

## Zhao Zhang

---

Room 3.226, Advanced Computing Building  
10100 Burnet Road  
Austin, Texas 78758  
zzhang@tacc.utexas.edu

**CURRENT POSITION**      **Texas Advanced Computing Center**      July 2016 - present  
Research Associate, Data Intensive Computing Group

**EDUCATION**      **University of Chicago**      September 2009 - June 2014  
*Ph.D.* Computer Science

**University of Chicago**      January 2007 - December 2007  
*M.S.* Computer Science

**Beijing University of Posts and Telecommunications** August 2002 - June 2006  
*B.E.* Software Engineering

**EXPERIENCE**      **Postdoc Researcher and Data Science Fellow**      July 2014 - June 2016  
AMPlab and Berkeley Institute for Data Science      Berkeley, CA  
Advisor: Michael J. Franklin

**Graduate Student Researcher**      September 2009 - June 2014  
Department of Computer Science, University of Chicago      Chicago, IL  
Advisor: Ian T. Foster

**Scientific Application Specialist**      January 2008 - July 2009  
Computation Institute, University of Chicago      Chicago, IL

**Senior Technical Associate**      July 2006 - December 2006  
Lucent Technology      Beijing, China

### ACTIVE PROJECTS

- PI, NSF OAC-2106661 “Collaborative Research: OAC Core: ScaDL: New Approaches to Scaling Deep Learning for Science Applications on Supercomputers” (10/1/21-9/30/24)
- TACC PI, NSF OAC-2112606 “AI Institute for Intelligent CyberInfrastructure with Computational Learning in the Environment (ICICLE)” (11/1/21-10/31/26)
- PI, NSF OAC-2008388 “Collaborative Research: OAC Core: Small: Efficient and Policy-driven Burst Buffer Sharing” (10/1/20-9/30/22)
- co-PI, NSF OAC-1931537 “Collaborative Research: Frameworks: Designing Next-Generation MPI Libraries for Emerging Dense GPU Systems” (11/1/19-10/31/22)

## CONFERENCE PUBLICATIONS

- [SC'21] J. G. Pauloski, Q. Huang, L. Huang, K. Chard, I. T. Foster, **Z. Zhang**. KAISA: An Adaptive Second-order Optimizer Framework for Deep Neural Networks. *to appear in Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis, 2021 (SC)*.
- [ISC'21] Evans, Richard Todd, Matthew Cawood, Stephen Lien Harrell, Lei Huang, Si Liu, Chun-Yaung Lu, Amit Ruhela, Yinzhi Wang, and **Zhao Zhang**. "Optimizing GPU-Enhanced HPC System and Cloud Procurements for Scientific Workloads." In International Conference on High Performance Computing, pp. 313-331. Springer, Cham, 2021.
- [SC'20] J. G. Pauloski, **Z. Zhang**, L. Huang, W. Xu, I. T. Foster. Convolutional Neural Network Training with Distributed K-FAC. *In Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis, 2020 (SC)*.
- [IPDPS'20] **Z. Zhang**, L. Huang, J. G. Pauloski, I. T. Foster. Efficient I/O for Neural Network Training with Compressed Data. *In 2020 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*.
- [CLUSTER'19] **Z. Zhang**, L. Huang, R. Huang, W. Xu, D. S. Katz. Quantifying the Impact of Memory Errors in Deep Learning. *In Proceedings of 2019 IEEE International Conference on Cluster Computing (CLUSTER), p.1-12, 2019*.
- [ICPP'18] Y. You, **Z. Zhang**, J. Demmel, K. Keutzer, C. Hsieh. ImageNet Training in Minutes. *In Proceedings of the 47th International Conference on Parallel Processing, p. 0. ACM, 2018. Best Paper*
- [HPDC'17] **Z. Zhang**, E. Sparks, M. J. Franklin. Diagnosing Machine Learning Pipelines with Fine-grained Lineage. *In Proceedings of the 26th international symposium on High-performance parallel and distributed computing (HPDC' 17)*
- [IPDPS'16] M. Turilli, F. Liu, **Z. Zhang**, A. Merzky, M. Wilde, J. Weissman, D. S. Katz, and S. Jha. Integrating abstractions to enhance the execution of distributed applications. *In 2016 IEEE International Parallel and Distributed Processing Symposium (IPDPS), pp. 953-962. IEEE, 2016*
- [CIDR'15] D. Crankshaw, P. Bailis, J. E. Gonzalez, H. Li, **Z. Zhang**, M. J. Franklin, A. Ghodsi, M. I. Jordan. The Missing Piece in Complex Analytics: Low Latency, Scalable Model Management and Serving with Velox. *7th Biennial Conference on Innovative Data Systems Research (CIDR), 2015*.
- [SIGMOD'15] F. A. Nothaft, M. Massie, T. Danford, **Z. Zhang**, U. Laserson, C. Yeksigian, J. Kottalam et al. Rethinking data-intensive science using scalable analytics systems. *In Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, pp. 631-646. ACM, 2015*.
- [eScience'14] **Z. Zhang**, D. S. Katz. Using Application Skeletons to Improve eScience Infrastructure. *In e-Science (e-Science), 2014 IEEE 10th International Conference on, vol. 1, pp. 111-118. IEEE, 2014*.
- [SC'13] **Z. Zhang**, D. S. Katz, T. Armstrong, J. Wozniak, I. Foster. Parallelizing the Execution of Sequential Scripts.

*In Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis (Supercomputing), p. 31. ACM, 2013.*

- [IPDPS'13] T. Li,, X. Zhou, K. Brandstatter, D. Zhao, K. Wang, A. Rajendran, **Z. Zhang**, and I. Raicu.  
ZHT: A light-weight reliable persistent dynamic scalable zero-hop distributed hash table.  
*In 2013 IEEE 27th International Symposium on Parallel and Distributed Processing, pp. 775-787. IEEE, 2013.*
- [HPDC'12] **Z. Zhang**, D. S. Katz, M. Wilde, J. Wozniak, I. Foster.  
MTC Envelope: Defining the Capability of Large Scale Computers in the Context of Parallel Scripting Applications.  
*In Proceedings of the 22nd international symposium on High-performance parallel and distributed computing (HPDC), pp. 37-48. ACM, 2013.*
- [SC'12] **Z. Zhang**, D. S. Katz, J. Wozniak, A. Espinosa, I. Foster.  
Design and Analysis of Data Management in Scalable Parallel Scripting.  
*In Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis (Supercomputing), p. 85. IEEE Computer Society Press, 2012.*
- [SC'08] I. Raicu, **Z. Zhang**, M. Wilde, I. Foster, P. Beckman, K. Iskra, B. Clifford.  
Toward Loosely Coupled Programming on Petascale Systems.  
*In Proceedings of the 2008 ACM/IEEE conference on Supercomputing, p. 22. IEEE Press, 2008.*

## JOURNAL PUBLICATIONS

- [TPDS'22] J. G. Pauloski, L. Huang, W. Xu, I. T. Foster, **Z. Zhang**.  
Deep Neural Network Training with Distributed K-FAC. *to appear in IEEE Transactions on Parallel and Distributed Systems (TPDS), 2022*
- [Nature Methods'21] Fang, Linjing, Fred Monroe, ..., **Z. Zhang**, et al.  
Deep learning-based point-scanning super-resolution imaging.  
*In Nature Methods (2021): 1-11.*
- [TPDS'19] Y. You, **Z. Zhang**, J. Demmel, K. Keutzer, C. Hsieh.  
Fast Deep Neural Network Training on Distributed Systems and Cloud TPUs  
*IEEE Transactions on Parallel and Distributed Systems 30, no. 11 (2019): 2449-2462.*
- [TBD'16] **Z. Zhang**, K. Barbary, F. A. Nothaft, E. Sparks, O. Zahn, M. J. Franklin, D. A. Patterson, S. Perlmutter.  
Kira: Processing Astronomy Imagery Using Big Data Technology.  
*IEEE Transactions on Big Data, Volume: 3. 2016*
- [CCPE'16] T. Li, X. Zhou, K. Wang, D. Zhao, I. Sadooghi, **Z. Zhang**, and I. Raicu.  
A convergence of key-value storage systems from clouds to supercomputers.  
*Concurrency and Computation: Practice & Experience 28, no. 1 (2016): 44-69.*
- [FGCS'15] D. S. Katz, A. Merzky, **Z. Zhang**, S. Jha.  
Application Skeletons: Construction and Use in eScience.  
*In Press, Future Generation Computer Systems (FGCS), 2015.*
- [CLUSTER'10] I. Raicu, I. Foster, M. Wilde, **Z. Zhang**, K. Iskra, P. Beckman, Y. Zhao et al.  
Middleware support for many-task computing.  
*Cluster Computing 13, no. 3 (2010): 291-314.*

- [COMPUTER'09] M. Wilde, I. Foster, K. Iskra, P. Beckman, **Z. Zhang**, A. Espinosa, M. Hategan, B. Clifford, and I. Raicu.  
Parallel scripting for applications at the petascale and beyond.  
*Computer 42, no. 11 (2009): 50-60.*

## ACADEMIC TALKS

- Efficient and Scalable Deep Learning on Supercomputers, EECS Seminar, UC Merced, September 2020, Zoom
- Enabling Scalable and Efficient Deep Learning on Supercomputers, Guest Lecture, UChicago, Nov 2018, Chicago, IL
- Enabling Scalable and Efficient Deep Learning on Supercomputers, Argonne National Laboratory, Nov 2018, Lemont, IL
- Scientific Computing Meets Big Data Technology: An Astronomy Use Case, Big Data, October 2015, San Jose, CA
- Flexible Astronomy Image Processing in Clouds, SLAC National Accelerator Laboratory, Stanford University, July 2015, Palo Alto, CA
- Kira: Flexible Astronomy Image Processing in Clouds, AMP Retreat, May 2015, Santa Cruz, CA
- Parallelizing the Execution of Sequential Scripts, SC13, November 2013, Denver, CO
- Enabling Parallel Scripting on Large Scale Computers, AMPLab, UC Berkeley, August 2013, Berkeley, CA
- MTC Envelope: Defining the Capability of Large Scale Computers in the Context of Parallel Scripting Applications, HPDC13, June 2013, New York, NY
- Parallel Programming on Clouds, Grids, and Supercomputers, Center for Earth System Science, Tsinghua University, December 2012, Beijing, China
- Design and Analysis of Data Management in Scalable Parallel Scripting, SC12, November 2012, Salt Lake City, UT

## PROFESSIONAL SERVICE

- **Technical Program Committee Member**, Machine Learning and HPC Track, SC'20, SC'21.
- **Co-chair**, Deep Learning on Supercomputers Workshop Series, SC and ISC, 2018-2021.
- **Review Committee**, Learning and AI Panel for INCITE, DOE, 2020, 2021.
- **Technical Program Committee Member**, International Symposium on Big Data Computing, 2014
- **Guest Co-editor**, Future Generation Computer Systems Special Issue on eScience Applications and Infrastructure, 2014.
- **Organizer**, Weekly System Research Seminar, UChicago Systems Group, 2010-2014.
- **Publicity Chair**, 5th Workshop on Many-Task Computing on Grids and Supercomputers (MTAGS) 2012, November 2012, Salt Lake City, UT.
- **Proceedings Chair**, IEEE International Conference on eScience, October 2012, Chicago, IL.
- **Organizer**, 1st Greater Chicago Area System Research Workshop, May 2012, Chicago, IL.

## TUTORIALS

- Tools and Best Practices for Distributed Deep Learning on Supercomputer, Tutorial, SC'18, SC'19, SC'20, SC'21
- Distributed Deep Learning, Tutorial, TACC Machine Learning Institute, 2018, 2019, 2020
- Introduction to ML/DL and its Applications in Natural Hazard, DesignSafe Bootcamp, 2020
- ML4GEO training for Petrobras, 2018
- Spark Internals, TACC Machine Learning Institute, 2017

## STUDENTS

- Qi Huang, 09/20-present, M.S. Student in CS at UT Austin
- Ishank Arora, 09/20-present, B.S. Student in CS at UT Austin
- J. Gregory Pauloski, 05/19-08/20, now a Ph.D Student in CS at UChicago

## REFERENCE

- **Niall Gaffney**  
Director of Data Intensive Computing, Texas Advanced Computing Center  
Email: ngaffney@tacc.utexas.edu
- **Michael J. Franklin**  
Chair, Department of Computer Science, UChicago  
Email: mjfranklin@cs.uchicago.edu
- **Saul Perlmutter**  
Director, Berkeley Institute for Data Science, UC Berkeley  
Franklin W. and Karen Weber Dabby Professor of Physics, UC Berkeley  
Email: saul@lbl.gov
- **Ian T. Foster**  
Director, Computation Institute, UChicago & ANL  
Arthur Holly Compton Distinguished Service Professor of Computer Science, UChicago  
Email: foster@anl.gov
- **Daniel S. Katz**  
Assistant Director, Scientific Software and Applications, National Center for Supercomputing Applications (NCSA)  
Email: d.katz@ieee.org