Yichang Hu

hyichang0510 | Yichang Hu | hyichang0510.github.io | huyich@cs.unc.edu

Education

University of North Carolina at Chapel Hill (UNC-CH) Chapel Hill, NC, US
Ph.D. Candidate in Computer Science *Aug. 2023 – Present*
- Advisor: Prof. Andrew Kwong

South China University of Technology (SCUT) Guangzhou, China
M.Eng. in Computer Science and Technology, GPA: 88.13 / 100 *Sept. 2020 – Jul. 2023*
- Advisor: Prof. Lu Lu

South China University of Technology (SCUT) Guangzhou, China
B.Eng. in Information Security, GPA: 3.71 / 4.0, Rank: Top 10% *Sept. 2016 – Jul. 2020*

Research Interest

My research focuses broadly on modern computer systems, including

· **Hardware Security:** Side-channel attacks, DRAM vulnerabilities, micro-architectural security
· **Parallel Computing:** High-performance computing (HPC), GPU computing
· **Heterogeneous Systems:** Hybrid architectures, concurrency, and parallelism

Research Experience

Side-Channel Attack against GPU (Under Embargo) *Sept. 2023 – Oct. 2025*

- Conducted an extensive investigation into DRAM behaviors on modern GPUs to understand side-channel leakage channels; specific methodologies and attack details remain under embargo.

Advanced RowPress Vulnerability in Modern CPU DRAM *May. 2024 - Present*

- Investigate advanced disturbance behaviors inspired by RowPress in commodity DDR memory.
- Early findings indicate amplified vulnerability under specific access patterns.

Highly Efficient Multi-GPU Framework for Distributed FFT *Jan. 2022 – Aug. 2022*

- Built a high-performance single-node FFT framework with optimized GPU kernels, exploiting memory hierarchy, batched execution, and computation–communication overlap to maximize throughput.
- Designed an efficient distributed FFT pipeline for clusters, incorporating topology-aware communication planning and data remapping strategies that significantly reduce inter-node transfer overhead.
- Achieved substantial performance gains demonstrated through extensive evaluations, showing improved scalability, lower communication cost, and higher utilization across heterogeneous systems.

Memory-Accelerated Multidimensional FFT on GPU *Mar. 2021 – Jul. 2021*

- Designed a GPU FFT framework with optimized non-strided memory access and shared-memory dataflow, reducing global memory transfer across multidimensional FFTs.
- Implemented mixed-radix and radix-16 incorporating kernels using Tensor Cores to combine multiple FFT stages, achieving significantly higher throughput for large and irregular-sized FFTs.

- Integrated four-step FFT reshuffling and in-place transposition to alleviate memory bottlenecks, boosting end-to-end performance and outperforming AMD FFT library by 25–250% on AMD GPUs.

High Performance LINPACK Simulation Model for HPC Systems *Jan. 2020 – Jun. 2020*

- Designed a single-node HPL simulation model that reproduces GPU behavior and CPU–GPU workload division, enabling fast estimation of floating-point performance.
- Developed a multi-node extension incorporating topology-aware communication modeling for Ethernet and InfiniBand, accurately capturing bandwidth, latency, and cross-node broadcast overhead.
- Validated the model against real heterogeneous clusters, achieving prediction error within 5%, demonstrating the framework's effectiveness in rapid performance estimation for large-scale HPC systems.

Publications

- **Under Embargo (in submission)**
  IEEE S&P
- **An Approach for Large-Scale Distributed FFT Framework on GPUs**
  Hu, Y., Zhou, B., Lu, L.
  SC (2022) paper link
- **Memory-Accelerated Parallel Method for Multidimensional FFT on GPU**
  Hu, Y., Lu, L., Li, C.
  Journal of Supercomputing (2022) paper link
- **Simulation Model for High Performance LINPACK in Hybrid CPU-GPU Systems**
  Hu, Y., Lu, L.
  Journal of Supercomputing (2021) paper link

Industry Experience

Advanced Micro Devices, Inc. (AMD) *Mar. 2022 - Aug. 2022*

Collaborator

- Migrated the LINPACK benchmark to AMD GPUs and optimized it for GPU-based HPC systems.
- Reached 2.5×10^4 TFLOPS across 256 computation nodes on LUMI supercomputer.

Tencent, WeChat Business Group *Jun. 2019 - Aug. 2019*

Engineer Intern

- Improved WeChat's content recommendation algorithms to enhance feed relevance and user engagement.

Technical Skills

Language C++, Rust, Java, Go
Programming CUDA, ROCm, OpenCL, MPI