

Junlin Chen

(+86)182-0080-2136 \diamond junlin.chen@buaa.edu.cn \diamond hiGiraffe.github.io

EDUCATION

Beihang University

Beijing, China

Bachelor of Computer Science

Sept. 2021 - Jun. 2025(Expected)

- \diamond GPA: 3.70/4
- \diamond Second Class Innovation and Entrepreneurship Scholarship, 2022-2023
- \diamond Second Class Outstanding Social Work Scholarship, 2022-2023
- \diamond Outstanding Student Cadre, 2022-2023

University of Macau

Macau, China

Exchange Program

Aug. 2024 - Dec. 2024(Expected)

RESEARCH INTERESTS

Performance Optimization, Machine Learning System and Distributed Heterogeneous System.

RESEARCH PUBLICATIONS

Junlin Chen*, Chaojing Liu*, Zhongzhi Luan, Ming Gong, Qingfeng Li, Depei Qian, "Large-Scale Parallelization and Optimization of Lattice QCD on Tianhe New Generation Supercomputer", The 25th IEEE High Performance Computing and Communications (HPCC 2023), Dec. 13-15, 2023, Melbourne, Australia.

RESEARCH EXPERIENCE

Research Intern

Mar. 2024 – Present

Cloud Computing and Distributed Systems Laboratory, University of Macau

Macau, China

- \diamond Conducted research on LLM inference system optimization and efficient scheduling under the guidance of Prof. Huanle Xu.
- \diamond Supplemented relevant cutting-edge knowledge, including existing LLM parallel inference methods, LLM serving system and efficient scheduling for mlsys.
- \diamond Engaged in a research project of LLM serving system(in progress).
- \diamond Familiar with the vLLM related code and made further modifications to it.

Research Intern

Sept. 2022 – Feb. 2024

Sino-German Joint Software Institute, Beihang University

Beijing, China

- \diamond Conducted research on performance optimization under the guidance of Prof. Zhongzhi Luan.
- \diamond Collaboratively completed a research project utilizing Tianhe new generation supercomputer to accelerate scientific computations, including parallel mode design and hardware optimization.
- \diamond Participated in the design of parallel computing patterns, experimentation, paper writing, scientific visualization and oral presentation.

PROJECTS

- A LLM serving system(In Progress)** | *Pytorch* Jun. 2024 – Present
- ◇ Modified the scheduler logic of vLLM to support more task features.
 - ◇ Modified the inference execution framework of vLLM to better utilize system resources.
- BattleByte: Online Programming Battle Platform** | *Spring Boot* Feb. 2024 – Jun. 2024
- ◇ Analyzed product requirements, designed in-game mechanics, authored product documentation and coordinated team efforts as a Product Manager.
 - ◇ Designed and developed the backend WebSocket real-time communication component.
- CME 213 Module(Finished 1,2,3)** | *OpenMP, MPI, CUDA* Oct. 2023 – Feb. 2024
- ◇ Studied the basic utilization of OpenMP, MPI, and CUDA.
- Online Flea Market Platform** | *Python, Flask* Sept. 2023 – Dec. 2023
- ◇ Utilized the Flask framework to complete the backend code for user center and flea market functionalities.
 - ◇ Integrated the backend with databases using GaussDB for MYSQL and MYSQL.
- SysY-to-LLVM Compiler Project** | *C++* Sept. 2023 – Dec. 2023
- ◇ Developed a compiler that translates SysY language into LLVM language, encompassing lexical analysis, syntax analysis, semantic analysis, LLVM intermediate code generation, and error handling.
- Accelerating Lattice QCD on Supercomputer** | *C, OpenMP, MPI* Dec. 2022 – Dec. 2023
- ◇ Accelerated communication between two computational processes through Global Shared Memory and Array Memory.
 - ◇ Accelerated vectorized calculations through the MT-3000 processor's Acceleration Array.
 - ◇ Conducted performance analysis on global reductions, identifying bottlenecks and proposing adaptive strategies to optimize reduction frequency.
- Multi-threaded Elevator Scheduling System** | *Java* Feb. 2023 – Jun. 2023
- ◇ Developed a multi-threaded elevator scheduling system supporting elevator maintenance and elevator accessibility.
 - ◇ Developed a local greedy approach to handle the addition of elevators and maintenance requests.
 - ◇ Completed the development using the principles of object-oriented programming.
- MIPS Pipeline Processor with Exception Handling Support** | *Verilog* Sept. 2022 – Dec. 2022
- ◇ Implemented a MIPS five-stage pipeline CPU that supports branch prediction and hazard handling.
 - ◇ Implemented external instruction memory and data memory.
 - ◇ Introduced CP0, Bridge, and Timer to support interrupt and exception handling.

OTHER INFORMATION

Language Proficiency: English, Mandarin, Cantonese

Programming Languages: C++, Python, Java, Spring Boot, Verilog, LLVM IR, MIPS Assembly Language, LaTeX

Frameworks and Tools: OpenMP, MPI, CUDA, PyTorch, Ray

Familiar Project Code: vLLM

Further personal skills are showcased on [my personal notes website](#).