HANQING ZHU

Graduate Research Assistant ⋄ ECE Department ⋄ University of Texas at Austin
hqzhu@utexas.edu ⋄ (512)200-6791 ⋄ https://zhuhanqing.github.io/

RESEARCH INTERESTS

Efficient and Robust AI Computing Systems with Emerging Technology (photonics, post-CMOS electronics)

Hardware-efficient ML, ML for Emerging AI Systems & Electronic Design Automation

EDUCATION

The University of Texas at Austin (UT-Austin), TX, USA *Aug. 2020 – - Present*
Ph.D. student, Department of Electrical and Computer Engineering
Advisor: David Z. Pan
Co-advisor: Ray T. Chen
(GPA: 3.93/4.00)

Shanghai Jiao Tong University (SJTU), Shanghai, China *Sept. 2016 - Jun. 2020*
B.E., Dept. of Microelectronics Science and Engineering
(GPA: 3.81/4.00)
(Rank: 2nd/57)

HONORS AND AWARDS

1st Place at IEEE/ACM MLCAD FPGA Macro-Placement Contest	MLCAD	2023
MLSys Student Travel Award	MLSys	2023
Winner at Robert S. Hilbert Memorial Optical Design Competition	Synopsys	2022
DAC Young Fellow	DAC	2021
Shanghai Outstanding Graduate	Shanghai City	2020
Departmental Excellent Undergraduate Thesis	Shanghai Jiao Tong University	2020
Hongyi Scholarship	Shanghai Jiao Tong University	2019
Outstanding Undergraduate Scholarship	Shanghai Jiao Tong University	2019
Samsung Scholarship	Shanghai Jiao Tong University	2018
Zhiyuan College Honors Scholarship	Shanghai Jiao Tong University	2018
1st Prize, National Mathematical Contest in Modeling	Shanghai Division	2018
Academic Excellence Scholarship	Shanghai Jiao Tong University	2017-2019

PROFESSIONAL EXPERIENCE

Lightelligence Inc., MA, USA *May. 2023 – Sept. 2023*
Software Research Intern, Software Team

- Build the low-bit noise-aware training framework for their emerging photonic AI computing chips.

Google LLC., CA, USA *Jul. 2022 – Nov. 2022*
Student Researcher, Google Brain

- Chip Placement with Reinforcement Learning. Integrate DREAMPlace to the RL chip placer.

SELECTED RESEARCH PROJECTS

- **Emerging Hardware/Accelerators for Efficient AI Computing**
  - Contribute to library for photonic AI computing Torch-ONN.
  - First photonic Transformer accelerator. [C14]
  - Electronic-photonic NN accelerator. [C1, C5, C6, J1, J3, J4]
  - Photonic in-memory computing accelerator. [J2, C5]
- **HW-SW Co-design & Optimization for Efficient and Reliable Emerging Hardware**
  - Model-circuit co-optimization for efficiency enhancement. [C9, J3]
    - Hardware-architecture-search for mixed-activation system [C9]

  - Reliability-driven optimization
    - Aging-aware optimization for Photonic in-memory computing. [J2, C5]
    - Customized quantization and variation-aware training for robustness enhancement. [C1, C5, J2]
  - On-chip/on-device learning for self-learnable emerging AI hardware. [C4]
- **Efficient Machine Learning**
  - Equivalent and efficient Pre-LN Transformer architecture [C13]
  - Circuit/System-aware Quantization and Compression [C4, C5, C9, C14, J2].
  - Efficient on-chip/on-device sparse training. [C4]
  - Memory-efficient neural network Design with multi-level low-rank decomposition. [C3]

INVITED TALKS

- "Towards Reliable and Self-Learnable Photonic Neural Network from the Lens of Software-Hardware Co-design", Lightelligence, May 25, 2023

PROFESSIONAL SERVICE

Reviewer

- *Journal*: Nature Photonics (Second reviewer), IEEE Transactions on Neural Networks and Learning Systems (TNNLS), Photonic Network Communications
- *Conference and Workshop*: ICLR(2024), NeurIPS(2023), DAC(2023), ICCAD(2022), AICAS(2022-2023), AAAI workshop on DL-Hardware Co-Design for AI Acceleration (2023)

RELATED COURSES

- EE381V: Combinatorial Optimization — *Prof. Constantine Caramanis*
- EE382M: VLSI CAD and Optimization — *Prof. David Z. Pan*
- EE382N: Computer Architecture: Parallelism/Locality — *Prof. Mattan Erez*
- EE381V: Advanced Topics in Computer Vision — *Prof. Zhangyang (Atlas) Wang*
- EE381K: Convex Optimization — *Prof. Constantine Caramanis*
- EE382M: VLSI I — *Prof. David Z. Pan*
- EE382M: VLSI Physical Design Automation — *Prof. David Z. Pan*
- EE382V: Cross-layer Machine Learning Algorithm/Hardware Co-design — *Prof. Mattan Erez and Prof. Michael Orshansky*
- EE382V: SysML: Computer Systems and Machine Learning Interplay — *Prof. Neeraja J. Yadwadkar*

SKILLS

Programming Languages
Python (PyTorch), C++, CUDA, Verilog

EDA Tools
Cadence Virtuoso, Synopsys Design Compiler, Hspice, Xilinx Vivado Design Suite, Synopsys Optodesigner

PUBLICATIONS

Journal Papers

[J4] Chenghao Feng*, Jiaqi Gu*, **Hanqing Zhu**, Zhoufeng Ying, Zheng Zhao, David Z. Pan , and Ray T. Chen, "A compact butterfly-style silicon photonic-electronic neural chip for hardware-efficient deep learning," in *ACS Photonics*, 2022..

[J3] Jiaqi Gu, Chenghao Feng, **Hanqing Zhu**, Zheng Zhao, Zhoufeng Ying, Mingjie Liu, Ray T. Chen and David Z. Pan, "SqueezeLight: A Multi-Operand Ring-Based Optical Neural Network with Cross-Layer Scalability," in *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)*, Jul., 2022.

[J2] **Hanqing Zhu**, Jiaqi Gu, Chenghao Feng, Mingjie Liu, Zixuan Jiang, Ray T. Chen, and David Z. Pan, "ELight: Towards Efficient and Aging-Resilient Photonic In-Memory Neurocomputing," in *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)*, Jun., 2022.

[J1] Jiaqi Gu, Chenghao Feng, **Hanqing Zhu**, Ray T. Chen and David Z. Pan, "Light in AI: Toward Efficient Neurocomputing with Optical Neural Networks - A Tutorial," in *IEEE Transactions on Circuits and Systems–II: Express Briefs (TCAS-II)*, Apr., 2022.

Conference Papers

[C14] **Hanqing Zhu**, Jiaqi Gu, Hanrui Wang, Rongxin Tang, Zhekai Zhang, Chenghao Feng, Song Han, Ray T. Chen, David Z. Pan, " DOTA: A Dynamically-Operated Photonic Tensor Core for Energy-Efficient Transformer Accelerator," in *IEEE International Symposium on High Performance Computer Architecture(HPCA)*,Mar.2024. (Accepted)

[C13] Zixuan Jiang, Jiaqi Gu, **Hanqing Zhu**and David Z. Pan, "Pre-RMSNorm and Pre-CRMSNorm Transformers: Equivalent and Efficient Pre-LN Transformers," in *Conference on Neural Information Processing Systems (NeurIPS)*, Dec 10 - Dec 16, 2023. **(Spotlight)**. (Accepted)

[C12] **Hanqing Zhu**, Jiaqi Gu, Hanrui Wang, Rongxin Tang, Zhekai Zhang, Chenghao Feng, Song Han, Ray T. Chen, David Z. Pan, " DOTA: A Dynamically-Operated Photonic Tensor Core for Energy-Efficient Transformer Accelerator," in *Conference on Machine Learning and Systems (MLsys), Workshop on Systems for Next-Gen AI Paradigms (SNAP)*, Jun 4 - Jun 8, 2023.

[C11] Jiaqi Gu, Zhengqi Gao, Chenghao Feng, **Hanqing Zhu**, Ray Chen, Duane S Boning, and David Z. Pan, " NeurOLight: A Physics-Agnostic Neural Operator Enabling Parametric Photonic Device Simulation," in *Conference on Neural Information Processing Systems (NeurIPS)*, Nov 26 - Dec 4, 2022.

[C10] Harrison Jin, **Hanqing Zhu**, Keren Zhu, Thomas Leonard, Mahshid Alamdar, David Z. Pan, and Jean Anne C. Incorvia, "Design of Domain Wall-Magnetic Tunnel Junction Analog Content Addressable Memory using Current and Projected Prototype Data," in *Annual Conference on Magnetism and Magnetic Materials (MMM)*, Minneapolis, MN, October 31 - November 4, 2022.

[C9] **Hanqing Zhu**, Keren Zhu, Jiaqi Gu, Harrison Jin, Ray Chen, Jean Anne Incorvia and David Z. Pan, "Fuse and Mix: MACAM-Enabled Analog Activation for Energy-Efficient Neural Acceleration" in *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, Oct., 2022

[C8] Chenghao Feng, Jiaqi Gu, **Hanqing Zhu**, Zhoufeng Ying, Zheng Zhao, David Z. Pan, and Ray T. Chen, "Optoelectronically Interconnected Hardware-Efficient Deep Learning using Silicon Photonic Chips," in *Smart Photonic and Optoelectronic Integrated Circuits (SPIE)*, Mar., 2022

[C7] Chenghao Feng, Jiaqi Gu, **Hanqing Zhu**, David Z. Pan, and Ray T. Chen, "Design and Experimental Demonstration of A Hardware-Efficient Integrated Optical Neural Network," in *Smart Photonic and Optoelectronic Integrated Circuits (SPIE)*, Mar., 2022

[C6] Jiaqi Gu, **Hanqing Zhu**, Chenghao Feng, Zixuan Jiang, Mingjie Liu, Shuhan Zhang, Ray T. Chen, and David Z. Pan, "ADEPT: Automatic Differentiable DEsign of Photonic Tensor Cores," in *ACM/IEEE Design Automation Conference (DAC)*, Jul., 2022

[C5] **Hanqing Zhu**, Jiaqi Gu, Chenghao Feng, Mingjie Liu, Zixuan Jiang, Ray T. Chen, and David Z. Pan, "ELight: Enabling Efficient Photonic In-Memory Neurocomputing with Life Enhancement," in *IEEE/ACM Asia and South Pacific Design Automation Conference (ASP-DAC)*, Jan., 2022.

[C4] Jiaqi Gu, **Hanqing Zhu**, Chenghao Feng, Zixuan Jiang, Ray T. Chen, and David Z. Pan, "L2ight: Enabling On-Chip Learning for Optical Neural Networks via Efficient in-situ Subspace Optimization," in *Conference on Neural Information Processing Systems (NeurIPS)*, Dec., 2021.

[C3] Jiaqi Gu, **Hanqing Zhu**, Chenghao Feng, Mingjie Liu, Zixuan Jiang, Ray T. Chen, and David Z. Pan, "Towards Memory-Efficient Neural Networks via Multi-Level in situ Generation," in *International Conference on Computer Vision (ICCV)*, Oct., 2021.

[C2] Chenghao Feng, Jiaqi Gu, **Hanqing Zhu**, David Z. Pan, and Ray T. Chen, "Experimental Demonstration of a WDM-based Integrated Optical Decoder for Compact Optical Computing," in *Conference on Lasers and Electro-Optics*, May, 2021.

[C1] Jiaqi Gu, Zheng Zhao, Chenghao Feng, **Hanqing Zhu**, Ray T. Chen, and David Z. Pan, "ROQ: A Noise-Aware Quantization Scheme Towards Robust Optical Neural Networks with Low-bit Controls," in *IEEE Design, Automation & Test in Europe Conference & Exhibition (DATE)*, Mar., 2020.