HARSH SHARMA

harshari.github.io • harsh.sharma@wsu.edu

RESEARCH SUMMARY

My general research interests are at the intersection of artificial intelligence (AI) and computing system with a focus on exploiting their synergistic strengths: AI for design and optimization of computing systems, and the design of optimized computing systems for AI applications. The current focus of my research is on AI-driven design and Optimization of *Chiplet-based Systems* for enabling high-performance and low-energy computing for various applications including training/inference of large AI models. Specific topics include:

- Enabling server-scale system design with low-latency interconnect networks.
- Hardware and software co-design to create chiplet systems for training/inference with large AI models including CNNs, GNNs, and Transformers.
- Design of high-performance and energy-efficent manycore systems to overcome Moore's law.
- Design of defect-aware chiplet-based systems to reduce carbon footprint at scale.
- Accelerating the design of robust, reliable, and environmentally sustainable paradigms.

EDUCATION

Ph.D. Candidate, Computer Engineering, 3.93 GPA **2021–Present**

Advisors: Partha Pande & Jana Doppa

Washington State University *Pullman, Washington*

Coursework: • Advanced Computer Architecture • Machine Learning • Computational Genomics • Neural Network Design & Analysis • SoC Design and Test • VLSI Systems Design

Bachelor of Engineering, Electronics and Communication Engineering **2017–2021**

NSIT, Delhi University *New Delhi, India*

Department ranker (Top 5%)

INDUSTRIAL EXPERIENCE

Machine Learning Research Intern **June 2020–December 2020**

Lenskart *New Delhi, India*

- The developed AR tools with vision model boosted online sales by 35%. 20% conversion rate.
- Project Involvement: Optimized a vision transformer model to overlay on the occluders(glasses here) in XR/Augmented Reality.
- Ccontributions to developing an Instagram toolchain using SparkAR that garnered over 200k impressions during the IPL cricket season in 2020.

SELECTED PUBLICATIONS

1. [Best Paper Award] **Harsh Sharma**, Lukas Pfromm, Rasit Topaloglu, Janardhan Rao Doppa, Umit Y. Ogras, Ananth Kalyanraman, Partha Pratim Pande. Florets for Chiplets: Data Flow-aware High-Performance and Energy-efficient Network-on-Interposer for CNN Inference Tasks. *ACM Transactions on Embedded Computing Systems, Hamburg*, 2023.
2. **Harsh Sharma**, Pratyush Dhingra, Janardhan Rao Doppa, Umit Ogras, Partha Pratim Pande. A Heterogeneous Chiplet Architecture for Accelerating End-to-End Transformer Models. *ACM TODAES*, Under Review.
3. **Harsh Sharma**, Gaurav Narang, Janardhan Rao Doppa, Umit Ogras, Partha Pratim Pande. Dataflow aware interconnect design for DNN accelerators. *Design Automation and Test in Europe DATE*, Spain, 2024.
4. **Harsh Sharma**, Sumit K. Mandal, Janardhan Rao Doppa, Umit Y. Ogras, Partha Pratim Pande. Achieving Datacenter-scale Performance through Chiplet-based Manycore Architectures. *Design Automation and Test in Europe DATE*, Belgium, 2023.

5. [Best Paper Award] **Harsh Sharma**, Sumit K. Mandal, Janardhan Rao Doppa, Umit Y. Ogras, Partha Pratim Pande. SWAP: A Server-Scale Communication-Aware Chiplet-Based Many-core PIM Accelerator. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, Phoenix/Shanghai*, 2022.
6. **Harsh Sharma**, Dhananjay Gadre, Sangeeta Gadre, Smriti Srivastava. Science on a stick: An experimental and demonstration platform for learning several physical principles. *American Journal of Physics*, 2022.

AWARDS AND HONORS

- Harvard Scholar at HPAIR Conference, Kazakhstan. Technology Track (**top 1%**)
- Best Paper Award at ACM/IEEE Embedded Systems Week Conference, 2023 *
- Best Paper Award at ACM/IEEE Embedded Systems Week Conference, 2022 †
- ACM SIGDA Richard Newton Young Fellowship, 2022

SELECTED PROFESSIONAL AND OUTREACH ACTIVITIES

Reviewer

- ESWEEK, ICCAD, DAC, DATE, AAAI

Invited Talks

- Talk on *'Accelerating the Future of Electronics and Beyond'* at IISc Bangalore, Jan 2024. ‡
- Talk on 'Florets for Chiplets for CNN Inference Tasks' at TUHH, Germany, Oct. 2023.
- Talk on *'Server-scale Communication Aware Chiplet-based Systems'* at Boston University, Apr 2023. §
- Talk on *AI-Driven SoC Design and Optimizations for more Moore* at NSIT Delhi, Apr 2023.
- *SWAP: A Server-scale Chiplet-PIM Accelerator* at Phoenix, Oct 2022.
- Talk on *AI-Driven Design and Optimization of Chiplet-based Systems* at Pullman, March 2023.

Panelist at JC Bose Science Dialogues at India Science Festival, IISER, Pune, Jan 2024

- Engaged in a selective, invitation-only roundtable designed to foster free and open discussion among science and technology ecosystem leaders from MIT, NYU, IISc, the Bill and Melinda Gates Foundation, former Principal Scientific Advisors, Pune Knowledge Cluster (Serum Institute of India), and the Indian Institute of Astrophysics.
- Focused on the evolving role of Indian universities in shifting from demographic to economic contributors, highlighting the importance of intellectual property development and industrial collaborations. That might call on India's academic institutions to evolve.
- **Outcome**: Contributed to a three-page document outlining key insights and strategies, submitted to the Prime Minister's Office and the Office of the Principal Scientific Advisor, influencing national science and technology policies and actions.

SKILLS

- **Programming Languages.** Python, Bash, C/C++, LaTeX, Java, MATLAB
- **Tools/Packages.** Git, PyTorch, Python data science tools

*https://news.wsu.edu/news/2023/10/11/researchers-receive-best-paper-award/
†https://school.eecs.wsu.edu/2022/10/14/cases-best-paper-award/
‡Link https://www.csa.iisc.ac.in/~skmandal/speakers.html
§Based on https://medium.com/@harshari/accelerating-the-future-of-electronics-e23cc42d9d39