HARSH SHARMA

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

SUMMARY

I am a Ph.D. Candidate in Computer Engineering. 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: Prof. Partha Pande and Prof. 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%)

AWARDS AND HONORS

- Voiland College of Engineering and Architecture, Outstanding RA in EECS, 2024.
- Harvard Scholar at HPAIR Conference, Kazakhstan. Technology Track (**top 1%**)
- Best Paper Award at ACM/IEEE Embedded Systems Week Conference, 2024 *
- 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 PhD Fellowship, 2022

PUBLICATIONS

1. **Harsh Sharma**, Alish Kanani, Janardhan Rao Doppa, Umit Ogras and Partha Pratim Pande. HeMu: Energy-Efficient DNN Inferencing via Heterogenous-Multi-Chiplet Architectures. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)*, 2025. Under Review.
2. **Harsh Sharma**, Janardhan Rao Doppa, Umit Ogras and Partha Pratim Pande. TAPAS: Thermally-aware Multi-Chiplet Architectures using Glass Interposers. *ACM Transactions on Embedded Computing Systems (TECS)*, 2025. Under Review.

*https://news.wsu.edu/news/2024/10/17/researchers-receive-cases-best-paper-award/
†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/

3. Lukas Pfromm, Alish Kanani, **Harsh Sharma**, Parth Solanki, Eric Tervo, Jaehyun Park, Janardhan Rao Doppa, Partha Pratim Pande, Umit Y Ogras. MFIT: Multi-Fidelity Thermal Modeling for 2.5 D and 3D Multi-Chiplet Architectures. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)*, 2025. Under Review.

4. **Harsh Sharma**, Pratyush Dhingra, Janardhan Rao Doppa, Umit Ogras, Partha Pratim Pande. A Heterogeneous Chiplet Architecture for Accelerating End-to-End Transformer Models. To appear in *ACM Transactions on Design Automation of Electronic Systems (TODAES)*, 2025

5. **Harsh Sharma**, Umit Y. Ogras, Ananth Kalyanraman, Partha Pratim Pande. A Dataflow-Aware Network-on-Interposer for CNN Inferencing in the Presence of Defective Chiplets. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD), 2024.* **[Best Paper Award @ IEEE/ACM Embedded Systems Week Conference]**

6. **Harsh Sharma**, Gaurav Narang, Janardhan Rao Doppa, Umit Ogras, Partha Pratim Pande. Dataflow-aware Interconnect design for DNN accelerators. *IEEE/ACM Design Automation and Test in Europe Conference (DATE)*, Spain, 2024.

7. Jaehyun Park, Alish Kanani, Lukas Pfromm, **Harsh Sharma**, Parth Solanki, Eric Tervo, Jana R. Doppa, Partha P. Pande, and Umit Y. Ogras. Thermal modeling and management challenges in heterogenous integration: 2.5 D chiplet platforms and beyond. *IEEE 42nd VLSI Test Symposium (VTS)*, 2024.

8. **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 (TECS)*, 2023. **[Best Paper Award @ IEEE/ACM Embedded Systems Week Conference]**

9. **Harsh Sharma**, Sumit K. Mandal, Janardhan Rao Doppa, Umit Y. Ogras, Partha Pratim Pande. Achieving Datacenter-scale Performance through Chiplet-based Manycore Architectures. *IEEE/ACM Design Automation and Test in Europe Conference (DATE)*, Belgium, 2023.

10. **Harsh Sharma**, Sumit K. Mandal, Janardhan Rao Doppa, Umit Y. Ogras, Partha Pratim Pande. SWAP: A Server-Scale Communication-Aware Chiplet-Based Manycore PIM Accelerator. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)*, 2022. **[Best Paper Award @ IEEE/ACM Embedded Systems Week Conference]**

11. **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.

INDUSTRIAL EXPERIENCE

Research Intern **August 2024–Present**

AMD Research Group *Austin, Texas*

- Black-Box design space exploration for enabling next-generation of chiplet systems. (under non-disclosure agreement).
- Interface for enabling end-to-end Traceability using graph structures.
- Key skills: Neural Networks, Machine Learning, GNN, Database Design, Snowflake, Statistics, Efficient Design-Space-Exploration, Optimization, BoTorch.
- *Rating: Exceeded expectations*

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 PROFESSIONAL AND OUTREACH ACTIVITIES

Invited Talks

- *Towards Large-Scale composable chiplet systems for Datacenters* at AMD, Austin, Nov. 24.
- *2.5D system design employing defective chiplets* at UNC, Raleigh, Oct. 2024.
- *Heterogenous architecture-package co-optimizations for Datacenters* at Microsoft, Bellevue, Sept. 24.
- *Accelerating the Future of Electronics and Beyond* at Indian Institute of Sciences (IISc), Bangalore, Jan 2024. [‡]
- *A Dataflow-Aware Network-on-Interposer for CNN Inferencing in the Presence of Defective Chiplets* at ESWEEK Conference, 2024.
- *Florets for Chiplets to Accelerate Deep Neural Network Inference Tasks* at ESWEEK Conference, Germany, Oct. 2023.
- *Server-scale Communication-Aware Chiplet Systems* at Boston University, Apr 2023. [§]
- *AI-Driven SoC Design and Optimizations for more than Moore* at NSIT Delhi, Apr 2023.
- *SWAP: A Server-scale Chiplet-PIM Accelerator* at ESWEEK Conference, Phoenix, Oct 2022.
- *AI-Driven Design and Optimization of Chiplet-based Systems* at Washington State University, 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.

Journal and Conference Reviewer

- ACM/IEEE Embedded Systems Week (ESWEEK) Conference
- ACM/IEEE Design Automation Conference (DAC)
- Embedded System Letters (ESL)
- ACM/IEEE Design Automation and Test in Europe (DATE) Conference
- AAAI Conference on Artificial Intelligence (AAAI)
- IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)
- ACM Transactions on Design Automation of Electronic Systems (TODAES)
- ACM Transactions on Embedded Computing Systems (TECS)

SKILLS

- **Programming Languages.** Python, C/C++, MATLAB
- **Tools/Packages.** BookSim2, COMPASS, MEMSYS, Gem5, HeteroGarnet, Ansys, State-Space modelling, ASTRA-SIM, RAMULATOR, Hotspot 6.0, NeuroSim, PyTorch, Python data science tools, ARM Microcontrollers, Xilinx FPGA, Deep Learning, Javascript, FrontEnd Design, Raspberry Pi, Embedded System Design, Design-space exploration, Topology construction, GNNs.

[‡]Link `https://www.csa.iisc.ac.in/~skmandal/speakers.html`
[§]Based on `https://medium.com/@harshari/accelerating-the-future-of-electronics-e23cc42d9d39`

REFERENCES

- Prof. Partha Pande, Boeing Centennial Chair Professor of Computer Engineering, WSU
- Prof. Jana Doppa, Huie-Rogers Chair Associate Professor of Computer Science, WSU
- Prof. Umit Ogras, Professor of ECE, University of Wisconsin-Madison