ZENG, ZHICHEN

• ✆: 607-327-2454 • ✉:zhichen.zengustc@gmail.com • Home: zhichenzzz.github.io

EDUCATION BACKGROUND

University of Science and Technology of China (USTC)

Bachelor's Degree of Science *09/2020 – 07/2024 (Expected)*

Yan Jici Talent Program in Physics, Major: Physical Electronics

GPA: **4.08/4.30**, Ranking: **1/31** (Major), **3/181** (School)

Core Courses: Computer Architecture (91), Principles of Microcomputer (91), Digital Logical Circuit (95), Electronic Circuits (95), Computational Physics A (95), Thermodynamics and Statistical Physics (100), Electromagnetism (100), Quantum Mechanics (95), Quantum Computing and Machine Learning (91), Computer Programming A (91), High-Level Digital Design Automation (audited, Cornell ECE 6775)

PUBLICATIONS

- H. Chen*, N. Zhang*, S. Xiang, **Zhichen Zeng**, M. Dai, Z. Zhang. *Allo: A Programming Model for Composable Accelerator Design.* Submmitted to **PLDI 2024, under review.**
- Q. Wu, Y. Gui, **Zhichen Zeng**, X. Wang, H. Liang, L. Tao, L. Zhao, Z. Zeng, X. Jin. *EN-Tensor: Advancing TensorCores Performance through Encoder-Based Methodology.* Submmitted to **DAC 2024, under review.**
- Z. Han*, M. Dai*, **Zhichen Zeng***, C. Ye, R. Dai, Z. Wang, X. Sun, Z. Zhang. *Highly Stable Photoelectric Detector Using Lead-Free Double Perovskite $Cs_2AgBiCl_6$ with Fast Response.* Submmitted to Journal of Materials Chemistry C, under review.

RESEARCH INTERESTS

- Computer architectures, Hardware accelerators, and Domain-specific compiler for hardware design (FPGA).

RESEARCH EXPERIENCES

Allo: A Programming Model for Composable Accelerator Design

Research Assistant, Supervisor: ***Prof. Zhiru Zhang****, Cornell University, NY, US* *07/2023 – 11/2023*

- Contributed to Allo, an MLIR-based programming infrastructure with decoupled hardware customizations targeting CPU, FPGA, and AI Engines.
- Developed new programming interface and frontend framework to support advanced Python-based features (supporting NumPy-like syntax, integrated linear algebra functions, and tensor slicing) which are based on MLIR Linalg dialect operations.
- Built a PyTorch model tracer frontend, supporting users porting PyTorch or Huggingface models directly to Allo, enabling an end-to-end compilation flow between high-level PyTorch models to hardware acclerators.
- Optimized KV cache technique on GPT2 module to reduce compilation time and realized end-to-end GPT2 computation on FPGA with **1.76X** faster and **5.6X** higher energy efficiency than the A100 GPU.
- Outperformed current state-of-the-art HLS tools and ADLs (like ScaleHLS, PyLog, Dahlia) on all test cases in the PolyBench.

EN-TensorCore: Advancing TensorCores Performance through Encoder-Based Methodology

Research Assistant, Supervisor: ***Prof. Xi Jin****, SoC Design Lab, USTC, Hefei, China* *03/2023– 07/2023*

- Proposed a novel architecture that extracts encoders from all the multipliers within Processing Element (PE) arrays to reduce redundancy of encoding in many existing micro-architectures.
- Contributed to develop a new radix-4 encoding method which substantially reduced encoded bit width from 3N/2 to just N+1 when computing N bits multiplication.
- Using number theory knowledge to validate the feasibility of the novel radix-4 encoding and expand to support both INT and FLOAT with N-bits matrix mulplication with higher performance.
- Reduced logic area and power consumption by **10%-20%** at computational scales of 256 GOPS, 1 TOPS, and 4 TOPS compared to Synopsys DesignWare standard multiplier IP core.

Useful Research Experiences *USTC, Hefei, China*

- **Highly Stable Photoelectric Detector Using Lead-Free Double Perovskite $Cs_2AgBiCl_6$ with Fast Response**
  Advisor: Prof. Zengming Zhang, Physics Experimental Center, USTC
  (1) Designed a highly stable photoelectric detector which has a fast response of light (~ms).
  (2) Explored the photoelectric properties of the detector which behaves excellent in high-temperature (473K) and vacuum ($8x10^{-5}$Pa) conditions.
- **Used Monte-Carlo and Metropolis Sampling Methods to Simulate Several Physical Problems**
  Advisor: Prof. Zejun Ding, Computational Physics, USTC
  (1) Simulated Brownian motion of particles in an electric field, Diffusion-limited aggregation (DLA), and Dielectric breakdown model (DBM) by using C++ and Monte-Carlo methods.
  (2) Percolation calculation, and 2D-Ising module by advance linear algebra methods, like Metropolis Sampling.
- **Designed Multi-Cycle Baseline CPU with five stages pipeline by Using Verilog HDL**

SCHOLARSHIPS

- USTC Fellowship **Top 1% for Research**, Chinese Academy of Sciences & USTC *08/2023*
- Outstanding Student Scholarship, **Gold (1%),** USTC *10/2022,2023*
- Scholarship for Talent Program in Basic Disciplines, **Class A**, USTC *09/2023, 2022, 2021*
- Yilin Chen Foundation Scholarship (2%) *09/2021*

REWARDS/COMPETITIONS

- USTC Physical College Physical Research-oriented, **grand prize (1%)** *04/2023*
- Atomic Physics Course Thesis Competition, **Second Prize**, USTC *01/2021*
- Mathematics Competition of Chinese College Students, **First Prize**, the Ministry of Education *10/2021*

TEACHING EXPERIENCES

- **Teaching Assistant** of Electrodynamics, by Professor Qing jia, *Spring 2023*

EXTRACURRICULAR ACTIVITY

- Member of the Association of Basketball of USTC *10/2020-present*
- Member of Volunteer Association of USTC *11/2021-present*

SKILLS/STANDARDIZED TESTS

- **Computer skills**: Python, C, C++, PyTorch, MLIR, Vitis HLS, Verilog HDL, Git, MATLAB, Mathematica, LATEX, Origin, Linux (Ubuntu)
- **TOEFL**: 99 (R:26, L:23, S:23, W:27)