# Bakshree Mishra

bakshree.github.io | bmishra3@illinois.edu | linkedin.com/in/bakshree

## Education

**PhD in Computer Science** **2021-Present**
University of Illinois, Urbana-Champaign GPA: **3.92/4**
Advisor: Prof. Sarita Adve ( *sadve@illinois.edu*)
Research Interests: Computer Architecture, Hardware-Software Codesign, Machine Learning & Systems

**M.Tech in Computer Science** **2015-2017**
National Institute Of Technology, Rourkela GPA: **9.69/10**
Advisors: Prof. Bansidhar Majhi (NIT Rourkela), Mr. Tarjinder Singh (Intel)

**B.Tech in Computer Science and Engineering** **2010-2014**
College Of Engineering and Technology, Bhubaneswar GPA: **8.85/10**

## Publications

**Mishra, B.** and Alsop, J. and Boyer, M. and Choi, J. and Adve, S. *A Stacked GPU-Memory Architecture for Flexible LLM Execution*. [Under Review].

Suresh, V. and **Mishra, B.** and Zhu, Z. and Jing, Y. and Jin, N. and Block, C. and Mantovani, P. and Giri, D. and Zuckerman, J. and Carloni, L. and Adve, S. *Mozart: Taming Taxes and Composing Accelerators with Shared-Memory*. In Proceedings of the 2024 International Conference on Parallel Architectures and Compilation Techniques (PACT '24). [Paper]

**Mishra, B.** and Chakraborty, D. and Makkadayil, S. and Patil, S. D. and Nallani, B. *Hardware Acceleration of Computer Vision and Deep Learning Algorithms on the Edge using OpenCL*. In Proceedings of EAI Endorsed Transactions on Cloud Systems, 2019 [Paper]

## Work Experience

**Research Associate** May 2025 – Nov 2025
Analytical modelling of LLM inference on future GPU architecture. *AMD, Bellevue*
**Graduate (PhD) Intern** May 2024 – Aug 2024
Analytical modelling of data movement optimizations. *AMD, Bellevue*
**Graduate Research Assistant** May 2022 – Present
Heterogeneous disaggregated accelerator systems *University of Illinois, Urbana Champaign*
**Graduate Teaching Assistant** August 2021 – May 2022
CS 233 Computer Architecture, CS 225 Data Structures *University of Illinois, Urbana Champaign*
**ML and IP Design Engineer** June 2017 – August 2021
Analysis and acceleration of machine learning algorithms *Intel Corporation, Bangalore*
**Graduate Technical Intern** May 2016 – May 2017
Acceleration of pedestrian detection and other ADAS algorithms *Intel Corporation, Bangalore*
**Assistant System Engineer** June 2014 – July 2015
Development of E-Municipality portal *Tata Consultancy Services, Bhubaneswar*
**Summer Intern** June 2013 – August 2013
Prototype modules for E-Municipality portal *Tata Consultancy Services, Bhubaneswar*

## Projects

**Anticipating bottlenecks for LLM inference on 3D stacked GPU-Memory architectures** May 2025 – Present
• Analyzed network on chip (NoC) bottlenecks for GPU architectures with very high bandwidth memory and proposed architectural changes.
• Proposed communication algorithms with insights from alternate domains such as spatial accelerators.
• Developed an analytical model for studying LLM inference workloads for different GPU architectural configurations and communication algorithms.
• Identified key trends and takeaways for the proposed architecture by evaluating LLM inference.
• Work under submission, internal paper accepted for AMD GTAC 2026 during internship at AMD.

**Disaggregation and Migration of Language Model Inference with Edge Devices** March 2024 – Present
• Developed an analytical-empirical model for predicting language inference performance on edge devices.
• Validated the performance model on multiple heterogeneous edge devices.
• Proposed a methodology for language model inference on multi-device edge setups.
• Work under submission, an early workshop paper accepted at MLArchSys 2025 (co-located with ISCA 2025).

**Mozart: Hardware acceleration with disaggregated accelerator systems** August 2021 – March 2024
• Evaluated the impact of accelerator invocation and data movement for different workloads (Mini-ERA, a fully connected network, 3D spatial audio) on a heterogeneous SoC instantiated on FPGA.
• Obtained thorough understanding by evaluating with different coherence protocols, and analyzing waveforms from FPGA as well as simulation.
• Designed and implemented a light-weight accelerator synchronization interface (ASI) in System-C, compatible with the accelerator suite from **ESP framework** to reduce the accelerator invocation overhead.
• Created a synthetic accelerator benchmark suite to evaluate trade-offs of 15 chained and pipelined monolithic and disaggregated accelerator systems for workloads with different compute patterns and intensities.
• Analyzed impact of ASI and disaggregated accelerator systems in accelerating complex workloads such as 3D spatial audio, Mini-ERA and FCNN.

## Select Projects at Intel India (2016 - 2021)

**Real-Time Computer Vision Tasks on Edge Devices**
• Worked on two projects: real-time barcode detection and optical character recognition (OCR).
• Created highly pipelined accelerators using OpenCL, implementing parallel convolution engines.
• Improved barcode detection from **19 FPS** to **104 FPS** and OCR from **10 FPS** to **50 FPS** with 2MP video.
• **Internal paper** and a live demo accepted at Intel Design and Test Technology Conference (DTTC), 2019.
• **Paper** presented at IEEE WinTechCon, Bangalore, India, 2019.

**Hardware Design for Functional Safety IP**
• Designed hardware for Fault Detector module for Functional Safety (FuSa).
• Analyzed High Level Architecture Specification (HAS) and created Micro Architecture Specifications (MAS)
• The IP achieved ISO26262 certification for Functional Safety.
• **Internal paper** on our work was accepted at Intel DTTC 2019, patent application filed.

**Real-Time Pedestrian Detection System Using OpenCL-Based FPGA Acceleration**
• Created a custom architecture for computer vision based Pedestrian Detection system for Master's research
• Deep-dived into FPGA OpenCL compiler optimization issues and reported to the compiler team
• Independently improved initial design to give **3x** performance while reducing area by **10x**

## Select Awards and Honors

• Among Teachers Ranked as Excellent for CS225 in Spring 2022 — 2022
• Co-authored 3 accepted internal papers, presented a demo at Intel DTTC, Portland, OR — 2019-2021
• Multiple Intel Divisional and Departmental Recognition Awards — 2017-2020
• Ranked 2nd among all Masters (~110) students in CS Department at NIT Rourkela — 2017
• CET Merit Scholarship (Undergrad scholarship 2010-2014) — 2010
• National Talent Search Examination Scholar and CBSE X board rank-holder in State — 2008

## Technical Skills

• **Languages** C/C++, Python, MATLAB, OpenCL, System-C, System Verilog
• **Frameworks** Stratus-HLS, Qualcomm SNPE, Pytorch, llama.cpp, .NET
• **Tools** Quartus, Design Compiler, Vivado, Stratus, V-Tune, NSight

## Volunteering and Service

• Planning committee member of GradWCS, hosting monthly Faculty/Student luncheons — 2023 - present
• Co-chair of the Systems and Architecture session for the 19th CSL Student Conference — 2024
• Co-started and run weekly coffee meet-ups for women in comp. arch in CS and ECE — 2022 - present
• Named one of Top 50 Volunteers in Intel India, for service at Cancer Hospice Karunashraya — 2020
• Won an Intel Seed Grant and oversaw renovation of **nurses' dining hall** at Karunashraya — 2019
• Co-founded the student e-zine **CET Rising**, and served as **Chief Editor** during undergrad — 2013 - 2014