# HARSH SHARMA

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

#### RESEARCH SUMMARY

My general research interests are in artificial intelligence (AI) and machine learning (ML) with a focus on Design and Optimization of *Chiplet-based systems* for enabling high-performance computing. The overarching goal of my research is to develop principled AI-driven Design Paradigms in such huge-search spaces for efficient chip design geared towards high-impact applications. Specific topics include:

- Enabling Server-Scale system design with low-latency interconnect networks.
- Knowledge distillation with data-flow aware high-performance computing for Large Language Models.
- Design of high-performance and energy-efficient manycore systems to overcome Moore's law.
- Defect-aware integration using 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

Washington State University

Pullman, Washington

# Bachelor of Engineering, Electronics and Communication Engineering

2017 - 2021

NSIT, Delhi University
Department ranker (Top 1%)

New Delhi, India

#### INDUSTRIAL EXPERIENCE

# Machine Learning Research Intern

June 2020-December 2020

Lenskart.com

New Delhi, India

Developed AR tools with vision model to boost online sales by 35% during COVID19 Pandemic.

### AWARDS AND HONORS

- Best Paper Candidate at ESWEEK(Hamburg, Germany), 2023
- Best Paper Award at ESWEEK(Phoenix/Shanghai), 2022 †
- DAC Richard Newton Young Fellow, 2022

# SELECTED PUBLICATIONS

- 1. **Harsh Sharma**, Lukas Pfromm, Janardhan Rao Doppa, Umit Y. Ogras, Ananth Kalyanraman, Partha Pratim Pande. Network-on-Interposer Design for CNN Inferencing in Presence of Defective Chiplets. *ICCAD*, 2023. Under Review
- 2. [Best Paper Candidate] 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. ESWEEK, 2023.
- 3. **Harsh Sharma**, Sumit K. Mandal, Janardhan Rao Doppa, Umit Y. Ogras, Ananth Kalyanraman, Partha Pratim Pande. Achieving Datacenter-scale Performance through Chiplet-based Manycore Architectures. *DATE*, 2023.
- 4. [Best Paper Award] Harsh Sharma, Sumit K. Mandal, Janardhan Rao Doppa, Umit Y. Ogras, Ananth Kalyanraman, Partha Pratim Pande. SWAP: A Server-Scale Communication-Aware Chiplet-Based Manycore PIM Accelerator. ESWEEK, 2022.
- 5. **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.

<sup>&</sup>lt;sup>†</sup>https://school.eecs.wsu.edu/2022/10/14/cases-best-paper-award/

#### SELECTED PROFESSIONAL AND OUTREACH ACTIVITIES

# Conferences and Invited Talks

- SWAP: A Server-scale Communication aware Chiplet-based PIM Accelerator at ESWEEK 2022.
- Achieving Datacenter-scale Performance through Chiplet-based Manycore Architectures at DATE 2023.
- Florets for Chiplets: Data Flow-aware High-Performance and Energy-efficient Network-on-Interposer for CNN Inference Tasks at ESWEEK-2023.
- Talk on AI-Driven Design and Optimization of Chiplet-based Manycore Systems for Server-Scale Applications at WSU Pullman-2023.
- Talk on AI-Driven Design and Optimization strategies for more Moore at NSIT Delhi (Virtual)-2023.
- Talk on Accelerating the Future of Electronics at Boston University (Virtual)-2023. ‡

#### Reviewer

• ESWEEK 2022-Present, ICCAD 2023-Present, DAC 2022-Present, DATE 2022-Present

#### SKILLS

- Programming Languages. Python, Bash, C/C++, HTML/CSS, LATEX, Java, MATLAB
- Tools/Packages. Git, SQL, PyTorch, TensorFlow, Python data science tools

<sup>&</sup>lt;sup>‡</sup>Based on https://medium.com/@harshari/accelerating-the-future-of-electronics-e23cc42d9d39