

# CURRICULUM VITAE FOR DI WU

Homepage: [diwu1990.github.io](http://diwu1990.github.io)

Phone: +1 608-886-2103

E-mail: [di.wu@ece.wisc.edu](mailto:di.wu@ece.wisc.edu)

## RESEARCH INTEREST

---

Low Power Computer Architecture, Neural Network Optimization

## EDUCATION

---

Sep 2017 – Now    **Ph.D Student in Department of ECE** at University of Wisconsin–Madison, USA

- **Advisor:** Joshua San Miguel; **GPA:** 3.82

Sep 2012 – Jan 2015    **M.E. in IC Engineering** at Fudan University, China

- **Academic Ranking:** 1/67; **GPA Ranking:** 3/67

Sep 2007 – Jun 2012    **B.S. in Micro-Electronics** at Fudan University, China

## HONOR

---

Jan 2021    uGEMM (ISCA'20) selected in **IEEE Micro Top Picks** for 2020

May 2020    **Gerald Holdridge Outstanding Teaching Assistant Awards** from UW–Madison

Jun 2019    **Chancellor's Opportunity Fellowship (COF)** from UW–Madison

Apr 2019    **ASPLOS 2019 SRC Grant** from ACM ASPLOS 2019, Providence, RI

Mar 2019    **Student Research Travel Grant–Conference** from UW–Madison

Mar 2019    **Student Research Competition (SRC) Candidate** in ACM ASPLOS 2019, Providence, RI

Mar 2019    **Finalist** in Qualcomm Innovation Fellowship 2019, UW–Madison

Mar 2019    **Winner** in Foxconn SmartCity Competition 2019, UW–Madison

Feb 2019    **Winner of Student Travel Grant** in ACM ASPLOS 2019, Providence, RI

Oct 2018    **The Hiran Mayukh Award** on Computer Architecture Affiliate Meeting 2018, UW–Madison

Sep 2015    **Excellent New Employee Prize** in Hisilicon Technologies Co., Ltd

Mar 2014    **National Scholarship** for Excellent Graduate Student of Fudan University (1/67)

Sep 2007    **Third Prize Scholarship** for Excellent Freshman of Fudan University (3/45)

Jun 2007    **Ranking 88/450,000** in National College Entrance Examination in Anhui Province, China

## PUBLICATION

---

### Journal:

**J4. D. Wu**, J. Li, R. Yin, H. Hsiao, Y. Kim and J. San Miguel, "uGEMM: Unary Computing for GEMM Applications," in *IEEE Micro*, 2021.

**J3. D. Wu**, R. Yin and J. San Miguel, "In-Stream Correlation-Based Division and Bit-Inserting Square Root in Stochastic Computing", in *IEEE Design & Test*, 2021.

**J2. D. Wu**, Y. Chen, Q. Zhang, Y. Ueng and X. Zeng, "Strategies for Reducing Decoding Cycles in Stochastic LDPC Decoders," in *TCASII*, 2016.

**J1. Y. Chen**, Q. Zhang, **D. Wu**, C. Zhou and X. Zeng, "An Efficient Multirate LDPC-CC Decoder With a Layered Decoding Algorithm for the IEEE 1901 Standard," in *TCASII*, 2014.

### Conference:

**C12. D. Wu**, etc., "uSystolic: Byte-Crawling Unary Systolic Array," *under review*.

- C11. V. Popescu, A. Venigalla, **D. Wu**, R. Schreiber, "Representation range needs for 16-bit neural network training," [under review](#).
- C10. **D. Wu**, etc., "UNO: Virtualizing and Unifying Nonlinear Operations for Emerging Neural Networks," [under review](#).
- C9. **D. Wu**, R. Yin and J. San Miguel, "Normalized Stability: A Cross-Level Design Metric for Early Termination in Stochastic Computing," in *ASP-DAC*, 2021.
- C8. **D. Wu**, J. Li, R. Yin, H. Hsiao, Y. Kim and J. San Miguel, "uGEMM: Unary Computing Architecture for GEMM Applications," in *ISCA*, 2020. *Selected to IEEE Micro Top Picks for 2020*.
- C7. Y. Kim, J. San Miguel, S. Behroozi, T. Chen, K. Lee, Y. Lee, J. Li, and **D. Wu**, "Approximate Hardware Techniques for Energy-Quality Scaling Across the System," in *ICEIC*, 2020. (Invited)
- C6. **D. Wu**, T. Chen, C. Chen, O. Ahia, J. San Miguel, M. Lipasti and Y. Kim, "SECO: A Scalable Accuracy Approximate Exponential Function Via Cross-Level Optimization," in *ISLPED*, 2019.
- C5. **D. Wu** and J. San Miguel, "In-Stream Stochastic Division and Square Root via Correlation," in *DAC*, 2019.
- C4. Q. Zhang, Y. Chen, **D. Wu**, X. Zeng and Y. Ueng, "Convergence-optimized variable node structure for stochastic LDPC decoder," in *ICASSP*, 2016.
- C3. Q. Zhang, Y. Chen, **D. Wu**, X. Zeng and Y. Ueng, "An area-efficient architecture for stochastic LDPC decoder," in *DSP*, 2015.
- C2. **D. Wu**, Y. C., Q. Zhang, L. Zheng, X. Zeng and Y. Ueng, "Latency-optimized stochastic LDPC decoder for high-throughput applications," in *ISCAS*, 2015.
- C1. **D. Wu**, Y. Chen, Y. Huang, Y. Ueng, L. Zheng and X. Zeng, "A high-throughput LDPC decoder for optical communication," in *ASICON*, 2013.

## RESEARCH

---

Feb 2018 – Now    Research Assistant in STACS Lab, Department of ECE, UW–Madison

### ♦ Approximate Computing

- **RAVEN: A Reconfigurable Architecture for Varying Emerging Neural Networks.** (*Qualcomm Innovation Fellowship 2019 Finalist*)
  - Linear-nonlinear compatibility.
  - Dense-sparse compatibility. ([in progress](#))
  - Task-aware resource partition. ([in progress](#))
- **Unified Nonlinear Operation (UNO)** ([under review](#))
  - Compatibility to nonlinearity using linear MAC.
  - Dynamic run-time accuracy-energy scaling using Horner's rule.
  - Accuracy validation on emerging DNNs, e.g. CapsNet, GCN and NMT.
  - Prototyping a UNO-based accelerator.
- **Approximate Exponential Function (C6: ISLPED 2019)**
  - Algorithm-level optimization for dynamic run-time accuracy-energy scaling.
  - Circuit-level error bound modeling static design-time accuracy-energy scaling.
  - Division simplification and multiplication skip.

### ♦ Unary Computing [Website]

- **General-purpose Simulator:** Open-source *UnarySim* [Github] to simulate unary computing.
  - PyTorch as backbone.
  - Cycle-accurate simulation.
  - Encoding compatibility, i.e., both rate and temporal codings.
  - Metric-based early termination for stochastic computing. (C9: *ASP-DAC 2021*)

- **Fully Streaming Unary Architecture**
  - Unary GEMM Architecture (uGEMM). (C8: ISCA 2020; J4: IEEE Micro 2021 Top Pick Issue)
  - In-stream Division and Square Root. (C5: DAC 2019; J3: IEEE D&T 2020)
- **Hybrid Unary-Binary Architecture**
  - Unary Systolic Array (uSystolic). (under review)
- **System-Level Performance Simulation**
  - Support for holistic memory hierarchy.
  - Computing scheme awareness.
  - Flexibility for arbitrary computing kernel.
- **Design Automation**
  - End-to-end mapping from algorithm simulation to hardware implementation. (in progress)
- **Stochastic CapsNet Architecture (ASPLOS 2019 SRC)**
  - Nonlinear stochastic computing units.
  - Dataflow-level feedback loops.
  - Logic sharing.
- **PIM-based Bitstream Generation**
  - Design processing in memory (PIM) architecture for generating bitstreams in DRAM.
  - Model the proposed architecture with an in-house simulator.
  - Test the bitstream randomness with NIST SP 800-22 statistical test suit.

#### ◇ Deep Learning

- **Security Side of Homomorphic Inference**
  - Successfully attack a DNN model using homomorphic encryption.
- **Style Transfer for Medical Images**
  - Modality transfer on medical images. [slides]
- **Automatic Quantization Platform**
  - Fast distribution-based fixed-point quantization for DNN.

Sep 2017 – Jan 2018 Research Assistant in WiCIL Lab, Department of ECE, UW–Madison

#### ◇ Hardware Architecture for Convolutional Neural Networks

- **Sparse Neural Network**
  - Resource estimation of a channel-level-pruned DNN.
  - Performance modeling for the DNN.

Aug 2012 – Jan 2015 State Key Laboratory of ASIC and System, Fudan University

#### ◇ High Throughput Low-Density Parity-Check (LDPC) Architecture Design

- **Latency Optimization in Stochastic LDPC decoding (C2: ISCAS 2015; J2: TCASII 2016)**
  - Fast channel information initialization.
  - Bit-flipping for early termination.
  - Posterior information recovering for accurate decision.
- **Dual clock edge triggered LDPC decoding (C1: AISCON 2013)**

## INDUSTRIAL EXPERIENCE

---

May 2020 – Aug 2020 Research Intern in Cerebras Systems.

#### ◇ Numerical Optimization for Deep Neural Network

- Explore novel 16-bit fp format for training both CV and language models. (under review)
- Simulate the training process on the real hardware with boundary casting in software simulator.

**May 2019 – Aug 2019   Research Intern in Facebook Inc.**

◊ **Lossless Quantization for Video Model**

- Post training quantization for ResNet-50 with Min-Max and distribution-based quantization.
- Perform quantization-aware training for ResNext-50 with channel-wise 3D convolution.

**Mar 2015 – May 2017   Digital Circuit Engineer in Hisilicon Technologies Co., Ltd.**

◊ **Algorithm Analysis for SoC**

- DSP resource estimation based on parallelism and ISA.
- Optimize task scheduling for both DSP and SoC.
- SoC bandwidth estimation.

◊ **Hardware Design for DSP**

- Customize the interface between DSP and SoC using ARM AXI-4 protocol.
- Unit-level formal verification using the Cadence Jaspergold.
- Power analysis for DSP system.

◊ **Design Automation**

- Automate place-and-route.
- Automate CAD report analysis.
- ISA auto-encoding using discrete gradient descent.

## SKILL

---

**Programming language:** Verilog HDL, Python, C++, Perl, MATLAB

**Hardware:** Modelsim, Design Compiler, VCS, Verdi, Quartus Prime, Vivado

**Software:** PyTorch, TensorFlow, Caffe2

## ACADEMIC ACTIVITY

---

**Apr 2021   Guest Lecture on Unary Computing** in ECE757 (Advanced Computer Architecture II), UW–Madison

**Jan 2021   Artifact Evaluation Committee** for ASPLOS 2021

**Seq 2020   External Reviewer** for HPCA 2021

**Sep 2020   Teaching Assistant** for ECE454 (Mobile Computing Lab), UW–Madison

**Feb 2020   Guest Lecture on Introduction to Unary Computing** in ECE757 (Advanced Computer Architecture II), UW–Madison

**Jan 2020   Teaching Assistant** for ECE554 (Digital Circuit Lab), UW–Madison

**Jan 2020   External Reviewer** for ISCA 2020

**Jan 2020   External Reviewer** for DAC 2020

**Jan 2020   Artifact Evaluation Committee** for ASPLOS 2020

**Seq 2019   Teaching Assistant** for ECE554 (Digital Circuit Lab), UW–Madison

**Apr 2019   External Reviewer** for MICRO 2019

**Apr 2019   ASPLOS 2019** attendee in Providence, RI, USA

**Mar 2019   Guest Lecture on Deep Neural Network and Stochastic Computing** in ECE752 (Advanced Computer Architecture I), UW–Madison

**Feb 2019   External Reviewer** for ISCA 2019

**Jan 2019   Research Assistant** with Prof. Joshua San Miguel

**Jan 2019   Teaching Assistant** for ECE554 (Digital Circuit Lab), UW–Madison

**Seq 2018**    **Teaching Assistant** for ECE552 (Introduction to Computer Architecture) and ECE554 (Digital Circuit Lab), UW-Madison  
**Jan 2018**    **External Reviewer** for DAC 2018  
**Seq 2017**    **Research Assistant** in WiCIL

## **Non-Academic Activity**

---

**Nov 2011**    **Campus Computer Gaming Champion of DOTA** in Fudan University  
**May 2010**    **Volunteer** for EXPO 2010 in Shanghai  
**Mar 2010**    **Excellent Student Leader** for Excellent Student Union Leader of Fudan University  
**Sep 2009**    **Minister of Sports** in the Student Union of Fudan University  
**2009,10,12**    **Campus Basketball Champion** in Zhangjiang Campus of Fudan University  
**May 2008**    **Volunteer** for helping disabled students of Guanxin School in Shanghai