Sian Jin

Assistant Professor

Temple University
Department of Computer & Information Sciences
sian.jin@temple.edu
www.sianjin.com

EDUCATION

Indiana University, Bloomington, IN   Aug. 2022 – Dec. 2023
Ph.D. in Computer Engineering
Thesis: advanced data movement techniques for HPC applications via lossy compression
Thesis Advisor: Dr. Dingwen Tao. Previous institutions:

    Washington State University, Pullman, WA   Aug. 2020 – Jul. 2022
    The University of Alabama, Tuscaloosa, AL   Aug. 2018 – Aug. 2020

Beijing Normal University, Beijing, China   Sep. 2014 - Jun. 2018
B.S. in Physics

RESEARCH INTERESTS

- Scientific data analytics & management
- Data reduction & lossy compression
- Parallel file system & I/O system

ACADEMIC EXPERIENCE

Temple University (Department of Computer & Information Sciences)   Jan. 2024 – Present
Assistant Professor

Indiana University (HiPDAC lab)   Sep. 2018 – Dec. 2023
Graduate Research Assistant

- Exploiting deep neural network (DNN) compression to significantly reduce model size for resource-limited scenarios, such as memory, I/O, network.
- Implementing parallel lossy compression algorithms on GPUs and FPGAs.
- Reducing GPU memory usage & improving distributed DNN training performance via data compression
- Developing online analysis & optimization tool for lossy compression in HPC applications.
- Designing compiler-assisted program analysis tool to understand impact of lossy compression and guide user to select the best-fit compression configuration.

Beijing Normal University, School of Physics   May 2016 – May 2017
Undergraduate Research Assistant

- Studying Zitterbewegung effect in graphite with scientific simulations.

PUBLICATION

- **[EuroSys'24] Sian Jin**, Sheng Di, Frédéric Vivien, Daoce Wang, Yves Robert, Dingwen Tao, and Franck Cappello."Concealing Compression-accelerated I/O for HPC Applications through In Situ Task Scheduling." *Proceedings of the Nineteenth European Conference on Computer Systems*, Athens, Greece, April 22–25, 2024.
- **[SC'22] Sian Jin**, Dingwen Tao, Houjun Tang, Sheng Di, Suren Byna, Zarija Lukic, and Franck Cappello."Accelerating Parallel Write via Deeply Integrating Predictive Lossy Compression with HDF5." *The International Conference for High Performance Computing, Networking, Storage, and Analysis*, Dallas, Texas, USA, Nov. 13—18, 2022.

- **[VLDB'22] Sian Jin**, Chengming Zhang, Xintong Jiang, Yunhe Feng, Hui Guan, Guanpeng Li, Shuaiwen Leon Song, and Dingwen Tao. "COMET: A Novel Memory-Efficient Deep Learning Training Framework by Using Error-Bounded Lossy Compression." *ACM International Conference on Very Large Data Bases*, Sydney, Australia, Sep. 5–9, 2022.

- **[ICDE'22] Sian Jin**, Di Sheng, Jiannan Tian, Suren Byna, Dingwen Tao and Franck Cappello. "Significantly Improving Prediction-Based Lossy Compression Via Ratio-Quality Modeling." *IEEE International Conference on Data Engineering*, Worldwide online event, May 9–12, 2022.

- **[HPDC'21] Sian Jin**, Jesus Pulido, Pascal Grosset, Jiannan Tian, Dingwen Tao, and James Ahrens. "Adaptive Configuration of In Situ Lossy Compression for Cosmology Simulations via Fine-Grained Rate-Quality Modeling." *ACM International Symposium on High-Performance Parallel and Distributed Computing*, Worldwide online event, June 21–25, 2021.

- **[PPoPP'21] Sian Jin**, Guanpeng Li, Shuaiwen Leon Song, and Dingwen Tao. "POSTER: A Novel Memory-Efficient Deep Learning Training Framework via Error-Bounded Lossy Compression." *ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming*. Worldwide online event, Feb. 27–Mar. 3, 2021.

- **[IPDPS'20] Sian Jin**, Pascal Grosset, Christopher M. Biwer, Jesus Pulido, Jiannan Tian, Dingwen Tao, and James Ahrens. "Understanding GPU-Based Lossy Compression for Extreme-Scale Cosmological Simulations." *IEEE International Parallel & Distributed Processing Symposium*, New Orleans, Louisiana, USA, May 18–22, 2020.

- **[HPDC'19] Sian Jin**, Sheng Di, Xin Liang, Jiannan Tian, Dingwen Tao, and Franck Cappello. "DeepSZ: A Novel Framework to Compress Deep Neural Networks by Using Error-Bounded Lossy Compression." *ACM International Symposium on High-Performance Parallel and Distributed Computing*, Phoenix, Arizona, USA, June 24–28, 2019.

- **[SIGMOD'24]** Jinyang Liu, Sheng Di, Kai Zhao, Xin Liang, **Sian Jin**, Zizhe Jian, Jiajun Huang, Shixun Wu, Zizhong Chen, and Franck Cappello. "High-performance Effective Scientific Error-bounded Lossy Compression with Auto-tuned Multi-component Interpolation." *The ACM Special Interest Group on Management of Data*, Santiago, Chile, June 9—15, 2024.

- **[BigData'23]** Jinyang Liu, Sheng Di, **Sian Jin**, Kai Zhao, Xin Liang, Zizhong Chen, and Franck Cappello. "Scientific Error-bounded Lossy Compression with Super-resolution Neural Networks." *2023 IEEE International Conference on Big Data*, Sorrento, Italy, December 15—18, 2023.

- **[SC'23]** Daoce Wang, Jesus Pulido, Pascal Grosset, Jiannan Tian, **Sian Jin**, Houjun Tang, Jean Sexton, Sheng Di, Zarija Lukić, Kai Zhao, Bo Fang, Franck Cappello, James Ahrens, and Dingwen Tao. "AMRIC: A Novel In Situ Lossy Compression Framework for Efficient I/O in Adaptive Mesh Refinement Applications." *The International Conference for High Performance Computing, Networking, Storage, and Analysis*, Denver, Colorado, USA, November 12—17, 2023.

- **[TPDS'23]** Haoyu Jin, Donglei Wu, Shuyu Zhang, Xiangyu Zou, **Sian Jin**, Dingwen Tao, Qing Liao, and Wen Xia. "Design of a Quantization-based DNN Delta Compression Framework for Model Snapshots and Federated Learning." *IEEE Transactions on Parallel and Distributed Systems*, Volume 23.

- **[TPDS'22]** Yuanjian Liu, Sheng Di, Kai Zhao, **Sian Jin**, Cheng Wang, Kyle Chard, Dingwen Tao, Ian Foster, and Franck Cappello. "Optimizing Error-Bounded Lossy Compression for Scientific Data with Diverse Constraints." *IEEE Transactions on Parallel and Distributed Systems*, Volume 22.

- **[ICS'22]** Chengming Zhang, **Sian Jin**, Tong Geng, Jiannan Tian, Ang Li, and Dingwen Tao. "CEAZ: Accelerating Parallel I/O via Hardware-Algorithm Co-Designed Adaptive Lossy Compression." *ACM International Conference on Supercomputing*, Worldwide online event, June 27–30, 2022.

- **[HPDC'22]** Daoce Wang, Jesus Pulido, Pascal Grosset, **Sian Jin**, Jiannan Tian, James Ahrens, and Dingwen Tao. "Optimizing Error-Bounded Lossy Compression for Three Dimensional Adaptive Mesh Refinement Simulations." *ACM International Symposium on High-Performance Parallel and Distributed Computing*, Minneapolis, Minnesota, USA, June 27–July 1, 2022.

- **[HiPC'21]** Yuanjian Liu, Sheng Di, Kai Zhao, **Sian Jin**, Cheng Wang, Kyle Chard, Dingwen Tao, Ian Foster, and Franck Cappello. "Optimizing Multi-Range based Error-Bounded Lossy Compression for Scientific

Datasets." *The IEEE International Conference On High Performance Computing, Data, & Analytics*, Bengaluru, India, Dec. 17–18, 2021.

- **[Cluster'21]** Bo Fang, Daoce Wang, **Sian Jin**, Quincey Koziol, Zhao Zhang, Qiang Guan, Suren Byna, Sriram Krishnamoorthy, Dingwen Tao."Characterizing Impacts of Storage Faults on HPC Applications: A Methodology and Insights." *The IEEE Cluster Conference*, Worldwide online event, September 7–10, 2021.
- **[Cluster'21]** Jinyang Liu, Sheng Di, Kai Zhao, **Sian Jin**, Dingwen Tao, Xin Liang, Zizhong Chen, Franck Cappello."Exploring Autoencoder-Based Error-Bounded Compression for Scientific Data." *The IEEE Cluster Conference*, Worldwide online event, September 7–10, 2021.
- **[Cluster'21]** Jiannan Tian, Sheng Di, Xiaodong Yu, Cody Rivera, Kai Zhao, **Sian Jin**, Yunhe Feng, Xin Liang, Dingwen Tao, Franck Cappello."cuSZ (x): Optimizing Error-Bounded Lossy Compression for Scientific Data on GPUs." *The IEEE Cluster Conference*, Worldwide online event, September 7–10, 2021.
- **[ICS'21]** Chengming Zhang, Geng Yuan, Wei Niu, Jiannan Tian, **Sian Jin**, Donglin Zhuang, Zhe Jiang, Yanzhi Wang, Bin Ren, Shuaiwen Leon Song, and Dingwen Tao. "ClickTrain: Efficient and Accurate End-to-End Deep Learning Training via Fine-Grained Architecture-Preserving Pruning." *ACM International Conference on Supercomputing*, Worldwide online event, June 14–17, 2021.
- **[PACT'20]** Jiannan Tian, Sheng Di, Kai Zhao, Cody Rivera, Megan Hickman Fulp, Robert Underwood, **Sian Jin**, and others. "cuSZ: An Efficient GPU-Based Error-Bounded Lossy Compression Framework for Scientific Data." *Proceedings of the ACM International Conference on Parallel Architectures and Compilation Techniques*. Virtual, October 2–7, 2020.
- **[ICPP'20]** Zhenbo Hu, Xiangyu Zou, Wen Xia, **Sian Jin**, Dingwen Tao, Yang Liu, Weizhe Zhang, and Zheng Zhang. "Delta-DNN: Efficiently Compressing Deep Neural Networks via Exploiting Floats." *The 49th International Conference on Parallel Processing*. Edmonton, Canada, August 17–20, 2020.
- **[PPoPP'20]** Jiannan Tian, Sheng Di, Chengming Zhang, Xin Liang, **Sian Jin**, Dazhao Cheng, Dingwen Tao, and Franck Cappello. "WAVESZ: A Hardware-Algorithm Co-Design of Efficient Lossy Compression for Scientific Data." *ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming*. San Diego, California, USA, February 22–26, 2020.

GRANTS & PROPOSALS

I contribute to the following grant and proposals by providing preliminary results, writing and research.

- **[Grant Writing]** 1/2023–12/2027, "CAREER: A Highly Effective, Usable, Performant, Scalable Data Reduction Framework for HPC Systems and Applications", NSF CAREER, funded, $450k.
- **[Grant Writing]** 9/2021–8/2024, "Collaborative Research: Elements: ROCCI: Integrated Cyberinfrastructure for In Situ Lossy Compression Optimization Based on Post Hoc Analysis Requirements", NSF ROCCI, funded, $280k.
- **[Grant Writing]** 08/2020–07/2023, "CDS&E: Collaborative Research: HyLoC: Objective-driven Adaptive Hybrid Lossy Compression Framework for Extreme-Scale Scientific Applications", NSF CDS&E, funded, $530k.
- **[Grant Writing]** 05/2020–04/2022, "CRII: OAC: An Efficient Lossy Compression Framework for Reducing Memory Footprint for Extreme-Scale Deep Learning on GPU-Based HPC Systems", NSF CRII, funded, $190k.
- **[Grant Assistant]** 08/2019–12/2019, "Improving Lossy Compression for Scientific Applications at Extreme Scale", DOE Argonne National Laboratory ECP, funded, $25k.

PRESENTATIONS & TALKS

- 11/2023, Poster Presentation, "Scaling HPC Applications Through Predictable and Reliable Data Reduction Methods." The International Conference for High Performance Computing, Networking, Storage, and Analysis. Denver, Colorado, USA.

- 08/2023, Invited talk at HDF5 User Group (HUG) Meeting 2023, "Accelerating Parallel Write via Deeply Integrating Predictive Lossy Compression with HDF5." Columbus, Ohio, USA.
- 06/2023, Paper Presentation, "Fast All-Pairs Shortest Paths Algorithm in Large Sparse Graph." ACM International Conference on Supercomputing. Orlando, Florida, USA.
- 01/2023, Poster Presentation, "Accelerating Parallel Write With Lossy Compression In HDF5." Exascale Computing Project Annual Meeting, Houston, Texas, USA.
- 11/2022, Paper Presentation, "Accelerating Parallel Write via Deeply Integrating Predictive Lossy Compression with HDF5." The International Conference for High Performance Computing, Networking, Storage, and Analysis, Dallas, Texas, USA.
- 09/2022, Paper Presentation, "COMET: A Novel Memory-Efficient Deep Learning Training Framework by Using Error-Bounded Lossy Compression." Worldwide online event.
- 06/2022, Paper Presentation, "Significantly Improving Prediction-Based Lossy Compression Via Ratio-Quality Modeling." IEEE International Conference on Data Engineering. Worldwide online event.
- 04/2022, Invited Talk at University of North Carolina at Charlotte, "Design And Use of Efficient Data Reduction Techniques For ML And HPC Applications."
- 06/2021, Paper Presentation, "Adaptive Configuration of In Situ Lossy Compression for Cosmology Simulations via Fine-Grained Rate-Quality Modeling." The 30th ACM International Symposium on High-Performance Parallel and Distributed Computing. Worldwide online event.
- 03/2021, Poster Presentation, "POSTER: A Novel Memory-Efficient Deep Learning Training Framework via Error-Bounded Lossy Compression." The 26th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming. Worldwide online event.
- 02/2021, Poster Presentation, "A Novel Memory-Efficient Deep Learning Training Framework via Error-Bounded Lossy Compression." The 26th Joint Laboratory for Extreme Scale Computing (JLESC). Worldwide online event.
- 05/2020, Paper Presentation, "Understanding GPU-Based Lossy Compression for Extreme-Scale Cosmological Simulations." 35th IEEE International Parallel & Distributed Processing Symposium. Worldwide online event.
- 06/2019, Paper Presentation, "DeepSZ: A Novel Framework to Compress Deep Neural Networks by Using Error-Bounded Lossy Compression." The 28th ACM International Symposium on High-Performance Parallel and Distributed Computing, Phoenix, AZ, USA.

TEACHING

- Fall 2020, Teaching Assistant. CptS/EE 455 "Introduction to Computer Networks". Washington State University.
- Fall 2019, Teaching assistant. CS470/570 "Computer Algorithms". University of Alabama

AWARD & HONOR

- Student Travel Grant $1500, ACM SC'19 — Nov. 2022
- Program Committee (PC) for IEEE IWBDR-3'23 — Nov. 2022
- Sixth cohort of the Distinguished Graduate Research Program (DGRP) — Mar. 2022
- Technical Program Committee (TPC) for IEEE HPCC'21 — Sep. 2021
- Reviewer for IEEE TPDS'21 — Sep. 2021
- Technical Program Committee (TPC) for IEEE HPCC'20 — Oct. 2020
- Mahmoud M. Dillsi Family Graduate Fellowship $1000 — Oct. 2020
- Outstanding Graduate Researcher, The University of Alabama — Apr. 2020
- Student Travel Grant $1,000, ACM HPDC'19 — Jun. 2019
- Graduate Conference Funding $1,000, The University of Alabama — Jun. 2019
- Graduate Research Funding $600, The University of Alabama — Mar. 2019

- National University Student Innovation Program 2017 – 2018
- Scholarships of Beijing Normal University OCT. 2015 & OCT. 2016
- National Olympiad in Information (Advanced Group), *First-Class Prize* DEC. 2012
- The 12th China Adolescent Robotics Competition, *First-Class Prize* JUL. 2012

Selected Software

- **Foresight**: A Compression Benchmark Suite for Visualization and Analysis of Simulation Data.
- **DeepSZ**: Lossy Compression Framework for Deep Neural Networks.
- **cuSZ**: A GPU Accelerated Error-Bounded Lossy Compressor for Scientific Data