

# YUFEI DING

Computer Science and Engineering,  
University of California, San Diego, CA  
Yufei's homepage

Phone: 757-634-1478  
Email: yufeiding@ucsd.edu  
Alt: yufeiding.picasso@gmail.com

## Employment

2023 – Now	<b>Associate Professor</b> Department of Computer Science and Engineering University of California, San Diego USA.
2017 – 2023	<b>Assistant Professor</b> Department of Computer Science University of California, Santa Barbara, USA.

## Education

2014 – 2017	<b>Ph.D. in Computer Science</b> , North Carolina State University, USA. Advisor: Dr. Xipeng Shen
2012 – 2014	<b>Ph.D. Candidate in Computer Science</b> , College of William and Mary, USA. Transfer to North Carolina State University with my advisor, Dr. Xipeng Shen.
2009 – 2011	<b>M.S. in Physics</b> , College of William and Mary, USA. Advisor: Dr. Gunter Luepke.
2005 – 2009	<b>B.S. in Physics</b> , University of Science and Technology of China, China. Advisor: Dr. Zejun Ding.

## Areas of Research

1. Our research area is grounded in programming systems, but we are been able to use this foundation as a bridge to impact these other important areas, such as machine learning and quantum computing.
2. Our research pioneers the efforts of building intelligent programming systems to automate the selection, mapping, optimization, and debugging of application implementations onto different hardware platforms.
3. Our efforts cut across multiple programming technologies, ranging from high-level algorithmic optimization and autotuning, to domain-specific programming language designs, kernel library implementations (especially for GPUs), advanced compilation constructions, and computer archiecture designs.

## Previous Professional Appointments

2014 – 2017	<b>Research Assistant</b> , Department of Computer Science North Carolina State University, USA
Summer 2015	<b>Research Intern</b> Microsoft Research, Redmond.
Summer 2012	<b>Visiting Research Student</b> Massachusetts Institute of Technology.
2012 – 2014	<b>Research Assistant</b> , Department of Computer Science College of William and Mary.

## Experience Highlights

Mentoring	<p>15 graduate students: 11 PhDs (5 coadvised w/ Dr. Yuan Xie) + 4 Masters;</p> <p>3 PhD graduated/graduating this year (with 2 in academia and 1 in industry);</p> <p>Ph.D. student, Yuke Wang, was awarded the NVIDIA Graduate Fellowship on AI-GPU research, 2022;</p> <p>Ph.D. student, Gushu Li, received the Quantum Information Science and Engineering Network (QISE-NET) Triplet Fellow Award, 2021;</p> <p>15 undergraduate summer interns.</p>
Teaching	(Re)designed 4 courses (2 at undergraduate level and 2 graduate level): CS160 Translation of Programming Language + CS165B Machine Learning + CS293 Advanced Compiler technology + CS 595C Neural Network Verification.
Research	<p>Over 50 top publications across several system technology stacks after joining UCSB (e.g., 5 ASPLOS, 4 ISCA, 4 MICRO, 1 OSDI, 2 USENIX ATC, 4 SC, 2 PPOPP, 3 DAC, 2 AAAI, ICML, ICLR, 2 ISSCC, etc.);</p> <p>First OSDI paper with all UCSB authors @ OSDI 2021;</p> <p>Best Paper Nominee @ ISCA 2022;</p> <p>Distinguished Paper Award @ OOPSLA 2020 (the first best paper award for quantum computing research in programming language conferences)</p>
Funding	<p>Large amount: total of about 5M with my share about 3M.</p> <p>Diverse resources: NSF, VMware, Cisco, SAMSUNG, Alibaba, etc.</p>
Awards	<p>NSF CAREER Award, 2021;</p> <p>VMWare Early Career Faculty Award, 2021;</p> <p>IEEE Computer Society TCHPC Early Career Researchers Award for Excellence in High Performance Computing, 2019;</p> <p>NCSU Computer Science Outstanding Dissertation Award, 2018;</p> <p>NCSU Computer Science Outstanding Research Award, 2016.</p>
Technology Transfer and Patents	<p>A series of our quantum compiler and debugging work have been incorporated to IBM's Qiskit, Amazon's Braket, Oak Ridge National Lab's qcor, and Cambridge Quantum Computing's tket infrastructure.</p> <p>Our proposed Yinyang clustering method has been adapted by Microsoft commercial interactive data visualization BI tools.</p> <p>4 Patents: Our proposed parallelized stochastic gradient has been used to optimize multiple machine learning applications and data analytics;</p>
Professional Services	<p>Served as Program committees for almost all top conferences in programming system areas (Architecture and Compiler: MICRO, ISCA, ASPLOS, HPCA; Programming Language: PLDI, OOPSLA; Computer System: OSDI, USENIX ATC; High-performance Computing: SC, PPOPP, IPDPS, etc.);</p> <p>PI and the co-chair of NSF Workshop on quantum software-hardware co-design, 2022;</p>

	Multiple panels (NSF and DoE).
Professional Services	Departmental level: Faculty Search Committee, Diversity, Equity & Inclusion Committee, Public Relations and Awards Committees, and Graduate Admissions Committee;  University level: University of California Presidential Working Group on AI (a 72-page UC AI Working Group Final Report was generated)

## Graduate Students (Init. Employment, Co-Advised\*)

	<b>Ph.D.</b>
<u>Gushu Li</u> *	Joined University of Pennsylvania as an Assistant Professor, 2023
<u>Liu Liu</u> *	Joined Rensselaer Polytechnic Institute (RPI) as an Assistant Professor, 2022
Boyuan Feng	Meta (with 8 first-author top conference publications), 2022
	<b>MS.</b>
Shu Yang	Google, 2022
Huake He	Oracle, 2022
Kun Wan	Adobe, 2020
Andrew Huang	Google, 2019

## Peer-Reviewed Conference Publications

Names of my Ph.D. students (including coadvised) are underlined.

	<b>QUANTUM COMPUTING SYSTEMS</b>
[MICRO'23]	Anbang Wu, <u>Yufei Ding</u> , Ang Li. “QuComm: Optimizing Collective Communication for Distributed Quantum Computing”, the IEEE/ACM International Symposium on Microarchitecture (MICRO), 2023
[ISCA'23]	Hezi Zhang, Anbang Wu, Yuke Wang, <u>Gushu Li</u> , Hassan Shapourian, Alireza Shabani and <u>Yufei Ding</u> , “A Compilation Framework for Photonic One-Way Quantum Computation”, the IEEE/ACM International Symposium on Computer Architecture (ISCA), 2023.
[ISCA'23]	Samuel Stein, Nathan Wiebe, <u>Yufei Ding</u> , James Ang, and Ang Li, “Q-BEEP: Quantum Bayesian Error Mitigation Employing Poisson Modeling over the Hamming Spectrum”, the IEEE/ACM International Symposium on Computer Architecture (ISCA), 2023.
[MICRO'22]	Anbang Wu, Hezi Zhang, <u>Gushu Li</u> , Alireza Shabani, and Yuan Xie, <u>Yufei Ding</u> , “AutoComm: A Framework for Enabling Efficient Communication in Distributed Quantum Programs”, the 55th IEEE/ACM International Symposium on Microarchitecture (MICRO) 2022.
[ISCA'22]	Anbang Wu, <u>Gushu Li</u> , Hezi Zhang, Gian Giacomo Guerreschi, <u>Yufei Ding</u> , and Yuan Xie, “A Synthesis Framework for Stitching Surface Code with Superconducting Quantum Devices”, the IEEE/ACM International Symposium on Computer Architecture (ISCA), 2022.
[ISCA'22]	Samuel Stein, Nathan Wiebe, <u>Yufei Ding</u> , Peng Bo, Karol Kowalski, Nathan Baker, James Ang, and Ang Li, “EQC : Ensembled Quantum Computing for Variational Quantum Algorithms”, the IEEE/ACM International Symposium on Computer Architecture (ISCA), 2022.

- [ASPLOS'22] Gushu Li, Anbang Wu, Yunong Shi, Ali Javadi-Abhari, **Yufei Ding**, and Yuan Xie, “Paulihedral: a generalized block-wise compiler optimization framework for quantum simulation kernels”, the International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS), 2022.
- [NanoCom'21] *(Invited paper)* Gushu Li, Anbang Wu, Yunong Shi, Ali Javadi-Abhari, **Yufei Ding**, Yuan Xie, “On the Co-Design of Quantum Software and Hardware”, the 8th ACM International Conference on Nanoscale Computing and Communication (NanoCom), 2021.
- [OOPSLA'20] Gushu Li, Li Zhou, Nengkun Yu, **Yufei Ding**, Mingsheng Ying, Yuan Xie, “Projection-based Runtime Assertions for Testing and Debugging Quantum Programs”, the ACM SIGPLAN Conference on Object-oriented Programming, Systems, Languages, and Applications (OOPSLA), 2020. **(Distinguished Paper Award)**
- [ASPLOS'20] Gushu Li, **Yufei Ding**, Yuan Xie, “Towards Efficient Superconducting Quantum Processor Architecture Design”, the International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS), 2020.
- [DAC'20] Gushu Li, **Yufei Ding**, Yuan Xie, “Eliminating Redundant Computation in Noisy Quantum Computing Simulation”, the 48th Design Automation Conference (DAC), 2020.
- [ASPLOS'19] Gushu Li, **Yufei Ding**, Yuan Xie, “Tackling the Qubit Mapping Problem for NISQ-Era Quantum Devices”, the International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS), 2019.

## AI GPU & AI COMPILER

- [ASPLOS'24] Boyuan Feng, Zheng Wang, Yuke Wang, Shu Yang, **Yufei Ding**. “Finding the Limit: Examining the Potential and Complexity of Compilation Scheduling for JIT-Based Runtime Systems”, the International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS), 2014.
- [OSDI'23] Yuke Wang, Boyuan Feng, Zheng Wang, Tong Geng, Ang Li, Kevin Barker, and **Yufei Ding**, “MGG: Accelerating Graph Neural Networks with Fine-grained intra-kernel Communication-Computation Pipelining on Multi-GPU Platforms”, the USENIX Symposium on Operating Systems Design and Implementation (OSDI), 2023.
- [MLSys'23] Guyue Huang, Yang Bai, Liu Liu, Yuke Wang, Bei Yu, **Yufei Ding**, Yuan Xie. “ALCOP: Automatic Load-Compute Pipelining in Deep Learning Compiler for AI-GPUs”, the Conference on Machine Learning and Systems (MLSys), 2023
- [PPoPP'23] Zhaodong Chen, Zheng Qu, Yuying Quan, Liu Liu, **Yufei Ding**, and Yuan Xie, “Dynamic N:M Fine-grained Structured Sparse Attention Mechanism”, the International Conference for High-Performance Computing, Networking, Storage, and Analysis (SC), 2023
- [SC'22] Zheng Wang, Yuke Wang, Boyuan Feng, Dheevatsa Mudigere, Bharath Muthiah, **Yufei Ding** “EL-Rec: Efficient Large-scale Recommendation Model Training via Tensor-train Embedding”, the International Conference for High Performance Computing, Networking, Storage, and Analysis (SC), 2022.
- [SC'22] Xiaohui Wang, Yang Wei, Guyue Huang, Ying Xiong, Xian Qian, **Yufei Ding**, Lei Li, Mingxuan Wang, “ATL: Accelerated Training for Transformer-based Models on GPUs”, the International Conference for High Performance Computing, Networking, Storage, and Analysis (SC), 2022.

- [ATC'22] Boyuan Feng, Tianqi Tang, Yuke Wang, Zhaodong Chen, Zheng Wang, Shu Yang, Yuan Xie, **Yufei Ding**. “Faith: An Efficient Framework for Transformer Verification on GPUs”, the USENIX Annual Technical Conference (ATC), 2022
- [PPoPP'22] Yuke Wang\*, Boyuan Feng\*, **Yufei Ding**, (\* co-primary authors), “TC-GNN: Accelerating Sparse Graph Neural Network Computation Via Dense Tensor Core on GPUs”, the ACM SIGPLAN Symposium on Principles & Practice of Parallel Programming (PPoPP), 2022.
- [DAC'22] Guyue Huang, Haoran Li, Minghai Qin, Fei Sun, **Yufei Ding**, Yuan Xie. “Shfl-BW: Accelerating Deep Neural Network Inference with Tensor-Core Aware Weight Pruning”, the Design Automation Conference (DAC), 2022.
- [DAC'22] Guohao Dai, Guyue Huang, Shang Yang, Zhongming Yu, Hengrui Zhang, **Yufei Ding**, Yuan Xie, Huazhong Yang, Yu Wang. “Heuristic Adaptability to Input Dynamics for SpMM on GPUs”, the Design Automation Conference (DAC), 2022.
- [MLSys'22] Hengrui Zhang, Zhongming Yu, Guohao Dai, Guyue Huang, Yufei Ding, Yuan Xie, Yu Wang. “Understanding GNN Computational Graph: A Coordinated Computation, IO, and Memory Perspective”, the Fifth Conference on Machine Learning and Systems (MLSys), 2022.
- [OSDI'21] Yuke Wang, Boyuan Feng, Gushu Li, Shuangchen Li, Lei Deng, Yuan Xie, **Yufei Ding**, “GN-Advisor: An Adaptive and Efficient Runtime System for GNN Acceleration on GPUs”, the USENIX Symposium on Operating Systems Design and Implementation (OSDI), 2021 .
- [USENIX ATC'21] Boyuan Feng, Yuke Wang, Gushu Li, Yuan Xie, **Yufei Ding**, “Palleon: A Runtime System for Efficient Video Processing toward Dynamic Class Skew”, the USENIX Annual Technical Conference (ATC), 2021.
- [IPDPS'21] Yuke Wang, Boyuan Feng, **Yufei Ding**, “DSXplore: Optimizing Convolutional Neural Networks via Sliding-Channel Convolutions”, the IEEE International Parallel & Distributed Processing Symposium (IPDPS), 2021.
- [PPoPP'21] Boyuan Feng, Yuke Wang, Guoyang Chen, Weifeng Zhang, Yuan Xie, **Yufei Ding**, “EGEMM-TC: Accelerating Scientific Computing on Tensor Cores with Extended Precision”, the ACM SIGPLAN Symposium on Principles & Practice of Parallel Programming (PPoPP), 2020.
- [SC'21] Boyuan Feng, Yuke Wang, Tong Geng, Ang Li, **Yufei Ding**, “APNN-TC: Accelerating Arbitrary-Precision Neural Networks on Tensor Cores”, the International Conference for High Performance Computing, Networking, Storage, and Analysis (SC), 2021.
- [SC'21] Zhaodong Chen, Zheng Qu, Liu Liu, **Yufei Ding**, Yuan Xie, “Efficient Tensor Core-based GPU Kernels for Structured Sparsity under Reduced Precision”, the International Conference for High Performance Computing, Networking, Storage, and Analysis (SC), 2021.
- [CIKM'21] Yuke Wang, Boyuan Feng, Xueqiao Peng, **Yufei Ding**, “An Efficient Quantitative Approach for Optimizing Convolutional Neural Network”, The ACM Conference on Information and Knowledge Management (CIKM), 2021 .
- [FCCM'19 Poster] Yuke Wang, Zhaorui Zeng, Boyuan Feng, Lei Deng, **Yufei Ding**, “KPynq: A Work-Efficient Triangle-Inequality based K-means on FPGA”, the 28th IEEE International Symposium On Field-Programmable Custom Computing Machines, 2019.
- [SysML'18] **Yufei Ding**, Xipeng Shen, Lin Ning, Hui Guan, Xipeng Shen, Madanlal Musuvathi, Todd Mytkowicz, “TOP: A Compiler-Based Framework for Optimizing Machine Learning Algorithms through Generalized Triangle Inequality”, the Conference on Systems and Machine Learning (SysML), 2018.

- [ICDE'18] Hui Guan, **Yufei Ding**, Xipeng Shen, Hamid Krim, “Reuse-Centric K-Means Configuration”, the IEEE International Conference on Data Engineering (ICDE), 2018.
- [OOPSLA'17] **Yufei Ding**, Xipeng Shen, “GLORE: Generalized Loop Redundancy Elimination upon LER-Notation”, the ACM SIGPLAN Conference on Object-oriented Programming, Systems, Languages, and Applications (OOPSLA), 2017.
- [PLDI'17] **Yufei Ding**, Ling Ning, Hui Guan, Xipeng Shen, “Generalizations of the Theory and Deployment of Triangular Inequality for Compiler-based Strength Reduction”, the ACM SIGPLAN conference on Programming Language Design and Implementation (PLDI), 2017.
- [ICDE'17] Guoyang Chen, **Yufei Ding**, Xipeng Shen, “Sweet KNN: An Efficient KNN on GPU through Reconciliation of Redundancy and Regularity”, the IEEE International Conference on Data Engineering (ICDE), 2017.
- [ICML'15] **Yufei Ding**, Yue Zhao, Xipeng Shen, Madan Musuvathi, Todd Mytkowicz, “Yinyang K-Means: A Drop-In Replacement of the Classic K-Means with Consistent Speedup”, the International Conference on Machine Learning (ICML), 2015.
- [VLDB'15] **Yufei Ding**, Xipeng Shen, Madan Musuvathi, Todd Mytkowicz, “TOP: A Framework for Enabling Algorithmic Optimizations for Distance-Related Problems”, the International Conference on Very Large Data Bases (VLDB), 2015.
- [PLDI'15] **Yufei Ding**, Jason Ansel, Kalyan Veeramachaneni, Xipeng Shen, Una-May O’Reilly, Saman Amarasinghe, “Autotuning Algorithmic Choice for Input Sensitivity”, the ACM SIGPLAN conference on Programming Language Design and Implementation (PLDI), 2015.

## AI HARDWARE ACCELERATOR

- [MICRO'23] Guyue Huang, Zhengyang Wang, Po-An Tsai, Chen Zhang, **Yufei Ding**, Yuan Xie. “RM-STC: Row-Merge Dataflow Inspired GPU Sparse Tensor Core for Energy-Efficient Sparse Acceleration”, the IEEE/ACM International Symposium on Microarchitecture (MICRO), 2023
- [ISCA'23] Siqi Li, Fengbin Tu, Liu Liu, Jilan Lin, Zheng Wang, Yangwook Kang, **Yufei Ding**, Yuan Xie. “ECSSD: Hardware/Data Layout Co-Designed In-Storage-Computing Architecture for Extreme Classification.”, 2023.
- [ISCA'22] Jilan Lin, Ling Liang, Zheng Qu, Ishtiyaque Ahmad, Liu Liu, Fengbin Tu, Trinabh Gupta, **Yufei Ding**, Yuan Xie. “INSPIRE: In-Storage Private Information Retrieval via Protocol and Architecture Co-design”, 2022.
- [ASPLOS'22] Zheng Qu\*, Liu Liu\*, Fengbin Tu, Zhaodong Chen, **Yufei Ding**, Yuan Xie, (\* co-primary authors), “DOTA: Detect and Omit Weak Attentions for Scalable Transformer Acceleration”, the International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS), 2022.
- [ISSCC '22] Fengbin Tu, Yiqi Wang, Zihan Wu, Ling Liang, **Yufei Ding**, Bongjin Kim, Leibo Liu, Shaojun Wei, Yuan Xie, Shouyi Yin, “A 28nm 29.2TFLOPS/W BF16 and 36.5TOPS/W INT8 Reconfigurable Digital CIM Processor with Unified FP/INT Pipeline and Bitwise in-Memory Booth Multiplication for Cloud Deep Learning Acceleration”, ISSCC, 2022
- [ISSCC '22] Fengbin Tu, Zihan Wu, Yiqi Wang, Ling Liang, Liu Liu, **Yufei Ding**, Leibo Liu, Shaojun Wei, Yuan Xie, Shouyi Yin, “A 28nm 15.59μJ/Token Full-Digital Bitline-Transpose CIM-based Sparse Transformer Accelerator with Pipeline/Parallel Reconfigurable Modes”, ISSCC, 2022

- [CCGrid'21] Yuke Wang, Boyuan Feng, Gushu Li, Georgios Tzimpragos, Lei Deng, Yuan Xie, **Yufei Ding**, “TiAcc: Triangle-inequality based Hardware Accelerator for K-means on FPGAs”, the IEEE International Symposium on Cluster, Cloud and Internet Computing (CCGrid), 2021.
- [MICRO'21] Liu Liu, Jilan Lin, Zheng Qu, **Yufei Ding**, Yuan Xie, “ENMC: Extreme Near-Memory Classification via Approximate Screening”, the IEEE/ACM International Symposium on Microarchitecture (MICRO), 2021.
- [MICRO'21] Abanti Basak, Zheng Qu, Jilan Lin, Alaa R. Alameldeen, Zeshan Chishti, **Yufei Ding**, Yuan Xie, “Improving Streaming Graph Processing Performance Using Input Knowledge”, the IEEE/ACM International Symposium on Microarchitecture (MICRO), 2021 .
- [ICCAD'21] Jilan Lin, Shuangchen Li, **Yufei Ding**, Yuan Xie, “Overcoming the Memory Hierarchy Inefficiencies in Graph Processing Applications”, the International Conference on Computer-Aided Design (ICCAD), 2021.
- [MICRO'20] Liu Liu, Zheng Qu, Lei Deng, Fengbin Tu, Shuangchen Li, Xing Hu, Zhenyu Gu, **Yufei Ding**, Yuan Xie, “DUET: Boosting Deep Neural Network Efficiency on Dual-Module Architecture”, the IEEE/ACM International Symposium on Microarchitecture (MICRO), 2020.
- [ISCA'20] Peng Gu, Xinfeng Xie, **Yufei Ding**, Guoyang Chen, Weifeng Zhang, Dimin Niu, Yuan Xie, “iPIM: Programmable In-Memory Image Processing Accelerator Using Near-Bank Architecture”, the IEEE/ACM International Symposium on Computer Architecture (ISCA), 2020.
- [ASPLOS'20] Xing Hu, Ling Liang, Lei Deng, Shuangchen Li, Xinfeng Xie, Yu Ji, **Yufei Ding**, Chang Liu, Timothy Sherwood, Yuan Xie, “Neural Network Model Extraction Attacks in Edge by Exploiting Architectural Hints”, the International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS), 2020.

## OTHERS

- [VLDB'23] Ling Liang, Jilan Lin, Zheng Qu, Ishtiyaque Ahmad, Liu Liu, Fengbin Tu, Trinabh Gupta, **Yufei Ding**, Yuan Xie, ”SPG: Structure-Private Graph Database via SqueezePIR”, the International Conference on Very Large Data Bases (VLDB), 2023.
- [NeurIPS'22] Biologically Inspired Dynamic Thresholds for Spiking Neural Networks. Jianchuan Ding, Bo Dong, Felix Heide, Yufei Ding, Yunduo Zhou, Baocai Yin, Xin Yang, Conference on Neural Information Processing Systems (NeurIPS), 2022.
- [AAAI'21] Boyuan Feng, Yuke Wang, **Yufei Ding**, “UAG: Uncertainty-aware Attention Graph Neural Network for Defending Adversarial Attacks”, the Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI), 2021.
- [ICASSP'21] Boyuan Feng, Yuke Wang, **Yufei Ding**, “SAGA: Sparse Adversarial Attack on EEG-based Brain Computer Interface”, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021.
- [ICTAI'20] Boyuan Feng, Yuke Wang, Xu Li, Shu Yang, Xueqiao Peng, **Yufei Ding**, “SGQuant: Squeezing the Last Bit on Graph Neural Networks with Specialized Quantization”, the IEEE International Conference on Tools with Artificial Intelligence (ICTAI), 2020.
- [NPC'20] Yitong Huang, Yu Zhang, Boyuan Feng, Xing Guo, Yanyong Zhang, **Yufei Ding**, “A Close Look at Multi-Tenant Parallel CNN Inference for Autonomous Driving”, the Annual IFIP International Conference on Network and Parallel Computing (NPC), 2020.

- [ICML'20] Liu Liu, Lei Deng, Zhaodong Chen, Yuke Wang, Shuangchen Li, Jingwei Zhang, Yihua Yang, Zhenyu Gu, Xing Hu, **Yufei Ding**, Yuan Xie, “Boosting Deep Neural Network Efficiency with Dual-Module Inference”, the International Conference on Machine Learning (ICML), 2020.
- [AAAI'20] Xiaolei Liu, Kun Wan, **Yufei Ding**, Xiaosong Zhang, Qingxin Zhu, “Weighted-Sampling Audio Adversarial Example Attack”, the Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI), 2020.
- [ICTAI'19] Kun Wan, Shu Yang, Boyuan Feng, Lingwei Xie, **Yufei Ding**, “Reconciling Feature-Reuse and Overfitting in DenseNet with Specialized Dropout”, the IEEE International Conference on Tools with Artificial Intelligence (ICTAI), 2019.
- [ICLR'19] Liu Liu, Lei Deng, Xing Hu, Maohua Zhu, Guoqi Li, **Yufei Ding**, Yuan Xie, “Dynamic Sparse Graph for Efficient Deep Learning”, the International Conference on Learning Representations (ICLR), 2019.
- [NSF Workshop Report'18] Albert Cohen, et. al., “Workshop on Inter-disciplinary Research Challenges in Computer Systems”, the Visioning Workshop at ASPLOS, 2018.
- [ASPLOS'14] **Yufei Ding**, Mingzhou Zhou, Zhijia Zhao, Sarah Eisenstat, Xipeng Shen, “Finding the Limit: Examining the Potential and Complexity of Compilation Scheduling for JIT-Based Runtime Systems”, the International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS), 2014.
- [OOPSLA'14] Zhijia Zhao, Bo Wu, Mingzhou Zhou, **Yufei Ding**, Jianhua Sun, Xipeng Shen, Youfeng Wu, “Call Sequence Prediction through Probabilistic Calling Automata”, the ACM SIGPLAN Conference on Object-oriented Programming, Systems, Languages, and Applications (OOPSLA), 2014.
- [CGO'13] Mingzhou Zhou, Bo Wu, **Yufei Ding**, Xipeng Shen, “ProfMig: The First Framework for Migrating Program Profiles Across Software Versions”, the International Symposium on Code Generation and Optimization (CGO), 2013.

## Journal Publications

- [TACO'23] Xinfeng Xie, Peng Gu, **Yufei Ding**, Dimin Niu, Hongzhong Zheng, Yuan Xie, “MPU: Memory-Centric SIMT Processor via In-DRAM Near-Bank Computing”, ACM Transactions on Architecture and Code Optimization.
- [JSSC'23] Fengbin Tu, Zihan Wu, Yiqi Wang, Ling Liang, Liu Liu, **Yufei Ding**, Leibo Liu, Shaojun Wei, Yuan Xie, Shouyi Yin, “TranCIM: Full-Digital Bitline-Transpose CIM-based Sparse Transformer Accelerator with Pipeline/Parallel Reconfigurable Modes.”, IEEE Journal of Solid State Circuits.
- [TCAD'22] Fengbin Tu, Yiqi Wang, Ling Liang, **Yufei Ding**, Leibo Liu, Shaojun Wei, Shouyi Yin, Yuan Xie, “SDP: Co-Designing Algorithm, Dataflow, and Architecture for in-SRAM Sparse NN Acceleration”, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems.
- [TC'22] Dynamic Sparse Attention for Scalable Transformer Acceleration. Liu Liu, Zheng Qu, Zhaodong Chen, Fengbin Tu, **Yufei Ding**, Yuan Xie, IEEE Transactions on Computers.
- [TC'22] A Systematic View of Model Leakage Risks in Deep Neural Network Systems. Xing Hu, Ling Liang, Xiaobing Chen, Lei Deng, Yu Ji, **Yufei Ding**, Zidong Du, Qi Guo, Timothy Sherwood, Yuan Xie. IEEE Transactions on Computers.



- [TCAD'21] Yuke Wang, Boyuan Feng, Gushu Li, Lei Deng, Yuan Xie, **Yufei Ding**, "STPAcc: Structural TI-based Pruning for Accelerating Distance-related Algorithms on CPU-FPGA Platforms", IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems.
- [TNNLS'21] Ling Liang, Xing Hu, Lei Deng, Yujie Wu, Guoqi Li, **Yufei Ding**, Peng Li, Yuan Xie, "Exploring Adversarial Attack in Spiking Neural Networks with Spike-Compatible Gradient", IEEE Transactions on Neural Networks and Learning Systems.
- [Information Systems'21] Lijun Zhang, Hui Guan, **Yufei Ding**, Xipeng Shen, Hamid Krim, "Reuse-centric K-means configuration". Information Systems.
- [TCAD'21] Xiaobing Chen, Yuke Wang, Xinfeng Xie, Xing Hu, Abanti Basak, Ling Liang, Mingyu Yan, Lei Deng, **Yufei Ding**, Zidong Du, Yuan Xie, "Rubik: A Hierarchical Architecture for Efficient Graph Neural Network Training", IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems.
- [[JAMT'20] Fei Shi, Hongrui Cao, Yuke Wang, Boyuan Feng, **Yufei Ding**, "Chatter detection in high-speed milling processes based on ON-LSTM and PBT", The International Journal of Advanced Manufacturing Technology.
- [BIOINF'20] Lingwei Xie, Song He, Zhongnan Zhang, Kunhui Lin, Xiaochen Bo, Shu Yang, Boyuan Feng, Kun Wan, Kang Yang, Jie Yang, **Yufei Ding**, "Domain-Adversarial Multi-Task Framework for Novel Therapeutic Property Prediction of Compounds", Bioinformatics.
- [TNNLS'20] Zhaodong Chen, Lei Deng, Guoqi Li, Jiawei Sun, Ling Liang, Xing Hu, **Yufei Ding**, Yuan Xie, "Effective and Efficient Batch Normalization Using Few Uncorrelated Data for Statistics Estimation", IEEE Transactions on Neural Networks and Learning Systems.
- [TNNLS'20] Lei Deng, Yujie Wu, Yifan Hu, Ling Liang, Guoqi Li, Xing Hu, **Yufei Ding**, Peng Li, Yuan Xie, "Comprehensive SNN Compression Using ADMM Optimization and Activity Regularization", IEEE Transactions on Neural Networks and Learning Systems.
- [JSSC'20] Li Deng, Guanrui Wang, Guoqi Li, Shuangchen Li, Ling Liang, Maohua Zhu, Yujie Wu, Jing Pei, Zhenzhi Wu, Xing Hu, **Yufei Ding**, Wei He, Yuan Xie, Luping Shi, "Tianjic: A Unified and Scalable Chip Bridging Neuroscience and Deep Learning", IEEE Journal of Solid State Circuits.
- [Neural Networks'19] Lei Deng, Yujie Wu, Xing Hu, Ling Liang, **Yufei Ding**, Guoqi Li, Guangshe Zhao, Peng Li, Yuan Xie, "Rethinking the Performance Comparison between SNNs and ANNs", the Journal of the International Neural Network Society, European Neural Network Society & Japanese Neural Network Society.
- [TVLSI'19] Liang Chang, Xin Ma, Zhaohao Wang, Youguang Zhang, **Yufei Ding**, Weisheng Zhao, Yuan Xie, "DASM: Data-streaming based Computing in Non-Volatile Memory Architecture for Embedded System", the IEEE Transactions on Very Large Scale Integration (VLSI) Systems.

## US Patent

- 2018, App.15/624,555 Madanlal S Musuvathi, Todd D Mytkowicz, Saeed Maleki, **Yufei Ding**, "Determining a likelihood of a user interaction with a content element".
- 2018, App.15/624,614 Madanlal S Musuvathi, Todd D Mytkowicz, Saeed Maleki, **Yufei Ding**, "Implementing network security measures in response to a detected cyber-attack".
- 2018, App.15/624,642 Madanlal S Musuvathi, Todd D Mytkowicz, Saeed Maleki, **Yufei Ding**, "Determining a course of action based on aggregated data".

2018, App.15/624,660 Madanlal S Musuvathi, Todd D Mytkowicz, Saeed Maleki, **Yufei Ding**, "Determining a likelihood of a resource experiencing a problem based on telemetry data".

## Professional Service

<b>Program Committee</b>	<p>ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS'24)</p> <p>International Symposium on Microarchitecture (MICRO'23)</p> <p>ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS'23)</p> <p>Conference on Machine Learning and Systems (MLSys'23)</p> <p>International Symposium on Microarchitecture (MICRO'22)</p> <p>IEEE International Symposium on High-Performance Computer Architecture (HPCA'22)</p> <p>Design Automation Conference (DAC'22)</p> <p>The USENIX Symposium on Operating Systems Design and Implementation (OSDI'22)</p> <p>The International Symposium on Computer Architecture (ISCA'22)</p> <p>International Symposium on Code Generation and Optimization (CGO'22)</p> <p>ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS'22)</p> <p>The International Conference for High Performance Computing, Networking, Storage, and Analysis (SC'21)</p> <p>International Symposium on Microarchitecture (MICRO'21)</p> <p>IEEE International Parallel &amp; Distributed Processing Symposium (IPDPS'21)</p> <p>ISC High Performance (ISC'21)</p> <p>ACM Asia-Pacific Workshop on Systems (APSys'21)</p> <p>International Symposium on Microarchitecture (MICRO'20)</p> <p>ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS'20)</p> <p>ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI'20)</p> <p>ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming (PPoPP'20)</p> <p>ACM SIGPLAN Conference on Object-oriented Programming, Systems, Languages, and Applications (OOPSLA'19)</p> <p>ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS'19)</p> <p>International Symposium on Code Generation and Optimization (CGO'19)</p> <p>ISC High Performance (ISC'19)</p>
--------------------------	---

	International Symposium on Memory Management (ISMM'19)
	ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI'19)
	International Conference on Computer Aided Verification (CAV'19)
	USENIX Annual Technical Conference (USENIX ATC'18).
	Annual IFIP International Conference on Network and Parallel Computing (NPC'18)
	High-Performance Power-Aware Computing (HPPAC'18) at IEEE International Parallel & Distributed Processing Symposium.
<b>Reviewer</b>	International Conference on Machine Learning (ICML'21)
	AAAI Conference on Artificial Intelligence (AAAI'21)
	International Conference on Learning Representations (ICLR'21)
	Conference on Neural Information Processing Systems (NeurIPS'19)
	International Conference on Machine Learning (ICML'19)
	Conference on Uncertainty in Artificial Intelligence (UAI'19)
	IEEE Transactions on Parallel and Distributed Systems (TPDS'18)
	International Conference on Parallel Architectures and Compilation Techniques (PACT'17)
	IEEE Transactions on Parallel and Distributed Systems (TPDS'17)
	Journal of Parallel and Distributed Computing (JPDC'17)
	Transactions on Pattern Analysis and Machine Intelligence (TPAMI'17)
<b>Organization Committee</b>	Co-chair of the Steering Committee of NSF Workshop on Software-Hardware Co-design for Quantum Computing, 2022.
	Associate Editor for ACM Transactions on Quantum Computing, 2022.
	Guest editor for IEEE Internet Computing Special Issue on Quantum and Post-Moore's Law, 2021.
	IEEE NTC Technical and Organization Committee on Quantum, Probabilistic and Neuro-morphic Computing, 2021.
	Organization Committee for Second Workshop on Quantum Tensor Networks in Machine Learning @ NeurIPS 2021.
	Web Chair for the 25h ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming (PPoPP'20)
	Co-chair for ACM Student Research Competition (SRC) at the 24th ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS'19)
	Co-chair for ACM Student Research Competition (SRC) at the 23rd ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS'18)
<b>Panelist</b>	NSF'2018, NSF'2021, 3 panels at NSF'2023;

DoE'2021, DoE'2022, DoE'2023.

## Teaching Experience

Fall 2023	CS 293S Introduction to Quantum Computing System, UCSB, CA
Fall 2022	CS 160 Translation of Programming Language, UCSB, CA
Spring 2022	CS 160 Translation of Programming Language, UCSB, CA
Spring 2022	CS 595C Neural Network Verification, UCSB, CA
Spring 2021	CS 160 Translation of Programming Language, UCSB, CA
Winter 2020	CS 293S Advanced Compiler Optimization, UCSB, CA
Spring 2019	CS 160 Translation of Programming Language, UCSB, CA
Winter 2019	CS 165B Machine Learning, UCSB, CA
Fall 2018	CS 293S Program Optimization, UCSB, CA
Fall 2016	Guest Lectures on CSC512 Compiler Construction Course, NCSU, NC
Spring 2016	Guest Lectures on CSC766 Code Optimization for Programs Course, NCSU, NC
Fall 2015	Guest Lectures on CSC512 Compiler Construction Course, NCSU, NC
Spring 2015	Guest Lectures on CSC766 Code Optimization for Programs Course, NCSU, NC

## Talks & Presentations

2023	Invited talk at Optics and Quantum Electronics Seminar Series at MIT.
2023	Invited talk at Quantum Science Center, ORNL.
2023	Invited talk at Google Quantum AI.
2023	Invited talk at American Physical Society's March Meeting on "Compiler for Distributed Quantum Computing".
2023	Invited talk at the International Conference on Compiler Construction (CC) on "Compiler for Quantum Computing".
2023	Invited talk at University of California San Diego on "Quantum Computing System: Challenges and Opportunities".
2023	Invited talk at University of Illinois at Urbana-Champaign on "Quantum Computing System: Challenges and Opportunities".
2023	Invited talk at University of Chicago on "Quantum Computing System: Challenges and Opportunities".
2023	Invited talk at University of Michigan on "Quantum Computing System: Challenges and Opportunities".

- 2022 Invited talk at Georgia Institute of Technology on "Quantum Computing System: Challenges and Opportunities".
- 2022 Invited talk at University of Columbia on "Quantum Computing System: Challenges and Opportunities".
- 2022 Tutorial at Design Automation Conference (DAC'2022) on "Full-stack System Optimization for Quantum Computing".
- 2022 Invited talk at Pacific Northwest National Laboratory (PNNL) on "Full-stack System Optimization for Quantum Computing".
- 2022 Keynote at the 3rd International Workshop on Quantum Software Engineering (Q-SE) on "Full-stack System Optimization for Quantum Computing".
- 2022 Invited talk at University of California, Los Angeles (UCLA) on "Full-stack System Optimization for Quantum Computing".
- 2022 Invited talk at Duke University on "Full-stack System Optimization for Quantum Computing".
- 2021 Invited talk at VMWare on "Software-Hardware Codesign to Explore the Dynamic Sparsity in Machine Learning Models".
- 2021 Invited talk at PlanQC @ PLDI'2021 on "Theory, Design, and Implementation of Projection-based Assertions for Quantum Debugging and Testing".
- 2021 Tutorial at STAQ Quantum Summer School on "Compiler Optimization and Testing in Quantum Computing".
- 2021 Invited talk at University of Science and Technology of China (USTC) on "Robust and Efficient Graph-based Machine Learning Systems".
- 2021 IEE Talk at UCSB on "Quantum Computing Engineering: Challenges and Opportunities".
- 2021 Invited talk at Hong Kong University of Science and Technology (HKUSTZ), "Quantum Computing Engineering: Challenges and Opportunities".
- 2020 Invited talk at Portland State University (PSU) on "Projection-based Runtime Assertions for Testing and Debugging Quantum Programs".
- 2020 Invited talk at Samsung on "Memory-centric Architecture Research Overview".
- 2020 Invited talk at Association of Chinese Scholars in Computing (ACSIC) on "Quantum Computing System Research: Challenges and Opportunities".
- 2020 Invited talk at University of California, Riverside (UCR) on "High-performance Programming System for Quantum Computing".
- 2019 UCSB Computer Science System Research Seminar on "A Compiler and Runtime System for Efficient Video Processing".
- 2018 Invited talk at Pacific Northwest National Laboratory on "Efficient Adversarial Example Generation toward Robust Machine Learning".
- 2018 UCSB Computer Science Faculty Research Seminar on "Faculty Job Preparation and Interview Questions".
- 2017 Invited talk at University of Southern California (USC) on "High-level Program Optimization for Big Data Analytics".

- 2017 OOPSLA'17, 32th Annual Conference for Object-oriented Programming, Systems, Languages, and Application.
- 2017 PLDI'17, 38th Annual ACM SIGPLAN Conference on Programming Language Design and Implementation, Barcelona, Spain.
- 2015 Intern Talk at Microsoft Research, Redmond, WA.
- 2015 VLDB'15, 41st International Conference on Very Large Data Bases, Kohala Coast, Hawai'i.
- 2015 PLDI'15, 36th Annual ACM SIGPLAN Conference on Programming Language Design and Implementation, Portland, OR.
- 2014 ASPLOS'14, Nineteenth International Conference on Architectural Support for Programming Languages and Operating Systems, Salt Lake City, Utah.

## Awards and Funds

- 2023 Samsung MSL Research Fund.
- 2021 New Cisco Research Grant.
- 2022 Alibaba Innovative Research Award, 2022.
- 2022 Amazon Research Awards, Fall 2022.
- 2022 Co-PI for Noyce Initiative gift fund for UCSB's Quantum Team.
- 2021 Samsung Research Grant, 2021.
- 2021 Cisco Research Grant, 2021.
- 2021 PI for "FET: NSF Workshop on Software-Hardware Co-design for Quantum Computing".
- 2021 VMware Early Career Faculty Grant, 2021.
- 2021 co-PI for NSF "FMitF: Track I: Scalable and Quantitative Verification for Neural Network Analysis and Design".
- 2021 NSF Faculty Early Career Development Program (CAREER) Award.
- 2020 Alibaba Gift Money on continuing "Domain Specific Accelerator Research".
- 2019 IEEE Computer Society TCHPC Early Career Researchers Award for Excellence in High Performance Computing.
- 2019 co-PI for NSF "CC\* Compute: A high-performance GPU cluster for accelerated research".
- 2019 Alibaba Gift Money on "Domain Specific Accelerator Research".
- 2019 Intel FPGA Grant.
- 2019 Xilinx FPGA Grant.
- 2018 NVIDIA GPU Grant.
- 2018 NCSU Computer Science Outstanding Dissertation Award.
- 2016 NCSU Computer Science Outstanding Research Award.

2005-2009 | Outstanding Undergraduate Student Scholarship Awarded Annually.