

Ritvik Sharma

+1-6509247092 | rsharma3@stanford.edu | [LinkedIn](#) | [Google Scholar](#)

Research Interests

Compiler tools for Optimization and Simulation of Quantum Computing and Sparse Tensor Algebra

Education

Stanford University PhD in Electrical Engineering Advisors: Prof. Sara Achour and Prof. Mark Horowitz	2021- 4.08/4.0
Indian Institute of Technology (IIT) Delhi B.Tech. in Electrical Engineering	2017-2021 9.704/10.000

Experience

Software Engineering Intern at Google • Developed KV cache pruning algorithms and system-level optimizations to enable infinite-context Large Language Models (LLMs) for on-device inference in the Litert-LM framework with the ODML Core Team at Google	June'25-Present
Research Intern at IonQ • Designed novel low-cost, non-fault-tolerant but error resilient gate implementations for gates in Bacon-Shor • The low-cost implementation retain some error-correction guarantees for near-term quantum computers	May'22-Aug.'22

Publications

Micro 2025: <i>A Probabilistic Perspective on Tiling Sparse Tensor Algebra.</i> Ritvik Sharma, Z. Y. Xue, N. Zhang, R. Lacouture, F. Kjolstad, S. Achour, and M. Horowitz	Paper , Code
PLDI 2025: <i>Optimizing Ancilla-Based Quantum Circuits with SPARE.</i> Ritvik Sharma, S. Achour	Paper , Code
PLDI 2024: <i>Compilation of Qubit Circuits to Optimized Qutrit Circuits</i> Ritvik Sharma, S. Achour	Paper , Code
In Review: <i>Fuseflow: A fusion-centric compilation framework for sparse ML on streaming dataflow.</i> R. Lacouture, O. Hsu, N. Zhang, Ritvik Sharma , M. Siracusa, F. Kjolstad, and K. Olukotun	
ASPLOS 2023: <i>The Sparse Abstract Machine</i> O. Hsu, M. Strange, Ritvik Sharma , J. Won, K. Olukotun, J. Emer, M. Horowitz, F. Kjolstad	Paper
TECS 2022: <i>CODEBench: Neural Architecture and Hardware Accelerator Co-Design Framework</i> S. Tuli, C.-H. Li, Ritvik Sharma , N. Jha	Paper
QIP 2022: <i>An Algorithm for Fast Supervised Learning in Variational Circuits through Simultaneous Processing of Multiple Samples</i> S. Dangwal, Ritvik Sharma , D. Bhowmik	Paper
ISCAS 2021: <i>A Crossbar Array of Analog-Digital-Hybrid Volatile Memory Synapse Cells for Energy-Efficient On-Chip Learning</i> J. Sharda, Ritvik Sharma , D. Bhowmik	Paper
ISCAS 2020: <i>A 4.2-pJ/Conv 10-b Asynchronous ADC with Hybrid Two-Tier Level-Crossing Event Coding</i> R. Kubendran, J. Park, Ritvik Sharma , C. Kim, S. Joshi, G. Cauwenberghs, S. Ha	Paper

Selected Research Projects

DSL and Compiler for Tensor Network based Quantum Simulation <i>Under Prof Sara Achour, NCS Lab, Stanford University</i> • Domain-specific language for Tensor Network-based Quantum physics and chemistry simulations with constraints. Optimizing compiler for compiling contained optimization problems on GPUs, CPUs with Sparse Tensor Algebra	Nov.'24-Present
--	-----------------

Rewrite-based Compilers for Quantum Circuits

Feb.'23-Present

Under Prof Sara Achour, NCS Lab, Stanford University

- Developed a quantum circuit compiler for control-flow-based high-level quantum computing languages (SPARE)
- Circuit optimization using rewrites that reduce the generated quantum circuit size by $2-7\times$. Published in PLDI 2025.
- Developed the DARE compiler to circuits that use qudit higher-energy states.
- DARE verified circuit rewrites to reduce the circuit size by up to $2.1\times$. Published in PLDI 2024.

Sparse Tensor Algebra Tiling Optimization

Sept.'23-Sept.'24

Under Prof Mark Horowitz, Stanford University

- Developed a performance model for sparse tensor algebra, which models data distribution probabilistically
- Identify and collect easy-to-access high-level statistics from compressed sparse data formats
- Implemented a static sparse tiling optimization tool (D2T2) based on the probabilistic traffic model, reducing traffic by 1-2 order of magnitude. Published in Micro 2025.

Architecture Simulations for Sparse Tensor Algebra accelerators

March'22-March.'23

Under Prof Mark Horowitz, AHA Project, Stanford University

- Built a cycle-approximate architecture simulator for Sparse Abstract Machine (published at ASPLOS 2023)
- Developing a calibration framework for architecture-level simulations against performance of taped-out hardware for cross-platform performance validation

Undergraduate Research Assistant at IIT Delhi

Sept.'18-May'21

- Designed CMOS-analog on-chip-memory based accelerators using RRAMs and other non-volatile memories
- Designed a novel Quantum Machine Learning algorithm that encodes and processes multiple data samples together providing exponential speedups for model training in Variational Classifiers

Undergraduate Research Intern at UCSD

May-August 2019

- Tested a monolithic CMOS-RRAM compute-in-memory accelerator for ML inference and training
- Tested a hybrid 10-bit resolution asynchronous ADC chip with 5-bits of level crossing and 5-bits with flash ADC

Undergraduate Research Assistant at Princeton University

May'20-Aug'21

- Created an accelerator-CNN co-design tool that explores a broad range of CNNs sampled from a family of model search space for Accelerator design
- Implemented RTL code for implementing the accelerator design and memory simulations

Achievements

- Awarded **Stanford Graduate Fellowship (SGF)** by Stanford University
- **Department rank 2** in a batch of 85+ students with GPA of 9.704/10 at IIT Delhi
- Awarded **semester merit award** for being in the top 7% of the class in 6 out of 8 semesters at IIT Delhi

Relevant Courses

Computer Science: Compilers for Domain-Specific Languages, Compilers, Parallel Processors Beyond Multi-Core Processing, Reinforcement Learning, Quantum Computing, Computer Architecture for HPC, Computer Architecture, Advanced Machine Learning, Data Structures, Computational Complexity

Electrical Engineering: VLSI Systems, Design Project in VLSI, Image Systems Engineering, Coding Theory, Digital Image Processing, Neuromorphic Engineering, Communication Engineering

Physics: Quantum Computing, Theory of General Relativity, Electromagnetics, Quantum Mechanics

Technical Skills

Languages: Python, C++, C, MATLAB, Rust, Java, Verilog, Assembly (x86-64 and MIPS)

Libraries and packages: Scipy, PyTorch, Tensorflow, Qiskit, Cirq, Taco Sparse Compiler, Eigen

Frameworks and Tools: HLS (catapult and Vivado), MLIR, Cadence and Synopsys EDA Tools

Teaching Experience

Teaching Assistant at IIT Delhi, for course EE 205, Signals and Systems for Fall and Spring Semester 2020-21.