



# Bin Ren@William & Mary

## PERSONAL INFORMATION

1. Name: Bin Ren  
Title: Associate Professor@Computer Science  
Email: bren [at] wm.edu  
Website: <https://www.cs.wm.edu/~bren/>  
Date: April, 2024
2. Research Interests:  
Parallel Computing and High-Performance Computing  
Compiler Techniques  
Real-time Machine Learning and Machine Learning Systems

## EDUCATION

Ph.D. Computer Science and Engineering, **The Ohio State University**, 9/2008 - 5/2014  
M.S. Computer Science and Engineering, **The Ohio State University**, 9/2008 - 8/2013  
M.S. Software Engineering, **Beihang University** (China), 9/2006 - 6/2008  
B.S. Software Engineering, **Beihang University** (China), 9/2002 - 6/2006

## ACADEMIC POSITIONS

8/2022- **The College of William & Mary**  
present Associate Professor

8/2016- **The College of William & Mary**  
7/2022 Assistant Professor

5/2014- **Pacific Northwest National Laboratory**  
7/2016 Postdoctoral Research Associate (Supervisor: Dr. Sriram Krishnamoorthy)

9/2008- **The Ohio State University**  
5/2014 Graduate Research/Teaching Associate (Ph.D. Advisor: Prof. Gagan Agrawal)

5/2013- **NEC Laboratories America**  
8/2013 Research Internship (Mentor: Dr. Nishkam Ravi and Dr. Yi Yang)

6/2011- **Microsoft Research, Redmond**  
9/2011 Research Internship (Mentor: Dr. Todd Mytkowicz)

## **HONORS, PRIZES AND AWARDS**

- NSF CAREER Award, 2021
- Best Paper Award, SC 2020
- Best Student Paper Nomination, SC 2020
- Jeffress Trust Award, 2020
- ISLPED Design Contest First Place, 2020
- Student Cluster Reproducibility Challenge Paper, SC 2019
- Best Paper Award, CGO 2013
- SIGPLAN Research Highlights, 2013
- Student Travel Awards, PACT, 2011, 2012
- University Fellowship, The Ohio State University, 2008-2009
- Outstanding Graduates Award, Beihang University, 2006

## **COURSES TAUGHT**

### *Instructor*

SP24 CS642 Compiler Techniques for High Performance Computing  
AU23 CS304 Computer Organization  
SP23 CS642 Compiler Techniques for High Performance Computing  
SP22 CS680 Compiler & Parallel Computing  
AU21 CS304 Computer Organization  
SP21 CS304 Computer Organization  
AU20 CS680 Compiler & Parallel Computing  
SP20 CS680 Compiler & Parallel Computing  
AU19 CS304 Computer Organization  
SP19 CS680 Compiler & Parallel Computing  
AU18 CS304 Computer Organization  
SP18 CS780 Compiler Optimization for High Performance Computing  
AU17 CS304 Computer Organization  
SP17 CS304 Computer Organization  
AU16 CS680 Compiler Optimization for High Performance Computing

## **FELLOWSHIPS AND GRANTS**

### *Awarded:*

- [1] [NSF-OAC'24] "Collaborative Research: OAC Core: CropDL - Scheduling and Checkpoint/Restart Support for Deep Learning Applications on HPC Clusters," **awarded**, leading **PI**, 10/2024-09/2027, NSF, \$224,984 (Total \$600,000).
- [2] [NIH-R01'23] "SCH: Novel and Interpretable Statistical Learning for Brain Images in AD/ADRDs," **awarded**, **Co-PI**, 09/2023-08/2027, NIH, \$217,887 (Total \$1.2M).
- [3] [NSF-CSR'23] "Collaborative Research: CNS Core: Small: A Compilation System for Mapping Deep Learning Models to Tensorized Instructions (DELITE)," **awarded**, leading **PI**,

10/2023-09/2026, NSF, \$299,999 (Total \$600,000).

[4] [NSF-SHF'22] "Collaborative Research: SHF: SMALL: Compile-Parallelize-Schedule-Retarget-Repeat (EASER) Paradigm for Dealing with Extreme Heterogeneity," **awarded, PI**, 06/2022-06/2025, NSF, \$250,000 (Total \$500,000).

[5] [NSF-CAREER'21] "CAREER: Achieving Real-Time Machine Learning with Sparsification-Compilation Co-design," **awarded**, sole **PI**, 10/2021-09/2026, NSF, \$493,685.

[6] [Jeffress-Memorial-Trust-Award'20] "A General, Real-time DNN Execution Framework on Mobile Devices," **awarded**, sole **PI**, 06/2020-06/2021, Jeffress Memorial Trust, \$104,500.

[7] [Kwai'20] "Mobile System Analysis and Optimization for Neural Network Execution," **awarded**, sole **PI**, 01/2021-12/2021, Kwai Inc. USA, \$50,000.

[8] [DoE-ECP/JLab'19/20/21] "Accelerating QCD with GPUs," **graduate student support**, 08/2019-08/2022, DoE, \$85,500.

[9] [NSF-SBIR'20] "CoCoPIE: Enabling Real-Time AI on End Devices through Compression-Compilation Co-Design," NSF issued to CoCoPIE co-founded by NEU, NCSU, and W&M, \$250,000.

[10] [NSF-III'20] "III: Small: Combinatorial Algorithms for High-dimensional Learning", **Co-PI**, transferred from Zhenming Liu, 10/2020-09/2023, NSF, \$395,403.

[11] [NSF-IIS'21] "EAGER: Collaborative Research: On the Theoretical Foundation of Recommendation System Evaluation", **PI**, transferred from Zhenming Liu, 09/2021-02/2023, NSF, \$19,999 (Total: \$100,000).

#### Internal Awards:

[1] [WM-Summer'18] "Exploiting Emerging Throughput-oriented Architectural Advances for Irregular Applications," summer grant, May, 2018

[2] [WM-Summer'17] "Improving Vectorization of Irregular Applications by Unified Static and Dynamic Analyses," summer grant, May, 2017

### **RESEARCH**

#### Peer-Reviewed Journals

[1] [ACM TORS'23] Dong Li, Ruoming Jin, Zhenming Liu, **Bin Ren**, Jing Gao, Zhi Liu, "On Item-Sampling Evaluation for Recommender System," ACM Transactions on Recommender Systems, 2023.

[2] [ACM Computing Surveys'22] Jou-An Chen, **Wei Niu**, **Bin Ren**, Yanzhi Wang, Xipeng Shen, "Survey: Exploiting Data Redundancy for Optimization of Deep Learning," ACM Computing Surveys, 2022.

[3] [TACO'22] **Qihan Wang**, **Zhen Peng**, **Bin Ren**, Jie Chen, Robert G. Edwards, "MemHC: An Optimized GPU Memory Management Framework for Accelerating Many-body Correlation," ACM Transactions on Architecture and Code Optimization (ACM TACO), June 2022. (**Original Work, Invited to HiPEAC'23**)

- [4] [TECS'22] Geng Yuan, Mengshu Sun, **Wei Niu**, Zhengang Li, Yuxuan Cai, Yanyu Li, Jun Liu, Weiwen Jiang, Xue Lin, **Bin Ren**, Xulong Tang, Yanzhi Wang, “*Mobile or FPGA? A Comprehensive Evaluation on Energy Efficiency and a Unified Optimization Framework*,” ACM Transactions on Embedded Computing Systems (TECS).
- [5] [TPDS'22] Mert Hidayetoglu, Tekin Bicer, Simon Garcia de Gonzalo, **Bin Ren**, Doga Gursoy, Rajkumar Kettimuthu, Ian T. Foster, Wen-Mei W. Hwu, “*MemXCT: Design, Optimization, Scaling, and Reproducibility of X-Ray Tomography Imaging*,” IEEE Transactions on Parallel and Distributed Systems. 33(9): 2014-2031 (2022).
- [6] [TODAES'21] Yifan Gong\*, Geng Yuan\*, Zheng Zhan, **Wei Niu**, Zhengang Li, Pu Zhao, Yuxuan Cai, Sijia Liu, **Bin Ren**, Xue Lin, Xulong Tang, and Yanzhi Wang, “*Automatic mapping of the best-suited DNN pruning schemes for real-time mobile acceleration*,” ACM Transactions on Design Automation of Embedded Systems.
- [7] [CACM'21] Hui Guan, Shaoshan Liu, Xiaolong Ma, **Wei Niu**, **Bin Ren**, Xipeng Shen, Yanzhi Wang, Pu Zhao (**author in alphabet order**), “*CoCoPIE: Making Mobile AI Sweet as PIE - Compression-Compilation Co-Design Goes a Long Way*,” Communications of the ACM (CACM), page 62-68, June 2021 (**flagship journal of ACM, featured with a video report**).
- [8] [TPAMI'21] **Wei Niu**, Zhenggang Li, Xiaolong Ma, Peiyan Dong, Gang Zhou, Xuehai Qian, Xue Lin, Yanzhi Wang, and **Bin Ren**, “GRIM: A General, Real-Time Deep Learning Inference Framework for Mobile Devices based on Fine-Grained Structured Weight Sparsity”, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2021 (**impact factor: 17.86**).
- [9] [Smart Health'21] Shuangquan Wang, Gang Zhou, **Jiexiong Guan**, Yongsen Ma, Zhenming Liu, **Bin Ren**, Hongyang Zhao, Amanda Watson, and Woosub Jung, “*Inferring Food Types through Sensing and Characterizing Mastication Dynamics*,” Elsevier Smart Health, page 1-15, January 2021.
- [10] [Smart Health'20] **Qihan Wang**, Gang Zhou, Zhenming Liu, and **Bin Ren**, “*Building a Skeleton-based 3D Body Model with Angle Sensor Data*,” Elsevier Smart Health, page 1-15, November 2020.
- [11][TPDS'20] Soklong Lim, Tyler Coy, Zaixin Lu, **Bin Ren**, and Xuechen Zhang, “*NVGRAPH: Enforcing Crash Consistency of Evolving Network Analytics in NVMM Systems*,” The IEEE Transactions on Parallel and Distributed Systems (TPDS), page 124-137, January, 2020.

[12][TOPC'19] **Bin Ren**, Shruthi Balakrishna, Youngjoon Jo, Sriram Krishnamoorthy, Kunal Agrawal, and Milind Kulkarni, “*Extracting SIMD Parallelism from Recursive Task-Parallel Programs*,” The ACM Transaction on Parallel Computing (TOPC), page 24:1-24:37, December, 2019.

[13][TACO'17] Mehmet Can Kurt, Sriram Krishnamoorthy, Gagan Agrawal, and **Bin Ren**, “*User-Assisted Store Recycling for Dynamic Task Graph Schedulers*,” ACM Transactions on Architecture and Code Optimization (TACO), page 55:1-55:24, December, 2016. (**Original work, invited to HiPEAC'17**).

[14][TACO'14] **Bin Ren**, Todd Mytkowicz, and Gagan Agrawal, “*A Portable Optimization Engine for Accelerating Irregular Data-Traversal Applications on SIMD Architectures*,” ACM Transactions on Architecture and Code Optimization (TACO), page 16:1-16:31, 2014.

#### Peer-Reviewed Conferences

[1] [ASPLOS'24] **Wei Niu**, Md Musfiqur Rahman Sanim, Zhihao Shu, **Jiexiong Guan**, Xipeng Shen, Miao Yin, Gagan Agrawal, **Bin Ren**, “*SmartMem: Layout Transformation Elimination and Adaptation for Efficient DNN Execution on Mobile*,” International Conference on Architectural Support for Programming Languages and Operating Systems, 2024.

[2] [ASPLOS'24] **Wei Niu**, Gagan Agrawal, **Bin Ren**, “*SoD<sup>2</sup>: Statically Optimizing Dynamic Deep Neural Network Execution*,” International Conference on Architectural Support for Programming Languages and Operating Systems, 2024.

[3] [IPDPS'2024] Malith Jayaweera, Yanyu Li, Yanzhi Wang, **Bin Ren**, David Kaeli, “*DEFCON: Deformable Convolutions Leveraging Interval Search and GPU Texture Hardware*,” 38th IEEE International Parallel & Distributed Processing Symposium (IPDPS), 2024

[4] [ICLR'2024] Gen Li, Lu Yin, Jie Ji, **Wei Niu**, Minghai Qin, **Bin Ren**, Linke Guo, Shiwei Liu, Xiaolong Ma, “*NeurRev: Train Better Sparse Neural Network Practically via Neuron Revitalization*,” The Twelfth International Conference on Learning Representations (ICLR), 2024.

[5] [USENIX ATC'23] Hsin-Hsuan Sung, Jou-An Chen, **Wei Niu**, **Jiexiong Guan**, **Bin Ren**, Xipeng Shen, “*Decentralized Application-Level Adaptive Scheduling for Multi-Instance DNNs on Open Mobile Devices*,” USENIX Annual Technical Conference, 2023. (Acceptance rate: 65/353=18.4%).

[6] [PPoPP'23] **Zhen Peng**, Minjia Zhang, Kai Li, Ruoming Jin, **Bin Ren**, “*iQAN: Fast and Accurate Vector Search with Efficient Intra-Query Parallelism on Multi-Core Architectures*,” The 28th ACM SIGPLAN Annual Symposium on Principles and Practice of Parallel Programming, 2023 (Acceptance rate: 23.7% = 31/131)

[7] [ICDM'2023] Hao Zhang, Malith Jayaweera, **Bin Ren**, Yanzhi Wang, Sucheta Soundarajan, “*Unfairness in Distributed Graph Frameworks*,” IEEE International Conference on Data Mining (ICDM), 2023 (Short Paper)

[8] [CVPR'23-Highlight] Gen Li, Jie Ji, Minghai Qin, **Wei Niu**, **Bin Ren**, Fatemeh Afghah, Linke Guo, Xiaolong Ma, “*Towards High-Quality and Efficient Video Super-Resolution via Spatial-Temporal Data Overfitting*,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR, 2023, Acceptance rate: 25.78%, Highlight: top 2.5%).

[9] [CVPR'23] Changdi Yang, Pu Zhao, Yanyu Li, **Wei Niu**, **Jiexiong Guan**, Hao Tang, Minghai Qin, **Bin Ren**, Xue Lin, Yanzhi Wang, “*Pruning Parameterization with Bi-level Optimization for Efficient Semantic Segmentation on the Edge*,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR, 2023, Acceptance rate: 25.78%).

[10] [AAAI'23] Dong Li, Ruoming Jin, Zhenming Liu, **Bin Ren**, Jing Gao, and Zhi Liu, “*Towards Reliable Item Sampling for Recommendation Evaluation*,” Thirty-Seventh AAAI Conference on Artificial Intelligence (AAAI), 2023 (Acceptance Rate: 19.6%)

[11] [AAAI'23] Yanyu Li, Changdi Yang, Pu Zhao, Geng Yuan, **Wei Niu**, **Jiexiong Guan**, Hao Tang, Minghai Qin, **Bin Ren**, Xue Lin, and Yanzhi Wang, “*Towards Real-Time Segmentation on the Edge*,” Thirty-Seventh AAAI Conference on Artificial Intelligence (AAAI), 2023 (Acceptance Rate: 19.6%).

[12] [MICRO'22] **Wei Niu**, **Jiexiong Guan**, Xipeng Shen, Yanzhi Wang, Gagan Agrawal, **Bin Ren**, “*GCD<sup>2</sup>: A Globally Optimizing Compiler for Mapping DNNs to Mobile DSPs*,” 55th IEEE/ACM International Symposium on Microarchitecture (MICRO), 2022. (Acceptance rate: 83/369 = 22%)

[13] [NeurIPS'22] Zifeng Wang, Zheng Zhan, Yifan Gong, Geng Yuan, **Wei Niu**, Tong Jian, **Bin Ren**, Stratis Ioannidis, Yanzhi Wang, Jennifer Dy, “*SparCL: Sparse Continual Learning on the Edge*,” Thirty-sixth Conference on Neural Information Processing Systems (NeurIPS), December 2022. (Acceptance rate: 25.6%)

[14] [ECCV'22] Yushu Wu, Yifan Gong, Pu Zhao, Yanyu Li, Zheng Zhan, **Wei Niu**, Hao Tang, Minghai Qin, **Bin Ren**, and Yanzhi Wang, “*Compiler-Aware Neural Architecture Search for On-Mobile Real-time Super-Resolution*,” in Proc. of European Conference on Computer Vision (ECCV), 2022. (Acceptance rate: 25%)

[15] [ECCV'22] Zhenglun Kong, Peiyan Dong, Xiaolong Ma, Xin Meng, **Wei Niu**, Mengshu Sun, Xuan Shen, Geng Yuan, **Bin Ren**, Hao Tang, Minghai Qin, and Yanzhi Wang, “*SPViT: Enabling Faster Vision Transformers via Soft Token Pruning*,” in Proc. of European Conference on Computer Vision (ECCV), 2022. (Acceptance rate: 25%)

[16] [MSN'22] **Junjie Wang, Jiexiong Guan**, Y.Alicia Hong, Hong Xue, Shuangquan Wang, Zhenming Liu, **Bin Ren** and Gang Zhou, “*Towards socially acceptable food type recognition*,” The 18th International Conference on Mobility, Sensing and Networking (MSN), 2022.

[17] [IPDPS'22] **Qihan Wang, Bin Ren**, Jie Chen, Robert G. Edwards, “*MICCO: An Enhanced Multi-GPU Scheduling Framework for Many-Body Correlation Functions*,” The 36th IEEE International Parallel & Distributed Processing Symposium (IPDPS), June 2022. (Acceptance rate of the 1st round: 9.7% = 46/474).

[18][ISQED'22–invited] Xiaolong Ma, Geng Yuan, Zhengang Li, **Wei Niu**, Yifan Gong, Tianyun Zhang, Zheng Zhan, Pu Zhao, Jian Tang, Xue Lin, **Bin Ren**, Yanzhi Wang, “*A General Pruning Framework Enabling Real-Time Inference on Resource-Limited Mobile Devices*,” in Proceedings of the 23rd International Symposium on Quality Electronic Design (ISQED), April 2022.

[19][NeurIPS'21 Spotlight: Top 3%] Geng Yuan, Xiaolong Ma, **Wei Niu**, Zhengang Li, Zhenglun Kong, Ning Liu, Yifan Gong, Zheng Zhan, Chaoyang He, Qing Jin, Siyue Wang, Minghai Qin, **Bin Ren**, Yanzhi Wang, Sijia Liu, Xue Lin, “*MEST: Accurate and Fast Memory-Economic Sparse Training Framework on the Edge*,” Thirty-fifth Conference on Neural Information Processing Systems (NeurIPS), December 2021. (Acceptance rate: 2344/9122 = 26%, Spotlight 3%)

[20][HiPC'21] **Qihan Wang, Wei Niu**, Li Chen, Ruoming Jin, **Bin Ren**, “*HEALS: A Parallel eALS Recommendation System on CPU/GPU Heterogeneous Platforms*,” The 28th IEEE International Conference on High Performance Computing, Data, & Analytics (HiPC), December 2021. (Acceptance rate: 23%)

[21][ICCV'21] Zheng Zhan, Yifan Gong, Pu Zhao, Geng Yuan, **Wei Niu**, Yushu Wu, Tianyun Zhang, Malith Jayaweera, David Kaeli, **Bin Ren**, Xue Lin, Yanzhi Wang, “*Achieving On-Mobile Real-Time Super-Resolution With Neural Architecture and Pruning Search*,” International Conference on Computer Vision (ICCV), October 2021. (Acceptance rate: 25.9%)

[22][PLDI'21] **Wei Niu, Jiexiong Guan**, Yanzhi Wang, Gagan Agrawal, and **Bin Ren**, “*DNNFusion: Accelerating Deep Neural Networks Execution with Advanced Operator Fusion*,” 42nd ACM SIGPLAN Conference on Programming Language Design and Implementation, page 883-898, June 2021. (Acceptance rate: 87/320 = 27%)

[23][ESEC/FSE'21] Jialiang Tan\*, **Yu Chen\***, Zhenming Liu, **Bin Ren**, Shuaiwen Leon Song, Xipeng Shen, and Xu Liu (\*equal contribution), "Toward Efficient Interactions between Python and Native Libraries," 29th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering, August 2021. (Acceptance rate: 97/396 = 24.5%)

[24][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*", 35th ACM International Conference on Supercomputing, page 266-278, June 2021. (Acceptance Rate: 38/157 = 24%)

[25][CVPR'21 Oral Paper: top 5%] Zhengang Li\*, Geng Yuan\*, **Wei Niu\***, Pu Zhao\*, Yanyu Li, Yuxuan Cai, Xuan Shen, Zheng Zhan, Zhenglun Kong, Qing Jin, Zhiyu Chen, Sijia Liu, Kaiyuan Yang, Yanzhi Wang, **Bin Ren**, and Xue Lin (\*equal contribution), "*NPAS: A compiler-aware framework of unified network pruning and architecture search for beyond real-time mobile acceleration*," IEEE Conference on Computer Vision and Pattern Recognition, page 14255-14266, June 2021. (Acceptance rate: 1593/7500 = 21.2%)

[26][DAC'21] Pu Zhao, Geng Yuan, Yuxuan Cai, **Wei Niu**, Qi Liu, Wujie Wen, **Bin Ren**, Yanzhi Wang, and Xue Lin, "*Neural Pruning Search for Real-Time Object Detection of Autonomous Vehicles*," 57th Annual Design Automation Conference (DAC), December 2021. (Acceptance rate: TBA, around 20%)

[27][ASP-DAC'21] Hongjia Li, Geng Yuan, **Wei Niu**, Yuxuan Cai, Mengshu Sun, Zhengang Li, **Bin Ren**, Xue Lin, and Yanzhi Wang, "*Real-Time Mobile Acceleration of DNNs: From Computer Vision to Medical Applications*," in Proceeding of Asia and South Pacific Design Automation Conference (ASP-DAC), page 581-586, January 2021.

[28][AAAI'21] **Wei Niu\***, Mengshu Sun\*, Zhengang Li\*, Jou-An Chen, Jiexiong Guan, Xipeng Shen, Yanzhi Wang, Xue Lin, and **Bin Ren**, "*Achieving Real-Time Execution of 3D Convolutional Neural Networks on Mobile Devices*," Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI), February 2021 (Acceptance Rate: 20.9%).

[29][AAAI'21] Yuxuan Cai, Hongjia Li, Geng Yuan, **Wei Niu**, Yanyu Li, Xulong Tang, **Bin Ren**, and Yanzhi Wang, "*YOLOmobile: Real-Time Object Detection on Mobile Devices via Compression-Compilation Co-Design*," Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI), February 2021 (Acceptance Rate: 20.9%).



[30] [SC'20] Mert Hidayetoglu, Tekin Bicer, Simon Garcia de Gonzalo, **Bin Ren**, Vincent De Andrade, Doga Gursoy, Rajkumar Kettimuthu, Ian Foster, and Wen-mei Hwu, "*Petascale XCT: 3D Image Reconstruction with Hierarchical Communications on Multi-GPU Nodes*," The 2020 ACM International Conference for High Performance Computing, Networking, Storage and Analysis (SC), November, 2020 (**Best Paper Award and Best Student Paper Finalist**). (Acceptance Rate: TBA, around 20%).

[31][ASE'20] Hongyu Liu\*, **Ruiqin Tian\***, **Bin Ren**, and Tongping Liu (\*equal contribution), "*Prober: Practically Defending Overflows with Page Protection*," The 35th IEEE/ACM International Conference on Automated Software Engineering (ASE), page 1116-1128, September, 2020. (Acceptance Rate: 93/414 = 22.5%).

[32][ECCV'20] Xiaolong Ma\*, **Wei Niu\***, Tianyun Zhang, Sijia Liu, Sheng Lin, Hongjia Li, Xiang Chen, Jian Tang, Kaisheng Ma, **Bin Ren**, and Yanzhi Wang (\*equal contribution), "*An Image Enhancing Pattern-based Sparsity for Real-Time Inference on Mobile Devices*," The 16th European Conference on Computer Vision (ECCV), page 1-16, August, 2020. (Acceptance Rate: 1360/5150 = 26%).

[33][GLSVLSI'20] Yifan Gong, Zheng Zhan, Zhengang Li, **Wei Niu**, Xiaolong Ma, Wenhao Wang, **Bin Ren**, Caiwen Ding, Xue Lin, Xiaolin Xu, and Yanzhi Wang, "*A Privacy-Preserving-Oriented DNN Pruning and Mobile Acceleration Framework*," in Proceedings of the 2020 on Great Lakes Symposium on VLSI (Special Session Paper), September. 2020.

[34][ICML'20] Yu Chen, Zhenming Liu, **Bin Ren**, and Xin Jin, "*On Efficient Constructions of Checkpoints*," The 37th International Conference on Machine Learning (ICML), July, 2020. (Acceptance Rate: 1,088/4,990 = 21.8%).

[35][ICS'20] Ruoming Jin\*, **Zhen Peng\***, Wendell Wu, Feodor Dragan, Gagan Agrawal, and **Bin Ren** (\*equal contribution), "*Parallelizing Pruned Landmark Labeling: Dealing with Dependencies in Graph Algorithms*," The 34th ACM International Conference on Supercomputing (ICS), June, 2020. (Acceptance rate: 40/132 = 30%).

[36][ICS'20] Tyler Coy, Shuibing He, **Bin Ren**, and Xuechen Zhang, "*Compiler Aided Checkpointing using Crash-Consistent Data Structures in NVMM Systems*," The 34th ACM International Conference on Supercomputing (ICS), June, 2020. (Acceptance rate: 40/132 = 30%).

[37][DAC'20] Peiyan Dong, Siyue Wang, **Wei Niu**, Chengming Zhang, Sheng Lin, Zhengang Li, Yifan Gong, **Bin Ren**, Xue Lin, and Dingwen Tao, "*RTMobile: Beyond Real-Time Mobile Acceleration of RNNs for Speech Recognition*," The 57th Annual Design Automation Conference (DAC), July, 2020. (Acceptance rate: TBA, around 20%).

[38][ASPLOS'20] **Wei Niu**, Xiaolong Ma, Sheng Lin, Shihao Wang, Xuehai Qian, Xue Lin, Yanzhi Wang, and **Bin Ren**, “*PatDNN: Achieving Real-Time DNN Execution on Mobile Devices with Pattern-based Weight Pruning*,” The 25th International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS), page 907-922, March, 2020 (Acceptance Rate:  $86/476 = 18\%$ ).

[39][CGO'20] **Yu Chen**, Ivy B. Peng, Zhen Peng, Xu Liu, and **Bin Ren**, “*ATMem: Adaptive Data Placement in Graph Applications on Heterogeneous Memories*,” The 2020 International Symposium on Code Generation and Optimization (CGO), page 293-304, February, 2020 (Acceptance Rate:  $26/95 = 27\%$ ).

[40][AAAI'20] Xiaolong Ma, Fu-Ming Guo, **Wei Niu**, Xue Lin, Jian Tang, Kaisheng Ma, **Bin Ren**, and Yanzhi Wang, “*PCONV: The Missing but Desirable Sparsity in DNN Weight Pruning for Real-Time Execution on Mobile Device*,” The 34th AAAI Conference on Artificial Intelligence (AAAI), February, 2020 (Acceptance Rate:  $1,591/7,737 = 20.6\%$ ).

[41][SC'19] Mert Hidayetoglu, Tekin Bicer, Simon Garcia de Gonzalo, **Bin Ren**, Doga Gursoy, Rajkumar Kettimuthu, Ian T. Foster, and Wen-Mei W. Hwu, “*MemXCT: Memory-Centric X-Ray CT Reconstruction with Massive Parallelization*,” The International Conference for High Performance Computing, Networking, Storage, and Analysis (SC), November, 2019 (**SC20 Student Cluster Reproducibility Challenge Paper**). (Acceptance Rate:  $87/344 = 25\%$ ).

[42][PACT'19] Soklong Lim, Zaixin Lu, **Bin Ren**, and Xuechen Zhang, “*Enforcing Crash Consistency of Evolving Network Analytics in Non-Volatile Main Memory Systems*,” The 28th International Conference on Parallel Architecture and Compilation Techniques (PACT), page 124-137, September, 2019. (Acceptance Rate:  $26/126 = 21\%$ ).

[43][CLOUD'19] Jared Polonitza, David Chiu, and **Bin Ren**, “*A Transactional Framework for Broadening Access to Geo-Diversification*,” The 12th IEEE International Conference on Cloud Computing (CLOUD), page 229-233, July, 2019. (Short Paper).

[44][CGO'19] **Ruiqin Tian**, Junqiao Qiu, Zhijia Zhao, Xu Liu, and **Bin Ren**, “*Transforming Query Sequences for High-Throughput B+ Tree Processing on Many-core Processors*,” The 2019 International Symposium on Code Generation and Optimization (CGO), page 2-14, February, 2019. (Acceptance Rate:  $21/69 = 31\%$ ).

[45][PACT'18] **Zhen Peng**, **Alexander Powell**, Bo Wu, Tekin Bicer, and **Bin Ren**, “*GraphPhi: Efficient Parallel Graph Processing on Emerging Throughput-oriented Architectures*,” The 27th International Conference on Parallel Architecture and Compilation Techniques (PACT), page 1-14, November, 2018. (Acceptance Rate:  $36/126 = 29\%$ ).

[46][eScience'17] Tekin Bicer, Doga Gursay, Rajkumar Kettimuthu, Ian T. Foster, **Bin Ren**, Vincent De Andrede, and Francesco De Carlo, "*Real-Time Data Analysis and Autonomous Steering of Synchrotron Light Source Experiments*," The 13th IEEE International Conference on e-Science (eScience), page 59-68, October, 2017.

[47][PPoPP'17] **Bin Ren**, Sriram Krishnamoorthy, Kunal Agrawal, and Milind Kulkarni, "*Exploiting Vector and Multicore Parallelism for Recursive, Task-Parallel Programs*," The 22nd ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming (PPoPP), page 117-130, February, 2017. (Acceptance Rate:  $29/132 = 22\%$ ).

[48][PACT'16] Junqiao Qiu, Zhijia Zhao, and **Bin Ren**, "*MicroSpec: Fine-Grained Speculative Parallelization for FSM Computations*," The 25th International Conference on Parallel Architecture and Compilation Techniques (PACT), page 221-233, September, 2016. (Acceptance Rate:  $31/119 = 26\%$ ).

[49][ICPP'16] Jeff Daily, Sriram Krishnamoorthy, Ananth Kalyanaraman, and **Bin Ren**, "*On the Impact of Widening Vector Registers on Sequence Alignment*," The 45th International Conference on Parallel Processing (ICPP), page 506-515, August, 2016. (Acceptance Rate:  $53/251 = 21.1\%$ ).

[50][PLDI'15] **Bin Ren**, Youngjoon Jo, Sriram Krishnamoorthy, Kunal Agrawal, and Milind Kulkarni, "*Efficient Execution of Recursive Programs on Commodity Vector Hardware*," The 36th annual ACM SIGPLAN conference on Programming Language Design and Implementation (PLDI), page 509-520, June, 2015. (Acceptance Rate:  $58/303 = 19\%$ ).

[51][IPDPS'15] Linchuan Chen, Xin Huo, **Bin Ren**, Surabhi Jain, and Gagan Agrawal, "*Efficient and Simplified Parallel Graph Processing over CPU and MIC*," The 29th IEEE International Parallel & Distributed Processing Symposium (IPDPS), page 819-828, May, 2015. (Acceptance Rate:  $108/496 = 21\%$ ).

[52][ICS'14] Xin Huo, **Bin Ren**, and Gagan Agrawal, "*A Programming System for Xeon Phis with Runtime SIMD Parallelization*," The 28th ACM International Conference on Supercomputing (ICS), page 283-292, June, 2014. (Acceptance Rate:  $34/162 = 21\%$ ).

[53][CGO'13] **Bin Ren**, Gagan Agrawal, James R. Larus, Todd Mytkowicz, Tomi Poutanen, and Wolfram Schulte, "*SIMD Parallelization of Applications that Traverse Irregular Data Structures*," The 2013 International Symposium on Code Generation and Optimization (CGO), February, 2013 (**Best Paper Award, SIGPLAN Research Highlight, Nominated for CACM Research Highlight**). (Acceptance Rate:  $33/117 = 28\%$ ).

[54][PACT'11] **Bin Ren**, and Gagan Agrawal, “*Compiling Dynamic Data Structure in Python to Enable the Use of Multi-core and Many-core Libraries*,” The 20th International Conference on Parallel Architectures and Compilation Techniques (PACT), page 68-77, October, 2011. (Acceptance Rate:  $36/221 = 16\%$ ).

Peer-Reviewed Demo-track Papers

[1] [AAAI'21 Demo track] Yuxuan Cai, Hongjia Li, Geng Yuan, **Wei Niu**, Yanyu Li, Xulong Tang, **Bin Ren**, and Yanzhi Wang, “*A Compression-Compilation Co-Design Framework Towards Real-Time Object Detection on Mobile Devices*,” in AAAI (Demonstration Paper), 2021.

[2] [IJCAI'20 Demo track] **Wei Niu\***, Pu Zhao\*, Zheng Zhan, Xue Lin, Yanzhi Wang, and **Bin Ren** (\*equal contribution), “*Towards Real-Time DNN Inference on Mobile Platforms with Model Pruning and Compiler Optimization*,” in Proc. of IJCAI, 2020. (Acceptance Rate: 15.8%).

[3] [ECCV'20 Demo track] **Wei Niu**, Mengshu Sun, Zhengang Li, Geng Yuan, Pu Zhao, Xue Lin, Yanzhi Wang, and **Bin Ren**, “*Real-time 3D CNN Inference for Action Recognition on Mobile Devices*,” in European Conference on Computer Vision (ECCV), 2020.

[4] [ECCV'20 Demo track] Zheng Zhan, Pu Zhao, Geng Yuan, **Wei Niu**, **Bin Ren**, Xue Lin, and Yanzhi Wang, “*Real-time DNN Accelerations on Mobile Devices for Versatile Practical Deep Learning Applications*,” in European Conference on Computer Vision (ECCV), 2020.

Peer-Reviewed Workshops and Posters

[1] [IPDPS'23 Workshop] Daniel T Chen, Ethan H Hansen, Xinpeng Li, Vinooth Kulkarni, Vipin Chaudhary, **Bin Ren**, Qiang Guan, Sanmukh Kuppannagari, Ji Liu, and Shuai Xu, “*Efficient quantum circuit cutting by neglecting basis elements*,” arXiv preprint arXiv:2304.04093, 2023.

[2] [NIPS'20 Workshop] Yuxuan Cai, Hongjia Li, Geng Yuan, **Wei Niu**, Yanyu Li, Xulong Tang, **Bin Ren**, and Yanzhi Wang, “*YOLObile: Real-Time Object Detection on Mobile Devices via Compression-Compilation Co-Design*,” in NeurIPS 2020 Workshop on Machine Learning for Autonomous Driving.

[3] [WHPC'20 Poster] Ruiqin Tian, Jiajia Li, **Bin Ren**, and Gokcen Kestor. “*High-Performance Sparse Tensor Algebra Compiler*,” The 11th International Women in High Performance Computing Workshop (WHPC), in conjunction with ACM/IEEE International Conference for High-Performance Computing, Networking, Storage, and Analysis (SC). 2020.

[4] [LCPC'20] Erdal Mutlu, **Ruiqin Tian**, **Bin Ren**, Sriram Krishnamoorthy, Roberto Gioiosa, Jacques Pienaar, and Gokcen Kestor. “*COMET: A Domain-Specific Compilation of High-Performance Computational Chemistry*,” The 33rd Workshop on Languages and Compilers for Parallel Computing (LCPC), October, 2020.

[5] [ICML-Workshop'19] **Wei Niu**, Xiaolong Ma, Yanzhi Wang, and **Bin Ren**, “*26ms Inference Time for ResNet-50: Towards Real-Time Execution of all DNNs on Smartphone*,” The 36th International Conference on Machine Learning (Joint Workshop on On-Device Machine Learning and Compact Deep Neural Network Representations), June, 2019.

[6] [PPoPP'16 Poster] Mehmet Can Kurt, **Bin Ren**, Sriram Krishnamoorthy, and Gagan Agrawal, “*An Automated Store Recycling Approach for Dynamic Task Graph Schedulers*,” The 21st ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming, March, 2016.

[7] [LCPC'15] **Bin Ren**, Nishkam Ravi, Yi Yang, Min Feng, Gagan Agrawal, and Srimat Chakradhar, “*Automatic and Efficient Data Host-Device Communication for Many- Core Coprocessors*,” The 28th International Workshop on Languages and Compilers for Parallel Computing, September, 2015.

[8] [LCPC'15] Mehmet Can Kurt, **Bin Ren**, and Gagan Agrawal, “*Low-Overhead Fault-Tolerance Support using DISC Programming Model*,” The 28th International Workshop on Languages and Compilers for Parallel Computing, September, 2015.

[9] [ICS'14 Poster] **Bin Ren**, Nishkam Ravi, Yi Yang, Min Feng, Gagan Agrawal, and Srimat Chakradhar, “*Automating and Optimizing Data Transfers for Many-core Coprocessors*,” The 28th ACM International Conference on Supercomputing, June, 2014..

[10] [PACT'12 Poster] **Bin Ren**, Gagan Agrawal, James R. Larus, Todd Mytkowicz, Tomi Poutanen, and Wolfram Schulte, “*Fine-Grained Parallel Traversals of Irregular Data Structures*,” The 21th International Conference on Parallel Architectures and Compilation Techniques, October, 2012.

[11][IPDPSW'11] **Bin Ren**, Gagan Agrawal, Brad Chamberlain, Steve Deitz, “*Translating Chapel to Use FREERIDE: A Case Study in Using an HPC language for Data- intensive Computing*,” The 16th International Workshop on High-Level Parallel Programming Models and Supportive Environments(HIPS) held in conjunction with IPDPS, May, 2011.

### Invited Talks

- “*Real-Time DNN Execution on Mobile Devices with Compiler Optimizations*”,

Guest lecture, North Carolina State University, 2024, 2023

Keynote Speak, PMAM@PPoPP, 2022  
IEEE MIPR Panel, 2022  
University of California Merced, 2022  
University of Massachusetts Amherst, 2021  
Rutgers University, 2021

•“*PatDNN: Achieving Real-Time DNN Execution on Mobile Devices with Pattern-based Weight Pruning*”,

ROAD4NN Workshop@DAC, 2020  
ASPLOS, Lausanne, Switzerland March, 2020  
Northeastern University, Real-Time Machine Learning Colloquium, December, 2019

•“*ATMem: Adaptive Data Placement in Graph Applications on Heterogeneous Memories*”,  
CGO, San Diego, USA, February, 2020

•“*Transforming Query Sequences for High-Throughput B+ Tree Processing on Many-core Processors*”,  
CGO, Washington DC, USA, February, 2019

•“*GraphPhi: Efficient Parallel Graph Processing on Emerging Throughput-oriented Architectures*”,  
PACT, Limassol, Cyprus, November, 2018

•“*Exploiting Vector and Multicore Parallelism for Recursive, Task-Parallel Programs*”,  
PPoPP, Austin, Texas, USA, February, 2017

•“*Challenging the Irregularity: Unleashing the Power of Modern SIMD Architectures*”,  
1) Computer Science Department at Missouri University of Science and Technology, Rolla, Missouri, USA, January, 2016;  
2) Department of Computer Science and Software Engineering at Auburn University, Auburn, Alabama, USA, February, 2016;  
3) Department of Electrical and Computer Engineering at University of Delaware, Newark, Delaware, USA, February, 2016;  
4) Computer Science Department at The College of William & Mary, Williamsburg, Virginia, USA, February, 2016;  
5) Department of Electrical, Computer & Biomedical Engineering at University of Rhode Island, Kingston, Rhode Island, USA, March, 2016;  
6) Department of Computer Science at Illinois Institute of Technology, Chicago, Illinois, USA, March, 2016;  
7) Computer Science Department at Rensselaer Polytechnic Institute, Troy, New York, USA, March, 2016;  
8) Pacific Northwest National Laboratory, Richland, Washington, USA, April, 2016;

- 9) Kent State University, Kent, Ohio, December, 2017;
- 10) Jefferson Lab, Newport News, Virginia, October, 2018

- “*Low-Overhead Fault-Tolerance Support using DISC Programming Model*”,  
LCPC, Raleigh, North Carolina, USA, September, 2015
- “*Automatic and Efficient Data Host-Device Communication for Many-Core Coprocessors*”,  
LCPC, Raleigh, North Carolina, USA, September, 2015
- “*Efficient Execution of Recursive Programs on Commodity Vector Hardware*”,
  - 1) PNNL, Richland, Washington, USA, June, 2015
  - 2) PLDI, Portland, Oregon, USA, June, 2015
- “*Automatic and Efficient Data Host-Device Communication for Many-Core Coprocessors*”,  
NEC Lab, Princeton, New Jersey, USA, August, 2013
- “*SIMD Parallelization of Applications that Traverse Irregular Data Structures*”,  
MSR, Redmond, Washington, USA, September, 2011
- “*Compiling Dynamic Data Structure in Python to Enable the Use of Multi-core and Many-core Libraries*”,
  - 1) MSR, Redmond, Washington, USA, August, 2011,
  - 2) PACT, Galveston Island, Texas, USA, October, 2011

#### Software Release

- [1] CoCoPIE demo for ASPLOS’20 is publicly available at:  
<https://www.youtube.com/channel/UCCKVDtg2eherTEuqIJ5cD8A>
- [2] The code “QTrans” for CGO’19 is publicly available at:  
<https://zenodo.org/record/1486393>
- [3] The code “GraphPhi” for PACT’18 is publicly available at:  
<https://zenodo.org/record/1318418>
- [4] The code “VectorCilk” for PLDI’15 is publicly available at:  
<https://engineering.purdue.edu/plcl/vectorcilk/index.php>

#### Patent

**[US20150067225 A1]** Automatic Communication and Optimization of Multi-dimensional Arrays for Many-core Coprocessor using Static Compiler Analysis

**[No.: 62/965,275]** RTMobile: A mobile acceleration framework of RNNs for beyond real-time speech recognition

[No.: 62/976,595] PatDNN: Achieving real-time DNN execution on mobile devices with pattern-based weight pruning.

[No.: 62/976,577] BPDNN: A general, real-time DNN execution framework on mobile devices with block-based column-row pruning

[US-2022-0413862-A1] DNNFusion: Accelerating Deep Neural Networks Execution with Advanced Operator Fusion

## **PROFESSIONAL SERVICE**

### a) *College committee service*

- 2023-2024 Personnel Committee, Computer Science, William & Mary
- 2023-2024 Faculty Recruiting Committee, Computer Science, William & Mary
- 2023-2024 Graduate Admissions, Computer Science, William & Mary
- 2023-2024 Quantum Computing, Computer Science, William & Mary
- 2023-2024 Undergraduate Pre-Major Advisor, William & Mary
- 2021-2022 Faculty Recruiting Committee, Computer Science, William & Mary
- 2021-2022 Diversity, Equity, and Inclusion Committee, Computer Science, William & Mary
- 2021-2022 Graduate Curriculum Committee, Computer Science, William & Mary
- 2021-2022 System Committee, Computer Science, William & Mary
- 2021-2022 Undergraduate Pre-Major Advisor, William & Mary
- 2020-2021 Graduate Admissions, Computer Science, William & Mary
- 2020-2021 Colloquium Chair, Computer Science, William & Mary
- 2020-2021 Thomas Jefferson Prize in Natural Philosophy Committee, William & Mary
- 2020-2021 Undergraduate Pre-Major Advisor, William & Mary
- 2019-2020 Colloquium Chair, Computer Science, William & Mary
- 2019-2020 Undergraduate Pre-Major Advisor, William & Mary
- 2018-2019 Zable Recruitment Fellowship, William & Mary



- 2018-2019 Awards and Prizes, Computer Science, William & Mary
- 2018-2019 Colloquium, Computer Science, William & Mary
- 2018-2019 Graduate Curriculum, Computer Science, William & Mary
- 2018-2019 Undergraduate Pre-Major Advisor, William & Mary
- 2017-2018 Awards and Prizes, Computer Science, William & Mary
- 2017-2018 Undergraduate Pre-Major Advisor, William & Mary
- 2016-2017 Graduate Admissions, Computer Science, William & Mary
- 2016-2017 Zable Recruitment Fellowship, William & Mary
- 2016-2017 Awards and Prizes, Computer Science, William & Mary

b) Other professional service

- Ph.D. Thesis Committee Member:
  - Qing Yang
  - Bin Nie
  - Shasha Wen
  - Probir Roy
  - Qingsen Wang
  - Yongsen Ma
  - Haonan Wang
  - Mohamed Ibrahim
  - Hao Xu
  - Xiaodan Zhu
  - Du Shen
  - Gurunath Kadam
  - Heather Switzer
  - Pengfei Su
  - Qiong Wu
  - Zheng Zhang
  - Wei Han (Colorado Mines)
  - Zeyi Tao
  - David Chris Thames
  - Tao Zhang
  - Yidong Gong
  - Yiyang Lu
  - Jiangtao Kong
  - Jialiang Tan

Minglong Sun  
Jeremy Myers  
Sabila Al Jannat  
Saima Afrin  
Zhenyu Zong

- Undergraduate Honor Thesis Committee Member:
  - Yanxin Xu
  - Xinzhi (Lulu) Zhang
- Program Committee Member:
  - The International Conference for High Performance Computing, Networking, Storage, and Analysis (SC), 2024, 2023
  - ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming (PPoPP), 2024 (PC Member & AE Chair), 2023 (AE Chair), 2022 (PC & Session Chair)
  - The International Symposium on Computer Architecture (ISCA), 2024 (ERC), 2022 (ERC)
  - The ACM SIGPLAN International Conference on Compiler Construction (CC), 2024, 2023, 2022
  - ACM International Conference on Supercomputing (ICS), 2021 (PC & Publication Chair), 2018 (Submission Chair)
  - IEEE International Conference on Distributed Computing Systems (ICDCS), 2021
  - IEEE Languages, Compilers, Tools and Theory of Embedded Systems (LCTES), 2024, 2021, 2020
  - International Workshop on High-Level Parallel Programming Models and Supportive Environments (HIPS), 2021 (Workshop Co-Chair)
  - IEEE International Parallel & Distributed Processing Symposium (IPDPS), 2024, 2023 (Track Co-Chair), 2022, 2021
  - IEEE International Symposium on Workload Characterization (IISWC), 2020 (Registration Chair)
  - IEEE International Conference on Computer Design (ICCD), 2023, 2022, 2021, 2020
  - ACM International Symposium on High-Performance Parallel and Distributed Computing (HPDC), 2019
  - ACM/IEEE International Conference on Parallel Architectures and Compilation Techniques (PACT), 2019, 2018
  - International Conference on Parallel Processing (ICPP), 2024 (Track Chair), 2020 (PC & Session Chair), 2015
  - IEEE International Conference on High Performance Computing, Data, and Analytics (HiPC), 2024 (Track Chair), 2023, 2022, 2021, 2019, 2018, 2017, 2015
  - ACM Asia-Pacific Workshop on Systems (APSys), 2022 (PC & Session Chair)
  - Annual IFIP International Conference on Network and Parallel Computing (NPC),

2020, 2019, 2018

- IEEE International Conference on Computer Communications and Networks (ICCCN), 2018
- IEEE International Conference on Future Internet of Things and Cloud (FiCloud), 2018
- International Symposium on Benchmarking, Measuring and Optimizing (Bench), 2023, 2022, 2021, 2020, 2018
- IEEE International Conferences on High Performance Computing and Communications (HPCC), 2016, 2015
- IEEE International Conference on Parallel and Distributed Systems (ICPADS), 2016
- International Workshop on High-level Parallel Programming Models and Supportive Environments (HIPS), 2020, 2019, 2016, 2015
- Workshop on General Purpose Processing Using GPU (GPGPU), 2020, 2019
- International Workshop on Domain-Specific Languages and High-Level Frameworks for High Performance Computing (WOLFHPC), 2017, 2016
- Annual Workshop on Emerging Parallel and Distributed Runtime Systems and Middleware (IPDRM), 2020, 2017

- Journal Reviewer:

- Topic Editor of Frontiers in High Performance Computing with the topic of “Real-Time Machine Learning on Edge Devices with HPC Support”, 2023
- IEEE Transactions on Computers (TC), 2020, 2019
- IEEE Transactions on Parallel and Distributed Systems (TPDS), 2021, 2020, 2019, 2017
- ACM Transactions on Architecture and Code Optimization (TACO), 2024 (Distinguished Reviewer), 2023, 2017
- ACM Computing Surveys (CSUR), 2017
- IEEE Transactions on Smart Grid (SmartGrid), 2017
- International Journal of Parallel Programming (IJPP), 2019, 2017, 2016
- Journal of Parallel and Distributed Computing (JPDC), 2018, 2017, 2016, 15, 14, 13
- IEEE Transactions on Knowledge and Data Engineering (TKDE), 2014

- Conference Reviewer:

- The International Conference on Machine Learning (ICML), 2024, 2023, 2022
- Conference on Neural Information Processing Systems (NeurIPS), 2023
- International Conference on Learning Representations (ICLR), 2024
- International Conference on Computer Vision (ICCV), 2023, 2022, 2021
- IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024, 2023, 2021
- AAAI Conference on Artificial Intelligence (AAAI), 2024, 2022, 2021
- International Conference on Parallel Processing (ICPP), 2016
- IEEE International Parallel & Distributed Processing Symposium (IPDPS), 2016, 15
- ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming

- (PPoPP), 2015
  - International Conference on Parallel Architectures and Compilation Techniques (PACT), 2015
  - IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid), 2015
  - International Conference on Network and Parallel Computing (NPC), 2014, 2015
- ACM SRC Reviewer/Judger:
  - ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS), 2018 (Judger)
  - International Symposium on Code Generation and Optimization (CGO), 2020
  - ACM/IEEE International Conference on Parallel Architectures and Compilation Techniques (PACT), 2021, 2020
- Proposal Reviewer:
  - NSF Panel, 2023, 2021, 2020, 2017
  - DoE Advanced Scientific Computing Research (ASCR), 2015
- Outreach Activities:
  - Guest Speaker, W&M Chapter of the Association of Computing Machinery
  - Judge, W&M Hackathon 2019
  - Presenter, The Annual Focusing on the Future Conference (for high-ability middle and high school students and their parents and counselors), The College of William & Mary's Center for Gifted Education, 2021, 2020, 2019, 2018

## **STUDENTS ADVISED:**

### a) Ph.D. students:

Ruiqin Tian (2015-2021, Graduated, First Job: Post-doc@PNNL)  
 Zhen Peng (2016-2023, Graduated, First Job: Post-doc@PNNL)  
 Qihan Wang (2017-present, Graduating soon, First Job: SE@Cruise LLC)  
 Wei Niu (2018-2023, Graduated, First Job: Assistant Professor@UGA)  
 Yu Chen (2018-2023, co-advised with Zhenming Liu and Andreas Stathopoulos,  
 Graduated, First job: Senior SE@Citadel LLC)  
 Junjie Wang (2020-2023, co-advised with Gang Zhou)  
 Jiexiong Guan (2020-present)  
 Sam Ma (2021-present)  
 Zhenqing Hu (2021-present)  
 Jiase E (2022-present)

### b) M.S. students:

Alexander Powell (graduated in Spring 2017)  
 Eunyoung Cho (graduated in Spring 2017)

Shuxin Zou (graduated in Spring 2019)  
Xiaoying Zhai (graduated in Spring 2019)

c) Undergraduate students:

*Major advisees:*

Hanqiu Peng (2017-2020)  
Zihan Yang (2017-2020)  
Kristine A. Tseng (2018-2020)  
Sejin Park (2018-2020)  
Aaron M. Klein (2018-2021)  
Frank Ding (2019-2020)  
Kaitlyn N. Huynh (2019-2021)  
Dieu Thy L. Ngo (2019-2021)  
Junzhao Sun (2019-2021)  
Charles J. Wang (2019-2022)  
Anneliese Brei (2020-2022)  
Jiaying Chen (2020-2022)  
Caitlyn P. Marat (2020-2022)  
Xianglu Peng (2020-2022)  
Sarah Wang (2020-2023)  
Shuhong Wang (2020-2023)  
Steven Z. Jia (2021-present)  
Taylor Liegel (2021-present)  
Gabriella K. Shelton (2021-present)  
Kyle L. Chen (2021-present)  
Jacob C. Feldman (2021-present)  
Abdimalik M. Hassan (2021-2023)  
Stephen D. Hoag (2021-2023)  
Daniel C. Lee (2021-2023)  
Alex Mui (2021-2023)  
Melody Vu (2021-present)  
Yichan Zhang (2021-2023)  
Yifei Zhang (2021-2023)  
Samuel P. Borak (2022-present)  
Dillon P. Buyrn (2022-present)  
Connor D. MacKinnon (2022-present)  
Rhea Malhotra (2022-present)  
David N. Montenegro (2022-present)  
Sayyed Hadi Razmjo (2022-2023)  
Quanhao Zhou (2021-2022)  
Jiarui Qi (2023-present)  
Huizhi Zhao (2023-present)

*Pre-major advisees:*

Michael G. Fairbanks (2017-2019)  
Stephen D. Hoag (2017-2021)  
Ashley R. Robinson (2017-2019)  
Anneliese D. Brei (2018-2020)  
Kaitlyn N. Huynh (2018-2019)  
Maheen Khan (2018-2020)  
Mariano A. Leyva Merino (2018-2020)  
Charles J. Wang (2018-2019)  
Bat-Enkh BaatarhUU (2019-present)  
Grayson Hoy (2019-2020)  
Guanyu Jiang (2019-2020)  
Blake Brown (2019-2020)  
Sarah Wang (2019-2020)  
Julia Butler (2019-2020)  
Megan Sierzega (2020-2020)  
Ryan T. Gainor (2020-2021)  
Steven Z. Jia (2020-2021)  
Taylor Liegel (2020-2021)  
Gabriella K. Shelton (2020-2021)  
Jonathon B. West (2020-2021)  
Dillon P. Buyrn (2021-2022)  
Francis Lorenzo G. Dayan (2021-2023)  
Manel Leong (2021-2023)  
Connor D. MacKinnon (2021-2022)  
Rhea Malhotra (2021-2022)  
Kenneth F. Begasse (2023-present)  
Owen C. Chang (2023-present)  
Zaid M. Dib (2023-present)  
Bracen J. Hart (2023-present)  
Kristina Knauss (2023-present)  
Zachery E. LeMay (2023-present)  
Abilasha Suri (2023-present)

d) Undergraduate research students:

Hanqiu Peng (2019, Graduated, PhD@National University of Singapore)  
Ruoyu Li (2019)  
Anneliese D. Brei (2019, Monroe Scholar, Graduated, PhD@UNC Chapel Hill)  
John Svoboda (2020)  
Charles J. Wang (2020-2022, Honor Thesis, Graduated, MS@UC Berkeley)  
Amanda Michel (2022 Summer, Monroe Scholar)  
Kyle Chen (2023-present, Research Project)