

Xuehai Qian

CONTACT INFORMATION	3740 McClintock Avenue, EEB 204 Los Angeles, CA 90089-2562, USA http://alchem.usc.edu/~xuehaiq/ http://alchem.usc.edu	Tel: 213-740-4459 Fax: 213-740-9803 Email: xuehai.qian@usc.edu
---------------------	--	--

RESEARCH INTERESTS	<ul style="list-style-type: none"> ○ Domain-specific system and architecture with focuses on graph processing and machine learning ○ Non-volatile memory system and architecture ○ Performance optimization with machine learning ○ Quantum computing
--------------------	---

EDUCATION	<ul style="list-style-type: none"> ○ University of Illinois at Urbana-Champaign (UIUC) Urbana, IL Ph.D. in Computer Science Aug. 2007 – Aug. 2013 • Dissertation title: <i>Scalable and Flexible Bulk Architecture</i> ○ Institute of Computing Technology, Chinese Academy of Sciences Beijing, China M.S. in Computer Science Sept. 2004 – Jul. 2007 ○ Beihang University Beijing, China B.S. in Computer Science Sept. 2000 – Jul. 2004
-----------	---

PROFESSIONAL EXPERIENCE	<ul style="list-style-type: none"> ○ Assistant Professor, <i>University of Southern California.</i> Aug. 2015 – Present ○ Postdoctoral Researcher, <i>UC Berkeley.</i> Oct. 2013 – Jun. 2015 ○ Research Intern, <i>Microsoft Research, Silicon Valley</i> Summer 2011 ○ Research Intern, <i>Microsoft Research, Redmond</i> Summer 2008
-------------------------	---

AWARDS AND HONORS	<ul style="list-style-type: none"> ○ Inducted into MICRO Hall of Fame 2021 ** For publishing 8 papers in MICRO. ○ Inducted into ISCA Hall of Fame 2021 ** For publishing 9 papers in ISCA. ○ Inducted into Computer Architecture Aggregated Hall-of-Fame 2020 ** For publishing 44 papers in ISCA, ASPLOS, MICRO, and HPCA—the four top computer architecture conferences. ** The only tenure-track assistant professor in the list and the "Hall of Fame" in all the four conferences. ○ ACSIC (American Chinese Scholar In Computing) Rising Star Award (2 awardees this year) 2019 ○ IEEE Senior Member 2019 ○ Inducted into HPCA Hall of Fame 2019 ** For publishing 12 papers in HPCA. ○ Inducted into ASPLOS Hall of Fame 2018 ** For publishing 15 papers in ASPLOS. ** Ranked 2nd in the list. ○ NSF CAREER Award 2018 ** Support my research on Algorithm-Centric High Performance Graph Processing. ○ ACM Distinguished Speaker 2017 ○ NSF CRII Award 2017 ○ One of 200 selected to participate Heidelberg Laureate Forum (HLF) 2015 ○ W.J. Poppelbaum Memorial Award, UIUC 2013 ○ Andrew and Shana Laursen Fellowship, UIUC 2007
-------------------	---

CONFERENCE
PUBLICATIONS

- ** My name is boldfaced; senior authors of each paper are underlined; my student names are highlighted with asterisks.
- Gengyu Rao*, Jingji Chen*, Jason Yik, Xuehai Qian. *SparseCore: Stream ISA and Processor Specialization for Sparse Computation*. In Proceedings of 27th International Conference on Architectural Support for Programming Languages and Operating Systems (**ASPLOS'22**)
 - Lutan Zhao, Peinan Li, Rui Hou, Michael Huang, Xuehai Qian, Lixin Zhang, Dan Meng. *HyBP: Hybrid Isolation-Randomization Secure Branch Predictor*. In Proceedings of The 28th IEEE International Symposium on High-Performance Computer Architecture (**HPCA'22**)
 - Shiyu Li, Edward Hanson, Xuehai Qian, Xuehai Qian, Hai Li, Yiran Chen. *ESCALATE: Boosting the Efficiency of Sparse CNN Accelerator with Kernel Decomposition*. In Proceedings of The 54th IEEE/ACM International Symposium on Microarchitecture (**MICRO'21**)
 - Jingji Chen*, Xuehai Qian. *Kudu: An Efficient and Scalable Distributed Graph Pattern Mining Engine*. arXiv:2105.03789
 - Chunhua Deng, Siyu Liao, Yang Sui, Xuehai Qian, Bo Yuan. *GoSPA: An Energy-efficient High-performance Globally Optimized Sparse Convolutional Neural Network Accelerator*. In Proceedings of The the 48th International Symposium on Computer Architecture (**ISCA'21**)
 - Geng Yuan, Payman Behnam, Zhengang Li, Ali Shafiee, Sheng Lin, Xiaolong Ma, Hang Liu, Xuehai Qian, Mahdi Nazm Bojnordi, Yanzhi Wang, Caiwen Ding. *FORMS: Fine-grained Polarized ReRAM-based In-situ Computation for Mixed-signal DNN Accelerator*. In Proceedings of The the 48th International Symposium on Computer Architecture (**ISCA'21**)
 - Qingcheng Xiao, Size Zheng, Bingzhe Wu, Pengcheng Xu, Xuehai Qian, Yun Liang. *HASCO: Towards Agile HARDWARE and SOFTWARE CO-design for Tensor Computation*. In Proceedings of The the 48th International Symposium on Computer Architecture (**ISCA'21**)
 - Linghao Song, Fan Chen, Xuehai Qian, Hai Li, Yiran Chen. *Low-Cost Floating-Point Processing in ReRAM for Scientific Computing*. arXiv:2011.03190
 - Sung-En Chang, Yanyu Li, Mengshu Sun, Runbin Shi, Hayden K.-H. So, Yanzhi Wang, Xuehai Qian, Xue Lin. *Mix and Match: A Novel FPGA-Centric Deep Neural Network Quantization Framework*. In Proceedings of The 27th IEEE International Symposium on High-Performance Computer Architecture (**HPCA'21**)
 - Jingji Chen*, Xuehai Qian. *DwarvesGraph: A High-Performance Graph Mining System with Pattern Decomposition*. arXiv:2008.09682
 - You Wu*, Xuehai Qian. *Comprehensive Elimination of Speculative Side-Channels and Interference without Blocking*. arXiv:2006.16535
 - Sheng Xu, Xiaoming Chen, Xuehai Qian, Yinhe Han. *TUPIM: A Transparent and Universal Processing-in-memory Architecture for Unmodified Binaries*. In Proceedings of the 30th edition of the ACM Great Lakes Symposium on VLSI (**GLSVLSI'20**)
 - Jinglei Cheng*, Haoqing Deng*, Xuehai Qian. *AccQOC: Accelerating Quantum Optimal Control Based Pulse Generation*. In Proceedings of the 47th International Symposium on Computer Architecture (**ISCA'20**)
 - Youwei Zhuo*, Jingji Chen*, Qinyi Luo*, Yanzhi Wang, Hailong Yang, Depei Qian, Xuehai Qian. *SympleGraph: Distributed Graph Processing with Precise Loop-carried Dependency Guarantee*. In Proceedings of 41st ACM SIGPLAN Conference on Programming Language Design and Implementation (**PLDI'20**)
 - Qinyi Luo*, Jiaao He*, Youwei Zhuo*, Xuehai Qian. *Prague: High-Performance Heterogeneity-Aware Asynchronous Decentralized Training*. In Proceedings of 25th International Conference on Architectural Support for Programming Languages and Operating Systems (**ASPLOS'20**)
 - Xuan Peng, Xuanhua Shi, Hulin Dai, Hai Jin, Weiliang Ma, Fan Yang, Xuehai Qian. *Capuchin: Tensor-based GPU Memory Management for Deep Learning*. In Proceedings of 25th International Conference on Architectural Support for Programming Languages and Operating Systems (**ASPLOS'20**)
 - Teng Ma, Mingxing Zhang, Kang Chen, Zhuo Song, Yongwei Wu, Xuehai Qian. *AsymNVM: An Efficient Framework for Implementing Persistent Data Structures on Asymmetric NVM Architecture*. In Proceedings of 25th International Conference on Architectural Support for Programming Languages and Operating Systems (**ASPLOS'20**)

- Xingbin Wang, Rui Hou, Boyan Zhao, Fengkai Yuan, Jun Zhang, Dan Meng, **Xuehai Qian**. *DNN-Guard: An Elastic Heterogeneous Architecture for DNN Accelerator against Adversarial Attacks*. In Proceedings of 25th International Conference on Architectural Support for Programming Languages and Operating Systems (**ASPLOS'20**)
- Wei Niu, Xiaolong Ma, Sheng Lin, Shihao Wang, **Xuehai Qian**, Xue Lin, Yanzhi Wang, Bin Ren. *PatDNN: Achieving Real-Time DNN Execution on Mobile Devices with Pattern-based Weight Pruning*. In Proceedings of 25th International Conference on Architectural Support for Programming Languages and Operating Systems (**ASPLOS'20**)
- Linghao Song, Fan Chen, Youwei Zhuo*, **Xuehai Qian**, Hai Li, Yiran Chen. *AccPar: Tensor Partitioning for Heterogeneous Deep Learning Accelerator Arrays*. In Proceedings of The 26th IEEE International Symposium on High-Performance Computer Architecture (**HPCA'20**)
- Youwei Zhuo*, Chao Wang*, Mingxing Zhang, Rui Wang, Dimin Niu, Yanzhi Wang, **Xuehai Qian**. *GraphQ: Scalable PIM-based Graph Processing*. In Proceedings of The 52nd IEEE/ACM International Symposium on Microarchitecture (**MICRO'19**)
- Yimin Jiang, Yong Cui, Wenfei Wu, Zhe Xu, Jiahan Gu, K. K. Ramakrishnan, Yongchao He, **Xuehai Qian**. *SpeedyBox: Low-Latency NFV Service Chains with Cross-NF Runtime Consolidation*. In Proceedings of The 39th IEEE International Conference on Distributed Computing Systems (**ICDCS'19**)
- Yuzhao Wang, Lele Li, You Wu*, Junqing Yu, Zhibin Yu, **Xuehai Qian**. *A Time-Space Sharing Scheduling Abstraction for Next Generation of Shared Cloud via Vertical Labels*. In Proceedings of The 46th International Symposium on Computer Architecture (**ISCA'19**)
- Chunhua Deng, Fangxuan Sun, **Xuehai Qian**, Jun Lin, Zhongfeng Wang, Bo Yuan. *TIE: Energy-Efficient Tensor Train-Based Inference Engine for Deep Neural Network*. In Proceedings of The 46th International Symposium on Computer Architecture (**ISCA'19**)
- Ruizhe Cai, Ao Ren, Olivia Chen, Ning Liu, Caiwen Ding, **Xuehai Qian**, Jie Han, Wenhui Luo, Yoshikawa Nobuyuki, Yanzhi Wang. *A Stochastic-Computing based Deep Learning Framework using Adiabatic Quantum-Flux-Parametron Superconducting Technology*. In Proceedings of The 46th International Symposium on Computer Architecture (**ISCA'19**)
- Qinyi Luo*, Jinkun Lin*, Youwei Zhuo*, **Xuehai Qian**. *HOP: Heterogeneity-Aware Decentralized Training*. In Proceedings of 24th International Conference on Architectural Support for Programming Languages and Operating Systems (**ASPLOS'19**)
- Xiongchao Tang, Jidong Zhai, **Xuehai Qian**, Wenguang Chen. *SW-Lock: A Fast Lock for Sunway Taihulight*. In Proceedings of 24th International Conference on Architectural Support for Programming Languages and Operating Systems (**ASPLOS'19**)
- Ao Ren, Jiayu Li, Tianyun Zhang, Shaokai Ye, Wenyao Xu, **Xuehai Qian**, Xue Lin, Yanzhi Wang. *ADMM-NN: An Algorithm-Hardware Co-Design Framework of DNNs Using Alternating Direction Methods of Multipliers*. In Proceedings of 24th International Conference on Architectural Support for Programming Languages and Operating Systems (**ASPLOS'19**)
- Linghao Song, Jiachen Mao, Youwei Zhuo*, **Xuehai Qian**, Hai Li, Yiran Chen. *HyPar: Towards Hybrid Parallelism for Deep Learning Accelerator Array*. In Proceedings of The 25th IEEE International Symposium on High-Performance Computer Architecture (**HPCA'19**)
- Zhe Li, Caiwen Ding, Siyue Wang, Wujie Wen, Youwei Zhuo*, Chang Liu, Qinru Qiu, Wenyao Xu, Xue Lin, **Xuehai Qian**, Yanzhi Wang. *E-RNN: Design Optimization for Efficient Recurrent Neural Networks in FPGAs*. In Proceedings of The 25th IEEE International Symposium on High-Performance Computer Architecture (**HPCA'19**)
- Xiebing Wang, Kai Huang, Alois Knoll, **Xuehai Qian**. *A Hybrid Framework for Fast and Accurate GPU Performance Estimation through Source-Level Analysis and Trace-Based Simulation*. In Proceedings of The 25th IEEE International Symposium on High-Performance Computer Architecture (**HPCA'19**)
- Youwei Zhuo*, Jinglei Cheng*, Qinyi Luo*, Jidong Zhai, Yanzhi Wang, Zhongzhi Luan, **Xuehai Qian**. *CSE: Parallel Finite State Machines with Convergence Set Enumeration*. In Proceedings of The 51st IEEE/ACM International Symposium on Microarchitecture (**MICRO'18**)
- Yirong Lv, Bin Sun, Qinyi Luo*, Jing Wang, Zhibin Yu, **Xuehai Qian**. *CounterMiner: Mining Big Performance Data from Hardware Counters*. In Proceedings of The 51st IEEE/ACM International Symposium on Microarchitecture (**MICRO'18**)

- Chunhua Deng, Siyu Liao, Yi Xie, Keshab K. Parhi, **Xuehai Qian**, Bo Yuan. *PermDNN: Efficient Compressed DNN Architecture with Permuted Diagonal Matrices*. In Proceedings of The 51st IEEE/ACM International Symposium on Microarchitecture (**MICRO'18**)
- Xiongchao Tang, Jidong Zhai, **Xuehai Qian**, Bingsheng He, Wei Xue and Wenguang Chen. *vSensor: Leveraging Fixed-Workload Modules of Programs for Performance Variance Detection*. In Proceedings of 23rd ACM SIGPLAN Annual Symposium on Principles and Practice of Parallel Programming (**PPOPP'18**)
- Mingxing Zhang, Yongwei Wu, Youwei Zhuo*, **Xuehai Qian**, Chenying Huan, Kang Chen. *Wonderland: A Novel Abstraction-Based Out-Of-Core Graph Processing System*. In Proceedings of 23rd International Conference on Architectural Support for Programming Languages and Operating Systems (**ASPLOS'18**)
- Ruizhe Cai, Ao Ren, Ning Liu, Caiwen Ding, Luhao Wang, **Xuehai Qian**, Massoud Pedram, Yanzhi Wang. *VIBNN: Hardware Acceleration of Bayesian Neural Networks*. In Proceedings of 23rd International Conference on Architectural Support for Programming Languages and Operating Systems (**ASPLOS'18**)
- Zhibin Yu, Zhendong Bei, **Xuehai Qian**. *DAC: Data-Aware Auto-Tuning High Dimensional Configurations of In-Memory Cluster Computing*. In Proceedings of 23rd International Conference on Architectural Support for Programming Languages and Operating Systems (**ASPLOS'18**)
- Bing Li, Linghao Song, Fan Chen, **Xuehai Qian**, Yiran Chen, Hai Li. *ReRAM-Based Accelerator for Deep Learning*. In Proceedings of 2018 Design, Automation & Test in Europe (**DATE'18**)
- Linghao Song, Youwei Zhuo*, **Xuehai Qian**, Miao Hu, Hai Li, Yiran Chen. *GraphR: Accelerating Graph Processing Using ReRAM*. In Proceedings of The 24th IEEE International Symposium on High-Performance Computer Architecture (**HPCA'18**)
- *Mingxing Zhang, Youwei Zhuo* (equal contribution), Chao Wang*, Mingyu Gao, Yongwei Wu, Kang Chen, Christos Kozyrakis, Xuehai Qian*. *GraphP: Reducing Communication of PIM-based Graph Processing with Efficient Data Partition*. In Proceedings of The 24th IEEE International Symposium on High-Performance Computer Architecture (**HPCA'18**)
- Abdulaziz Tabbakh, **Xuehai Qian**, Murali Annavaram. *G-TSC: Timestamp Based Coherence for GPUs*. In Proceedings of The 24th IEEE International Symposium on High-Performance Computer Architecture (**HPCA'18**)
- Yanzhi Wang, Caiwen Ding, Geng Yuan, Siyu Liao, Zhe Li, Xiaolong Ma, Bo Yuan, **Xuehai Qian**, Jian Tang, Qinru Qiu, Xue Lin. *Towards Ultra-High Performance and Energy Efficiency of Deep Learning Systems: An Algorithm-Hardware Co-Optimization Framework*. In Proceedings of the 32nd AAAI Conference on Artificial Intelligence (**AAAI'18**)
- Xiaoxiao Liu, Wei Wen, **Xuehai Qian**, Hai Li, Yiran Chen. *Neu-NoC: A High-Efficient Interconnection Network for Accelerated Neuromorphic Systems*. In Proceedings of The 23rd Asia and South Pacific Design Automation Conference (**ASP-DAC'18**)
- Caiwen Ding, Yanzhi Wang, Siyu Liao, Zhe Li, Yu Bai, Youwei Zhuo*, Chao Wang*, **Xuehai Qian**, Ning Liu, Geng Yuan, Xiaolong Ma, Yipeng Zhang, Xue Lin, Jian Tang, Qinru Qiu, Bo Yuan. *Cir-CNN: Accelerating and Compressing Deep Neural Networks Using Block-Circulant Weight Matrices*. In Proceedings of The 50th IEEE/ACM International Symposium on Microarchitecture (**MICRO'17**)
- Zhiyuan Ai, Mingxing Zhang, Yongwei Wu, **Xuehai Qian**, Kang Chen, Weimin Zheng. *Squeezing out All the Value of Loaded Data: An Out-Of-Core Graph Processing System with Reduced Disk I/O*. In Proceedings of 2017 USENIX Annual Technical Conference (**ATC'17**)
- Abdulaziz Tabbakh, Murali Annavaram and **Xuehai Qian**. *Power Efficient Sharing-Aware GPU Data Management*. In Proceedings of 31st IEEE International Parallel & Distributed Processing Symposium (**IPDPS'17**)
- Ao Ren, Ji Li, Zhe Li, Caiwen Ding, **Xuehai Qian**, Qinru Qiu, Bo Yuan and Yanzhi Wang. *SC-DCNN: Highly-Scalable Deep Convolutional Neural Network using Stochastic Computing*. In Proceedings of 22nd International Conference on Architectural Support for Programming Languages and Operating Systems (**ASPLOS'17**)

- Mengxing Liu, Mingxing Zhang, Kang Chen, **Xuehai Qian**, Yongwei Wu, Weimin Zheng and Jinglei Ren. *DudeTM: Building Durable Transactions for Persistent Memories with Decoupling*. In Proceedings of 22nd International Conference on Architectural Support for Programming Languages and Operating Systems (**ASPLOS'17**)
- Linghao Song, **Xuehai Qian**, Hai Li and Yiran Chen *PipeLayer: A Pipelined ReRAM-Based Accelerator for Deep Learning*. In Proceedings of The 23rd IEEE International Symposium on High-Performance Computer Architecture (**HPCA'17**)
- Mingxing Zhang, Yongwei Wu, Kang Chen, **Xuehai Qian**, Xue Li and Weimin Zheng *Exploring the Hidden Dimension in Graph Processing*. In Proceedings of 12th USENIX Symposium on Operating Systems Design and Implementation (**OSDI'16**)
- Xuehai Qian, Koushik Sen, Paul Hargrove and Costin Iancu. *SReplay: Deterministic Group Replay for One-Sided Communication*. In Proceedings of 2016 International Conference on Supercomputing (**ICS'16**)
- Xuehai Qian, Benjamin Sahelices and Depei Qian. *Pacifier: Record and Replay for Relaxed-Consistency Multiprocessors with Distributed Directory Protocol*. In Proceedings of The 41st International Symposium on Computer Architecture (**ISCA'14**)
- Xuehai Qian, Benjamin Sahelices and Josep Torrellas. *OmniOrder: Directory-Based Conflict Serialization of Transactions*. In Proceedings of The 41st International Symposium on Computer Architecture (**ISCA'14**)
- Xuehai Qian, Benjamin Sahelices, Josep Torrellas and Depei Qian. *BulkCommit: Scalable and Fast Commit of Atomic Blocks in a Lazy Multiprocessor Environment*. In Proceedings of The 46th IEEE/ACM International Symposium on Microarchitecture (**MICRO'13**)
- Xuehai Qian, Benjamin Sahelices, Josep Torrellas and Depei Qian. *Volition: Precise and Scalable Sequential Consistency Violation Detection*. In Proceedings of The 18th International Conference on Architectural Support for Programming Languages and Operating Systems (**ASPLOS'13**)
- Xuehai Qian, He Huang, Benjamin Sahelices, and Depei Qian. *Rainbow: Efficient Memory Dependence Recording with High Replay Parallelism for Relaxed Memory Model*. In Proceedings of The 19th IEEE International Symposium on High-Performance Computer Architecture (**HPCA'13**)
- Xuehai Qian, Benjamin Sahelices and Josep Torrellas. *BulkSMT: Designing SMT Processors for Atomic-Block Execution*. In Proceedings of The 18th IEEE International Symposium on High-Performance Computer Architecture (**HPCA'12**)
- Xuehai Qian, Wonsun Ahn and Josep Torrellas. *ScalableBulk: Scalable Cache Coherence for Atomic Blocks in a Lazy Environment*. In Proceedings of The 43rd IEEE/ACM International Symposium on Microarchitecture (**MICRO'10**)

JOURNAL
PUBLICATIONS

** My name is boldfaced; senior authors of each paper are underlined; my student names are highlighted with asterisks.

- Chao Wang*, **Xuehai Qian**. *RDMA-enabled Concurrency Control Protocols for Transactions in the Cloud Era*. To appear in **IEEE Transactions on Cloud Computing**
- Youwei Zhuo*, Jingji Chen*, Gengyu Rao*, Qinyi Luo*, Yanzhi Wang, Hailong yang, Depei Qian, **Xuehai Qian**. *Distributed Graph Processing System and Processing-In-Memory Architecture with Precise Loop-Carried Dependency Guarantee*. **ACM Transactions on Computer Systems** To appear.
- Xue Lin, Mingxing Zhang, Kang Chen, Yongwei Wu, **Xuehai Qian**, Weimin Zheng. *3-D Partitioning for Large-Scale Graph Processing*. **IEEE Transactions on Computers** 70(1): 111-127 (2021)
- Xiebing Wang, **Xuehai Qian**, Alois Knoll, Kai Huang *Efficient Performance Estimation and Work-Group Size Pruning for OpenCL Kernels on GPUs*. **IEEE Transactions on Parallel and Distributed Systems** 31(5): 1089-1106 (2020)
- Hao Yan, Hebin R. Cherian, Ethan C. Ahn, **Xuehai Qian**, Lide Duan *iCELIA: A Full-Stack Framework for STT-MRAM-Based Deep Learning Acceleration*. **IEEE Transactions on Parallel and Distributed Systems** 31(2): 408-422 (2020)
- Zhe Li, Ji Li, Ao Ren, Ruizhe Cai, Caiwen Ding, **Xuehai Qian**, Jeffrey Draper, Bo Yuan, Jian Tang, Qinru Qiu, Yanzhi Wang. *HEIF: Highly Efficient Stochastic Computing Based Inference Framework*

- for Deep Neural Networks*. **IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems** 38(8): 1543-1556 (2019)
- Sheng Xu, Xiaoming Chen, Ying Wang, Yinhe Han, **Xuehai Qian**, Xiaowei Li. *PIMSim: A Flexible and Detailed Processing-in-Memory Simulator*. **Computer Architecture Letters** 18(1): 6-9 (2019)
 - Zhiyuan Ai, Mingxing Zhang, Yongwei Wu, **Xuehai Qian**, Kang Chen, Weimin Zheng. *CLIP: A Disk I/O Focused Parallel Out-Of-Core Graph Processing System*. **IEEE Transactions on Parallel and Distributed Systems** 30(1): 45-62 (2019)
 - Mengxing Liu, Mingxing Zhang, Kang Chen, **Xuehai Qian**, Yongwei Wu, Weimin Zheng and Jinglei Ren. *DudeTX: Durable Decoupled Transaction*. **ACM Transactions on Storage** 14(1): 7:1-7:28 (2018)

STUDENTS

- *Ph.D Students*
 - Youwei Zhuo* (<http://alchem.usc.edu/~youwei/>). Ph.D Candidate of CS, Fall 2016 – Present
 - Nominated for **Microsoft Research Ph.D Fellowship**, 2018
 - Invited to attend Heidelberg Laureate Forum (HLF), 2020
 - Passed screening exam and written qual exam
 - Passed oral qualifying exam
 - Plans to defend thesis in Dec. 2021
 - Chao Wang* (<http://alchem.usc.edu/~chao/>). Ph.D Candidate of ECE, Fall 2016 – Present
 - Passed screening exam
 - Passed the qualifying exam
 - Plans to defend thesis in Spring 2022
 - Qinyi Luo* (<http://alchem.usc.edu/~qinyi/>). Ph.D Candidate of CS, Fall 2017 – Present
 - Nominated for **Microsoft Research Lovelace Fellowship**, 2018
 - Nominated **Microsoft Research Ph.D Fellowship**, 2019
 - Qualcomm Top-Off Fellow, 2020
 - Invited to attend Heidelberg Laureate Forum (HLF), 2020
 - Passed screening exam and written qualifying exam
 - Passed the oral qualifying exam
 - Plans to defend thesis in Spring 2023
 - **Facebook Fellowship Finalist** (3.5%) among 2100 applicants, 2021
 - **MHI Ph.D Scholar**, 2021
 - You Wu*. Ph.D Candidate of ECE, Fall 2017 – Present
 - Passed screening exam
 - Plans to have qualifying exam in Fall 2021
 - Jinglei Cheng*. Ph.D Candidate of ECE, Fall 2018 – Present
 - Passed screening exam
 - Jingji Chen*. Ph.D Candidate of ECE, Fall 2019 – Present
 - Take screening exam in 4/2021
 - Gengyu Yao*. Ph.D Candidate of ECE, Fall 2020 – Present
- *USC Undergraduate Mentoring*
 - Haoqing Deng*. ECE@USC, Spring 2019 – Present
 - Working with my Ph.D student Jinglei Cheng on accelerating quantum optimal control. **Paper appeared in ISCA'20**
 - Jason Yik*. ECE@USC, Spring 2020 – Present
 - Working with my Ph.D student Jingji Chen and Gengyu Yao on accelerating graph mining applications. **Paper submitted to MICRO'21**
 - Received **USC Provosts Research Fellowship** in 2021
 - Jocelyn Liang*. ECE@USC, Spring 2020 – Present
 - Working with my Ph.D student Jinglei Cheng on accelerating quantum optimal control.
 - Sean Syed*. ECE@USC, Summer 2020
 - Working with my Ph.D student Jingji Chen on distributed graph mining.

- *Undergraduate Internship Students*
- Dingyuan Cao.* Undergraduate Intern (remote) from Tsinghua University Summer 2020
 - Worked on distributed graph mining
 - Will start Ph.D at UIUC in Fall 2021
- Gengyu Rao.* Undergraduate Intern from HUST Summer 2019
 - Worked on PIM-based graph processing accelerator, paper in submission
 - Offered fellowship at USC
- Kezhao Huang.* Undergraduate Intern from Tsinghua Summer 2019
 - Worked on concurrency control protocol evaluation with RDMA, paper in submission
 - Currently Ph.D student at Tsinghua
- Jiaao He.* Undergraduate Intern from Tsinghua Summer 2019
 - Worked on distributed machine learning, paper to appear in **ASPLOS'20**
 - Offered RA at USC
- Jinkun Lin.* Undergraduate Intern from Tsinghua Summer 2018
 - Worked on distributed machine learning, paper appeared in **ASPLOS'19**
 - Currently Ph.D student at NYU
- Jingji Chen.* Undergraduate Intern from Tsinghua Summer 2018
 - Worked on distributed graph processing, paper to appear in **PLDI'19**
 - Currently my Ph.D student at USC
- Jinglei Cheng.* Undergraduate Intern from Tsinghua Summer 2017
 - Worked on data-parallel finite state machine, paper appeared in **MICRO'18**
 - Currently my Ph.D student at USC
- Fangke Ye.* Undergraduate Intern from Tsinghua Summer 2016
 - Worked on compiler optimization for HPC applications
 - Currently Ph.D student at Gatech

TALKS

-
- *High Performance Graph Mining Systems*
 - UC Berkeley PL seminar, Apr. 2021
 - Duke ECE Seminar, Feb. 2021
 - Keynote at HotDC (Hot Topics in Data Center), Jan. 2021
 - MIT Fast Code Seminar (<http://fast-code.csail.mit.edu>), Jan. 2021
 - *AccQOC: Accelerating Quantum Optimal Control Based Pulse Generation*
 - HPC China 2020: Quantum Computing Advanced Forum
 - *Communication-Efficient Heterogeneity-Aware Machine Learning System and Architecture*
 - Stony Brook University, Dec. 2019
 - New York University, Nov. 2019
 - Columbia University, Nov. 2019
 - Beijing Academy of Artificial Intelligence Conference, 2020 (29,225 registrations) (<https://2020.baai.ac.cn>)
 - *Structured and Systematic Approach for Energy Efficient DNN Acceleration*
 - Keynote at The 4th Workshop on Energy Efficient Machine Learning and Cognitive Computing co-located with ISCA 2019
 - *Efficient Graph Processing Through Cross-Stack Co-Design*
 - Alibaba Group, Nov. 2017
 - Brown University, Oct. 2017
 - *Reducing Data Movements in Graph Processing: from Distributed System to Emerging Technology*
 - University of Texas, Austin, Feb. 2017
 - Rice University, Feb. 2017
 - Cornell University, Dec. 2016
 - University of Illinois, Urbana-Champaign, Dec. 2016
 - University of Pennsylvania, Dec. 2016
 - VMware Research, Dec. 2016
 - Google, Dec. 2016
 - University of California, Santa Barbara, Jan. 2017

- *Breaking the Myth of "Think as a Vertex"*.
– University of California, Irvine, Oct. 2016
- *Taming Relaxed-Consistency and Non-determinism in Parallel Systems*.
– CALCM Seminar, Carnegie Mellon University, Jan. 2015.
– Georgia Tech, Jan. 2015.
– University of Pennsylvania, Feb. 2015
– Syracuse University, Feb. 2015
– Purdue University (CS), Feb. 2015
– University of California Riverside, Mar. 2015
– University of Virginia, Mar. 2015
– Boston University, Mar. 2015
– University of Southern California, Mar. 2015
– Purdue University (ECE), Mar. 2015
- *Clearing Chaos: Taming Relaxed-Consistency in Multiprocessors*.
– Intel Labs, Santa Clara, Jun. 2014.
– AMD Research, Sunnyvale, Jun. 2014.
- *Scalable and Flexible Bulk Architecture*.
– CECA, Peking University, Dec. 2012.
– University at Buffalo, The State University of New York, Feb. 2013.
– Washington University in St. Louis, Feb. 2013.
– IIS, Tsinghua University, Feb. 2013.
– Rutgers University, Mar. 2013.
– Florida International University, Mar. 2013.
– Penn State University, Mar. 2013.
– The George Washington University, Mar. 2013.
– Oracle, Apr. 2013.
– Drexel University, Apr. 2013.
– Microsoft Research, Asia (MSRA), May. 2013.
– The University of Edinburgh, May. 2013.
– The University of Manchester, May. 2013.
– AMD Research Lab, June. 2013.
– UC Berkeley, Jun. 2013.
– Intel Labs, Santa Clara, Dec. 2013.

SERVICE

- *External Service*
Program Chair of the 8th Workshop on Architecture and Systems for Big Data (ASBD 2018) at ISCA 2018 (<http://acs.ict.ac.cn/asbd2018/>).
Program Co-Chair of the 1st International Workshop on Architecture for Graph Processing (AGP-1) Workshop at ISCA 2017. (<https://sites.google.com/view/agp2017/home>).
Program Co-Chair of The 14th CCF International Symposium on Advanced Parallel Processing Technology (APPT'21)
PC member of ASPLOS 2022, HPCA 2022, MICRO 2021, ISCA 2021, HPCA 2021, PACT 2020, IPDPS 2020, ASPLOS 2020, ISPASS 2020, MICRO 2019, ISCA 2019, HPCA 2019, MICRO 2018, HPCA 2018, ISCA 2017, SOCC 2017, HiPC 2017, ICPP 2016, IPDPS 2015, HiPC 2016.
External Review Committee of ASPLOS 2021, ISCA 2020, ASPLOS 2018, MICRO 2017, ASPLOS 2017, HPCA 2017, MICRO 2016, MICRO 2015, PACT 2016.
PC member of *IEEE Transactions on Parallel and Distributed System* Special Issue on Parallel and Distributed Computing Techniques for AI, ML and DL
Guest Editor of *IEEE Transactions on Computers* Special Issue on Machine-Learning Architectures and Accelerators.
Web Co-Chair of ASPLOS 2018, 2019.
Sponsorship Chair of HPCA 2018, 2019, 2020.
Local Arrangement Chair of ISCA 2018.
Publication Chair of HPCA 2017.

Travel Grant Chair of ASPLOS 2017.

Submission Co-chair of ISCA 2012.

Thesis evaluation committee of Xiebing Wang, advised by Prof. Alois Knoll (Technische Universität München)

◦ *Internal Service*

Organizer of Computer Engineering Seminar at USC

– All talk information (with videos of some) can be found at: <http://alchem.usc.edu/ceng-seminar/>

Qualifying Exam committee of Qiumin Xu (2016), Abdulaziz Tabbakh (2017), Aditya Deshpande (2017), Ji Li (2017), Haonan Lu (2017), Praveen Sharma (2018), Chung Ming Cheung (2018), Hanqing Zeng (2019), Mingyang Zhang (2019), Yitao Hu (2020)

Thesis committee of Qiumin Xu (2016), Abdulaziz Tabbakh (2018), Haonan Lu (2019), Mingyang Zhang (2020)

Department committee of improving Ph.D student study environment

Judge of Research Festival in 2018

GRANTS

** I am the lead PI of all NSF grants. The total amount of funding (my portion) is 2.325 million.*

- CRII: SHF: Improving Programmability of GPGPU/NVRAM Integrated Systems with Holistic Architectural Support (NSF: CCF-1657333), *sole PI*, 2/1/2017 — 8/31/2020 (after extensions), \$175,000.
- SHF: Small: Accelerating Graph Processing with Vertically Integrated Programming Model, Runtime and Architecture (NSF: CCF-1717754), *sole PI*, 7/15/2017 — 7/14/2021 (1-year no-cost extension included), \$450,000.
- CSR: Small: Collaborative Research: GAMBIT: Efficient Graph Processing on a Memristor-based Embedded Computing Platform (NSF: CNS-1717984), *lead PI (co-PI Hai Li from Duke)*, 10/1/2017 — 9/30/2021 (1-year no-cost extension included), \$350,000 (my portion is \$250,000).
- CAREER: Algorithm-Centric High Performance Graph Processing (NSF: CCF-1750656), *sole-PI*, 2/1/2018 — 1/31/2023, \$450,000.
- SPX: Collaborative Research: FASTLEAP: FPGA based compact Deep Learning Platform (NSF: CCF-1919289), *lead PI at USC (co-PI Viktor Prasanna)*, 10/01/2019 — 9/30/2023 (1-year no-cost extension included), \$848,718 (my portion is \$450,000).
- Distributed Machine Learning: From Stragglers to Security (Facebook), *Collaborative project with Murali Annavaram, Salman Avestimehr, Viktor Prasanna*, \$100,000 (my portion is \$25,000).
- Intel Mindshare Curriculum Program (Intel), *Collaborative project with Viktor Prasanna to enhance ECE451 and ECE653*, \$100,000 (my portion is \$25,769).
- SHF: Small: High Performance Graph Pattern Mining System and Architecture (NSF: CCF-2127543), *sole PI*, 8/1/2021 — 7/31/2024, \$500,000.
- PPOSS: Planning: Streamware: A Scalable Framework for Accelerating Streaming Data Science (NSF: CCF-2119816), *co-PI*.