

PROFESSOR DI WU

✉ di.wu@ucf.edu

🏠 www.unarylab.com

📍 4328 Scorpius Street, Orlando, FL, 32816-2362, US

FACULTY APPOINTMENT

| | |
|--|---|
| Assistant Professor <i>Electrical and Computer Engineering</i> University of Central Florida | Aug. 2023 – Present Orlando, FL, USA |
| Joint Assistant Professor <i>Computer Science</i> University of Central Florida | Aug. 2023 – Present Orlando, FL, USA |

EDUCATION




| | |
|---|---|
| Doctor of Philosophy <i>Electrical and Computer Engineering</i> University of Wisconsin–Madison <ul style="list-style-type: none">• Advisor: Joshua San Miguel• Thesis: Power-Efficient Computer Architecture via Unary and Approximate Computing<ul style="list-style-type: none">🏆 Harold Peterson Outstanding Dissertation Award | Sep. 2017 – Jul. 2023 Madison, WI, USA |
| Master of Engineering <i>Microelectronics</i> Fudan University | Sep. 2012 – Jan. 2015 Shanghai, China |
| Bachelor of Science <i>Microelectronics</i> Fudan University | Sep. 2007 – Jul. 2012 Shanghai, China |

RESEARCH INTEREST

- Computer architecture
— efficiency, etc.
- Emerging computing
— neuromorphic, quantum, etc.
- Domain specific acceleration
— AI, BCI, cryptography, etc.
- Heterogeneous system
— CXL, etc.



HONORS AND AWARDS

| | |
|---|------|
| Harold Peterson Outstanding Dissertation Award 🔗 UW–Madison | 2024 |
| Distinguished Artifact Award 🔗 ASPLOS | 2024 |
| Machine Learning and Systems Rising Star MLCommons | 2023 |
| Capstone PhD Teaching Award Nomination UW–Madison | 2022 |
| Grainger Wisconsin Distinguished Graduate Fellowship 1 of the 3 PhD students selected from College of Engineering, UW–Madison | 2022 |
| Student Travel Award ISCA | 2022 |

| | |
|---|------|
| Dissertator Travel Award (twice) UW–Madison | 2022 |
| Ph.D. Forum Invitation DAC | 2021 |
| IEEE Micro Top Pick  1 of the 12 publications selected from all computer architecture publications in 2020 | 2021 |
| Gerald Holdridge Outstanding Teaching Assistant Award UW–Madison | 2020 |
| Chancellor’s Opportunity Fellowship UW–Madison | 2019 |
| Student Research Travel Award UW–Madison | 2019 |
| Student Research Competition Travel Award ASPLOS | 2019 |
| Student Travel Award ASPLOS | 2019 |
| Qualcomm Innovation Fellowship Finalist  Qualcomm | 2019 |
| Foxconn SmartCity Competition Winner Foxconn | 2019 |
| Hiran Mayukh Award  UW–Madison | 2018 |
| Rising Star Award HiSilicon | 2015 |
| National Scholarship (ranking 1/67) Fudan University | 2015 |
| Excellent Student Union Leader Fudan University | 2010 |
| Third Prize Freshman Scholarship (ranking 3/45) Fudan University | 2007 |

PUBLICATIONS


Conference

- [1] 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 (ISCA), 2024
- [2] Carat: Unlocking Value-Level Parallelism for Multiplier-Free GEMMs [Acceptance rate=18.6%]
Zhewen Pan, Joshua San Miguel, **Di Wu**
International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS), 2024
[Open-source artifact](#) 
 **Distinguished Artifact Award**

- [3] uBrain: A Unary Brain Computer Interface [Acceptance rate=16.8%]
Di Wu, Jingjie Li, Zhewen Pan, Younghyun Kim, Joshua San Miguel
International Symposium on Computer Architecture (ISCA), 2022
- [4] uSystolic: Byte-Crawling Unary Systolic Array [Acceptance rate=30.5%]
Di Wu, Joshua San Miguel
International Symposium on High-Performance Computer Architecture (HPCA), 2022
[Open-source software: uSystolic-Sim](#) 
- [5] When Dataflows Converge: Reconfigurable and Approximate Computing for Emerging Neural Networks [Invited]
Di Wu, Joshua San Miguel
International Conference on Computer Design (ICCD), 2021
- [6] UNO: Virtualizing and Unifying Nonlinear Operations for Emerging Neural Networks [Acceptance rate=26.4%]
Di Wu, Jingjie Li, Setareh Behrooz, Younghyun Kim, Joshua San Miguel
International Symposium on Low Power Electronics and Design (ISLPED), 2021
- [7] Normalized Stability: A Cross-Level Design Metric for Early Termination in Stochastic Computing [Acceptance rate=30.2%]
Di Wu, Ruokai Yin, Joshua San Miguel
Asia and South Pacific Design Automation Conference (ASP-DAC), 2021
- [8] uGEMM: Unary Computing Architecture for GEMM Applications [Acceptance rate=18.3%]
Di Wu, Jingjie Li, Ruokai Yin, Hsuan Hsiao, Younghyun Kim, Joshua San Miguel
International Symposium on Computer Architecture (ISCA), 2020
[Open-source software: UnarySim](#) 
 **Awarded 1 out of 12 IEEE Micro Top Picks 2021 from all computer architecture publications in 2020**
- [9] Approximate Hardware Techniques for Energy-Quality Scaling Across the System [Invited]
Younghyun Kim, Joshua San Miguel, Setareh Behrooz, Tianen Chen, Kyuin Lee, Yongwoo Lee, Jingjie Li, **Di Wu**
International Conference on Electronics, Information, and Communication (ICEIC), 2020
- [10] SECO: A Scalable Accuracy Approximate Exponential Function Via Cross-Layer Optimization [Acceptance rate=35.2%]
Di Wu, Tianen Chen, Chienfu Chen, Oghenefego Ahia, Joshua San Miguel, Mikko Lipasti, Younghyun Kim
International Symposium on Low Power Electronics and Design (ISLPED), 2019
- [11] In-Stream Stochastic Division and Square Root via Correlation [Acceptance rate=18.9%]
Di Wu, Joshua San Miguel
Design Automation Conference (DAC), 2019
- [12] 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 (ICASSP), 2016
- [13] 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 (DSP), 2015
- [14] 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 (ISCAS), 2015

- [15] A High-Throughput LDPC Decoder for Optical Communication
Di Wu, Yun Chen, Yuebin Huang, Yeongluh Ueng, Lirong Zheng, Xiaoyang Zeng
International Conference on ASIC (ASICON), 2013


Journal

- [1] uGEMM: Unary Computing for GEMM Applications [Acceptance rate=9.9%]
Di Wu, Jingjie Li, Ruokai Yin, Hsuan Hsiao, Younghyun Kim, Joshua San Miguel
IEEE Micro 41.3 (2021), pp. 50–56
 IEEE Micro Top Pick Issue 2021
- [2] 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
- [3] 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
- [4] 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

Workshop

- [1] 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 (WUC), collocated with International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS) (2024)
- [2] xBrain: Brain-Like Computing for Explainable Brain-Computer Interfaces
Queenly Xie, Prabhu Vellaisamy, **Di Wu**
Young Architect Workshop (YArch), collocated with International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS) (2024)
- [3] T-MAC: Temporal Multiplication with Accumulation
Zhewen Pan, **Di Wu**, Joshua San Miguel
Young Architect Workshop (YArch), collocated with International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS) (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)
[Resource: Mixed-Precision Training – Data Formats – CB16 Half-Precision](#) 

INVITED TALKS

| | |
|---|-----------------------------------|
| Salvage Deep Learning Efficiency: A Unary Computing Approach University of Minnesota Twin Cities | Mar. 2024 Minneapolis, MN, USA |
| Salvage Deep Learning Efficiency: A Unary Computing Approach University of Louisiana at Lafayette | Nov. 2023 Virtual |
| Unary Computing for Power-Efficient Computer Architecture AMD Research | Jul. 2023 Virtual |
| Unary Computing for Power-Efficient Computer Architecture University of Central Florida | Feb. 2023 Orlando, FL, USA |
| Unary Computing for Power-Efficient Computer Architecture University of California, Los Angeles | Nov. 2022 Virtual |
| uBrain: A Unary Brain Computer Interface University of Central Florida | Nov. 2022 Virtual |
| Unary Computing for Power-Efficient Computer Architecture University of California, Santa Barbara | Oct. 2022 Virtual |

EMPLOYMENT

| | |
|--|--|
| Research Assistant Department of Electrical and Computer Engineering, UW–Madison | Sep. 2017 – Jul. 2023 Madison, WI, USA |
| Research Intern Cerebras Systems | May 2022 – Sep. 2022 Sunnyvale, CA, USA |
| Research Intern Cerebras Systems | May 2020 – Sep. 2020 Sunnyvale, CA, USA |
| Research Intern Meta (Formerly Facebook) | May 2019 – Sep. 2019 Palo Alto, CA, USA |
| Digital Circuit Engineer HiSilicon | Mar. 2015 – May 2017 Shanghai, China |
| Research Assistant State Key Laboratory of ASIC and System, Fudan University | Sep. 2012 – Jan. 2015 Shanghai, China |

TEACHING AND MENTORING

| | |
|--|-------------|
| Instructor EEL5796 (Big Data Computer Architecture and Systems), UCF | Spring 2024 |
| ECE697 (Capston Project in Machine Learning and Signal Processing), UW–Madison | Summer 2023 |
| Teaching Assistant ECE554 (Digital Engineering Lab), UW–Madison | Spring 2022 |
| ECE454 (Mobile Computing Lab), UW–Madison | Fall 2021 |
| ECE454 (Mobile Computing Lab), UW–Madison | Fall 2020 |
| ECE554 (Digital Engineering Lab), UW–Madison | Spring 2020 |
| ECE554 (Digital Engineering Lab), UW–Madison | Fall 2019 |
| ECE554 (Digital Engineering Lab), UW–Madison | Spring 2019 |
| ECE554 (Digital Engineering Lab), UW–Madison | Fall 2018 |

| | |
|--|-------------|
| ECE552 (Introduction to Computer Architecture), UW–Madison | Fall 2018 |
| Guest Lecturer | |
| ECE757 (Advanced Computer Architecture II), UW–Madison | Spring 2021 |
| ECE757 (Advanced Computer Architecture II), UW–Madison | Spring 2020 |
| ECE752 (Advanced Computer Architecture I), UW–Madison | Spring 2019 |

PROFESSIONAL SERVICE

Panelist

NSF Medium Panel in Division of Computer and Network Systems (CNS) 2024

Committee

Organizing Committee of Workshop on Unary Computing (WUC) 2024

Organizing Committee of Quantum Computing for Drug Discovery Challenge at ICCAD 2023

Program Committee of PhD Forum at DAC 2024

Program Committee of Young Architect Workshop (YArch) 2024

Program Committee of ISCA 2024

Program Committee of ISPASS 2024

Program Committee of HPCA 2024

Program Committee of ICA3PP 2023

Program Committee of ICCD 2023

Program Committee of IISWC 2023

Program Committee of Young Architect Workshop (YArch) 2023

Program Committee of ASPLOS Artifact Evaluation 2021

Program Committee of MICRO Artifact Evaluation 2021

Program Committee of ASPLOS Artifact Evaluation 2020

Mentor

Computer Architecture Long-term Mentoring (CALM) 2024

Young Architect Workshop (YArch) 2023

Undergrad Architecture Mentoring Workshop (uArch) 2023

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 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)

FORMER STUDENTS

| | |
|---|-----------|
| Tyler Goldsmith | 2024 |
| AMD Orlando | Full-time |
| Mustafa Nisar | 2024 |
| AMD Toronto | Co-op |
| James Chen | 2024 |
| Department of Electrical and Computer Engineering, UCF | Undergrad |
| Francisco Soriano | 2024 |
| Department of Electrical and Computer Engineering, UCF | Undergrad |
| Zhewen Pan | 2022 |
| Department of Electrical and Computer Engineering, UW–Madison | PhD |
| Ruokai Yin | 2021 |
| Department of Electrical Engineering, Yale University | PhD |

REFERENCE

Professor Joshua San Miguel

jsanmiguel@wisc.edu

Professor Younghyun Kim

younghyun.kim@wisc.edu

Professor John Paul Shen

jpshen@cmu.edu

Professor Timothy Sherwood

sherwood@cs.ucsb.edu

Professor Ulya R. Karpuzcu

ukarpuzc@umn.edu

PhD advisor

University of Wisconsin–Madison

PhD committee member

University of Wisconsin–Madison

Carnegie Mellon University

University of California, Santa Barbara

University of Minnesota, Twin Cities