Ruokai Yin

PhD candidate, ECE, Yale University

ruokai.yin@yale.edu | Google Scholar | Github | Personal Website

Research Interests

- Computer Architecture
  - → Accelerator Design
    - → Systolic-Array
    - → Sparse Tensor Accelerator
    - → SIMT Architecture
  - → Performance Modeling
- Neuromorphic Computing
  - → Spiking Neural Networks
- Stochastic Computing
  - → Unary Computing
- Efficient Machine Learning Co-Design
  - → Network Compression
    - → Quantization
    - → Pruning

Education

Ph.D., Electrical and Computer Engineering, Yale University **Sep. 2021 — Current**

Advisor: Prof. Priyadarshini Panda

M.S. & M.Phil., Electrical and Computer Engineering, Yale University **Sep. 2021 — May. 2024**

en route

B.S., Electrical Engineering & Computer Science & Math, University of Wisconsin - Madison **Sep. 2018 — May. 2021**

Graduate with the highest honor, GPA: 3.98/4.00

Experience

Research Intern, ASIC team, Cerebras Systems, **May. 2024 — Aug. 2024**

- advisor: Vipin Sharma

- Architecture design and modeling for Cerebras's next-generation wafer-scale engine, with a focus on:

- Exploring PE scalability and projecting performance.

- Modeling and projecting performance for several new architectural features.

- Modeling inter-PE fabric movement for both dense and sparse LLM workloads.

- Modeling the checkpoints evict/refill process at both PE and IO levels.

Research Assistant, ICL Lab, Yale University, **July. 2021 — Current**

- advisor: Prof. Priyadarshini Panda

- Computer architectures, systems, and algorithm co-design for neural network acceleration & neuromorphic computing.

Research Assistant, UW STACS Lab, University of Wisconsin - Madison, **June. 2019 — May. 2021**

- advisor: Prof. Joshua San Miguel

- Computer architectures & systems for unary & stochastic computing.

Publications [Conference]

Neuromorphic Computing:

LoAS: Fully Temporal-Parallel Dataflow for Dual-Sparse Spiking Neural Networks.
Ruokai Yin, Youngeun Kim, Di Wu, and Priyadarshini Panda
International Symposium on Microarchitecture (MICRO) 2024.

MINT: Multiplier-less Integer Quantization for Spiking Neural Networks.
Ruokai Yin, Yuhang Li, Abhishek Moitra, and Priyadarshini Panda
Asia and South Pacific Design Automation Conference (ASP-DAC) 2024, **Nomination of Best Paper (2% of submitted papers).**

TT-SNN: Tensor Train Decomposition for Efficient Spiking Neural Network Training.
Donghyun Li, **Ruokai Yin**, Youngeun Kim, Abhishek Moitra, Yuhang Li, and Priyadarshini Panda
Design Automation and Test in Europe (DATE) 2024.

Are SNNs Truly Energy-efficient? − A Hardware Perspective.
Abhiroop Bhattacharjee*, **Ruokai Yin***, Abhishek Moitra, and Priyadarshini Panda
IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2024.

Lottery Ticket Hypothesis for Spiking Neural Networks.
Youngeun Kim, Yuhang Li, Hyoungseob Park, Yeshwanth Venkatesha, **Ruokai Yin**, and Priyadarshini Panda
European Conference on Computer Vision (ECCV) 2022, **Oral Presentation (2.7% of submitted papers).**

Wearable-based Human Activity Recognition with Spatio-Temporal Spiking Neural Networks.
Yuhang Li, **Ruokai Yin**, Hyoungseob Park, Youngeun Kim, and Priyadarshini Panda
Conference on Neural Information Processing Systems (NeurIPS) 2022 Workshop, **Spotlight Paper.**

Stochastic Computing:

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 (ISCA) 2020, **IEEE Top-pick 2020.**

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 (ASP-DAC) 2021

Publications [Journal]

Neuromorphic Computing:

Workload-balanced Pruning for Sparse Spiking Neural Networks.
Ruokai Yin, Youngeun Kim, Yuhang Li, Abhishek Moitra, Nitin Satpute, Anna Hambitzer, Priyadarshini Panda
IEEE Transactions on Emerging Topics in Computational Intelligence (TETCI), 2024.

Rethinking Skip Connections in Spiking Neural Networks with Time-To-First-Spike Coding.
Youngeun Kim, Adar Kahana, **Ruokai Yin**, Yuhang Li, Panos Stinis, George Em Karniadakis, Priyadarshini Panda
Frontiers in Neuroscience, 2024.

Efficient Human Activity Recognition with Spatio-Temporal Spiking Neural Networks.
Yuhang Li, **Ruokai Yin**, Youngeun Kim, and Priyadarshini Panda
Frontiers in Neuroscience, 2023.

Sharing Leaky-Integrate-and-Fire Neurons for Memory-Efficient Spiking Neural Networks.
Youngeun Kim, Yuhang Li, Abhishek Moitra, **Ruokai Yin**, and Priyadarshini Panda
Frontiers in Neuroscience, 2023.

SATA: Sparsity-Aware Training Accelerator for Spiking Neural Networks.
Ruokai Yin, Abhishek Moitra, Abhiroop Bhattacharjee, Youngeun Kim, and Priyadarshini Panda
IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD), 2022.

Stochastic Computing:

uGEMM: Unary Computing for GEMM Applications.
Di Wu, Jingjie Li, **Ruokai Yin**, Hsuan Hsiao, Younghyun Kim, Joshua San Miguel
IEEE Micro, 2021.

In-Stream Correlation-Based Division and Bit-Inserting Square Root in Stochastic Computing.
Di Wu, **Ruokai Yin**, Joshua San Miguel
IEEE Design & Test, 2021.

Talks

LoAS: Fully Temporal-Parallel Dataflow for Dual-Sparse Spiking Neural Networks
57th MICRO (Austin, USA), Nov 2024

MINT: Multiplier-less Integer Quantization for Energy Efficient Spiking Neural Networks
29th ASP-DAC (Incheon, South Korea), Jan 2024

SATA: Sparsity-Aware Training Accelerator for Spiking Neural Networks
Center for Brain-Inspired Computing (C-BRIC, SRC), Nov 2022

UnarySim and Characterizing Early Termination in Stochastic Computing
2020 UW Computer Architecture Industrial Affiliates (Madison, WI, USA), Sep 2020

Awards & Honors

Research

- **Best Paper Award Nomination, Asia-South Pacific Design Automation Conference (ASP-DAC)**, 2024
- **Spotlight Paper, NeurIPS Workshop on Learning from Time Series for Health**, 2022
- **IEEE Micro Top Pick, Computer Architecture,** 2020

Academic

- **John Bennett Fenn Fellowship Fund,** Fall 2021 – Spring 2022, Yale University
- **Dean's Honor List,** Fall 2018 – Spring 2021, University of Wisconsin - Madison

Teaching Experience

TA - EENG 439, Neural Networks & Learning Systems, Fall 2023
Instructor: Prof. Priya Panda

TA - EENG 348, Digital Systems, Spring 2023
Instructor: Prof. Rajit Manohar

Academic Activities

Reviewer

- IEEE Transactions on Neural Networks and Learning Systems
- 2024 IEEE International Symposium on Circuits and Systems
- IEEE Journal on Emerging and Selected Topics in Circuits and Systems
- IEEE Transactions on Very Large Scale Integration Systems
- IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems
- AI Communications