Rishi Nandha Vanchinathan https://rishinandha.github.io rishinandhav@smail.iitm.ac.in

Research Statement

Motivation and Vision: To reform the current paradigm of computing, to reduce every watt our deep learning models burn, is indispensable to the pursuit of fair and accessible technology today. In a world constrained by resources, powering such models with a disproportionate amount of energy is unjust. The desire for justice in technology drives my focus on in-memory and neuromorphic computing. My research interests can be broadly framed as:

1. Neuromorphic accelerators that reduce power demands for both edge inference and large-scale server-side deployment.
2. Analog/Mixed-Signal In-Memory Computing architectures that generalize well for modern layers (such as Grouped-Query Attention and Mixture-of-Experts)

Grounded in an RF IC design background, I also am excited to explore neuromorphic computing from an RF-informed and RF-inspired perspective for addressing energy bottlenecks

Origin of Interest: In 2023, as part of a five-member team, I co-developed a pre-incubated assistive mobility device for visually-impaired individuals navigating complex environments in India. Our system relied on deep learning inference at the edge. However, the power consumption in the current paradigm of GPU-based inference made it infeasible to meet our energy budget. This challenge sparked my exploration towards emerging computing paradigms.

Exploration of Emerging Paradigms: To pursue this goal, I explored emerging paradigms of computing through coursework at IIT Madras. At first, as part of a collaboration with IBM Research, India, our team contributed open-source quantum computing datasets to the IBM Qiskit framework.

Then, I also undertook the course "Devices and Technologies for AI and Neuromorphic Computing" (offered by Prof. Bhaswar Chakrabarti), which introduced me to emerging non-volatile memory technologies and their applications in computing. In the course project about passive RRAM arrays for neural inference, we explored co-designing quantization-aware networks as a replacement for post-training quantization, thereby improving hardware-software consistency.

I currently serve as a teaching assistant in the subsequent offering of the course, wherein I'm also conducting tutorials on design of NVM synaptic arrays and CMOS spiking neurons. These experiences have collectively strengthened my belief that in-memory and neuromorphic computing can reform the energy landscape of deep learning inference.

Foundation in RF IC Design: Coming from a background in RF integrated circuit design, I have learned to operate at the edge of technological limits: perspectives I aim to carry forward in my approach towards energy-aware architectures. My ongoing master's thesis, under the supervision of Prof. Aniruddhan Sankaran, involves the tape-out of a 7 GHz transceiver in 65 nm CMOS technology. Being responsible for the complete design flow, I have gained hands-on experience across the IC design lifecycle. This has provided me with the discipline I'd like to utilize in designing energy-efficient neuromorphic chips.

Future Direction: Motivated by my pursuit of accessible technology and informed by my experiences, I aim to pursue doctoral studies with focus in energy-aware analog design for neuromorphic and in-memory computing systems to develop sustainable AI hardware.