

Shiyang Li

☎ (+86) 183-3567-0051 ✉ nkulsy0815@163.com 📍 NO.38, Tongyan Road, Tianjin, China 300350 🌐 github.com/NKU-Yang

Interests: Computer Architecture and System, GPU, MLSys

Develop Skill: C/C++, CUDA, GitHub, Linux TOEFL Score :102

🎓 Education

- > Nankai University 2022 – now
Master
Major in Computer Science & Technology Embedded Systems and Information Security Lab **Ranking 4/17**
Advisor: Prof. Gong Xiaoli
🔗 scholar.google.com/citations?hl=en&user=KZDQTBQAAAAJ&view_op=list_works&sortby=pubdate
- > Nankai University 2018 – 2022
Bachelor of Engineering
Major in Internet of Things Engineering Colledge of Cyber Security **Ranking 7/49**

📖 Publication

- > **Li Shiyang**, Tang Ruiqi, Zhu Jingyu, Zhao Ziyi, Gong Xiaoli, Wang Wenwen, Zhang Jin, and Yew Pen-Chung(Fellow, IEEE),
“**Liberator : A Data Reuse Framework for Out-of-Memory Graph Computing on GPUs**”
IEEE Transactions on Parallel and Distributed Systems, vol.34, no.6, pp.1954-1967, June 2023, doi :10.1109/TPDS.2023.3268662.
Graph Computing NVIDIA GPU Parallel processing Data reuse Zero-copy
Optimized large-scale graph computing on GPUs by data reuse and concurrent processing of CPU and GPU, achieved an average acceleration of 2.7x across various graph algorithms, and reduced data transfer by 22% compared to the state-of-the-art (SOTA).
- > **Li Shiyang**, Zhu Jingyu, Han Jiaxun, Peng Yuting, Wang Zhuoran, Gong Xiaoli, Wang Gang, Zhang Jin, Wang Xuqiang,
“**OneGraph : a cross-architecture framework for large-scale graph computing on GPUs based on oneAPI**”
CCF Transactions on High Performance Computing, Nov. 2023, doi : 10.1007/s42514-023-00172-w.
CPU-GPU Heterogeneous Computing Intel oneAPI Cross-Architecture Migration
Developed an efficient framework with Intel OneAPI to accelerate graph computing with heterogeneous devices. It can be seamlessly ported to NVIDIA, Intel, and AMD GPU platforms without code modifications. The performance loss compared to a dedicated programming model, such as CUDA, is less than 1%.

🏢 Research Experience

- > **Large-scale Dynamic Temporal Graph Computing on Distributed Muti-GPUs**(Current Project) May.2024 – now
Funding by the Ant Group, Alibaba
GPU Temporal Graph Computing Distributed and parallel computing Memory management
 - High-performance graph computing framework based on CPU-GPU systems.
 - Design of efficient incremental computing method and storage format for the dynamic temporal graph.
 - Design of temporal graph partitioning method for Distributed muti-GPUs system.
- > **Design of Serverless Framework for Heterogeneous Embedded Cluster** (Current Project) DEC.2023–now
Heterogeneous Computing Embedded cluster Mobile cloud computing Serverless
 - Integrated multiple heterogeneous SoCs onto a single PCB to form a portable, mobile embedded cluster.
 - Develop a serverless framework for the heterogeneous embedded cluster.
 - Dynamic scheduling and auto-scaling of heterogeneous computing resources, including NPUs, GPUs, and CPUs.
 - Collaborative large model inference on multiple SoCs.

🔧 Work & Programm Experience

- Zhiyuan Research Institute, Hangzhou, China** System Engineer SEPT.2023 – DEC.2023
Deployment and optimization of object detecting algorithm on embedded platform
 - > Based on HUAWEI Atlas 200I A2 platform, developed inference application of yolov5 with NPU.
 - > Analysed the program performance bottleneck with the profiling tools.
 - > Improved end-to-end performance from 20FPS to 30FPS on the embedded platform with multi-threaded pipeline.
- Alibaba Cloud, Shanghai, China** System Engineer JUL.2021 – SEPT.2021
Security container and OS kernel development with Rust
 - > Developed Linux kernel with Rust. Rewrote the PCIe device registry process in Hypervisor.
 - > Realized the passthrough of the PCIe devices and hotplug of NVIDIA GPUs in the security container.

🏆 Awards

- China National Scholarship (only one in college per year) DEC.2023
- Outstanding student leader of Nankai University SEPT.2023
- First-class Scholarship for graduate students of Nankai University (top 10% of college) OCT.2023
- Nankai University Special Scholarship (3 times) 2022-2024