Professor Di Wu

di.wu@ucf.edu
www.unarylab.com
https://scholar.google.com/citations?user=v6DNkTAAAAAJ

Appointment

Assistant Professor | ***Electrical and Computer Engineering*** 08/2023 – Present
University of Central Florida
Joint Assistant Professor | ***Computer Science*** 08/2023 – Present
University of Central Florida

Research Interest

- Computer architecture
  — performance, efficiency, etc.
- Emerging computing
  — unary, neuromorphic, quantum, etc.
- Domain specific acceleration
  — artificial intelligence, brain computer interface, etc.
- Heterogeneous system
  — Machine learning systems, multi-GPU systems, etc.

Education

Doctor of Philosophy | ***Electrical and Computer Engineering*** 09/2017 – 07/2023
University of Wisconsin–Madison
- Advisor: Prof. Joshua San Miguel
- Thesis: Power-Efficient Computer Architecture via Unary and Approximate Computing
  Harold Peterson Outstanding Dissertation Award

Master of Engineering | ***Integrated Circuit Engineering*** 09/2012 – 01/2015
Fudan University
Bachelor of Science | ***Microelectronics*** 09/2007 – 07/2012
Fudan University

Employment

Research Assistant 09/2017 – 07/2023
Department of Electrical and Computer Engineering, UW–Madison
Research Intern 05/2022 – 09/2022, 05/2020 – 09/2020
Cerebras Systems
Research Intern 05/2019 – 09/2019
Meta (Formerly Facebook)
Digital Circuit Engineer 03/2015 – 05/2017
HiSilicon
Research Assistant 09/2012 – 01/2015
State Key Laboratory of ASIC and System, Fudan University

Honors and Awards

Amar Mukherjee Best Paper Award of ISVLSI 2025
IEEE Micro Top Pick Honorable Mention 2025
AMD Faculty Award 2024
Harold Peterson Outstanding Dissertation Award at UW–Madison 2024
Distinguished Artifact Evaluation Award of ASPLOS 2024
MLCommons Machine Learning and Systems Rising Star 2023
Capstone PhD Teaching Award Nomination at UW–Madison 2022
Grainger Wisconsin Distinguished Graduate Fellowship at UW–Madison 2022
Ph.D. Forum of DAC 2021
IEEE Micro Top Pick 2021
Gerald Holdridge Outstanding Teaching Assistant Award at UW–Madison 2020
Chancellor's Opportunity Fellowship at UW–Madison 2019
Qualcomm Innovation Fellowship Finalist 2019
Foxconn SmartCity Competition Winner 2019
Hiran Mayukh Award at UW–Madison 2018
Rising Star Award at HiSilicon 2015
National Scholarship at Fudan University (1/67) 2015
Excellent Student Union Leader at Fudan University 2010
Third Prize Freshman Scholarship at Fudan University (3/45) 2007

PUBLICATIONS

$^{+}$ – Student * – Collaborator $^{\times}$ – Advisor – Top tier – Award

Conference

[1] Lit Silicon: A Case Where Thermal Imbalance Couples Concurrent Execution in Multiple GPUs [acceptance rate: 19.1%]
Marco Kurzynski^{+}, Shaizeen Aga*, **Di Wu**
International Symposium on Computer Architecture, 2026
Open-source software: Lit Silicon

[2] CryptOracle: A Modular Framework to Characterize FHE
Cory Brynds^{+}, Parker McLeod^{+}, Lauren Caccamise^{+}, Asmita Pal, Dewan Saiham, Sazadur Rahman*, Joshua San Miguel$^{\times}$, **Di Wu**
International Symposium on Performance Analysis of Systems and Software, 2026
Open-source software: CryptOracle

[3] Mugi: Value Level Parallelism For Efficient LLMs [acceptance rate: 14.5%]
Daniel Price^{+}, Prabhu Vellaisamy, John Shen*, **Di Wu**
International Conference on Architectural Support for Programming Languages and Operating Systems, 2026, DOI: 10.1145/3779212.3790189
Open-source software: artifact

[4] PIM-SUM: Fast and Reliable In-Memory Summation for Recommendation Systems
Fan Li, Ruizhi Zhu, Huize Li, **Di Wu**, Xin Xin*
International Conference on Computer Design, 2025, DOI: 10.1109/ICCD65941.2025.00042

[5] Can Photonic Interconnects be used for High-Throughput Memory Access in FHE Accelerators?
Dewan Saiham, Mariam Rabadi, **Di Wu**, Sazadur Rahman*
International Symposium on Low Power Electronics and Design, 2025, DOI: 10.1109/ISLPED65674.2025.11261747

[6] Catwalk: Unary Top-K for Efficient Ramp-No-Leak Neuron Design for Temporal Neural Networks
Devon Lister^{+}, Prabhu Vellaisamy, John Shen*, **Di Wu**
IEEE Computer Society Annual Symposium on VLSI, 2025, DOI: 10.1109/ISVLSI65124.2025.11130314
Amar Mukherjee Best Paper Award

[7] Leveraging Photonic Interconnects for Scalable and Efficient Fully Homomorphic Encryption
Dewan Saiham, **Di Wu**, Sazadur Rahman*
Government Microcircuit Applications & Critical Technology Conference, 2025

[8] LoAS: Fully Temporal-Parallel Datatflow for Dual-Sparse Spiking Neural Networks [acceptance rate: 22.7%]
Ruokai Yin, Youngeun Kim, **Di Wu**, Priyadarshini Panda*
International Symposium on Microarchitecture, 2024, DOI: 10.1109/MICRO61859.2024.00084
Open-source software: artifact

[9] Exploration of Unary Arithmetic-Based Matrix Multiply Units for Low Precision DL Accelerators
Prabhu Vellaisamy, Harideep Nair, **Di Wu**, Shawn Blanton*, John Paul Shen*
IEEE Computer Society Annual Symposium on VLSI, 2024, DOI: 10.1109/ISVLSI61997.2024.00126

[10] ALISA: Accelerating Large Language Model Inference via Sparsity-Aware KV Caching [acceptance rate: 19.6%]
Youpeng Zhao, **Di Wu**, Jun Wang*
International Symposium on Computer Architecture, 2024, DOI: 10.1109/ISCA59077.2024.00077

[11] Carat: Unlocking Value-Level Parallelism for Multiplier-Free GEMMs [acceptance rate: 20.9%]
Zhewen Pan, Joshua San Miguel$^{\times}$, **Di Wu**
International Conference on Architectural Support for Programming Languages and Operating Systems, 2024, DOI: 10.1145/3620665.3640364
IEEE Micro Top Pick Honorable Mention (24 from all computer architecture papers)
Distinguished Artifact Evaluation Award
Open-source software: artifact

[12] uBrain: A Unary Brain Computer Interface [acceptance rate: 16.6%]
Di Wu, Jingjie Li, Zhewen Pan, Younghyun Kim*, Joshua San Miguel$^{\times}$
International Symposium on Computer Architecture, 2022, DOI: 10.1145/3470496.3527401

[13] uSystolic: Byte-Crawling Unary Systolic Array [acceptance rate: 30.5%]
Di Wu, Joshua San Miguel$^{\times}$
International Symposium on High-Performance Computer Architecture, 2022, DOI: 10.1109/HPCA53966.2022.00010
Open-source software: uSystolic-Sim

[14] When Dataflows Converge: Reconfigurable and Approximate Computing for Emerging Neural Networks
Di Wu, Joshua San Miguel[×]
International Conference on Computer Design, 2021, DOI: 10.1109/ICCD53106.2021.00014

[15] UNO: Virtualizing and Unifying Nonlinear Operations for Emerging Neural Networks
Di Wu, Jingjie Li, Setareh Behrooz, Younghyun Kim[*], Joshua San Miguel[×]
International Symposium on Low Power Electronics and Design, 2021, DOI: 10.1109/ISLPED52811.2021.9502473

[16] Normalized Stability: A Cross-Level Design Metric for Early Termination in Stochastic Computing
Di Wu, Ruokai Yin, Joshua San Miguel[×]
Asia and South Pacific Design Automation Conference, 2021, DOI: 10.1145/3394885.3431549

[17] uGEMM: Unary Computing Architecture for GEMM Applications [acceptance rate: 18.0%]
Di Wu, Jingjie Li, Ruokai Yin, Hsuan Hsiao, Younghyun Kim[*], Joshua San Miguel[×]
International Symposium on Computer Architecture, 2020, DOI: 10.1109/ISCA45697.2020.00040
IEEE Micro Top Pick (12 from all computer architecture papers)
Open-source software: UnarySim

[18] Approximate Hardware Techniques for Energy-Quality Scaling Across the System
Younghyun Kim[*], Joshua San Miguel[×], Setareh Behroozi, Tianen Chen, Kyuin Lee, Yongwoo Lee, Jingjie Li, **Di Wu**
International Conference on Electronics, Information, and Communication, 2020, DOI: 10.1109/ICEIC49074.2020.9051208

[19] SECO: A Scalable Accuracy Approximate Exponential Function Via Cross-Layer Optimization
Di Wu, Tianen Chen, Chienfu Chen, Oghenefego Ahia, Joshua San Miguel[×], Mikko Lipasti[*], Younghyun Kim[*]
International Symposium on Low Power Electronics and Design, 2019, DOI: 10.1109/ISLPED.2019.8824959

[20] In-Stream Stochastic Division and Square Root via Correlation [acceptance rate: 18.9%]
Di Wu, Joshua San Miguel[×]
Design Automation Conference, 2019, DOI: 10.1145/3316781.3317844

[21] Convergence-Optimized Variable Node Structure for Stochastic LDPC Decoder
Qichen Zhang, Yun Chen, **Di Wu**, Xiaoyang Zeng, Yeong-luh Ueng
International Conference on Acoustics, Speech and Signal Processing, 2016, DOI: 10.1109/ICASSP.2016.7472936

[22] An Area-Efficient Architecture for Stochastic LDPC Decoder
Qichen Zhang, Yun Chen, **Di Wu**, Xiaoyang Zeng, Yeong-luh Ueng
International Conference on Digital Signal Processing, 2015, DOI: 10.1109/ICDSP.2015.7251868

[23] Latency-Optimized Stochastic LDPC Decoder for High-Throughput Applications
Di Wu, Yun Chen, Qichen Zhang, Lirong Zheng, Xiaoyang Zeng, Yeong-luh Ueng
International Symposium on Circuits and Systems, 2015, DOI: 10.1109/ISCAS.2015.7169329

[24] A High-Throughput LDPC Decoder for Optical Communication
Di Wu, Yun Chen, Yuebin Huang, Yeongluh Ueng, Lirong Zheng, Xiaoyang Zeng
International Conference on ASIC, 2013, DOI: 10.1109/ASICON.2013.6811973

Journal

[1] OptoLink: Breaking Memory Bandwidth Bottlenecks in FHE Accelerators with Photonic Interconnects
Dewan Saiham, **Di Wu**, Sazadur Rahman[*]
ACM Transactions on Design Automation of Electronic Systems (2026)

[2] Towards Plug & Play Myoelectric Control via One-Shot Latent Representation Alignment
Zubaidah Al-Mashhadani[+], Trevor Overton, **Di Wu**, Mohsen Rakhshan[*]
IEEE Transactions on Neural Systems & Rehabilitation Engineering 33 (2025), pp. 4409–4421, DOI: 10.1109/TNSRE.2025.3626255

[3] Synergizing Quantum Techniques with Machine Learning for Advancing Drug Discovery Challenge
Zhiding Liang, Zichang He, Yue Sun, Dylan Herman, Qingyue Jiao, Yanzhang Zhu[+], Weiwen Jiang[*], Xiaowei Xu[*], **Di Wu**, Marco Pistoia[*], Yiyu Shi[*]
Scientific Reports 14.1 (2024), p. 31216, DOI: 10.1038/s41598-024-82576-4

[4] uGEMM: Unary Computing for GEMM Applications
Di Wu, Jingjie Li, Ruokai Yin, Hsuan Hsiao, Younghyun Kim[*], Joshua San Miguel[×]
IEEE Micro 41.3 (2021), pp. 50–56, DOI: 10.1109/MM.2021.3065369
IEEE Micro Top Pick (12 from all computer architecture papers)

[5] In-Stream Correlation-Based Division and Bit-Inserting Square Root in Stochastic Computing
Di Wu, Ruokai Yin, Joshua San Miguel[×]
IEEE Design & Test 38.6 (2021), pp. 53–59, DOI: 10.1109/MDAT.2021.3050716

[6] Strategies for Reducing Decoding Cycles in Stochastic LDPC Decoders
Di Wu, Yun Chen, Qichen Zhang, Yeong-luh Ueng, Xiaoyang Zeng
IEEE Transactions on Circuits and Systems II: Express Briefs 63.9 (2016), pp. 873–877, DOI: 10.1109/TCSII.2016.2535038

[7] An Efficient Multirate LDPC-CC Decoder With a Layered Decoding Algorithm for the IEEE 1901 Standard
Yun Chen, Qichen Zhang, **Di Wu**, Changsheng Zhou, Xiaoyang Zeng
IEEE Transactions on Circuits and Systems II: Express Briefs 61.12 (2014), pp. 992–996, DOI: 10.1109/TCSII.2014.2362721

Workshop and Poster

[1] PyTorch-Based Implementation of Relay-BP
Thomas Danielsen[+], Caden Brock[+], Yanzhang Zhu[+], **Di Wu**
Florida Quantum Conference (2026)

[2] Mugi: Value Level Parallelism For Nonlinear Operations in LLMs
Daniel Price[+], Prabhu Vellaisamy, John Shen[*], **Di Wu**
Workshop on Unary Computing, collocated with ASPLOS (2026)

[3] Agraph: A unified Graph Representation for At-Will Simulation of Emerging Stacks
Daniel Price[+], Prabhu Vellaisamy, Patricia Gonzalez[*], George Michelogiannakis[*], John Shen[*], **Di Wu**
Workshop on Unary Computing, collocated with ASPLOS (2026)

[4] Sense-as-You-Go: A Neuromorphic Framework for Efficient Edge Sensing and Processing
Di Wu, Mohsen Rakhshan[*]
Energy Consequences of Information Workshop (2026)

[5] A-Graph: A Unified System Representation for Automated Cross-Stack Optimization
Di Wu
Energy Consequences of Information Workshop (2026)

[6] Catwalk Neuron for Efficient Temporal Neural Networks
Di Wu
Energy Consequences of Information Workshop (2026)

[7] EnerQy: Energy Estimation for Quantum Computing
Di Wu, Siyuan Niu[*], Zhipeng Deng[*]
Energy Consequences of Information Workshop (2026)

[8] Syndrilla: Simulating Decoders for Quantum Error Correction using PyTorch
Yanzhang Zhu[+], Chen-Yu Peng, Yun Hao Chen, Siyuan Niu[*], Yeong-Luh Ueng[*], **Di Wu**
IEEE International Conference on Quantum Computing and Engineering, 2025, DOI: 10.1109/QCE65121.2025.10470
Open-source software: Syndrilla

[9] Are We Scaling the Right Thing? A System Perspective on Test-Time Scaling
Youpeng Zhao, Jinpeng LV[*], **Di Wu**, Jun Wang[*], Christopher Gooley[*]
NeurIPS Workshop on Efficient Reasoning (2025)

[10] Synergizing Error Suppression, Mitigation and Correction for Fault-Tolerant Quantum Computing
Yanzhang Zhu[+], Siyuan Niu[*], **Di Wu**
Workshop on Quantum Intelligence, Learning & Security, collocated with IEEE TPS (2024), DOI: 10.1109/TPS-ISA62245.2024.00065

[11] Exploration of Unary Arithmetic-Based Matrix Multiply Units for Low Precision DL Accelerators
Prabhu Vellaisamy, Harideep Nair, **Di Wu**, Shawn Blanton[*], John Paul Shen[*]
Workshop on Unary Computing, collocated with ASPLOS (2024)

[12] xBrain: Brain-Like Computing for Explainable Brain-Computer Interfaces
Queenly Xie[+], Prabhu Vellaisamy, **Di Wu**
Young Architect Workshop, collocated with ASPLOS (2024)

[13] T-MAC: Temporal Multiplication with Accumulation
Zhewen Pan, **Di Wu**, Joshua San Miguel[×]
Young Architect Workshop, collocated with ASPLOS (2022)

Pre-Print

[1] Lottery BP: Unlocking Quantum Error Decoding at Scale
Yanzhang Zhu[+], Chen-Yu Peng, Yun Hao Chen, Yeong-Luh Ueng[*], **Di Wu**
arXiv Pre-print (2026)
Open-source software: Syndrilla

[2] uSense: Unary-Computing-based Stochastic Edge Neuromorphic Sensing
Zubaidah Al-Mashhadani[+], Mohsen Rakhshan[*], **Di Wu**
Research Square Pre-print (2026)

[3] A-Graph: A Unified Graph Representation for At-Will Simulation across System Stacks
Daniel Price[+], Prabhu Vellaisamy, Patricia Gonzalez[*], George Michelogiannakis[*], John Shen[*], **Di Wu**
arXiv Pre-print (2026)
Open-source software: Archx

[4] Chopper: A Multi-Level GPU Characterization Tool & Derived Insights Into LLM Training Inefficiency
Marco Kurzynski[+], Shaizeen Aga[*], **Di Wu**
arXiv Pre-print (2025)
Open-source software: Chopper

[5] Unleashing The Potential of LLMs for Quantum Computing: A Study in Quantum Architecture Design
Zhiding Liang, Jinglei Cheng, Rui Yang, Hang Ren, Zhixin Song, **Di Wu**, Tongyang Li[*], Yiyu Shi[*]
arXiv Pre-print (2023)

[6] Representation Range Needs for 16-Bit Neural Network Training
Valentina Popescu[*], Abhinav Venigalla[*], **Di Wu**, Robert Schreiber[*]
arXiv Pre-print (2021)
Open-source software: production-level cbfloat16

Patent

[1] Computer Architecture with Value-Level Parallelism
Joshua San Miguel, Di Wu, Zhewen Pan
US Patent Application 18/221,127, 2025

Invited Talks

Lit Silicon: How Thermal Variation Impairs AI Performance via Concurrent Execution in GPU Nodes and Its Simple Repairs
AMD Research and Advanced Development 05/2026

Value-Level Parallelism: New Opportunities for Parallel Computing from Fewer Bits in the AI Era
University of Pittsburgh 03/2026
University of Wisconsin–Madison 12/2025
Lawrence Berkeley National Laboratory 10/2025
NVIDIA Research 10/2025

The Phantom of the Datacenter: Unmasking the Culprit behind Performance Variation in GenAI Training
AMD Research and Advanced Development 10/2025

Salvage Deep Learning Efficiency: A Unary Computing Approach
University of California, Santa Cruz 02/2025
ShanghaiTech University 12/2024
Fudan University 12/2024
Case Western Reserve University 10/2024
Peking University 08/2024
University of Minnesota Twin Cities 03/2024
University of Louisiana at Lafayette 11/2023

Unary Computing for Power-Efficient Computer Architecture
AMD Research and Advanced Development 07/2023
University of Central Florida 02/2023
University of California, Los Angeles 11/2022
University of California, Santa Barbara 10/2022

Teaching and Mentoring

Instructor
EEL5796 (Big Data Computer Architecture and Systems), UCF SP 2026
EEL4768 (Computer Architecture), UCF SP 2026
EEE3342C (Digital Systems), UCF FA 2025
EEL4742C (Embedded Systems), UCF SP 2025
EEE3342C (Digital Systems), UCF FA 2024
EEL5796 (Big Data Computer Architecture and Systems), UCF SP 2024
ECE697 (Capstone Project in Machine Learning and Signal Processing), UW–Madison SU 2023

Teaching Assistant
ECE554 (Digital Engineering Lab), UW–Madison SP 2022
ECE454 (Mobile Computing Lab), UW–Madison FA 2021
ECE454 (Mobile Computing Lab), UW–Madison FA 2020
ECE554 (Digital Engineering Lab), UW–Madison SP 2020
ECE554 (Digital Engineering Lab), UW–Madison FA 2019
ECE554 (Digital Engineering Lab), UW–Madison SP 2019
ECE554 (Digital Engineering Lab), UW–Madison FA 2018
ECE552 (Introduction to Computer Architecture), UW–Madison FA 2018

Guest Lecturer
ECE18743 (Neuromorphic Computer Architecture & Processor Design), CMU SP 2025
ECE757 (Advanced Computer Architecture II), UW–Madison SP 2021
ECE757 (Advanced Computer Architecture II), UW–Madison SP 2020
ECE752 (Advanced Computer Architecture I), UW–Madison SP 2019

PROFESSIONAL SERVICE

Panelist
- NSF Medium Panel in Division of Computer and Network Systems (CNS) 2024
- IEEE Workshop on Quantum Intelligence, Learning & Security (QUILLS) 2024

Committee
- Organizing Chair of Artifact Evaluation at ISPASS 2026
- Organizing Chair of Workshop on Architecting Error Corrected Quantum Computers (ARQTEC) at HPCA 2025
- Organizing Chair of Undergrad Panel on "Charging STEM Career" at UCF 2024
- Organizing Chair of Workshop on Unary Computing (WUC) at ASPLOS 2024, 2026
- Organizing Committee of Quantum Computing for Drug Discovery Challenge at ICCAD 2023
- Program Committee of ASPLOS 2026
- Program Committee of HPCA 2024, 2025, 2026
- Program Committee of ISCA 2024, 2025, 2026
- Program Committee of MICRO 2025, 2026
- Program Committee of ICS 2026
- Program Committee of IISWC 2023, 2025
- Program Committee of ISPASS 2024, 2025, 2026
- Program Committee of IPDPS 2025
- Program Committee of DAC 2025
- Program Committee of ICCAD 2024
- Program Committee of DAC PhD Forum 2024
- Program Committee of ICCD 2023, 2024
- Program Committee of GLSVLSI 2026
- Program Committee of ICRC 2024
- Program Committee of ICA3PP 2023
- Program Committee of IEEE Workshop on Quantum Intelligence, Learning & Security (QUILLS) 2024
- Program Committee of Young Architect Workshop (YArch) 2023, 2024, 2025, 2026
- Program Committee of ASPLOS Artifact Evaluation 2020, 2021
- Program Committee of MICRO Artifact Evaluation 2021

Mentor
- UCF IEEE Engineering in Medicine and Biology Society (EMBS) Student Branch Chapter 2024
- Computer Architecture Long-term Mentoring (CALM) 2024
- Young Architect Workshop (YArch) 2023
- Undergrad Architecture Mentoring Workshop (uArch) 2023, 2024
- MICRO – "Meet a Senior PhD Student" 2020

Journal Reviewer
- ACM Transactions on Architecture and Code Optimization (TACO)
- ACM Transactions on Embedded Computing Systems (TECS)
- ACM Transactions on Reconfigurable Technology and Systems (TRETS)
- IEEE Computer Architecture Letter (CAL)
- IEEE Transactions on Circuits and Systems I (TCAS-I)
- IEEE Transactions on Computers (TC)
- IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)
- IEEE Transactions on Emerging Topics in Computing (TETC)
- IEEE Transactions on Very Large Scale Integration (VLSI) Systems (TVLSI)
- IEEE Signal Processing Letters (SPL)
- Journal of Network and Computer Applications (JNCA)
- npj Quantum Information
- npj Unconventional Computing

FUNDING

NSF EAGER: SENSE: National Security Evaluation of Neurotechnology Systems and Emerging Tools 2025
Co-PI: $100k/$300k

AMD Developer Cloud Credit: $2k 2025

Quantum Computing Access at NERSC (QCAN) Program 2025

AMD Faculty Award 2024
PI: $100k

AMD AI & HPC Cluster Award 2024, 2025, 2026

STUDENTS

Current PhD Students

Yang Yu	2026 – Present
Chetan Choudhary	2025 – Present
Daniel Price	2024 – Present
Marco Kurzynski	2024 – Present
Yanzhang Zhu	2024 – Present
Zubaidah Al-Mashhadani	2024 – Present

Former Students

Cory Brynds (BS/MS@UCF → full-time@AMD)	2023 – 2026
Lauren Caccamise (BS@UCF → PhD@Purdue)	2024 – 2025
Parker McLeod (BS@UCF → full-time@AMD)	2023 – 2024
Tyler Goldsmith (BS@UCF → full-time@AMD)	2023 – 2024
Mustafa Nisar (BS@UoT → co-op@AMD)	2023 – 2024
Zhewen Pan (MS@UW–Madison → PhD@UW–Madison)	2020 – 2022
Ruokai Yin (BS@UW–Madison → PhD@Yale)	2019 – 2021