

## education

### Washington State University, Pullman, WA

**2020-present** PH.D. CANDIDATE in Computer Science PH.D. ADVISOR: Dr. Dingwen Tao  
**laboratory** High Performance Data Analytics & Computing Lab (HiPDAC)

### The University of Alabama, Tuscaloosa, AL

**2019-2020** PH.D. CANDIDATE in Computer Science PH.D. ADVISOR: Dr. Dingwen Tao  
**laboratory** High Performance Data Analytics & Computing Lab (HiPDAC)  
**affiliation** Remote Sensing Center, Alabama Water Institute

### University of Massachusetts Amherst, Amherst, MA

**2017** MASTER OF SCIENCE in Computer Engineering

### Dalian Maritime University, Dalian, China

**2013** BACHELOR OF SCIENCE in Electrical & Computer Engineering

## research

**stack** C/C++, CUDA, Kokkos, OpenMP, MPI; Python, R; Scientific Computing, Data Analytics  
**interest** **HPC** Heterogeneous Computing (GPU, FPGA), Extreme-Scale Scientific Computing, Data Management, Lossy Compression, Big Data Analytics  
**cloud** HPC on Cloud, Virtualization and Containerization for Scientific Applications

## publication

### first-author paper : as presenter

- CLUSTER '21** **Jiannan Tian** , Sheng Di, Xiaodong Yu, Cody Rivera, Kai Zhao, Sian Jin, Yunhe Feng, Xin Liang, Dingwen Tao, Franck Cappello. "Optimizing Error-Bounded Lossy Compression for Scientific Data on GPUs." *Proceedings of the 2021 IEEE International Conference on Cluster Computing*, (Virtual Event) Portland, OR, September 7–10, 2021. [\[arXiv:2105.12912\]](#)
- IPDPS '21** **Jiannan Tian** , Cody Rivera, Jieyang Chen, Dingwen Tao, Sheng Di, and Franck Cappello. "Revisiting Huffman Coding: Toward Extreme Performance on Modern GPU Architectures." *IEEE International Parallel & Distributed Processing Symposium*, (Virtual Event) Portland, OR, May 17–21, 2021. [\[arXiv:2010.10039\]](#)
- PACT '20** **Jiannan Tian** , Sheng Di, Kai Zhao, Cody Rivera, Megan Hickman Fulp, Robert Underwood, Sian Jin, Xin Liang, Jon Calhoun, Dingwen Tao, and Franck Cappello. "CUSZ: A High-Performance GPU Based Lossy Compression Framework for Scientific Data." *The 29th International Conference on Parallel Architectures and Compilation Techniques*, (Virtual Event) Atlanta, GA, October 3–7 2020. [\[arXiv:2007.09625\]](#)
- 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." *Proceedings of the 25th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming*, San Diego, CA, February 22–26, 2020. [\[doi:10.1145/3332466.3374525\]](#)

### coauthored paper

- 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." *Proceedings of the 36th ACM International Conference on Supercomputing*, Virtual Event, June 27–30, 2022. [\[arXiv:2106.13306\]](#)
- HPDC '22** Xiaodong Yu, Sheng Di, Kai Zhao, **Jiannan Tian**, Dingwen Tao, Xin Liang, and Franck Cappello. "Ultra-fast Error-bounded Lossy Compression for Scientific Dataset." (*Preprint*) *The 31st International ACM Symposium on High-Performance Parallel and Distributed Computing*, Minneapolis, Minnesota, United States, June 27–July 1, 2022. [\[arXiv:2204.00711\]](#)

- 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.” (Preprint) *The 31st International ACM Symposium on High-Performance Parallel and Distributed Computing*, Minneapolis, Minnesota, United States, June 27–July 1, 2022. [[arXiv:2201.13020](#)]
- ICDE '22** Sian Jin, Sheng Di, **Jiannan Tian**, Suren Byna, Dingwen Tao, and Franck Cappello. “Significantly Improving Prediction-Based Lossy Compression Via Ratio-Quality Modeling.” (Preprint) *IEEE International Conference on Data Engineering 2022*, (Virtual) Kuala Lumpur, Malaysia, May 9–May 12, 2022. [[arXiv:2111.09815](#)]
- IPDPS '22** Cody Rivera, Sheng Di, **Jiannan Tian**, Xiaodong Yu, Dingwen Tao, and Franck Cappello. “Optimizing Huffman Decoding for Error-Bounded Lossy Compression on GPUs.” (Preprint) *IEEE International Symposium on High-Performance Parallel and Distributed Computing*, Ecole Normale Supérieure de Lyon, Lyon, France, May 30–June 3, 2022. [[arXiv:2201.09118](#)]
- 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.” *Proceedings of the 30th International Symposium on High-Performance Parallel and Distributed Computing*, Virtual Event, Sweden, June 21–25, 2021. [[arXiv:2104.00178](#)]
- 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.” *2021 International Conference on Supercomputing*, Virtual Event, USA, June 14–17, 2021. [[arXiv:2011.10170](#)]
- IPDPS '20** Sian Jin, Pascal Grosset, Christopher M. Biber, 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, LA, May 18–22, 2020. [[arXiv:2004.00224](#)]
- 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.” *Proceedings of the 28th ACM International Symposium on High-Performance Parallel and Distributed Computing*, Phoenix, AZ, June 24–28, 2019. [[arXiv:1901.09124](#)]

## workshop

- BigData '20** Baodi Shan, Aabid Shamji, **Jiannan Tian**, Guanpeng Li, and Dingwen Tao. “LCFI: A fault injection tool for studying lossy compression error propagation in HPC programs.”, *International Workshop on Big Data Reduction (IWBDP)*, (Virtual Event) Atlanta, GA, December 10–13, 2020. [[arXiv:2010.12746](#)]
- BigData '19** Donglin Yang, Wei Rang, Dazhao Cheng, Yu Wang, **Jiannan Tian**, and Dingwen Tao. “Elastic Executor Provisioning for Iterative Workloads on Apache Spark.” *Proceedings of 2019 IEEE International Conference on Big Data*, Los Angeles, CA, December 9–12, 2019. [[doi:10.1109/BigData47090.2019.9006021](#)]

## release

- CUSZ** Error-Bounded Lossy Compressor for Scientific Data (CUDA Implementation) [[GitHub repo](#)]

## activity

- 2021** ACM Student Research Competition (Finalist) SC '21, St. Louis, Nov. 2020
- 2021** Argonne 2021 GPU Hackathon Digital Event, Apr. 2021
- 2020** Student Volunteer SC '20, Virtual Event, Nov. 2020

## experience

### Graduate Research Assistant

*HiPDAC Lab, Washington State University, 2020 August-present*

- Research in designing efficient accelerator-based (GPU, FPGA, etc.) high-efficiency lossy compressor
- On development of GPU accelerated SZ compression, CUSZ [[anl.gov page](#)]

### Research & Development Internship

*Argonne National Laboratory, Lemont, IL, 2020 May-present*

- INTERNSHIP MENTOR: Dr. Sheng Di
- MAJOR DEVELOPER of CUSZ
- Research in data reduction and codesign of large-scale scientific applications

### Research & Development Internship

*Argonne National Laboratory, Lemont, IL, 2019 August-2020 August*

- INTERNSHIP MENTOR: Dr. Sheng Di
- Research in designing GPU-based high-throughput lossy compressor

### Graduate Research Assistant

*HiPDAC Lab, The University of Alabama, 2019 January-2020 August*

- Research in designing efficient accelerator-based (GPU, FPGA, etc.) high-efficiency lossy compressor
- On accelerator based optimization of SZ compression [\[anl.gov page\]](#)

### Graduate Student Researcher

*Advanced Human & Health Analytics Laboratory, University of Massachusetts Amherst, 2017 January-2017 August*

- Research on Mobile & Personalized Health, focusing on developing wearable sensors and data analytic methodologies to understand the health conditions

### Graduate Research Assistant

*Sustainable Computing Lab, University of Massachusetts Amherst, {2015 summer, 2016 spring}*

- Research on building robust and scalable computing system on top of existing cloud infrastructure.
- Thesis: “Analyzing Spark Performance on Spot Instances”

### Graduate Teaching Assistant

*University of Massachusetts Amherst, 2015 spring*

- ASSISTANT to Dr. David Irwin on *System Software Design*

## project

### cusZ API

*Jan. 2022-present (ongoing)*

- Toward releasing CUSZ API.

### Modularized cusZ: Toward Portability Performance

*Dec. 2021-Jan. 2022*

- Code refactoring & modularization.
- In-depth study of source portability and interoperability.

### cusZ+: Enhanced cusZ

*Jan. 2021-present (ongoing)*

- High-throughput Lorenzo reconstruction in decompression
- Sparsity-aware methods for beyond-32× reduction rate

### huffre: Huffman Coding Toward Extreme Performance on Modern GPU Architectures

*Jun. 2020-Oct. 2021*

- Parallel Huffman codebook construction
- REDUCE-MERGE based Huffman encoder for multihundred-GBps throughput

### cusZ: GPU-Based High-Throughput Lossy Compressor for Scientific Data

*Aug. 2019-present (long-term)*

- ECP Software Track, Data & Visualization
- MAJOR DEVELOPER. Cooperate with EXAFEL team for data reduction.
- Proposed fine-granularity SIMT-based high-throughput SZ implementation called CUSZ.
- Also included high-throughput GPU based Huffman codec.
- Future work includes porting to Kokkos for *Performance Portability*.

### WAVESZ: HLS-Based Hardware-Algorithm Co-Designed Framework

*May 2019-Aug. 2019*

- Adopted a wavefront memory layout to fit into SZ algorithm to alleviate data dependency during the prediction-quantization procedure (compute kernel).
- Proposed a co-design framework for SZ lossy compression called WAVESZ, implemented it on FPGA by using High-

Level Synthesis (HLS), and evaluated on three real-world HPC simulation datasets from Scientific Data Reduction Benchmarks (SDRB) suite.

### **Optimizing SZ: Architecture-Specific Performance Improvement**

*Dec. 2018-Jan. 2019*

- On profiling and optimizing SZ, the error-bounded lossy compressor (testbed: x86-based system, with floating-point and integer typed datasets).

### **Wearable Sensor for Joint-Angle Estimation**

*Feb. 2017-Aug. 2017*

- Built proof-of-concept wearable sensor system to estimate knee angle based on data-driven methodology.
- Data mining on morphological information.

### **Analysis of Anonymized Solar Power Data**

*Oct. 2016-Dec. 2016*

- Data mining on anonymized solar power data to estimate power utilization and market capacity based on GIS data and history weather data.
- Utilized map API of Google Cloud to collect layered geographic data; employed OpenCV to filter satellite images with objects.

### **Thesis: Analyzing Spark Performance on Spot Instances**

*master thesis, dissertation*

- Proposed optimal scheduling scheme for distributed computing on cost-reducing and resource-stringent cloud infrastructure on top of Mesos and Hadoop.
- Built prototype with Apache Spark on AWS EC2, contributing to ~20% shorter completion time and ~23% less cost with test cases of history records.

### **Design and Verification of Dual-Priority Scheduling System (FPGA)**

*undergraduate project*

- Proposed real-time scheduling method in the context of building infrastructures for vehicle network.
- Proposed dual-priority featured real-time scheduling framework based on multiprocessor systems and FPGA.