

# Robert Underwood

✉ rr.underwood94@gmail.com • 🌐 robertu94.github.io  
🔗 github.com/robertu94

## Education

---

Clemson University	Clemson, SC
PhD in Computer Science, GPA 3.92/4.0	December 2021
Dissertation: Approachable Error Bounded Lossy Compression	
Passed Qualifying Exam: May 2018	
Co-Advisers: Dr. Amy Apon, Dr. Jon Calhoun, and Dr. Franck Cappello	
Clemson University	Clemson, SC
Master of Science in Computer Science, GPA 4.0/4.0	August 2018
Concentration: Systems and Implementation	
Clemson University, Calhoun Honors College	Clemson, SC
Bachelor of Science, Suma Cum Laude in Computer Science, GPA 4.0/4.0	December 2016
Honors Thesis: Automation in the Classroom, Adviser: Dr. Jacob Sorber	

## Research Experience

---

Argonne National Laboratory	Lemont, IL
Assistant Computer Scientist	2024-
○ Researched Multi-modal data and systems software for AI	
- coLead of AuroraGPT Data Team with Ian Foster 2024-present	
○ Researched applications and approaches of lossy compression to ensure data integrity.	
○ Researched limits of compress-ability of lossy compression.	
○ Researched data movement and check-pointing systems	
Argonne National Laboratory	Lemont, IL
Post Doctoral Appointee	2022-2024
○ Researched Multi-modal data and systems software for AI	
- coLead of AuroraGPT Data Team with Ian Foster 2024-present	
○ Researched applications and approaches of lossy compression to ensure data integrity.	
○ Researched data movement and checkpointing systems	
Clemson University	Clemson, SC
Clemson Data Intensive Computing Environments	2016-2021
○ Researched and modeled reliability and performance of applications using lossy compression	
○ Developed techniques to understand the impacts of lossy compression on AI applications	
○ Designed experiments to analyze performance of high performance computing systems	
○ Designed models to understand and improve the reliability of computer infrastructure	
○ Researched trade-offs of compressor agnostic tooling for scientific applications	
Argonne National Laboratory	Lemont, IL
Under Dr. Franck Cappello	Summer-Fall 2019

- Researched the design of optimization based techniques for enforcing user-level error bounds
- Designed and implemented LibPressio – a generic abstraction between compression libraries
- Contributed to the design and implementation of SZ – a lossy compression framework – for CPUs and GPUs

Clemson University

Clemson, SC

Clemson PERSIST Lab

2015-2016

- Designed and developed an automated grading framework using Python, C, Raspberry Pi, and Docker
- System used modular design, supports process isolation, and multiple test formats

## Publications

---

- [1] Cappello, F., Underwood, R., Alexeev, Y., Baker, A., Bozdag, E., Burtscher, M., Chard, K., Di, S., Felker, K. G., O’Grady, P. C., Guo, H., Huang, Y., Jiang, P., Li, S., Liang, X., Lindahl, E., Lindstrom, P., Lukic, Z., Lundborg, M., Lykov, D., Nagaso, M., Sato, K., Sign, A., Son, S. W., Tang, W., Tao, D., Tian, J., Yoshii, K., Zaho, K., “What to Support When You’re Compressing: The State of Practice, Gaps, and Opportunities for Scientific Data Compression”. In: SC’25. Author, to Appear (authoring, editing, methods, investigation, data curation, and visualization). St. Louis, MO: IEEE/ACM, 2025. DOI: N/A.
- [2] Gokdemir, O., Getty, N., Underwood, R., Madireddy, S., Cappello, F., Ramanathan, A., Foster, I., Stevens, R., “Automated MCQA Benchmarking at Scale: Evaluating Reasoning Traces as Retrieval Sources for Domain Adoption of Small Language Models”. In: Frontiers in Generative AI for HPC Science and Engineering: Foundations, Challenges, and Opportunities. TPC@SC’25. coAuthor, to Appear, (editing, methodology, investigation). St. Louis, MO: IEEE/ACM, 2025. DOI: N/A.
- [3] Huang, Y., Di, S., Underwood, R., Myint, P., Chu, M., Li, G., Schwarz, N., Cappello, F., “lsCOMP: Efficient Light Source Compression”. In: Frontiers in Generative AI for HPC Science and Engineering: Foundations, Challenges, and Opportunities. SC’25. coAuthor, to Appear, (authoring, editing, methodology, artifact, investigation). St. Louis, MO: IEEE/ACM, 2025. DOI: N/A.
- [4] Lattanzio, E. E., Ranjan, S., Underwood, R., Saupe, T., Speck, R., Calhoun, J. C., “Performance Analysis of Inline Compression in pySDC”. In: Frontiers in Generative AI for HPC Science and Engineering: Foundations, Challenges, and Opportunities. High Performance Extreme Computing Conference. coAuthor, to Appear, (editing, methodology, software, investigation). Virtual: IEEE, 2025. DOI: N/A.
- [5] Liu, Y., Di, S., Getty, N., Mallick, T., Underwood, R., Jin, S., “Beyond End-to-End: Understanding the Limits of LLMs in Scientific Problem Solving”. In: Frontiers in Generative AI for HPC Science and Engineering: Foundations, Challenges, and Opportunities. TPC@SC’25. coAuthor, to Appear, (authoring, editing, methodology, advising, investigation). St. Louis, MO: IEEE/ACM, 2025. DOI: N/A.
- [6] Ockerman, S., Gueroudji, A., Oh, S. Y., Underwood, R., Chia, N., Chard, K., Ross, R., Venkataraman, S., “Exploring Distributed Vector Databases Performance on HPC Platforms: A Study with Qdrant”. In: Frontiers in Generative AI for HPC Science and

- Engineering: Foundations, Challenges, and Opportunities. TPC@SC'25. coAuthor, to Appear, (authoring, editing, methodology, investigation). St. Louis, MO: IEEE/ACM, 2025. DOI: N/A.
- [7] Qiu, Z., Liu, J., Zhao, K., Underwood, R., Di, S., “Benchmarking Cutting-Edge Scientific Error-Bounded Lossy Compressors on Correlation-Based Rate-Distortion”. In: 2025 IEEE/ACM 11th International Workshop on Data Analysis and Reduction for Big Scientific Data. DRBSD'25. co-Author (investigation, methodology, advising, authoring, and editing), Best Paper Runner Up, To Appear. St. Louis, MO: IEEE/ACM, 2025. DOI: N/A.
  - [8] Zhang, L., Li, R., Ren, C., Di, S., Liu, J., Huang, J., Underwood, R., Grosset, P., Tao, D., Liang, X., Guo, H., Cappello, F., Zhao, K., “LCP: A High-Performance Scientific Lossy Compressor for Particles”. In: Proceedings of the ACM on Management of Data. DRBSD'25 (2025). co-Author (investigation, editing). DOI: 10.1145/3709700.
  - [9] Di, S., Liu, J., Zhao, K., Liang, X., Underwood, R., Zhang, Z., Shah, M., Huang, Y., Huang, J., Yu, X., Ren, C., Guo, H., Wilkins, G., Tao, D., Tian, J., Jin, S., Jian, Z., Wang, D., Rahman, M. H., Zhang, B., Song, S., Calhoun, J., Li, G., Yoshii, K., Alharthi, K., Cappello, F., “A Survey on Error-Bounded Lossy Compression for Scientific Datasets”. In: ACM Comput. Surv. ACM Computing Surveys (June 2025). co-Author (authoring, and editing). ISSN: 0360-0300. DOI: 10.1145/3733104.
  - [10] Poulous, A., Underwood, R., Calhoun, J. C., Di, S., Cappello, F., “Sensitivity of Black-Box Statistical Prediction of Lossy Compression Ratios for 3D Scientific Data”. In: 39th IEEE International Parallel & Distributed Processing Symposium. IPDPS25. co-Author (concept, investigation, methodology, advising, software design, authoring, and editing). Milian, Italy: IEEE/ACM, June 2025. DOI: 10.1109/IPDPS64566.2025.00102.
  - [11] Song, S., Underwood, R., Huang, Y., Di, S., Jiang, P., Cappello, F., “An Efficient and Computation-balanced Lossy Compressor on Wafer-Scale Engine”. In: 39th IEEE International Parallel & Distributed Processing Symposium. IPDPS25. co-Author (concept, investigation, methodology, advising, software design, authoring, and editing). Milian, Italy: IEEE/ACM, June 2025. DOI: 10.1109/IPDPS64566.2025.00084.
  - [12] Wilkins, G., Di, S., Calhoun, J. C., Underwood, R., Cappello, F., “To Compress or Not To Compress: Energy Trade-Offs and Benefits of Lossy Compressed I/O”. In: 39th IEEE International Parallel & Distributed Processing Symposium. IPDPS25. co-Author (investigation, editing). Milian, Italy: IEEE/ACM, June 2025. DOI: 10.1109/IPDPS64566.2025.00082.
  - [13] Siebenschuh, C., Hippe, K., Gokdemir, O., Brace, A., Khan, A. M., Hossain, K., Babuji, Y., Chia, N., Vishwanath, V., Ramanathan, A., Stevens, R. L., Foster, I., Underwood, R., “AdaParse: An Adaptive Parallel PDF Parsing and Resource Scaling Engine”. In: Eighth Conference on Machine Learning and Systems. MLSYS '25. co-Author (concept, investigation, methodology, advising, authoring, and editing) <https://openreview.net/forum?id=wYzc2UnQKt>. Santa Clara, CA, USA: IEEE/ACM, May 2025. DOI: NA.

- [14] Underwood, R., Liu, J., Zhao, K., Di, S., Cappello, F., “Evaluating Advanced Scientific Compressors on Climate Datasets”. In: European Geospatial Union. EGU-25. Author (concept, investigation, methodology, software design, authoring, and editing). Vienna, Austria: IEEE/ACM, Apr. 2025. DOI: 10.5194/egusphere-egu25-7371.
- [15] Agarwal, T., Di, S., Huang, J., Huang, Y., Gopalakrishnan, G., Underwood, R., Zhao, K., Liang, X., Li, G., Cappello, F., “SZOps: Scalar Operations for Error-bounded Lossy Compressor for Scientific Data”. In: 2024 IEEE/ACM 10th International Workshop on Data Analysis and Reduction for Big Scientific Data (DRBSD). DRBSD’24. co-Author (investigation, methodology, and editing), Published. Atlanta, GA, USA: IEEE/ACM, Apr. 2024. DOI: 10.1109/SCW63240.2024.00042.
- [16] Cappello, F., Di, S., Underwood, R., Tao, D., Calhoun, J., Kazutomo, Y., Sato, K., Singh, A., Giraud, L., Agullo, E., Yepes, X., Acosta, M., Jin, S., Tian, J., Vivien, F., Zhang, B., Sano, K., Ueno, T., Grützmacher, T., Anzt, H., “Multifacets of lossy compression for scientific data in the Joint-Laboratory of Extreme Scale Computing”. In: Future Generation Computer Systems. FGCS’24 (Apr. 2024). co-Author (authoring, and editing). ISSN: 0167-739X. DOI: <https://doi.org/10.1016/j.future.2024.05.022>.
- [17] Grützmacher, T., Underwood, R., Di, S., Cappello, F., Anzt, H., “FRSZ2 for In-Register Block Compression Inside GMRES on GPUs”. In: 2025 IEEE/ACM 10th International Workshop on Data Analysis and Reduction for Big Scientific Data. DRBSD’24. co-Author (concept, investigation, methodology, advising, software design, authoring, and editing), Best Paper, Published. Atlanta, GA, USA: IEEE/ACM, Apr. 2024. DOI: 10.1109/SCW63240.2024.00038.
- [18] Liu, J., Tian, J., Wu, S., Di, S., Zhang, B., Underwood, R., Huang, Y., Huang, J., Zhao, K., Li, G., Tao, D., Chen, Z., Cappello, F., “High-ratio Scientific Lossy Compression on GPUs with Optimized Multi-level Interpolation”. In: The International Conference for High Performance Computing, Networking, Storage, and Analysis. SC’24. co-Author (background, motivation, and editing). Atlanta, GA, USA: IEEE/ACM, Nov. 2024. DOI: 10.1109/SC41406.2024.00019.
- [19] Maurya, A., Underwood, R., Rafique, M. M., Cappello, F., Nicolae, B., “DataStates-LLM: Lazy Asynchronous Checkpointing for Large Language Models”. In: International Symposium on High-Performance Parallel and Distributed Computing. HPDC’24. co-Author (investigation and editing), Best Paper. Pisa, Italy: ACM, June 2024. DOI: 10.1145/3625549.3658685.
- [20] Underwood, R., Madhyastha, M., Burns, R., Nicolae, B., “Evostore: Towards Scalable Storage of Evolving Learning Models”. In: International Symposium on High-Performance Parallel and Distributed Computing. HPDC’24. First Author. Pisa, Italy: ACM, June 2024. DOI: 10.1145/3625549.3658679.
- [21] Jian, Z., Di, S., Liu, J., Zhao, K., Liang, X., Xu, H., Underwood, R., Wu, S., Chen, Z., “CliZ: Optimizing Lossy Compression for Climate Datasets with Adaptive Fine-tuned Data Prediction”. In: 38th IEEE International Parallel & Distributed Processing Symposium. IPDPS’24. co-Author (concept, methodology). San Francisco, CA, USA: IEEE, May 2024. DOI: 10.1109/IPDPS57955.2024.00044.

- [22] Wilkins, G., Di, S., Calhoun, J., Kim, K., Underwood, R., Mortier, R., Cappello, F., “FedSZ: Leveraging Floating-Point Lossy Compression for Federated Learning Communications”. In: 44th IEEE International Conference on Distributed Computing Systems. ICDCS’24. co-Author (methodology, editing, resources). Jersey City, NJ, USA: IEEE, May 2024. DOI: 10.1109/IPDPSW63119.2024.00201.
- [23] Underwood, R., Di, S., Jin, S., Rahman, M. H., Khan, A., Cappello, F., “LibPressio-Predict: Flexible and Fast Infrastructure For Inferring Compression Performance”. In: Proceedings of the SC ’23 Workshops of The International Conference on High Performance Computing, Network, Storage, and Analysis. SC-W ’23. co-Author (methodology, editing, resources). Denver, CO, USA: Association for Computing Machinery, May 2023. ISBN: 9798400707858. DOI: 10.1145/3624062.3625124.
- [24] Underwood, R., Madhyastha, M., Burns, R., Nicolae, B., “Understanding Patterns of Deep Learning Model Evolution in Network Architecture Search”. In: 30th IEEE International Conference on High Performance Computing, Data, and Analytics. HiPC’23. First Author. Goa, India: IEEE, Dec. 2023. ISBN: 9798400707858. DOI: 10.1109/HiPC58850.2023.00025.
- [25] Ganguli, A., Underwood, R., Bessac, J., Krasowska, D., Calhoun, J. C., Di, S., Cappello, F., “A Lightweight, Effective Compressibility Estimation Method for Error-Bounded Lossy Compression”. In: 2023 IEEE International Conference on Cluster Computing. CLUSTER’23. First two authors contributed equally. Santa Fe, NM, USA: IEEE, Oct. 2023. ISBN: 9798400707858. DOI: 10.1109/CLUSTER52292.2023.00028.
- [26] Liang, X., Zhao, K., Di, S., Li, S., Underwood, R., Gok, A. M., Tian, J., Deng, J., Calhoun, J. C., Tao, D., Chen, Z., Cappello, F., “SZ3: A Modular Framework for Composing Prediction-Based Error-Bounded Lossy Compressors”. In: IEEE Transactions on Big Data. CLUSTER’23 9.2 (Aug. 2023). Co-author (software design, editing), pp. 485–498. ISSN: 0167-739X. DOI: 10.1109/TBDATA.2022.3201176.
- [27] Underwood, R., Yoon, C., Gok, A., Di, S., Cappello, F., “ROIBIN-SZ: Fast and Science-Preserving Compression for Serial Crystallography”. In: Synchrotron Radiation News. SRN 9.2 (July 2023). First Author, Invited, Not Peer Reviewed, pp. 485–498. ISSN: 0167-739X. DOI: 10.1080/08940886.2023.2245722.
- [28] Madhyastha, M., Underwood, R., Burns, R., Nicolae, B., “DStore: A Lightweight Scalable Learning Model Repository with Fine-Grain Tensor-Level Access”. In: Proceedings of the 37th International Conference on Supercomputing. Vol. 9. ICS’23 2. Co-Author (editing, software, data curation, and visualization). Orlando, FL, USA: IEEE, June 2023, pp. 133–143. ISBN: 9798400707858. DOI: 10.1145/3577193.3593730.
- [29] Underwood, R., Bessac, J., Krasowska, D., Calhoun, J. C., Di, S., Cappello, F., “Black-box statistical prediction of lossy compression ratios for scientific data”. In: The International Journal of High Performance Computing Applications. IJHPCA 9.2 (June 2023). First two authors contributed equally, p. 10943420231179417. ISSN: 0167-739X. DOI: 10.1177/10943420231179417.

- [30] Underwood, R., Nicolae, B., “MPIGDB: A Flexible Debugging Infrastructure for MPI Programs”. In: 13th Workshop on AI and Scientific Computing at Scale using Flexible Computing Infrastructures. Vol. 9. FlexScience’23 2. First Author Presented at FCRC/HPDC’23. Orlando, FL, USA: ACM, June 2023, p. 10943420231179417. ISBN: 9798400707858. DOI: 10.1145/3589013.3596675.
- [31] Bessac, J., Underwood, R., Di, S., “Discussion on “Saving Storage in Climate Ensembles: A Model-Based Stochastic Approach””. In: Journal of Agricultural, Biological and Environmental Statistics. JABES 28.2 (May 2023). Co-author (investigation, editing), pp. 358–364. ISSN: 0167-739X. DOI: 10.1007/s13253-023-00540-7.
- [32] Underwood, R., Di, S., Calhoun, J. C., Apon, A., Cappello, F., “OptZConfig: Using Optimization Configure Error Bounded Lossy Compressors”. In: Transactions on Parallel and Distributed Systems. TPDS 28.2 (Dec. 2022). First Author, pp. 358–364. ISSN: 0167-739X. DOI: 10.1109/TPDS.2022.3154096.
- [33] Underwood, R., Bessac, J., Di, S., Cappello, F., “Understanding the Effects of Modern Compressors on the Community Earth Science Model”. In: 2022 IEEE/ACM 8th International Workshop on Data Analysis and Reduction for Big Scientific Data (DRBSD). Vol. 28. DRBSD’22 2. First Author, Best Paper. Dalas, TX, USA: ACM, Nov. 2022, pp. 1–10. ISBN: 9798400707858. DOI: 10.1109/DRBSD56682.2022.00006.
- [34] Bessac, J., Krasowska, D., Calhoun, J. C., Underwood, R., Di, S., Cappello, F., “Exploring Lossy Compressibility through Statistical Correlations of Scientific Datasets”. In: The 7th International Workshop on Data Analysis and Reduction for Big Scientific Data. Vol. 28. DRBSD’21 2. Co-Author (advising, design, software, editing), Presented at SC21. Saint Louis, MO, USA: IEEE, Nov. 2021, pp. 1–10. ISBN: 9798400707858. DOI: 10.1109/DRBSD754563.2021.00011.
- [35] Underwood, R., Malvoso, V., Di, S., Calhoun, J. C., Apon, A., Cappello, F., “Productive and Performant Generic Lossy Data Compression with LibPressio”. In: The 7th International Workshop on Data Analysis and Reduction for Big Scientific Data. Vol. 28. DRBSD’21 2. First Author, Presented at SC21. Saint Louis, MO, USA: IEEE, Nov. 2021, pp. 1–10. ISBN: 9798400707858. DOI: 10.1109/DRBSD754563.2021.00005.
- [36] Fulp, D., Poulos, A., Underwood, R., Calhoun, J. C., “ARC: An Automated Approach to Resiliency for Lossy Compressed Data via Error Correcting Codes”. In: Proceedings of 30th International ACM Symposium on High-Performance Parallel and Distributed Computing. Vol. 28. HPDC’21 2. Co-Author (software, methodology, editing). Stockholm, Sweden: ACM, June 2021, pp. 1–10. ISBN: 9798400707858. DOI: 10.1145/3431379.3460638.
- [37] Tian, J., Di, S., Zhao, K., Rivera, C., Hickman, M., Underwood, R., Jin, S., Liang, X., Calhoun, J., Tao, D., Cappello, F., “cuSZ: An Efficient GPU Based Error-Bounded Lossy Compression Framework for Scientific Data”. In: Proceedings of 29th International Conference on Parallel Architectures and Compilation Techniques. Vol. 28. PACT’20 2. Co-Author (design, editing). Atlanta, Georgia (virtual): ACM, Oct. 2020, pp. 1–10. ISBN: 9798400707858. DOI: 10.1145/3410463.3414624.

- [38] Underwood, R., Di, S., Calhoun, J. C., Cappello, F., “FRaZ: A Generic High-Fidelity Fixed-Ratio Lossy Compression Framework for Scientific Floating-point Data”. In: proceedings of the 9th international conference on performance engineering. Vol. 28. IPDPS’20 2. First Author, Presented virtually at IPDPS 2020. New Orleans, Louisiana (virtual): IEEE, May 2020, pp. 1–11. ISBN: 9798400707858. DOI: 10.1109/IPDPS47924.2020.00065.
- [39] Underwood, R., Anderson, J., Apon, A., “Measuring Network Latency Variation Impacts to High Performance Computing Application Performance”. In: Proceedings of the 9th International Conference on Performance Engineering. Vol. 28. IPDPS’20 2. First Author, Presented at ICPE 2018. Berlin, Germany: ACM/SPEC, Apr. 2018, pp. 1–12. ISBN: 9798400707858. DOI: 10.1145/3184407.3184427.

## Peer-Reviewed Academic Poster Presentations

---

Integrating TEZip into LibPressio	Denver, Colorado
Supercomputing 2023	November 2023
Isita Talukdar, Amarjit Singh, Robert Underwood, Kento Sato, Weikuan Yu	
Understanding the Effects of Sampling on Lossy CR Estimation	Denver, Colorado
Supercomputing 2023	November 2023
Alexandra Poulos, Robert Underwood, Jon Calhoun, Sheng Di, Franck Cappello	
Statistical Prediction of Lossy Compression Ratios for 3D Scientific Data	Dallas, Texas
Supercomputing 2022, Best Poster ACM Student Research Competition	November 2022
David Krasowska, Robert Underwood, Julie Bessac, Sheng Di, Frank Cappello	
LibPressio: A Unifying Data Compression Interface for Users and Developers	Virtual
United States Department of Energy Data Days	May 2022
Robert Underwood, Sheng Di, and Franck Cappello	
Approachable Error Bounded Lossy Compression	Virtual
Supercomputing 2021	November 2021
Robert Underwood	
Predicting Optimal E.B.L.C. Configuration for Sampled Data	Spokane, WA
S.I.A.M. Conference on Computer Science and Engineering	February 2019
Robert Underwood, Jon Calhoun, and Amy Apon	

## Significant Software

---

AuroraGPT-ANL	Team co-Lead
<a href="https://github.com/auroraGPT-ANL/">https://github.com/auroraGPT-ANL/</a>	2024-present
○ Scalable Data Preparation Pipelines for PRO, ACM, CORE, and other datasets (core developer)	
○ Evaluation Question Authoring and Review Platform used for the 1000 JAM (core developer)	
○ AdaParse: Scalable and Accurate PDF parsing (aurora port)	
○ LSHBloom: Scalable text deduplication (core developer)	
○ Various Repositories, Python, Rust, Javascript, SQL, CSS	

DataStates-AI	Core Contributor
<a href="https://github.com/DataStates/datastates-ai">https://github.com/DataStates/datastates-ai</a>	2023-present

- Fast and Productive Check-pointing for AI models to accelerate various AI for Science workflows
- Contributed initial implementation and integration with DeepHyper Network Architecture Search framework
- Released in June 2024:  $\approx 1.6k$  LoC C++,  $\approx 3.1k$  LoC Python

LibPressio	Lead Developer
<a href="https://github.com/robertu94/libpressio">https://github.com/robertu94/libpressio</a>	2019-present

- High-performance generic abstraction for compression of dense tensors
- Supports 69+ plugins for compressors and analysis in collaboration with 17 institutions world-wide
- Significant plugins include: LibPressio-Opt (automatic configuration of compression), a parallel compression runtime, a compression performance prediction system, and the external metrics and compressors framework
- Significant integrations include: LCLS-II (light source), Python bindings, HDF5-filters, R bindings, ADIOS2, Spack, Z-checker
- $\approx 275$  unique monthly downloads from Github;  $\approx 58k$  LoC mostly in C++
- Primary Code Contributors from Argonne National Laboratory, National Center for Atmospheric Research, Texas State University, University of Chicago, University of Iowa, Clemson University, Michigan State University, and Indiana University

SZ	Core Contributor
<a href="https://szcompressor.org/">https://szcompressor.org/</a>	2019-present

- One of the leading open and transparent Lossy Compression Frameworks for scientific data
- the SZ framework is a R&D 100 Award Winner for 2021
- Contributed an early design of SZ for GPUs and the design of the modular SZ-3 and SZ-4
- Implemented the Python bindings for SZ
- $\approx 420$  unique monthly downloads from Github;  $\approx 9.9k$  LoC of C++
- Primary Code Contributors from Argonne National Laboratory, University of Chicago, University of Iowa, Clemson University, The Ohio State University, Florida State University, and Indiana University

mpigdb	Sole Developer
<a href="https://github.com/robertu94/mpigdb">https://github.com/robertu94/mpigdb</a>	2023-present

- Open Source Parallel Debugger for MPI based codes that scales to 64 ranks
- Exposes capabilities of underlying GDB with extensions for MPI including
  - Extensions for parallel debugging, scripting
  - Cross-language debugging with C++ and Python
  - CUDA kernel debugging with CUDA-GDB
- $\approx 190$  unique monthly downloads from Github and Crates.io;  $\approx 0.4k$  Rust,  $\approx 0.1k$  Python

## Funding

---

Funded Proposals.....

US Dept. of Energy	11.25M Node Hours
INCITE '24, Data Team co-Lead	2024
AuroraGPT: A Large-Scale Foundation Model for Advancing Science	



US Dept. of Energy	\$3m
Data Reduction '24, co-Author	2024
ZF: A novel framework to design trustworthy lossy compressors for scientific data approaching lossy compressibility limits	
National Science Foundation, Office of Advanced Cyber Infrastructure	\$3.4m
Collaborative Research: Frameworks, Proposal Co-Author and Postdoctoral Researcher	2023
FZ: A fine-tunable cyberinfrastructure framework to streamline specialized lossy compression development (2311875)	
US Dept. of Energy	\$50k
Laboratory Directed Research and Development, Principal Investigator	2023
Science preserving data approximation and I/O optimization using AI for real-time high-resolution tomography on integrated research infrastructure (ANL-2023-0318)	
US Dept. of Energy	2M/yr
Competitive Programs, Senior Personnel	2025
TIDES4AI: Transformative Integration of Scientific Data for AI	
Laboratory Directed Research and Development, Collaborator	2025
Multimodal AI Foundation Modeling of Turbulent, Multiphase, and Reacting Flows for Propulsion and Power Applications	
Laboratory Directed Research and Development, Investigator	2025
Supply Chain Optimizer for Understanding Trends in Critical Minerals (SCOUT-CM)	
700	
High Performance Software Foundation, Presenter	2025
Travel Grant	
300	
ISC-High Performance Computing, Presenter	2025
Travel Grant	
Proposals Under Review.....	
Other Proposal Contributions.....	
Argonne National Laboratory	
Contributions	2022-present
<ul style="list-style-type: none"> <li>○ Intelligent Learning for Light Source and Neutron Source User Measurements Including Navigation and Experiment Steering (PI: Jana Thayer, ASCR-BES, contributor, \$10m, funded)</li> <li>○ Autonomous Active Learning for Inter-atomic Potentials using Large Language Models (PI: Murat Keçeli, LDRD-Seed'24, co-PI, \$50k, declined)</li> <li>○ Co-Scientist: An Assistant Synthesizing Knowledge Graphs and Simulations With Large Language Models (PI: Murat Keçeli, AI For Science'24, co-Author, \$1.1m, not encouraged by DoE)</li> <li>○ <i>DRD<sup>3</sup>R</i>: Data Reduction for Derivatives, Derivatives for Data Reduction (PI: Paul Hovland, Data Reduction '24, co-Author, not encouraged by ANL)</li> <li>○ SciForge: AI-assisted Capture of Science Processes as Programs (PI: Ian Foster, AI For Science '24, Contributor, declined)</li> </ul>	

# Teaching and Mentoring Experience

---

Argonne National Laboratory

Mentoring

2022-present

- (2025) Co-Mentored Yuxiao Li (PhD Student at the Ohio State University) with Hanqi Guo on topology preserving compression for light sources
- (2025) Co-Mentored Congrong Ren (PhD Student at the Ohio State University) with Hanqi Guo on FFT preserving compression for light sources
- (2025) Co-Mentored ZiWei Qiu (PhD Student at the University of Houston) with Jinyang Liu on benchmarking lossy compressors
- (2025) Co-Mentored Youyuan Liu (PhD Student at the Temple University) with Sian Jin on data adaptation for AI
- (2025) Co-Mentored Seth Ockerman (PhD Student at the University of Wisconsin Madison) with Amal Gueroudji, Rob Ross, Shivram Venkataram on distributed vector databases
- (2024-2025) Co-Mentored Arham Kahn (PhD Student at the University of Chicago) with Yadu N. Babuji on scalable text deduplication for AI
- (2024) Co-Mentored Carlo Siebenschuh (PhD Student at the University of Chicago) with Ian Foster scalable PDF parsing
- (2024-2025) Co-Mentored Sujata Sighn (PhD Student at Virginia Tech) with Lingja Liu on Lossy Compressability
- (2024-2025) Co-Mentored Shihui Song (PhD Candidate at University of Iowa) with Peng Jiang and Sheng Di on translating compression techniques to Cerebras CS-2 – a dataflow architecture – and using compiler techniques for lossy compression.
- (2024-2025) Co-Mentored Yafan Huang (PhD Candidate at University of Iowa) with Sheng Di on compression for light sources and network architecture search.
- (2024-2025) Co-Mentored Jianman Tian (PhD Candidate at Indiana University, then Postdoc at University of Kentucky) with Dingwen Tao (until August 2024) and Xin Liang (after August 2024) on compression for light sources and project management.
- (2023-2025) Mentored Alexandra Poulos (PhD Candidate at Clemson University) on aspects of software development, debugging, and research communication in the context of performing sensitivity analysis on compression estimation techniques.
- (2025) Co-Mentored Youyuan Liu (PhD Student at Temple University) with Sian Jin on Data Reduction for AI
- (2023-2024) Co-Mentored Thomas Grüzmacher (PhD Candidate at Technische Universität München) with Hartwig Anzt on lossy compression on GPUs for Linear Algebra.
- (2022-2024) Co-Mentored Meghana Madhyastha (PhD Candidate in Computer Science at Johns Hopkins University) with Bogdan Nicolae on various aspects of the Datastates project from designing lock-free concurrent data structures to accessing data pointers on GPUs, to navigating subtleties using HPC machines. She is now a postdoc at META FAIR.
- (2023) Co-Mentored Isita Talukdar (Undergraduate Electrical Engineering and Computer Science at University of California Berkley) with Amrajit Singh and Kento Sato of Riken Center for Computational Science on aspects of software development, debugging, and research communication in the context of integrating LibPressio with TeZIP (a complex, heterogenous, AI-based, compressor implemented in Python)
- (2022) Co-Mentored David Krasowska (Undergraduate Computer Engineering at Clemson University) with Julie Bessac on aspects of software development, research communication, and how to quickly understand large code bases in the context of extending a compressor agnostic predictor compress-ability to 3d data sets. David took 1<sup>st</sup> in the ACM student poster competition at SC22 on a project I mentored him on and supervised the writing of a workshop paper and journal article. He is now a graduate student at Northwestern University under Peter Dinda.
- (2022) Co-Mentored Arkaprabha Ganguli (PhD Candidate in Statistics at Michigan State University) with Julie Bessac on how to design and implementation of statistical measures to estimate lossy compress-ability without compressors. He is now a Postdoc at Argonne National Laboratory.

Clemson University  
Mentoring

Clemson, SC  
Summer 2021

- Mentored one female and one male, undergraduate student on projects that led to two ACM student research poster submissions and later journal submission.
- Provided training on git, python, C++, lossy compression, and scientific experiment design

Clemson University  
CPSC/ECE 3220: Operating Systems

Clemson, SC  
Fall 2018

- Graduate Teacher of Record, produced all lectures and most materials
- Junior/Senior level course - 50 Students enrolled, Completed (78%), Course GPA (2.42)
- Course materials <https://robertu94.github.io/cpsc3220-f18/>
- Anonymous Student Assessment Responses:
  - Response Rate (92.3%), Would Recommend (72.2%)
  - Median Results: Effective Instructor (4/5), Helpful Feedback (4/5), Relative Difficulty (5/5)
  - Selected Student Comments:
    - “Definitely. One of the best professors I’ve had at Clemson.”
    - “Yes. He is very knowledgeable [sic] and works very hard to impart that knowledge to others.”
    - “Yes, it is obvious that Mr. Underwood is passionate about operating systems and is extensively knowledgeable about computer science in general. This course felt overwhelming at times, but I definitely learned a lot through it.”

## Work Experience

---

The Boeing Company  
Information Technology Intern

Charleston, SC  
Summer 2016, 2017

- Developed improvements for a web based portal system in HTML, Python, and JavaScript
- Developed the user interface for a materials database using HTML and JavaScript
- Designed, developed, and led development on a resource management tool using C#, HTML, and JavaScript.
- Worked on the Network Automation, Tooling, and Standards Integration Team

Unitrends, Inc  
Software Development Intern

Columbia, SC  
2014-2016

- Developed GPU offloading for AES encryption using Nvidia CUDA.
- Designed and developed automated configuration scripts for testing environments using Ansible.
- Designed and developed new cloud infrastructure using LVM, Linux, and Docker
- Designed and developed a Dynamic Alert System in Python
- Worked on the Alerts System in PHP, BASH, C, PERL, SQL
- Worked on the internal Customer Incident Analysis web portal using Django, Postgresql, HTML, CSS, and JavaScript

## Academic Presentations

---

Presentations associated with peer reviewed (co-)authored papers omitted

Lessons Learned From Developing and Distributing Advanced Scientific Compressors with SpackCh  
High Performance Software Foundation Meeting

Preparing Data at Scale: The Data Pipeline for AuroraGPT National Cancer Institute	Virtual January, 2025
Parsing and Duplicating Scientific Text at Scale for AuroraGPT International Council of Scientific and Technical Information Forum	Virtual November, 2024
Scientific Data At Scale Mathematics and Computer Science Division Seminar	Lemont, Illinois August, 2024
Preparing Data at Scale for AuroraGPT Mathematics and Computer Science Division Seminar	Virtual June, 2024
Compression for Light Sources Johns Hopkins University	Baltimore, Maryland August, 2024
CereSZ: Enabling and Scaling Error -bounded Lossy Compression on Cerebras CS-2 International Symposium High-Performance Parallel and Distributed Computing for Shihui Song	Pisa, Italy June, 2024
Project Talk on Compression for Instruments 16th JLESC Workshop	Kobe, Japan April, 2024
Compression for Light Sources George Mason University	Fairfax, Virginia March, 2024
Project Talk on Compression for Instruments 15th JLESC Workshop	Bordeaux, France March 2023
LibPressio 15th JLESC Workshop	Bordeaux, France March 2023
The Benefits of JLESC: An Early Career Prospective JLESC Evaluation Meeting	Bordeaux, France March 2023
Coping with Large Scientific Training Datasets in Deep and Machine Learning Using Lossy Compression AI+Science Summer School at the University of Chicago Data Science Institute	Chicago, Illinois August 2022
LibPressio Argonne Postdoctoral Symposium	Lemont, Illinois October 2022
Project Talk on Compression for Instruments 14th JLESC Workshop	Urbanna, Illinois September 2022
LibPressio 14th JLESC Workshop	Urbanna, Illinois September 2022
Hands on With Lossy Compression Tutorial entitled “Compression for Scientific & Engineering Data” at Super Computing	Dallas, Texas November 2022
Hands on With Lossy Compression Tutorial entitled “Compression for Scientific & Engineering Data” at ISC High Performance	Hamburg, Germany May 2022

LibPressio	Virtual
Session entitled “Compression for Scientific Data and ECP Applications”	May 2022
Exascale Computing Project Annual Meeting	
Understanding the Effects of Modern Lossless and Lossy Compressors on the Community Earth Science Model	Virtual
Session entitled “ Meeting Exascale Computing Challenges with Compression and Pangeo”	May 2022
European Geospatial Union General Assembly	
Improving Lossy Compression for Climate Datasets with SZ3	Virtual
Session entitled “ Meeting Exascale Computing Challenges with Compression and Pangeo”	May 2022
European Geospatial Union General Assembly	
Exploring Lossy Compressibility through Statistical Correlations of Geophysical Datasets	Virtual
Session entitled “ Meeting Exascale Computing Challenges with Compression and Pangeo”	May 2022
European Geospatial Union General Assembly	
LibPressio	Virtual
Session entitled “Lossy Data Reduction for ECP Applications”	April 2021
Exascale Computing Project Annual Meeting	
Approachable Error Bounded Lossy Compression	Virtual
An Interface, Automated Tuning, and analysis for lossy compression	November 2020
Super Computing 2020, Doctoral Showcase	
Lossy Compression for AI	Virtual
An overview of how to use lossy compression to reduce storage needs for AI	September 2020
Joint Laboratory for Extreme Scale Computing	
FRaZ	Virtual
A Generic High Fidelity Fixed Ratio Lossy Compression Framework for Floating Point Scientific Data; IPDPS 2020	May 2020
LibPressio: A Generic Abstraction for Compression	Huston, TX
Session entitled “Lossy Data Reduction/Compression for ECP Applications”	February 2020
Exascale Computing Project Annual Meeting	
Approachable Error Bounded Lossy Compression	Lemont, IL
Overview of tools and techniques for using error bounded lossy compression	December 2019
Argonne National Laboratory Mathematics and Computer Science Division Seminar	
Predicting Optimal Error-Bounded-Lossy-Compression Configuration	Huston, TX
Techniques for predicting error bounded lossy compression ratios	November 2018
Supercomputing 18 Student Volunteer Talks	

## Computer Skills

---

Advanced: Bash, Bourne Shell, C, C++, Linux Containers (podman, docker, ...), Linux Kernel and Userspace, Python, Vim

Intermediate: Ansible, Cuda, Git, Hadoop, JAVA, JavaScript, Keras, HDF5, Julia, L<sup>A</sup>T<sub>E</sub>X, Linux Profiling (perf), LLVM-libtooling, MPI, OpenCL, OpenMP, SQL, Salt-

Stack, SciKit Learn, Spack, Systemd, Tensorflow 2.x, SWIG, TypeScript, vLLM, PyTorch

Basic: Apache Spark, ARM assembly, CSS3 C#, Haskell, HTML5, Perl, Puppet, RCpp/RInside, Rust, SNMP, SVN, FreeBSD, PHP

## Professional Affiliations

---

Association for Computing Machinery (ACM): Professional Member 2022-Present

Association for Computing Machinery (ACM): Student Member 2014-2021

Institute of Electrical and Electronics Engineers (IEEE): Professional Member 2023-Present

Joint Laboratory for Extreme Scale Computing (JLESC): 2018-Present

## Professional Service

---

JLESC Early Career Ambassador for Argonne National Laboratory: 2023-present

Program Chair: HPAI4S 2025

Session Chair: JLESC 2022-2024

Reviewer: ICPE 2017, ICCCN 2017, PABS 2017, SC17, IEEE CLOUD 2018, IEEE TSE 2018, IPDPS 2018, IPDPS 2019, IEEE CLUSTER 2020, Sustainable Computing 2021, IEEE Big Data 2021, PacificVis 2022, HPDC 2023, IEEE TC 2023, SC23-Posters 2023, FGCS 2024, ATPETSC 2024, ATPETSC 2025, ISC 2025, ACM ICS 2025, IEEE Cluster 2025, et al

Volunteer: SC2018, SC21

## Conference/Workshop/Tutorial/BoF Organization

---

Workshops: Joint Laboratory for Extreme Scale Computing (JLESC) Breakout Session Organizer on AI '24-'25, Trillion Parameter Consortium (Oct'24, July'25), Program Chair High Performance AI for Science (HPAI4S) (Jun'25)

Tutorials: Supercomputing (SC) Data Reduction Tutorial Presenter, (Nov'22-'25), Super Computing AI Evaluation (Nov '25)

Events: Argonne AI Evaluation Hackathons (Jan'24, July'24, Aug'24), Postdoctoral Research and Career Symposium – Program Committee Member (Nov 2023), 1000 Scientist JAM – Data Collection (Feb'25)

## Honors

---

- Impact Argonne Award for contributions to the TPC effort led by Charlie Catlett, with other vital contributions by Franck Cappello and Arvind Rathmanaman
- Outstanding Paper, IEEE HPEC 2025
- Best Paper Runner Up, (ACM SC-W, DRBSD) 2025

- Outstanding Postdoctoral Performance Award in Basic Science, Argonne National Laboratory 2024
- Best Paper, Data Reduction for Big Scientific Data Workshop (ACM SC-W, DRBSD) 2024
- Best Paper, The International ACM Symposium on High-Performance Parallel and Distributed Computing (ACM HPDC), 2024
- Best Poster, ACM Student Research Competition at Supercomputing (ACM SC), 2022
- Best Paper, Data Reduction for Big Scientific Data Workshop (ACM SC-W, DRBSD) 2022
- Clemson Outstanding Ph.D. in Computer Science Award, 2021
- Graduate Student Research Lighting Talk Competition Faculty Award, 2020
- Department of Energy Office of Science Graduate Student Research Award, 2019
- Fellowship, National Research Traineeship: Resilient Infrastructure Systems 2017-2020
- National Science Foundation Graduate Research Fellowship Honorable Mention 2017
- Faculty Scholarship Award, Clemson University 2016
- Benefitfocus Scholarship 2015-2016
- McAlister Scholarship 2015-2016
- Palmetto Fellows Recipient 2013-2016
- President's List at Clemson University 2013-2016
- Outstanding Sophomore in Computer Science at Clemson University 2015
- Order of the Arrow, Vigil Honor 2013
- Eagle Scout 2010