RuQing G. Xu

https://github.com/xrq-phys/
https://qsl.r-xu.dns-cloud.net/

+81 070 3760 4882
rxu.sol@gmail.com

Education

The University of Tokyo Oct. 2021 – Sept. 2024
Ph.D. in Physics expected in 2024 *Tokyo, JP*

The University of Tokyo Sept. 2019 – Sept. 2021
M.S. in Physics *Tokyo, JP*

University of Science and Technology of China Sept. 2015 – Jul. 2019
B.S. in Physics *Hefei, CN*

Industrial Experience

Deep Learning Algorithm Engineering Intern (Team undisclosed) Jan. 2023 – Present
NVIDIA *Tokyo, JP*

- Performance optimizations for various deep learning tasks.

Deep Learning Algorithm Engineering Intern (cuTENSOR) Sep. 2021 – Jan. 2022
NVIDIA *Tokyo, JP*

- Special techniques for medium-size performance improvements.
- Tackled multiple (unusual) underperforming cases.
- In-depth L2 bandwidth analysis & optimizations.

Selected Research Projects

Speeding up Markov Chain Samplings on Fermionic Pair Products Jul. 2020 - Present
Optimizations for the many-variable Variational Monte Carlo *Univ. Tokyo & Waseda Univ.*

- $6\times$ speed up for sampling phases via blocked updating techniques.
- $3\times$ speed up for estimation phase via query batching.

GEMMFIP Nov. 2022 - Present
Unifying GEMM implementations in BLIS *Collab. w/ the University of Texas at Austin*

- Fuse packing with the first computational pass over each input tile.
- A unified technique for implementing matrix operations that can achieve high performance across the problem size spectrum.
- Prototype implementation (available to all on GitHub) outperforms OpenBLAS, AOCL, and ArmPL.
- *WIP: Automatic kernel generation via ExoLang.*

Arm SVE Kernels for BLIS Jul. 2020 - Sept. 2021
Tuned for Fujitsu A64FX on Supercomputer Fugaku *Collab. w/ Forschungszentrum Jülich & The University of Texas at Austin*

- BLIS outperforms vendor libraries on Supercomputer Fugaku (as of Jan. 2023).
- Talks presented in the Arm HPC User Group Workshop.
- Same set of kernels available also in TBLIS, provisional BLIS extension for direct tensor contraction.

Customizing BLAS on Apple's Matrix Coprocessor Jun. 2021 – Aug. 2022
Personal work

- Crafted AMX2 machine code for BLIS & TBLIS.
- BLIS' peak performance is above Accelerate framework.

Publications

*ArXiv:**2302**.08417* (ISC2023 submitted), **RuQing G. Xu**, Field G. Van Zee, Robert A. van de Geijn, *GEMMFIP: Unifying GEMM in BLIS*

*Comput. Phys. Commun. **277**, 108375*, **RuQing G. Xu**, Tsuyoshi Okubo, Synge Todo, Masatoshi Imada, *Optimized Implementation for Calculation and Fast-Update of Pfaffians Installed to the Open-Source Fermionic Variational Solver mVMC*

*Phys. Rev. Research **3**, 023048*, Xinliang Lyu, **RuQing G. Xu**, Naoki Kawashima, *Scaling dimensions from linearized tensor renormalization group transformations*

*J. Chem. Theory Comput. 2019, **15**, 3, 1728-1742*, James S. Spencer, Nick S. Blunt, ..., William A. Vigor, **RuQing Xu**, Alex J. W. Thom, *The HANDE-QMC project: open-source stochastic quantum chemistry from the ground state up*

Skills & Hobbies

Programming Languages: C, C++, CUDA, x86 Assembly, Arm64 Assembly, NVPTX, Julia, Python
Natural Languages: Chinese (native), English (GRE: 330; TOEFL: 108), Japanese (JLPT N1)