Kuai Yu

Huazhong University of Science and Technology, Wuhan, China, 430074

Email: kuaiyubilly@gmail.com | Tel: +86-18358046758

Education

09/2021 – Present Huazhong University of Science and Technology (HUST), Wuhan, China

- Undergraduate of Computer Science, Dept. of Computer Science and Technology
- GPA: 4.52/5.00 (90.3/100) Integrated Rank: 45/340(top 15%)

Awards: Merit Student Scholarship (Top 10% of Dept)

Enterprise Scholarship (certificate by HUAWEI)

2022 Chinese Mathematics Competitions – First Prize of Hubei Competition Area

2022 Massive Storage Code Challenge - Winning Prize in Final

Research Experiences

01/2023 – present Prof. Long Zheng's group, HUST, China

Project: **Heterogeneous Graph Processing System with GPUs/FPGAs**

- Read related papaers in graph processing system.
- Learn CUDA programming and Verilog HDL.

Project: **Vision Transformer Acceleration under graph processing perspective**

- Read papers in LLM system optimization and acceleration, especially the acceleration techniques in attention mechanism;
- Discuss with Ph.D graduate Yu Huang. Search and read papers in transformer optimization with Graph methods. In order to solve the memory bound problem, we propose "Img2Graph" and "Reorder" operations, which make it easy to perform differentiated sparsity/quantization in matrix multiplication in the following progress.
- Do experiments to verify ideas

10/2022 – 11/2022 Under Prof. You Zhou's assistance, HUST, China

Project: **Scheduling of Data Retrival Tasks in Tape Storage System(HUAWEI codiing challenge)**

- A short time research experience.
- Finlly we got a winning prize in the final competition(rank 10/74 in teams)

Language Skills

04/2023 IELTS - R: 7.5 L: 8.0 S: 6.0 W: 6.0 Total: 7.0

Research Skills

- Professional Knowledge:
  - Computer Architecture, OS and Compilers
  - Deep Learning Models, including Transformer/LLM/GNNs
- Engineering Skills:
  - **Coding languages**: C++/Python/Java, familiar with modern C++ and system design patterns
  - **Deep learning framework**: Pytorch
  - **Heterogeneous computing** tools: CUDA programming, Verilog HDL and knowledge in Xilinx HLS framework
  - git/make/cmake/linux basic usage.

Importantly, my ability to learn swiftly enables me to efficiently acquire and apply knowledge or a new tool in a short period .

Interests

I mainly focus on Computer Architecture, including Heterogeneous Computing and next generation architecture(CPUs, Compilers, In-Memory Computing and so on). Recently, I am also interested in Deep Learning System, like LLM acceleration techniques and AI Accelerators.