ROHAN JUNEJA

COM3, 11 Research Link ◇ Singapore 119391

+65 8132 4894 ◇ rohan@comp.nus.edu.sg ◇ https://rohanjuneja.github.io/

Research Interests

I work at the intersection of VLSI design, computer architecture, and artificial intelligence, co-designing hardware-software solutions to address challenges from the end of Dennard Scaling. My focus lies on developing reconfigurable architectures that support irregularity and mixed precision demands arising from sparsity and quantization - key techniques for AI model compression. By optimizing performance, efficiency, and reliability, I aim to reduce the carbon footprint of modern computing systems, validating my designs through real-world ASIC tape-outs.

Education

- **National University of Singapore**
  Ph.D. in Computer Science (CGPA: 4.58/5) *January 2021 - Present*

  Advisors: Prof Peh Li Shiuan, Prof Tulika Mitra

  Thesis: Scalable Architectures for Sparse and Quantized AI Models: Bridging Efficiency and Computational Complexity

- **IIIT Delhi**
  B.Tech in Electronics and Communications Engineering (CGPA: 8.41/10) *May 2014 - May 2018*

  Advisors: Prof Lam Siew Kei, Prof Sujay Deb

  Thesis: Securing untrusted memories in embedded systems

Professional Experience

- **Advanced Micro Devices**
  PhD Research Intern *May 2022 - July 2022*
  - Worked as a PhD Research Intern, responsible for designing an accelerator for Ethereum's Beacon Chain (based on Proof-of-Stake).

- **Renesas Electronics Corporation**
  PhD Engineering Intern *Jan 2022 - April 2022*
  - Worked as a PhD Engineering Intern on the Renesas' Dynamically Reconfigurable Processor (DRP).

- **Qualcomm**
  CPU Design Engineer *July 2018 - January 2021*
  - Worked as a CPU design engineer for Qualcomm Snapdragon Processors.
  - Delivered multi-clock domain and Low Power (UPF) RTL for ARM Kryo cores in Snapdragon 765G, as well as medium-, high-tier, and compute chips.
  - Responsible for restructuring memory model RTL to support partial power gating.
  - Gained experience with Power Manager IP, DCVS, Low Power Modes using ARM's P-channel, and boot RTL in Snapdragon CPUs.
  - Experienced in writing SystemVerilog assertions, code coverage and functional coverage closure.
  - Experienced in Synthesis flows, reviewing Design Constraints, timing arcs, and optimised registers.

Publication Record

Conferences

1. HALO: Hardware-aware quantization with low critical-path-delay weights for LLM acceleration *DAC 2025 [Under Review]*
   Rohan Juneja, Shivam Aggarwal, Safeen Huda, Tulika Mitra, Li-Shiuan Peh
2. A Data-Driven Dynamic Execution Orchestration Architecture *ISCA 2025 [Under Review]*
   Pranav Dangi, Zhenyu Bai, **Rohan Juneja**, Zhaoying Li, Zhanglu Yan, Huiying Lan, Tulika Mitra
3. Nexus Machine: An Active Message Inspired Reconfigurable Architecture for Irregular Workloads *ISCA 2025 [Under Review]*
   Rohan Juneja, Thilini Kaushalya, Pranav Dangi, Zhaoying Li, Tulika Mitra, Li-Shiuan Peh
4. Enhancing CGRA Efficiency through Aligned Compute and Communication Provisioning *ASPLOS 2025*
   Zhaoying Li, Pranav Dangi, Chenyang Yin, Thilini Kaushalya, **Rohan Juneja**, Cheng Tan, Zhenyu Bai, Tulika Mitra
5. ZeD: A Generalized Accelerator for Variably Sparse Matrix Computations in ML *PACT 2025*
   Pranav Dangi, Zhenyu Bai, **Rohan Juneja**, Dhananjaya Wijerathne, Tulika Mitra
6. NOVA: NoC-based Vector Unit for Mapping Attention Layers on a CNN Accelerator *DATE 2024*
   Mohit Upadhyay, **Rohan Juneja**, Weng-Fai Wong, Li-Shiuan Peh
7. FLEX: Introducing FLEXible Execution on CGRA with Spatio-Temporal Vector Dataflow *ICCAD 2023*
   Thilini Kaushalya, Dan Wu, **Rohan Juneja**, Dhananjaya Wijerathne, Tulika Mitra, Li-Shiuan Peh
8. REACT: A Heterogeneous Reconfigurable Neural Network Accelerator with Software Configurable NoCs for Training and Inference on Wearables *DAC 2022*
   Mohit Upadhyay, **Rohan Juneja**, Bo Wang, Jun Zhou, Weng-Fai Wong, Li-Shiuan Peh
9. Cache-Aware Dynamic Skewed Tree for Fast Memory Authentication *ASP-DAC 2021*
   Saru Vig, **Rohan Juneja**, Siew Kei Lam
10. DISSECT: Dynamic Skew-and-Split Tree for Memory Authentication *DATE 2020*
    Saru Vig, **Rohan Juneja**, Siew Kei Lam, Guiyuan Jian
11. Dynamic NoC Platform for Varied Application Needs *ISQED 2018*
    Sidhartha Shankar, Hemanta K. Mondal, **Rohan Juneja**, Sri Harsha Gade, Sujay Deb

Journals

1. CTScan: A CGRA-based Platform for Emulation of Power Side-Channel Attacks on Edge CPUs *TRETS 2025 [Minor Revision]*
   Yaswanth Tavva, **Rohan Juneja**, Trevor E. Carlson, Li-Shiuan Peh
2. Framework for Fast Memory Authentication using Dynamically Skewed Integrity Tree *TVLSI 2019*
   Saru Vig, **Rohan Juneja**, Guiyuan Jiang, Siew Kei Lam, Changhai Ou

Chip Tapeouts

1. A 360 GOPS/W CGRA in a RISC-V SoC with Multi-Hop Routers and Idle-State Instructions for Edge Computing Applications *ISOCC 2024*
   Vishnu Nambiar, Yi Sheng Chong, Thilini Kaushalya, Dhananjaya Wijerathne, Zhaoying Li, **Rohan Juneja**, Li-Shiuan Peh, Tulika Mitra, Anh Tuan Do
2. PACE: A Scalable and Energy Efficient CGRA in a RISC-V SoC for Edge Computing Applications *HotChips 2024*
   Vishnu Nambiar, Yi Sheng Chong, Thilini Kaushalya, Dhananjaya Wijerathne, Zhaoying Li, **Rohan Juneja**, Li-Shiuan Peh, Tulika Mitra, Anh Tuan Do

Patents

1. A Reconfigurable Execution Unit for Collective Routing and Computation of Multiple Operations for Hardware Acceleration *Patent 10202402819X, 2025*
   Zhaoying Li, **Rohan Juneja**, Tulika Mitra, Pranav Dangi

Teaching Experience

- **National University of Singapore (NUS)**
  Teaching Assistant, Introduction to Operating Systems *January 2023 – April 2023*
  - Independently led weekly tutorial sessions for a cohort of 48 undergraduate students, facilitating an in-depth understanding of core operating systems concepts.
  - Provided detailed feedback on student assignments and exams, fostering continuous improvement and strengthening analytical skills.
- **National University of Singapore (NUS)**
  Teaching Assistant, Introduction to Operating Systems *August 2022 – November 2022*
  - Delivered interactive tutorial sessions tailored to a class of 48 students, emphasizing problem-solving and practical application of OS principles.
  - Supported Prof. Weng-Fai Wong in streamlining assessments, including the design, evaluation, and grading of assignments and examinations.
- **IIIT Delhi**
  Teaching Assistant, GPU Computing *January 2018 – April 2018*
  - Conducted weekly lab sessions on CUDA programming, equipping students with practical skills in GPU computing and parallel programming.
  - Designed and graded hands-on programming assignments to evaluate comprehension and encourage problem-based learning.
  - Collaborated with Prof. Ojaswa Sharma to develop supplemental course materials that bridged theory with real-world GPU applications.
- **IIIT Delhi**
  Teaching Assistant, Digital Circuits *January 2017 – April 2017*
  - Led tutorials and hands-on lab sessions on SystemVerilog and FPGA-based design, providing students with foundational experience in digital logic implementation.
  - Created and assessed assignments to rigorously evaluate student understanding and skill development.
  - Assisted Prof. Sumit Darak in delivering a comprehensive curriculum focused on digital circuit design methodologies and tools.