SAION KUMAR ROY

411 Coordinated Science Laboratory, 1308 W Main St, Urbana, IL, 61801
Ph.D., University of Illinois at Urbana-Champaign
(217)-898-2084 ⋄ saionkr2@illinois.edu

WORK EXPERIENCE

Northeastern University *Sept 2024 onwards*
Incoming Postdoctoral Researcher, Electrical and Computer Engineering
Advisor: Professor Yunsi Fei

EDUCATION

University of Illinois at Urbana-Champaign *2018 - 2024*
PhD student, Electrical and Computer Engineering GPA: 3.88/4.0
Advisor: Professor Naresh Shanbhag

Indian Institute of Technology Kharagpur *2013 - 2018*
Bachelor and Masters of Technology GPA: 9.26/10
Electronics and Electrical Communication Engineering

RESEARCH EXPERIENCE

University of Illinois at Urbana-Champaign *Urbana, IL*
Research Assistant *August 2018 - July 2024*

- **Advised by Professor Naresh Shanbhag**
- **Research Interests**: **Energy-efficient** IC design focusing on resistive in-memory computing (IMC) architectures for **machine learning** (ML) applications. Additionally, exploring the **security** vulnerabilities of deep neural network (DNN) implementations.
- **Research Projects - Energy Efficient IC Design for DNNs:**
  - **Benchmarking ML Accelerators** (**OJ-SSCS**'22, **CICC**'22)
    - Proposed a comprehensive benchmarking methodology leveraging a compositional view.
    - Benchmarked > 70 ML IC designs published in major circuit conferences (ISSCC, VLSI, CICC) since 2018 to comprehend the limits and identify challenges in ML accelerator design space.
    - SRAM-based IMCs have higher energy efficiency (TOPS/W) and compute density (TOPS/mm^2) over digital accelerators at the bank level, but the gap reduces dramatically at the processor level.
    - Resistive IMCs currently lag behind SRAM IMCs and digital accelerators in both TOPS/W and TOPS/mm^2 due to their low compute accuracy.
  - **Algorithmic Error Compensation for MRAM IMC** (**ESSCIRC**'23, **JSSC**'24 (accepted))
    - Characterized the compute SNDR and the energy-accuracy trade-offs for a 22 nm MRAM IMC chip.
    - Proposed SNDR-boosting techniques, including offset-compensated current sensing (OCCS) to minimize static ADC column mismatch and algorithmic statistical error correction (SEC) to reduce wire parasitic induced non-linearity.
    - Obtained a $5\times$ reduction in energy by trading-off SEC-enabled SNDR gain of 2.6 dB-to-6 db.
    - Demonstrated SEC-enabled increase in neural network (NN) accuracy from 74.8% to 82.0% for CIFAR-10 over ResNet-20 without resorting to noise-aware training.
  - **Energy-Accuracy Trade-offs in Resistive Parallel-bars** (**JxCDC**'24)
    - Formulated behavioral models for circuit non-idealities at array, sensing, and ADC levels and verified with chip measurements.
    - Evaluated the limits of compute-SNDR and its dependencies on device, circuit, and architecture parameters obtaining a maximum SNDR of 18 dB-to-23 dB barely meeting the requirement for achieving network-level accuracy close to digital.
    - Highlights need SNDR boosting methods such as SEC and noise-aware training.

- **Accuracy Limits of Resistive Crossbars** (**ISCAS**'22)
  - Evaluated the fundamental limits on compute-SNDR for MRAM, ReRAM, and FeFET-based crossbars using behavioral models verified in 22 nm commercial node.
  - Analyzed the dependence of compute SNDR on sensing resistance R_s, ADC precision B_{ADC} (circuit), dot-product dimension N (architecture), and the resistive contrast R_{off}/R_{on} (device).
  - Increasing device level resistive contrast provides diminishing returns in SNDR since the input DAC and conductance mismatch begin to dominate.
  - Demonstrated that network-level accuracy correlates with the bank-level SNDR, revealing the accuracy-maximizing parameters using SNDR analysis.
- **Security Vulnerabilities of Resistive IMCs** (**ICCAD**'24 (to appear), **IEDM**'24 (to appear))
  - Proposed a statistical framework for model extraction attacks (MEAs) on resistive IMCs, and use it to construct three query-based attacks: basis vector (BV), least squares (LS), and stochastic gradient descent (SGD) attack.
  - Attack on real-life MRAM-based IMC prototype revealed that it is indeed vulnerable to MEAs despite the presence of analog non-idealities.
  - Energy-accuracy-security trade-offs for eNVM-based IMCs reveal that they can simultaneously achieve high energy efficiency, inference accuracy, and enhanced resilience against MEAs by operating at low compute accuracy.

INTERNSHIP EXPERIENCE

Texas Instruments, Kilby Labs *Dallas, TX*
Supervised by: Dr. Mahesh Mehendale and Dr. Avishek Biswas *Summer 2021*

- **NAND Flash-based IMC Design**: Characterized NAND flash cell properties and designed sensing circuits for IMC operation.
- **Behavioral Modeling of ADC in PyTorch**: Analyzed the impact of ADC non-idealities on the DNN accuracy of IMC implementations using PyTorch models.

PUBLICATIONS

- **S. K. Roy**, and N. Shanbhag, "The Energy-Accuracy-Security Trade-off in Resistive In-memory Architectures", **IEDM**, 2024 [to appear].
- **S. K. Roy**, and N. Shanbhag, "On the Security Vulnerabilities of MRAM-based In-Memory Computing Architectures against Model Extraction Attacks", **ICCAD**, 2024 [to appear].
- **S. K. Roy**, H.-M. Ou, M. Ahmed, P. Deaville, B. Zhang, N. Verma, P. Hanumolu, and N. Shanbhag, "Compute SNDR-boosted 22nm MRAM-based In-memory Computing Macro using Statistical Error Compensation", **JSSC**, 2024 [accepted].
- **S. K. Roy**, and N. Shanbhag, "Energy-Accuracy Trade-offs for Resistive In-Memory Computing Architectures", **JxCDC**, 2024.
- **S. K. Roy**, H.-M. Ou, M. Ahmed, P. Deaville, B. Zhang, N. Verma, P. Hanumolu, and N. Shanbhag, "Compute SNR-boosted 22nm MRAM-based In-memory Computing Macro using Statistical Error Compensation", **ESSCIRC**, 2023.
- N. Shanbhag and **S. K. Roy**, "Benchmarking In-memory Computing Architectures", **OJ-SSCS**, 2022.
- N. Shanbhag and **S. K. Roy**, "Comprehending In-memory Computing Trends via Proper Benchmarking", **CICC**, 2022.
- **S. K. Roy**, A. Patil, and N. Shanbhag, "Fundamental Limits on the Computational Accuracy of Resistive Crossbar-based In-memory Architectures", **ISCAS**, 2022.

CONFERENCE TALKS

- "Compute SNR-boosted 22nm MRAM-based In-memory Computing Macro using Statistical Error Compensation", **ESSCIRC**, 2023.
- "Fundamental Limits on the Computational Accuracy of Resistive Crossbar-based In-memory Architectures", **ISCAS**, 2022.
- "Fundamental Limits on the Compute SNR of Resistive In-memory Architectures", **CSL Student Conference, Circuits and Systems Session (Best Speaker Award)**, 2023.
- "GPLeak: Software Power Side Channel Leakage in GPUs", **CSL Student Conference, Security and Privacy Session (Oral Presentation Award)**, 2024.

TEACHING EXPERIENCE

University of Illinois at Urbana-Champaign *Fall 2022*
ECE 342: Electronics Circuits, Teaching Assistant

- **Instructors**: Professor Pavan Hanumolu and Professor Chandrasekhar Radhakrishnan
- Helped prepare homework and exams, held office hours, and assisted students in understanding course material and solving problems.

TECHNICAL SKILLS

Programming	Python (PyTorch), Verilog, VerilogA, MATLAB, C
EDA Tools	Cadence Virtuoso, Genus, Innovus, SPICE, Xilinx Vivado
ASIC Testing	KiCAD (PCB design), worked with PYNQ, Arduino, and Atmel AVR Boards

GRADUATE COURSEWORK

Circuits: Digital IC Design, Analog IC Design
Signal Processing: Advanced DSP, Vector Space Signal Processing
Systems: Random Processes, Detection and Estimation Theory, Machine Learning
Computer Science: Computer Architecture, Advanced Computer Security