Jieyang Chen

(+1) (716)480-7217 ⋄ chenj3@ornl.gov
Computer Science and Mathematics Division
Oak Ridge National Laboratory, Oak Ridge, TN, 37830

EDUCATION

Ph.D. in Computer Science, advisor: Dr. Zizhong Chen — *09/2015 - 03/2019*
University of California, Riverside — GPA: 3.8/4.0

Master in Computer Science, advisor: Dr. Zizhong Chen — *09/2012 - 12/2014*
University of California, Riverside — GPA: 3.8/4.0

Bachelor in Computer Science and Technology, advisor: Dr. Dan Wang — *09/2008 - 07/2012*
Beijing University of Technology — GPA: 3.5/4.0

RESEARCH INTERESTS

- High Performance Computing
- Scientific Data Management
- Scientific Visualization/Data Analytics
- GPU Computing

WORK EXPERIENCE

Computer Scientist — 12/2020 - present
Oak Ridge National Laboratory

· *Scientific Visualization with Lossy Compression (ECP-ALPINE)*: Studied the performance-quality trade-offs in scientific visualization workflows with lossy compression.
· *GPU-accelerated Lossy Compression (ECP-CODAR)*: Designed and developed multigrid-based scientific data refactoring/reduction software for modern GPU accelerators.
· *Multi-/Many-core accelerated visualization toolkit (ECP-VTKM)*: Designed and developed scientific visualization tools for the next generations of multi-/many-core accelerated computing systems.

Postdoctoral Research Associate — 05/2019 - 11/2020
Oak Ridge National Laboratory

· *Scientific Visualization with Lossy Compression (ECP-ALPINE)*: Studied the performance-quality trade-offs in scientific visualization workflows with lossy compression.
· *GPU-accelerated Lossy Compression (ECP-CODAR)*: Designed and developed multigrid-based scientific data refactoring/reduction software for modern GPU accelerators.

Research Intern — 07/2018 - 09/2018
Pacific Northwest National Laboratory

· *Evaluating Multi-GPU Systems (ASCR-CENATE):* Developed benchmark suits for evaluating modern GPU interconnect: PCIe, NVLink, NVSwitch, and GPUDirect RDMA.
· *Sparse BLAS on Multi-GPU Systems (PNNL-HPDA):* Designed optimized data structure and scheduling strategies to speed up core sparse linear algebra kernels on multi-GPU systems.

Research Intern — 01/2017 - 06/2017
Los Alamos National Laboratory

· *Containerized In Situ Visualization Workflow System (ECP-BEE):* Designed and developed a workflow management system for launching containerized execution environment for scientific in situ visualization workflows on HPC and cloud infrastructure.

Research Assistant 11/2014 - 03/2019
University of California, Riverside

· *GPU Fault Tolerance:* Designed and developed fault tolerance matrix decomposition algorithms (Cholesky, LU, and QR) that can detect and correct silent data corruptions (SDCs) on heterogeneous systems with GPUs.
· *GPU Energy Efficiency Optimization:* Designed and developed energy efficient matrix decomposition algorithms based on Dynamic Voltage and Frequency Scaling (DVFS) and slack prediction on heterogeneous systems with GPUs.
· *GPU Performance Optimization:* Designed and developed the optimized matrix-matrix multiplication specialized for tall-and-skinny inputs on modern GPUs.

Teaching Assistant 01/2016-06/2016, 09/2018-12/2018
University of California, Riverside

· CS 211: High Performance Computing
· CS 180: Introduction to Software Engineering
· CS 8: Introduction to Computing
· CS 177: Modeling and Simulation

PUBLICATIONS (C/S: FULL/SHORT CONFERENCE PAPERS; J: JOURNAL PAPERS)

- **[J6]** Xin Liang, Ben Whitney, **Jieyang Chen**, Lipeng Wan, Qing Liu, Dingwen Tao, James Kress, Dave Pugmire, Matthew Wolf, Norbert Podhorszki, Scott Klasky. MGARD+: Optimizing Multilevel Methods for Error-bounded Scientific Data Reduction. *IEEE Transactions on Computers, 2021.*
- **[C23]** Jiannan Tian, Cody Rivera, Sheng Di, **Jieyang Chen**, Xin Liang, Dingwen Tao, and Franck Cappello. Revisiting Huffman Coding: Toward Extreme Performance on Modern GPU Architectures. *35th IEEE International Parallel and Distributed Processing Symposium, May 2021, Portland, OR.*
- **[C22]** Huizhang Luo, Junqi Wangy, Qing Liu, **Jieyang Chen**, Scott Klasky, Norbert Podhorszki, zMesh: Exploring Application Characteristics to Improve Lossy Compression Ratio for Adaptive Mesh Refinement. *35th IEEE International Parallel and Distributed Processing Symposium, May 2021, Portland, OR.*
- **[C21]** **Jieyang Chen**, Lipeng Wan, Xin Liang, Ben Whitney, Qing Liu, David Pugmire, Nicholas Thompson, Jong Youl Choi, Matthew Wolf, Todd Munson, Ian Foster, Scott Klasky. Accelerating Multigrid-based Hierarchical Scientific Data Refactoring on GPUs. *35th IEEE International Parallel and Distributed Processing Symposium, May 2021, Portland, OR.*
- **[J5]** Cody Rivera, **Jieyang Chen**, Nan Xiong, Jing Zhang, Shuaiwen Leon Song, Dingwen Tao. TSM2X: High-performance tall-and-skinny matrix–matrix multiplication on GPUs. *Elsevier Journal of Parallel and Distributed Computing (JPDC), 2021.*
- **[J4]** Kai Zhao, Sheng Di, Sihuan Li, Xin Liang, Yujia Zhai, **Jieyang Chen**, Kaiming Ouyang, Zizhong Chen, and Franck Cappello. FT-CNN: Algorithm-Based Fault Tolerance for Convolutional Neural Networks. *IEEE TPDS Special Section on Parallel and Distributed Computing Techniques for AI/ML/DL (2021).*
- **[C20]** Igor Yakushin, Kshitij Mehta, **Jieyang Chen**, Matthew Wolf, Ian Foster, Scott Klasky, Todd Munson. Feature-preserving Lossy Compression for In Situ Data Analysis. *49th International Conference on Parallel Processing-ICPP: Workshops, August 2020.*

- **[C19]** Lipeng Wan, Matthew Wolf, Feiyi Wang, Jong Youl Choi, George Ostrouchov, **Jieyang Chen**, Norbert Podhorszki, Jeremy Logan, Kshitij Mehta, Scott Klasky, Dave Pugmire. I/O Performance Characterization and Prediction through Machine Learning on HPC Systems. *Cray User Group Conference, 2020.*

- **[C18]** Bingbing Li, Santosh Pandey, Haowen Fang, Yanjun Lyv, Ji Li, **Jieyang Chen**, Mimi Xie, Lipeng Wan, Hang Liu, Caiwen Ding. FTRANS: energy-efficient acceleration of transformers using FPGA. *the ACM/IEEE International Symposium on Low Power Electronics and Design, 2020.*

- **[C17]** Zhenbo Qiao, Qing Liu, Norbert Podhorszki, Scott Klasky, **Jieyang Chen**. Taming I/O Variation on QoS-Less HPC Storage: What Can Applications Do? *Proceedings of the 32nd ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis, 2020.*

- **[C16]** David Pugmire, James Kress, **Jieyang Chen**, Hank Childs, Jong Choi, Dmitry Ganyushin, Berk Geveci, Mark Kim, Scott Klasky, Xin Liang, Jeremy Logan, Nicole Marsaglia, Kshitij Mehta, Norbert Podhorszki, Caitlin Ross, Eric Suchyta, Nick Thompson, Steven Walton, Lipeng Wan, and Matthew Wolf. Visualization as a Service for Scientific Data. *Smoky Mountains Computational Science and Engineering Conference (2020).*

- **[J3]** Zhenlu Qin, Jinzhen Wang, Qing Liu, **Jieyang Chen**, Dave Pugmire, Norbert Podhorszki, Scott Klasky. Estimating Lossy Compressibility of Scientific Data Using Deep Neural Networks. *IEEE Letters of the Computer Society (2020).*

- **[J2]** Jeremy Logan, Mark Ainsworth, Chuck Atkins, **Jieyang Chen**, Jong Choi, Junmin Gu, James Kress, Greg Eisenhauer, Berk Geveci, William Godoy, Mark Kim, Tahsin Kurc, Qing Liu, Kshitij Mehta, George Ostrouchov, Norbert Podhorzski, David Pugmire, Eric Suchyta, Nicolas Thompson, Ozan Tugluk, Lipeng Wan, Ruonan Wang, Ben Whitney, Matthew Wolf, Kesheng Wu, Scott Klasky. Extending the Publish/Subscribe Abstraction for High-Performance I/O and Data Management at Extreme Scale. *IEEE The Bulletin of the Technical Committee on Data Engineering (2020).*

- **[C15]** Sihuan Li, Hongbo Li, Xin Liang, **Jieyang Chen**, Elisabeth Giem, Kaiming Ouyang, Kai Zhao, Sheng Di, Franck Cappello, and Zizhong Chen. FT-iSort: Efficient Fault Tolerance for Introsort, *Proceedings of the 31st ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis*, Denver, Colorado, USA, Nov 17 - 22, 2019. Acceptance Rate: 20.9% (72/344).

- **[C14]** **Jieyang Chen**, Nan Xiong, Xin Liang, Dingwen Tao, Sihuan Li, Kaiming Ouyang, Nathan DeBardeleben, Qiang Guan, Zizhong Chen. TSM2: Optimizing Tall-and-skinny Matrix-Matrix Multiplication on GPUs. *33rd ACM International Conference on Supercomputing*, Phoenix, AZ, USA, June 26-28, 2019, Acceptance Rate: 23.3% (45/193).

- **[C13]** **Jieyang Chen**, David Pugmire, Matthew Wolf, Nicholas Thompson, Jeremy Logan,Kshitij Mehta, Lipeng Wan, Jong Youl Choi, Ben Whitney, Scott Klasky. A Understanding Performance-Quality Trade-offs in Scientific Visualization Workflows with Lossy Compression, *The 5th International Workshop on Data Reduction for Big Scientific Data*, Denver, CO, USA, Nov. 17, 2019.

- **[C12]** Jong Youl Choi, Jeremy Logan, Kshitij Mehta,Eric Suchyta, William Godoy, Nicholas Thompson, Lipeng Wan, **Jieyang Chen**, Norbert Podhorszki, Matthew Wolf, Scott Klasky, Julien Dominski and Choong-Seock Chang. A Co-Design Study Of Fusion Whole DeviceModeling Using Code Coupling, *The 5th International Workshop on Data Reduction for Big Scientific Data*, Denver, CO, USA, Nov. 17, 2019.

- **[S2]** Bo Fang, **Jieyang Chen**, Karthik Pattabiraman, Matei Ripeanu, Sriram Krishnamoorthy. Towards Predicting the Impact of Roll-Forward Failure Recovery for HPCApplications, *the 49th*

Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN 2019), Portland, Oregon, USA, June 24 – 27, 2019.

- **[J1]** Ang Li, Shuaiwen Leon Song, **Jieyang Chen**, Jiajia Li, Xu Liu, Nathan Tallent, and Kevin Barker. Evaluating Modern GPU Interconnect: PCIe, NVLink, NV-SLI, NVSwitch and GPUDirect. *IEEE Transactions on Parallel and Distributed Systems.*

- **[C11] Jieyang Chen**, Qiang Guan, Xin Liang, Paul Bryant, Patricia Grubel, Allen McPherson, Li-Ta Lo, Timothy Randles, Zizhong Chen and James Ahrens. Build and Execution Environment (BEE): an Encapsulated Environment Enabling HPC Applications Running Everywhere. *Proceedings of the 2018 IEEE International Conference on Big Data* Seattle, WA, USA, Dec 10-13, 2018. Acceptance Rate: 18.9% (98/518).

- **[C10]** Ang Li, Shuaiwen Leon Song, **Jieyang Chen**, Xu Liu, Nathan Tallent, Kevin Barker. Tartan: Evaluating Modern GPU Interconnect via a Multi-GPU Benchmark Suite, *2018 IEEE International Symposium on Workload Characterization*, Sep 30-Oct 2, 2018, Raleigh, NC, USA.

- **[C9] Jieyang Chen**, Hongbo Li, Sihuan Li, Panruo Wu, Xin Liang, Dingwen Tao, Kaiming Ouyang, Yuanlai Liu, Kai Zhao, Qiang Guan, and Zizhong Chen. Fault Tolerant Dense Matrix Decomposition on Heterogeneous Systems with GPUs, *Proceedings of the 30th ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis*, Dallas, Texas, USA, Nov 11 - 16, 2018. Acceptance Rate: 19.1% (55/288).

- **[C8] Jieyang Chen**, Qiang Guan, Zhao Zhang, Xin Liang, Louis James Vernon, Allen McPherson, Li-Ta Lo, Patricia Grubel, Tim Randles, Zizhong Chen, and James Paul Ahrens. `BeeFlow` : A Workflow Management System for In-Situ Processing Across HPC and Cloud Systems, *38th IEEE International Conference on Distributed Computing Systems*, July 2 – 5, 2018, Vienna, Austria. Acceptance Rate: 20.6% (78/378).

- **[C7]** Xin Liang, **Jieyang Chen**, Dingwen Tao, Sihuan Li, Panruo Wu, Hongbo Li, Kaiming Ouyang, Yuanlai Liu, Fengguang Song, and Zizhong Chen. Correcting Soft Errors Online in Fast Fourier Transform, *Proceedings of the 29th ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis*, Denver, Colorado, USA, Nov 12 - 17, 2017. Acceptance Rate: 18.6% (61/327).

- **[C6]** Panruo Wu, Qiang Guan, Nathan DeBardeleben, Sean Blanchard, **Jieyang Chen**, Dingwen Tao, Xin Liang, Sihuan Li, Kaiming Ouyang, and Zizhong Chen. Silent Data Corruption Resilient Two-sided Matrix Factorizations, *Proceedings of the 22nd ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming*, Austin, Texas, USA, February 4 - 8 2017. Acceptance Rate: 21.9% (29/132).

- **[C5] Jieyang Chen***, Li Tan*, Panruo Wu, Dingwen Tao, Hongbo Li, Xin Liang, Sihuan Li, Rong Ge, Laxmi Bhuyan, and Zizhong Chen. GreenLA: Green Linear Algebra Software for GPU-Accelerated Heterogeneous Computing, *Proceedings of the 28th ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis*, Salt Lake City, Utah, USA, Nov 13 - 18, 2016. Acceptance Rate: 18.4% (82/446). *Authors contributed equally.

- **[C4] Jieyang Chen**, Xin Liang, and Zizhong Chen. Online Algorithm-Based Fault Tolerance for Cholesky Decomposition on Heterogeneous Systems with GPUs, *Proceedings of the 30th IEEE International Parallel & Distributed Processing Symposium*, Chicago, Illinois, USA, May 23 - 27, 2016. Acceptance Rate: 22.98% (114/496).

- **[C3]** Panruo Wu, Nathan DeBardeleben, Qiang Guan, Sean Blanchard, Dingwen Tao, Xin Liang, **Jieyang Chen**, and Zizhong Chen. Towards Practical Algorithm Based Fault Tolerance in Dense Linear Algebra, *Proceedings of the 25th ACM International Symposium on High-Performance Parallel and Distributed Computing*, Kyoto, JAPAN, May 31- June 4, 2016. Acceptance Rate: 15.5% (20/129).

- **[S1] Jieyang Chen**, Sihuan Li, and Zizhong Chen. GPU-ABFT: Optimizing algorithm-based fault tolerance for heterogeneous systems with GPUs, *IEEE International Conference on Networking, Architecture and Storage (NAS)*, Long Beach, CA, August 8 - 10, 2016.
- **[C2]** Teresa Davies, Xin Liang, **Jieyang Chen**, Zizhong Chen. Simulated Annealing to Generate Numerically Stable Real Number Error Correction Codes, *Proceedings of the 2015 IEEE 17th International Conference on High Performance Computing and Communications, 2015 IEEE 7th International Symposium on Cyberspace Safety and Security, and 2015 IEEE 12th International Conference on Embedded Software and Systems*, New York, USA, August 24 - 26, 2015
- **[C1] Jieyang Chen** and Zizhong Chen. Cholesky Factorization on Heterogeneous CPU and GPU Systems, *9th IEEE International Conference on Foundations of Computer Science & Technology (FCST)*, Dailian, China, August 26 - 28, 2015

TALKS AND PRESENTATIONS

Accelerating Multigrid-based Hierarchical Scientific Data Refactoring on GPUs. 05/2021
35th IEEE International Parallel and Distributed Processing Symposium *Virtual*

Progressive Visualization via Hierarchical Scientific Data Refactoring on GPUs. 04/2021
US Department of Energy Computer Graphics Forum *Virtual*

Understanding Performance-Quality Trade-offs in Scientific Visualization Workflows with Lossy Compression, 11/2019
The 5th International Workshop on Data Reduction for Big Scientific Data *Denver, CO*

TSM2: Optimizing Tall-and-skinny Matrix-Matrix Multiplication on GPUs 06/2019
33rd ACM International Conference on Supercomputing *Phoenix, AZ*

Fault tolerant and Energy Efficient One-sided Matrix decompositions on heterogeneous systems with GPUs 03/2019
Ph.D. Dissertation Defence *Riverside, CA*

High Performance Computing at Extreme Scale: Resilience, Energy Efficiency, and Performance 02/2019
New Mexico State University *Las Cruces, NM*

High Performance Computing at Extreme Scale: Resilience, Energy Efficiency, and Performance 02/2019
Oak Ridge National Laboratory *Oak Ridge, TN*

High Performance Computing at Extreme Scale: Resilience, Energy Efficiency, and Performance 02/2019
Tennessee Tech University *Cookville, TN*

High Performance Computing at Extreme Scale: Resilience, Energy Efficiency, and Performance 02/2019
San Francisco State University *San Francisco, CA*

Build and Execution Environment (BEE): an Encapsulated Environment Enabling HPC Applications Running Everywhere. 12/2018
Proceedings of the 2018 IEEE International Conference on Big Data *Seattle, WA*

Fault Tolerant Dense Matrix Decomposition on Heterogeneous Systems with GPUs 12/2018

Proceedings of the 30th ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis *Dallas, TX*

High Performance Computing at Extreme Scale: Resilience, Energy Efficiency, and Performance 04/2018
Southern Illinois University *Carbondale, IL*

GreenLA: Green Linear Algebra Software for GPU-Accelerated Heterogeneous Computing 11/2016
Proceedings of the 28th ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis *Salt Lake City, UT*

Online Algorithm-Based Fault Tolerance for Cholesky Decomposition on Heterogeneous Systems with GPUs 05/2016
Proceedings of the 30th IEEE International Parallel & Distributed Processing Symposium Chicago, IL

AWARDED FUNDING

ESAMR: Enabling Scalable Analytics using Multiprecision Refactoring 10/2020-09/2022
Awarded amount: $640,000/2 yrs (Role: Co-PI) ORNL Laboratory Directed Research & Development

SOFTWARE

GPU-MGARD: GPU-accelerated multigrid-based lossy compression software

· Link: github.com/CODARCode/MGARD

Tartan: benchmark suite for evaluating modern GPU interconnect

· Link: github.com/uuudown/Tartan

BEE: containerized scientific in situ workflow management system for HPC applications

· Link: github.com/lanl/BEE

AWARDS

2017 Dissertation Year Program (DYP) Fellowship, University of California, Riverside

2015 Ph.D. Department Fellowship, University of California, Riverside

2011 Meritorious Winner in US Mathematical Contest in Modeling

SERVICES

Reviewer/Committee: IEEE IGSC'18, Elsevier SUSCOM'19, IEEE Access, SELSE'20, Journal of Systems Architecture, ACM Transactions on Knowledge Discovery from Data, China VIS'20, IEEE HPCC'20, IEEE IWBDR'20, ACM PPoPP'20 AE, IEEE TCAD'21, SELSE'21, IEEE ScalCom'21.

Subreviewer: ACM PACT'15, ACM/IEEE SC'15, IEEE IPDPS'16, ACM ICS'16, ACM/IEEE SC'16, IEEE IPDPS'17, ACM/IEEE SC'17, ACM HPDC'18, ACM/IEEE SC'18, IEEE VIS'20, IEEE Cluster'20, IEEE IPDPS'21, ACM ICS'21, IEEE ICDCS'21, ICPP'21.

Volunteer Work: ACM/IEEE SC'14 Student Volunteer, ACM/IEEE SC'15 Student Volunteer.

TECHNICAL SKILLS

Programming Languages	C/C++, CUDA (C/ptx), Python, Java, MATLAB, Shell Script
Software & Tools	Paraview, VisIt, VTK-m, ADIOS, MGARD, SZ, ZFP, Docker, QEMU/KVM, LAPACK, MAGMA, cuBLAS, cuLA.