2022  Outstanding Researcher at the Institute of Computing Technology, Chinese Academy of Sciences, (top 2%).
- 2020  Guanghua Scholarship, (top 2%).
- 2019  Tongyu Cai Scholarship, (top 2%).
- 2018  Scholarship for International Congress of Mathematicians, (top 1%).
- 2012  Gold medal in China High School Mathematics Olympiad, Inner Mongolia, China.
 Gold medal in China High School Physics Olympiad, Inner Mongolia, China.

Research Interests & Projects

Primary: Optimization in Deep Learning

1. I have studied how momentum affects generalization and its relationship with the sharpness of the landscape.
2. I explored why the training performance of deep neural networks deteriorates with larger batch sizes, and how momentum can be introduced to improve the performance. I designed adaptive momentum methods to enhance training with large batch sizes.
3. Many optimizers perform well on small models (parameters<100M), but their performance lags behind SGD on large models (parameters>100M), especially considering the memory and communication constraints unique to large-scale models. How can we design efficient and practical optimizers for large-scale models?
4. Kalman filtering is a classical algorithm in control domain, I adapt it to a novel optimizer. This optimizer has shown promising results in molecular potential surface fitting, earning praise from experts in the field. Moreover, this algorithm has been incorporated into the DeePMD library for molecular dynamics simulations.
5. I have also dabbled in training neural networks for PDEs (Partial Differential Equations). During my PhD in the Mathematics Department at Fudan University, I researched stochastic control and operations, focusing on stochastic partial differential equations and how to solve forward-backward stochastic differential equations and corresponding PDEs by deep learning.

Secondary: Reinforcement Learning

1. I designed a co-running scheduler for AI training tasks using reinforcement learning algorithms, fully utilizing the large scheduling window to achieve better performance than traditional (dynamic) schedulers.
2. I developed a distributed I/O scheduler, especially for training tasks, with reinforcement learning algorithms. It is an interference-aware scheduler between multiple reads and writes, using a neural network to predict the real bandwidth under interference.

Advantages for Why I am Suitable for our Group

Research interests align with the group's direction.

I am very interested in designing efficient optimizers for deep learning, including design, implementation, theoretical analysis, and experimentation on various pragmatic tasks independently. I have several papers on this topic, some of which are currently under review. Besides, I am currently working on developing a testing and analysis platform for the optimizer. This aligns perfectly with the research direction of our group.

Curiosity, Adaptability, and Self-Motivation.

I am capable of working independently and eager to learn new knowledge, even willing to adopt a doctoral student mindset. I aim to contribute to highly original research. Additionally, I am open to discussing ideas with group members and assisting in guiding PhD students.

Computational Resources.

Due to my professional experience in the High Performance Computer Research Center, I have established good collaborative relationships with some leading supercomputing companies, internet companies, as well as distributed training centers in universities and research institutes. If necessary, I can provide the required computational resources for our group.

Relevant Theoretical Foundation.

I obtained my Ph.D. in the School of Mathematical Sciences, with strong algorithm analysis skills and the ability to learn new theories. My background covers deep learning optimizers, reinforcement learning, stochastic analysis, control and game theory, distributed training, system scheduling, and PDEs. This not only aligns well with the group's expertise but also allows me to provide some support for our research.

Referees

Prof. Shanjian Tang

My Ph.D. supervisor,
Fudan University,
sjtang@fudan.edu.cn

Prof. Weinan E

Expert in our field who appreciates my work,
Princeton University,
weinan@math.princeton.edu

Prof Weile Jia

Principal Investigator of our research group,
High Performance Computer Research Center, Institute of Computing Technology, Chinese Academy of Sciences.
weilejia@ict.ac.cn