

Chendi Li

lichendi.cs@gmail.com

PhD student | AI+System

Personal website:
<https://www.lichendi.top>

I am currently a PhD student at the University of Utah, and supervised by Prof. **P. (Saday) Sadayappan**. I received my master's degree from the State Key Laboratory of Computer Architecture, Institute of Computing Technology, Chinese Academy of Sciences, and supervised by Prof. **Yunquan Zhang**. My research primarily focuses on software system design for accelerators.

RESEARCH INTERESTS

- High-Performance Computing
- High-Performance Machine learning
- Optimizing Tensor Computations

EDUCATION

PhD student in Computer Science, University of Utah	Sep 2022 — expected 2027
Master in Computer Science, University of Chinese Academy of Sciences	Sep 2019 — July 2022
Bachelor in Computer Science, Hunan Agricultural University	Sep 2014 — June 2018

RESEARCH EXPERIENCES

Graduate Student Research Assistant University of Utah	Sep 2022 — Present
Graduate Student Research Assistant State Key Laboratory of Computer Architecture, Institute of Computing Technology, Chinese Academy of Sciences	Sep 2019 — July 2022
Undergraduate Research Assistant State Key Laboratory of Computer Architecture, Institute of Computing Technology, Chinese Academy of Sciences	Jan 2018 — June 2019

PUBLICATIONS

[ICS 2024] **Chendi Li***, Yufan Xu*, Sina Mahdipour Saravani, Saday Sadayappan, Accelerated Auto-Tuning of GPU Kernels for Tensor Computations.

[TPDS 2024] Cunyang Wei, Haipeng Jia, Yunquan Zhang, Jianyu Yao, **Chendi Li**, Wenxuan Cao. IrGEMM: An Input-Aware Tuning Framework for Irregular GEMM on ARM and X86 CPUs.

[ICS 2023] T Chen, H Jia, Y Zhang, K Li, Z Li, X Zhao, J Yao, **Chendi Li**, et al. OpenFFT: An Adaptive Tuning Framework for 3D FFT on ARM Multicore CPUs.

[ISPA 2021] **Chendi Li**, Haipeng Jia, Hang Cao, et al. AutoTSMM: An Auto-tuning Framework for Building High-Performance Tall-and-Skinny Matrix-Matrix Multiplication on CPUs.

[ICPADS 2021] Jianyu Yao, Boqian Shi, Chunyang Xiang, Haipeng Jia, **Chendi Li**, et al. IAAT: An Input-Aware Adaptive Tuning framework for Small GEMM.

[HPCC 2021] Tun Chen, Haipeng Jia, Zhihao Li, **Chendi Li**, et al. A Transpose-free Three-dimensional FFT Algorithm on ARM CPUs.

[HPC China 2020] **Chendi Li**, Guangting Zhang, Haipeng Jia. Fast Computation of Elementary Functions on ARM Platforms

RESEARCH PROJECTS

Apache TVM, Contributor Feb 2023 — Feb 2024

- Designed and developed the dynamic gradient algorithm in TVM. Experimental evaluation on a number of matrix-matrix multiplication and 2D convolution kernels demonstrates an order-of-magnitude improvement in auto-tuning time to achieve the same level of code performance. The paper was published in ICS 2024.

OpenBLAS, Contributor**Nov 2020 — June 2022**

- Developed and optimized pre-packed matrix-matrix multiplication and triangular solve with multiple right-hand sides (TRSM) for ARMv8 and x86 platforms. OpenBLAS is one of the most widely used open-source BLAS libraries.

AutoTSMM, Author**Nov 2020 — June 2022**

- Designed and developed AutoTSMM, which is used to build high-performance tall-and-skinny matrix multiplication on mainstream CPUs. AutoTSMM is competitive with Intel OneMKL and outperforms all conventional GEMM implementations on ARMv8 platform. This work was published in IEEE ISPA 2021 and it was published in TPDS'24 as a part of IrGEMM framework.

IAAT, Contributor**Nov 2020 — Feb 2022**

- Launched the project and investigated JIT tools for small GEMM. IAAT is a template-driven, just-in-time small GEMM framework targeting CPUs. This work was accepted by IEEE ICPADS 2021.

AutoFFT and OpenFFT, Contributor**Jan 2018 — Feb 2021**

- Optimized small-scale FFT, and contributed to multi-threading and 2D-FFT. AutoFFT and OpenFFT is a template-based FFT codes auto-generation framework that contributes to many Chinese vendors' libraries. This work was published in SC'19, TPDS'20, HPC'21, and ICS'23.

OpenVML, Contributor**Jan 2020 — Oct 2020**

- Enhancing the mathematical functions by manipulating IEEE 754 floating-point representations, OpenVML, a high-performance vector math library, demonstrated significant performance improvements over both the C standard library and ARMPL. This work was accepted and presented at HPC China 2020.

AWARDS & HONORS

- 2022** Ph.D. Student Fellowship
- 2021** First-class scholarships
- 2020** Second-class scholarship
- 2019** Third-class scholarship, Outstanding intern in PerfXLab
- 2015** Collegiate programming contest first prize; Outstanding volunteer

TECHNICAL SKILLS

- Tools** Linux, Git, NVIDIA Nsight, Vim, GDB, CMake, OpenMP, Pthreads
- Programming/Scripting** Python, C, C++, X86 and ARM Assembly, X86 and NEON intrinsic, CUDA, Latex