Professor Di Wu

di.wu@ucf.edu
www.unarylab.com
https://scholar.google.com/citations?user=v6DNkTAAAAAJ
4328 Scorpius Street, Orlando, FL, 32816-2362, US

Appointment

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

Research Interest

- Computer architecture
  — performance, efficiency, etc.
- Emerging computing
  — unary, neuromorphic, quantum, photonic, etc.
- Domain specific acceleration
  — artificial intelligence, brain computer interface, etc.
- Heterogeneous system
  — Compute Express Link, multi-GPU, etc.

Education

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

Master of Engineering | ***Microelectronics*** 09/2012 – 01/2015
Fudan University Shanghai, China
Bachelor of Science | ***Microelectronics*** 09/2007 – 07/2012
Fudan University Shanghai, China

Employment

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

Honors and Awards

IEEE Micro Top Pick Honorable Mention 2025
1 of the 24 publications selected from all computer architecture publications in 2024
Harold Peterson Outstanding Dissertation Award 2024
UW–Madison
Distinguished Artifact Evaluation Award 2024
ASPLOS
Machine Learning and Systems Rising Star 2023
MLCommons
Capstone PhD Teaching Award Nomination 2022
UW–Madison
Grainger Wisconsin Distinguished Graduate Fellowship 2022
1 of the 3 PhD students selected from College of Engineering, UW–Madison
Student Travel Award 2022
ISCA
Dissertator Travel Award (twice) 2022
UW–Madison
Ph.D. Forum 2021
DAC

IEEE Micro Top Pick 2021
1 of the 12 publications selected from all computer architecture publications in 2020

Gerald Holdridge Outstanding Teaching Assistant Award 2020
UW–Madison

Chancellor's Opportunity Fellowship 2019
UW–Madison

Qualcomm Innovation Fellowship Finalist 2019
Qualcomm

Foxconn SmartCity Competition Winner 2019
Foxconn

Student Research Travel Award 2019
UW–Madison

Student Research Competition Travel Award 2019
ASPLOS

Student Travel Award 2019
ASPLOS

Hiran Mayukh Award 2018
UW–Madison

Rising Star Award 2015
HiSilicon

National Scholarship (ranking 1/67) 2015
Fudan University

Excellent Student Union Leader 2010
Fudan University

Third Prize Freshman Scholarship (ranking 3/45) 2007
Fudan University

PUBLICATIONS

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

Conference

[1] 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

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

[3] LoAS: Fully Temporal-Parallel Datatflow for Dual-Sparse Spiking Neural Networks
Ruokai Yin, Youngeun Kim, **Di Wu**, Priyadarshini Panda*
International Symposium on Microarchitecture, 2024, DOI: 10.1109/MICRO61859.2024.00084

[4] Evaluating Unary GEMM for Low-Precision AI: Toward Scalable Energy-Efficient 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

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

[6] Carat: Unlocking Value-Level Parallelism for Multiplier-Free GEMMs
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 2025 (24 from all computer architecture papers)
Distinguished Artifact Evaluation Award
Open-source artifact

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

[8] uSystolic: Byte-Crawling Unary Systolic Array
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

[9] 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

[10] 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

[11] 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

[12] uGEMM: Unary Computing Architecture for GEMM Applications
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 2021 (12 from all computer architecture papers)
Open-source software: UnarySim

[13] 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

[14] 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

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

[16] 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

[17] 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

[18] 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

[19] 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] 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 (2025), p. 31216, DOI: 10.1038/s41598-024-82576-4

[2] 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 Issue 2021

[3] 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

[4] 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

[5] 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

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

[2] 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 International Conference on Architectural Support for Programming Languages and Operating Systems (2024)

[3] xBrain: Brain-Like Computing for Explainable Brain-Computer Interfaces
Queenly Xie[+], Prabhu Vellaisamy, **Di Wu**
Young Architect Workshop, collocated with International Conference on Architectural Support for Programming Languages and Operating Systems (2024)

[4] T-MAC: Temporal Multiplication with Accumulation
Zhewen Pan, **Di Wu**, Joshua San Miguel[×]
Young Architect Workshop, collocated with International Conference on Architectural Support for Programming Languages and Operating Systems (2022)

Pre-Print

[1] 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)

[2] Representation Range Needs for 16-Bit Neural Network Training
Valentina Popescu[*], Abhinav Venigalla[*], **Di Wu**, Robert Schreiber[*]
arXiv Pre-print (2021)
Industry adoption: Automatic Mixed Precision – cbfloat16

Invited Talks

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 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
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 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
- Organizing Committee of Quantum Computing for Drug Discovery Challenge at ICCAD 2023
- Program Committee of ASPLOS 2026
- Program Committee of HPCA 2024, 2025
- Program Committee of ISCA 2024, 2025
- Program Committee of MICRO 2025
- Program Committee of IISWC 2023
- Program Committee of ISPASS 2024, 2025
- 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 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
- 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 Signal Processing Letters (SPL)
- Journal of Network and Computer Applications (JNCA)

Funding

Awarded
- AMD AI & HPC Fund (GPU node hours) 2024, 2025
- AMD Fund for Academic Research (Unrestricted Gift): $100k 2024
- Quantum Computing Access at NERSC (QCAN) Program 2025

Students

Current Students
- Daniel Price (PhD) 2024 – Present
- Marco Kurzynski (PhD) 2024 – Present
- Yanzhang Zhu (PhD) 2024 – Present
- Zubaidah Al-Mashhadani (PhD) 2024 – Present

Former Students
- Parker McLeod (UG, AMD, full-time) 2023 – 2024
- Tyler Goldsmith (UG, AMD, full-time) 2023 – 2024
- Mustafa Nisar (UG, AMD, full-time) 2023 – 2024
- Zhewen Pan (MS, UW–Madison, PhD) 2020 – 2022
- Ruokai Yin (UG, Yale, PhD) 2019 – 2021

Reference

Professor Joshua San Miguel — PhD advisor
jsanmiguel@wisc.edu — University of Wisconsin–Madison

Professor Younghyun Kim — PhD committee member
younghyun.kim@wisc.edu — University of Wisconsin–Madison

Professor John Paul Shen
jpshen@cmu.edu — Carnegie Mellon University

Professor Timothy Sherwood
sherwood@cs.ucsb.edu — University of California, Santa Barbara

Professor Ulya R. Karpuzcu
ukarpuzc@umn.edu — University of Minnesota, Twin Cities