Haichao Yang

Ph.D. Student · UC San Diego · Computer Engineering and VLSI

3869 Miramar St #1916, La Jolla, CA, 92092

+1 734-680-9178 | hay025@ucsd.edu | https://github.com/hcyang99

Education

University of California, San Diego *La Jolla, CA, U.S.*

Ph.D. Computer Science and Engineering, GPA: 3.96/4.00 *Sep. 2021 - Jun. 2026 (Expected)*

- Related Courses: Intro to Embedded Computing (CSE 237A), Parallel Computing *(CSE 260)*
  Parallel Computing in Bio-informatics *(ECE 284)*, Algorithms for VLSI CAD *(CSE 284, in progress)*
  Probabilistic Reasoning *(CSE 250A)*, Data Mining *(CSE 255)*, Recommender Systems *(CSE 258, in progress)*

University of Michigan, Ann Arbor *Ann Arbor, MI, U.S.*

B.S.E. Computer Engineering (Dual Degree, summa cum laude), GPA: 3.93/4.00 *Sep. 2019 - May 2021*

- Related Courses: Computer Architecture *(EECS 470, EECS 570, EECS 573)*, Compiler Construction *(EECS 483)*,
  VLSI Design *(EECS 312, EECS 427, EECS 627)*

Shanghai Jiao Tong University *Shanghai, China*

B.S.E. Electrical and Computer Engineering (Dual Degree), GPA: 3.71/4.00 *Sep. 2017 - Aug. 2021*

Professional Experience

Internship, GPU Architecture *San Diego, CA, U.S.*

Qualcomm Technologies, Inc. *Jun. 2023 - Sep. 2023*

- Developed a visualizer for ray tracing acceleration structures, showing the tree hierarchy and various heatmaps
- Tool deployed in GPU architecture team, has discovered a subtle issue increasing memory footprint, under fixing
- C++, wxWidgets, Vulkan

Selected Projects

ReRAM-based Hyper-dimensional Computing Few-shot Learning Accelerator *La Jolla, CA, U.S.*

UCSD (Advisor: Prof. Tajana Rosing) *Jan. 2023 - Current*

- Accelerator **in fabrication** using TSMC CLN40LP (40 nm)
- Responsible for RTL design and verification of the CNN part (added bf16 support and microcodes)
- Responsible for physical design of all digital modules

CNN and Hyper-dimensional Computing Accelerator *La Jolla, CA, U.S.*

UCSD (Advisor: Prof. Tajana Rosing) *Sep. 2021 - Current*

- Accelerator **fabricated** using TSMC CLN40LP (40 nm)
- Responsible for RTL design and verification of the CNN part
- Responsible for physical design and verification of the whole chip
- Responsible for post-silicon testing
- AMD Zynq FPGA based testbench, using Ethernet, AXI, and programmable logic to communicate between PC and chip

Performance Evaluation of CUDA K-mer Counting Algorithms *La Jolla, CA, U.S.*

UCSD (Advisor: Prof. Yatish Turakhia) *Feb. 2022 - Mar. 2022*

- https://github.com/hcyang99/cuda-kmer-counting
- K-mer counting using CUDA hashtables
- Re-implemented algorithms of Gerbil, SlabHash, and WarpCore for a fair comparison

Low-power Object Detection Accelerator *Ann Arbor, MI, U.S.*

U-M (Advisor: Prof. David Blauww) *Feb. 2021 - May 2021*

- VLSI II (EECS 627) Course Project
- Ultra-low-power attention unit activates classification circuit when necessary
- Responsible for RTL design and verification

RTL design of an Out-of-order RV32I Core with Configurable Superscalar Width *Ann Arbor, MI, U.S.*

U-M (Advisor: Dr. Mark Brehob) *Feb. 2020 - Apr. 2020*

- https://github.com/hcyang99/rv32-core
- Computer Architecture (EECS 470) Course Project, best-performing project of the semester
- Responsible for design and verification of RAT/RRAT, ROB, and data prefetcher
- Responsible for top-level verification

Publications

B. Khaleghi*, U. Mallappa*, D. Yaldiz, **H. Yang**, M. Shah, J. Kang, T. Rosing "PatterNet: Explore and Exploit Filter Patterns for Efficient Deep Neural Networks", DAC, 2022.

U. Mallappa, P. Gangwar, B. Khaleghi, **H. Yang**, T. Rosing "TermiNETor: Early Convolution Termination for Efficient Deep Neural Networks", ICCD, 2022.

Skills

Languages	**C/C++, SystemVerilog, CUDA, Golang, Python, Tcl, Shell Script**
EDA Tools	**VCS, DVE, Design Compiler, Formality, PrimeTime, Virtuoso, Innovus, Calibre, Vivado**
Software	**Git, Linux, Docker, CMake, vcpkg, Vitis, PyTorch, NGINX**