

Ximing Huang, Zihan Huo, Zihao Zhan, Shengjie Zhou, Runzhong Huang, Yuran Wang, Zhenxin Fu (advisor)

huo, hzlgpc, zhanshaia, dbblack, runzhonghuang, scxmqgl@pku.edu.cn, fushenxin@pku.edu.cn

About Peking University



Peking University (PKU) is a leading research university in China. Founded in 1898, it has been at the forefront of innovation and improvement and has played a significant role in shaping intellectual movements in the country. PKU is widely recognized as one of the top academic institutions in China, Asia, and worldwide, attracting top Chinese students who are dedicated to academic excellence and making a positive impact on society.

HPC at Peking University

- Peking University's high performance computing university-level public platform was officially launched in 2018.
- The total number of computing cores on the public platform reached **40500**, the peak computing capacity is **6638.3 TFLOPS**, and the storage capacity is **14783 TB**.
- Peking University develops SCOW, an open-source supercomputer management, portal and CraneSched, a supercomputer scheduler.
- The Peking University Supercomputing Team, established in 2015, achieved the **first place** in ASC 2023 and ASC 2024, and second place in SCC 2023(Highest Linpack) and SCC 2024.



Team Members

We have juniors and sophomores in our team, all obsessed with supercomputing.



Ximing Huang Team leader
Sophomore, College of Engineering
Major: Mechanics
Interest: Computer System, ML/Sys
Focus: HPL, SST, Cluster Management



Runzhong Huang
Sophomore, School of EECS
Major: Computer Science
Interest: Mathematics, ML
Focus: ECE



Zihan Huo
Junior, School of Foreign Languages
Major: Hindi
Interest: Computer System, Simulation
Focus: HPL, ECE, Cluster Management



Shengjie Zhou
Sophomore, School of EECS
Major: Computer Science
Interest: Parallel Programming, ML
Focus: ECE, Mystery Application



Zihao Zhan
Sophomore, School of EECS
Major: Electronic Engineering
Interest: Simulation, Architecture
Focus: SST, Poster



Yuran Wang
Sophomore, School of EECS
Major: Computer Science
Interest: Networking, Parallelization
Focus: SST, Mystery Application



FAN Chun Advisor
Chief Engineer, High Performance Computing Platform of Peking University
Research Interest: High Performance Computing, CFD.



FU Zhenxin Advisor
Engineer, High Performance Computing Platform of Peking University
Research Interest: High Performance Computing, Multimedia



Optimizations and Strategies

General Strategy

- Use profiler to identify and optimize hot routines, porting to GPU if necessary
- Compile with different compilers, compiler options and libraries
- Optimize IO and MPI bottlenecks by analyzing results from benchmarks
- Prioritize tasks by impact and dynamically allocate time to maximize score

HPL

- Use EPYC-optimized AOCC and BLIS library to get best performance
- Run HPL under different parameters thoroughly to gain experience about parameter tuning
- Conduct benchmarks on Jetstream 2 to find the best parameters

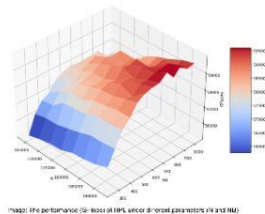
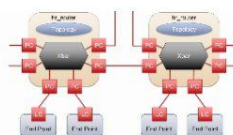


Figure 1: HPL performance (Gflop/s) as a function of parameters N and NB

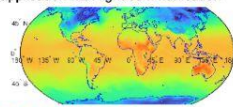
SST

- Profile the application to explore the impact of **synchronization** on performance
- Compare the scalability of different **parallel strategies** such as multiprocessing or multithreading
- Analyze the input graph to implement better **partitioner**



ECE

- Profile to figure out hotspots and memory behaviors and offload hotspots onto GPU
- Deduce that this is a memory-bounded application with light communication



Mystery Application

- Reserve resources for the mystery application according to the time budget
- Analyze the scalability and computation/communication-intensity

Hardware and Software Configuration

Cluster Architecture

- One login node with all storage quota attached.
- 3 CPU nodes(m3.2x) and 1 GPU node(g3.x) for compute, managed by Slurm.
- Operating System & Images
- Customized Rocky Linux 9 VM image includes all toolchains and optimizations.
- Cloud-init used for fast VM deployment.

Compilers & Toolchains

- GCC 15, Intel oneAPI, and AOCC compilers available in the image.
- Toolchain choice optimized by benchmarking each application.

Monitoring & Maintenance

- Resource usage logs maintained and visualized by Grafana for auditing and optimization.
- Manually monitor Jetstream 2 SUs consumption in the preparation phase.

Storage & File System

- NFS with optimized buffer and I/O settings used for shared storage.
- All storage accessed via login node.

login (m3.2x)	cpu-1 (m3.2x)	cpu-2 (m3.2x)	cpu-3 (m3.2x)	gpu (g3.x)
OS: Rocky Linux 9	OS: Rocky Linux 9	OS: Rocky Linux 9	OS: Rocky Linux 9	OS: Rocky Linux 9
16GB	16GB	16GB	16GB	16GB
100GB	100GB	100GB	100GB	100GB

Strength and Activities

Great team members

- In the process of organizing **HPCGame**, we are searching for and cultivating excellent team members.
- We carry out the **HPC from Scratch** project and periodically arrange lectures to rapidly develop new members for our team while promoting HPC
- Our team's spirit of collaboration extends beyond graduation, with alumni regularly returning to help train and mentor the next generation of members.

Team Diversity

- Diversity of genders.** We recruited members of different genders and actively encourage both females and males to develop interest in computer
- Diversity of majors and skills.** Our members come from various majors, including not only Computer Science students but also students from other majors that heavily utilize HPC
- Diversity of background.** Our team is composed of students from diverse ethnic backgrounds, with birthplaces covering almost three-fifths of China.

What Makes Our Team

- Open discussion.** The team meets every Wednesday evening face to face to discuss insights and issues.
- Clear responsibility.** Two to three members are in charge of one challenge together with efficient collaboration.
- Great team tools.** We use modern team management tools like Outline and Teams to boost performance.

HPCGame

Founded in 2023, PKU HPCGame is China's pioneering high-performance computing competition designed to welcome participants of all skill levels, from beginners to experts. Successfully held 3 times, the competition has attracted **1300+ students from 100+ universities** across the nation. PKU HPCGame serves as a comprehensive training ground where students from diverse academic majors gain hands-on experience with HPC systems, applications, and optimization techniques.



SCOW & CraneSched

- SCOW:** a supercomputer portal and management system. Users can efficiently use computing resources simply through a modern web browser conveniently.
- CraneSched:** a high performance HPC and Cloud computing fused Job Scheduling System



Sponsor



- Technology-driven investment firm, focused on artificial intelligence and machine learning in finance
- Core team from Stanford, Carnegie Mellon, Facebook, Google; 35% PhDs; medalists from NOI, CMO, CPhO; 25% of research team; over 10 ICPC World Finals contestants in team
- Providing mentorship and sharing expertise in performance optimization
- Registered as a private equity fund manager under the supervision of AMAC in Oct 2020
- First private investment fund issued Nov 2020
- Total AUM reached **CNY 10 billion** in April 2023
- Aiming to become a first-class hedge fund manager with international recognition