ZIYI GUO

E-mail: ehsanguo@whu.edu.cn
Homepage: Elana-Ehsan.github.io
+86 18737902777 ⋄ Wuhan, China

EDUCATION

Wuhan University, China **Sep 2022 - Jun 2026 (expected)**
Hongyi Honor College
Major: Microelectronics Science & Engineering (EE) | GPA: 3.94/4.00
Relevant Coursework:

- *Electronics*: Analog Circuits (95), Solid State Physics (95), Semiconductor Physics (94)
- *Computer Science*: Data Structures (97), Operating Systems (95), AI Fundamentals (95), Computer Architecture
- *Mathematics*: Numerical Linear Algebra (100), Adv. Mathematics A (96)

Honors & Awards:

- **National Scholarship** (Top 0.2% nationwide, 2024)
- Innova International Exchange Scholarship, $10,000 award (2025)
- University Outstanding Student Scholarship & Merit Student (2023 & 2024)

University of California, Berkeley **Jan 2025 - May 2025**
Visiting student, Berkeley International Study Program
Relevant Coursework: Digital Design and IC Circuits (A)

PUBLICATIONS

1. X. Zheng †, **Z. Guo** †, et al., **A Unified Function Processor with Integer Arithmetic Based on Piecewise Chebyshev Polynomial Approximation**. *Submitted to ISCAS 2026*
   †These authors contributed equally to this work.

RESEARCH EXPERIENCE

Energy-efficient Acceleration System on Chip (SoC) for Generative AI **Jun 2025 - Present**
Duke University, Durham, USA
Module leader| Supervised by Prof. Yiran Chen and Dr. Changchun Zhou

- Designed a **Unified Function Processor (UFP)** that accelerates nonlinear functions (ReLU, GeLU, Softmax, etc.) using polynomial/LUT fitting, achieving execution speed comparable to linear operations while supporting parallel 32-element processing and multi-bitwidth computation. The design reduced rMAE by $> 75\%$ (sigmoid), area by 93.6%, and energy by $\sim 79\%$ compared to prior approaches.
- Built a **Sample Processing Module (SMP)** for up/downsampling with compact memory layout, kernel-size independent reuse, and minimal write operations, effectively reducing memory overhead by 80%.
- Implemented a **Data Similarity Module (DSM)** with pipelined cosine similarity and configurable precision to enhance data reuse and reduce computational cost.
- Led module-level design and Verilog implementation, coordinating integration across three IPs to ensure energy-efficient acceleration for generative AI workloads.
- Performed **C–Verilog co-simulation** via DPI-C, enabling system-level verification of SoC and validating end-to-end functionality with C-based reference models.
- Preparing for tape-out in **16 nm CMOS process**; RTL freeze completed, pre-layout verification in progress.

Lightweight Neural Inference and Hardware Acceleration for Embedded Systems **Sep 2025 - Present**
University of Electronic Science and Technology of China (UESTC)
Undergraduate Research Assistant| Supervised by Prof. Yiwen Wang

- Developed an end-to-end embedded AI inference framework integrating model optimization, software deployment, and hardware acceleration based on open-source RISC-V architecture.
- Trained a lightweight MobileNetV2-based hand-gesture recognition model (12.4 K parameters, 36.6 KB after quantization) achieving 98.8% accuracy across 29 gesture classes; successfully deployed on STM32 MCU via TensorFlow Lite Micro (TFLM).
- Implemented a weight-stationary systolic array and modular dataflow (Data Setup → MAC Array → Accumulator → Quantization → Write-back); validated RTL modules with Python co-simulation.
- Built and verified a RISC-V co-processor interface (NICE), extending the ISA via .insn macros and successfully testing CSR read/write and instruction handshake in SoC simulation.
- Analyzed end-to-end performance (1.7 s @ 480 MHz on STM32; 17 s @ 70 MHz on FPGA) and identified memory and bus bottlenecks for future SIMD-based optimization toward a low-power AI SoC prototype.

RISC-V 3-Stage Pipelined CPU with UART on FPGA **Mar 2025 - May 2025**
UC Berkeley, USA
Project leader| Instructed by Prof. John Wawrzynek

- Designed and implemented a three-stage pipelined RISC-V CPU with complete I/O peripheral circuits, including FIFO, UART and related modules in Verilog, targeting the Xilinx PYNQ-Z1 platform; achieved full RV32I instruction set support.
- Built all pipeline components from scratch including hazard detection unit, forwarding logic, and branch prediction logic, reducing CPI from 1.60 to 1.12 on a matrix multiplication benchmark.
- Created testbenches to validate CPU and I/O modules at both unit and system levels.
- Optimized critical path and control logic to increase maximum frequency from 50 MHz to 90 MHz (top 10% in class) and reduced LUT-based area from 200,000 to 140,000.

OTHER PROJECTS

Analog Front-End Circuit Design for Mixed-Signal IoT Chip **Mar 2025 - May 2025**
Image Recognition in Artificial Intelligence Fundamentals **Nov 2024 - Jan 2025**
Thermal Dissipation Mechanism of High-Temperature Hydrogels **Jun 2024 - Jan 2025**

SKILLS

Digital IC Design: Front-end SoC IP design for AI accelerators, including RTL coding, verification, and DPI-C co-simulation; familiar with AXI bus protocol.
Tools & EDA Software: Cadence, Vivado, ModelSim, VCS.
Programming Languages: Verilog & SystemVerilog, Python, C/C++, SQL, Matlab.
Languages: English: Fluent, TOEFL iBT 106(Speaking 25, Writing 27); Chinese(Mandarin): Native

EXTRA-CURRICULAR ACTIVITIES

Head of Wuhan University Student Art Troupe **Sep 2024 - Aug 2025**
Invited to attend the 2024 WA Chinese New Year Ball and sing for Australian Prime Minister Albanese
Collaborated with Mr. Dai Yuqiang, one of the "Three Tenors of China"
Member of University Volunteer Center **Sep 2022 - Sep 2023**
Member of University Chinese Kungfu Association **Sep 2022 - Sep 2023**

Hobbies: Classical Tenor Singing, Ski, Tennis