# MUHAMMAD HUSNAIN MUBARIK

Website: https://husnainmubarik.github.io/
Address: 405 W Park APT 03, Urbana, IL, 61801
Email: mubarik3@illinois.edu, Phone: +1-217-778-8784

## RESEARCH INTERESTS

Computer Architecture, Hardware Acceleration of Machine Learning Applications, Hardware for Graphics (Conventional, Neural and Generative), Hardware Acceleration of Generative AI, Printed Computer Systems, Emerging Technologies and Applications.

## EDUCATION

**University of Illinois at Urbana Champaign (UIUC), USA.**          *Jan 2019 - May 2024*
PhD - Electrical and Computer Engineering (ECE)
Advisor: Prof. Rakesh Kumar

**National University of Sciences and Technology (NUST), Pakistan.**          *Sep 2013 - July 2017*
Bachelors of Science in Electrical Engineering.
CGPA: 3.95/4.00, **summa cum laude**.

## PUBLICATION SUMMARY AND IMPACT

- 5× (ISCA, MICRO), 1× (DATE)
- 1× ISCA Retrospective (1996-2020) Selection
- 2× IEEE Top Picks Honorable Mention
- Dan Vivoli Endowed Fellowship
- Emerging Vision and Graphics Systems and Architectures (EVGA), workshop organization, co-located with ISCA 2024.

## RESEARCH AND WORK EXPERIENCE

**Advanced Rendering Research (ARR), AMD, USA.**

- Senior Member of Technical Staff (SMTS), Graphics and ML          *June 2024 -* **_Present_**
  Full-time, Manager: Takahiro Harada

**Graduate Research Intern, Intel, USA.**

- **Emerging Visual-AI Systems Research Lab**          *Sep 2023 - May 2024*
  Part-time, Managers: Nilesh Jain, Ravishankar Iyer.

- **Graphics Research Organization (GRO)**          *May 2023 - Aug 2023*
  *Full-time, Managers: Tobias Zirr, Anton Sochenov, Anton Kaplanyan.*

- **Next-Gen Architecture (NGA)**          August 2022 - May 2023
  *Part-time, Manager: Maxim Kazakov.*

- **Graphics Research Organization (GRO)**          June 2022 - August 2022
  *Full-time, Managers: Rama Harihara, Anton Kaplanyan.*

- **Cloud Systems Research Lab (CSR)**          Dec 2021 - May 2022
  *Part-time, Managers: Nilesh Jain, Ravishankar Iyer.*

- **Heterogeneous Platforms Lab (HPL)**                     May 2021 - Aug 2021
  *Full-time, Manager: Tanay Karnik.*

**Coordinated Science Laboratory (CSL), UIUC, USA**            *Jan 2019 - May 2024*
Research Assistant
Advisor: Prof. Rakesh Kumar

- Hardware Acceleration of Latent Diffusion Models (Undersubmission)

- ML accelerators for accelerating gaming and vision applications (Ongoing)

- Hardware Acceleration of Neural Graphics (**ISCA-50th, 2023**)

  - Neural graphics as a potent replacement for traditional rendering algorithms.

  - In this paper we addressed the question does NG need HW support?

  - Disparity between desired performance and the current state of the art (1.5x-55x).

  - Identification of input encoding and MLP kernels as performance bottlenecks.

  - Proposal of NGPC  a hardware solution to accelerate input encoding and MLP.

  - The NGPC achieves up to 58X end-to-end application-level performance improvement.

  - NGPC enables rendering of 4k Ultra HD resolution frames at 30 FPS for NeRF and 8k Ultra HD frames at 120 FPS for all other neural graphics applications.

- Understanding Interactions Between Chip Architecture and Uncertainties in Semiconductor Supply and Demand (Under Submission - Preprint available on arXiv.)

- Exploiting Short Application Lifetimes for Low Cost Hardware Encryption in Flexible Electronics (**DATE 2023**)

- Rethinking Programmable Earable Processors (**ISCA-49th, 2022**)

  - We first explors the hardware requirements for earable computing platforms like earphones, hearing aids, and smart glasses.

  - we propose EarBench, a suite of earable applications, to analyze performance gaps between earable computational needs and the capabilities of existing microprocessors.

  - Analysis reveals a performance gap of 13.54x-3.97x on average, with more complex microprocessors being energy-inefficient for earable applications.

  - EarBench applications are found to be dominated by a few DSP and ML-based kernels with significant computational similarity.

  - The proposed solution is SpEaC, a reconfigurable spatial architecture optimized for energy-efficient execution of earable kernels at the cost of generality.

  - SpEaC outperforms modeled programmable cores (M4, M7, A53, HiFi4 DSP) by up to 99.3x, and offers substantial energy efficiency benefits, outperforming a low power Mali T628 MP6 GPU by 15.7x - 1087x.

- Model-specific Design of Deeply-Embedded Tiny Neural Network Accelerators (Ongoing).

  - Current edge NN accelerator hardware designs tend to be model-agnostic and are, therefore, over designed and expensive for a specific model.

  - For a suite of tiny neural networks, we show that customizing hardware to model structure can provide $101\times$ benefits in terms of inference throughput and $69\times$ benefits in terms of energy

efficiency over Eyeriss. Additional 4.3× and 2.4× benefits are achievable in terms of area and energy efficiency respectively if model data is also known during design time.

- Printed Machine Learning Classifiers (**MICRO-53rd, 2020**)

  – Printed electronics meet cost, conformity, and non-toxicity needs unfulfilled by silicon-based systems in many applications.

  – We first explore the accuracy and cost of classification algorithms for two printed technologies, EGT and CNT-TFT.

  – Our EGT-bespoke Decision Tree and SVM classifiers exhibit 4x, 75x, 48x and 1.4x, 12.7x, 12.8x improvements in delay, power, and area, respectively, over conventional systems.

  – Our lookup-based classifiers enhance area and power efficiency by 1.93x and 1.65x with a 50% delay overhead for EGT Decision Trees. SVMs present a 40% delay cut and 8% and 1% area and power improvements.

  – EGT analog Decision Trees and SVMs surpass their digital equivalents in area and power by 437x, 27x and 490x, 1212x, respectively, but are 1.63x slower.

  – We presented several functional printed prototypes.

- Printed Microprocessors (**ISCA-47th, 2020.**)

  – We first explore the design space for microprocessors implemented in printing technologies, also devising standard cell libraries for these technologies.

  – Our design space exploration of printed microprocessor architectures across various parameters reveals our best cores surpass pre-existing ones by at least an order of magnitude in power and area.

  – Introducing printed-specific optimizations and Program-specific ISA (TP-ISA), we achieve up to 4.18x and 1.93x improvements in power and area, respectively. A Cross point-based instruction ROM outperforms a RAM-based design by 5.77x, 16.8x, and 2.42x in power, area, and delay, respectively.

**National Electronics Complex of Pakistan (NECOP), Pakistan**          *Jul 2017 - Oct 2018*
RF and Microwave Design Engineer - Assistant Manager

- Designed and Developed IMF Receiver, Antennas for Phased Array Radar, Branched Line Hybrid Coupler, Power Dividers and Band Pass Filters for X-Band Applications.

**Research Internee at MERL - NUST, Pakistan**          *June 2016- July2017*

- Designed and Developed a "Low Cost Short Range Frequency Modulated Continuous Wave Radar (FMCW) Radar". (**Second Position in Rector's Gold Medal Competition - NUST.**)

## PUBLICATIONS

1 **M. Mubarik**, V Chua, J. Nilesh, R.Kumar "Hardware Acceleration of Latent Diffusion Models" - Undersubmission, MICRO 2024.

2 N. Bleier **M. Mubarik**, G. Swanson, R. Kumar "Space Microdatacenters" **MICRO-56th, 2023, IEEE Micro Top Picks - Honorable Mention 2024.**

3 **M. Mubarik**, R. Kanungo, T. Zirr, R.Kumar "Hardware Acceleration of Neural Graphics" **ISCA-50th, 2023.**

4 N. Bleier, **M. Mubarik**, S. Balaji, F. Rodriguez, A. Sou, S. White and R. Kumar, "Exploiting Short Application Lifetimes for Low Cost Hardware Encryption in Flexible Electronics," **DATE 2023.**

5 N. Bleier, **M. Mubarik**, S. Chakraborty, S. Kishore, R. Kumar, "Rethinking Programmable Earable Processors," **ISCA-49th, 2022.**

6 **M. Mubarik**, D. Weller, N. Bleier, M. Tomei, J. Aghassi-Hagmann, M. Tahoori, R. Kumar, "Printed Machine Learning Classifiers," **MICRO-53rd 2020, IEEE Micro Top Picks - Honorable Mention 2021.**

7 N. Bleier*, **M. Mubarik**\*, F. Rasheed*, J. Aghassi-Hagmann, M. Tahoori, R. Kumar, "Printed Microprocessors," **ISCA-47th, 2020.** [* Co-First Authors, Listed in Alphabetical Order] **Selected for the retrospective of the years 1996 through 2020 on the 50th anniversary of ISCA.**

8 R. Kanungo, S. Siva, N. Bleier, **M. Mubarik**, L. Varshney, R. Kumar "Understanding Interactions Between Chip Architecture and Uncertainties in Semiconductor Supply and Demand" - Under Submission, preprint arXiv:2305.11059, 2023.

## TALKS

- Delivered a guest lecture on "Tensor Processing Units" at Intel Labs, April 2024.

- Delivered a guest lecture on "Accelerators/Specialization/Emerging Architectures" in ECE 411, Spring 2024.

- Delivered a guest lecture on "GPU Architecture and Programming" in ECE 411, Spring 2024.

- Gave a talk titled "Hardware Acceleration Opportunities in Generative AI Based Graphics" at Intel [2024].

- Presented "Hardware Acceleration of Neural Graphics" paper at ISCA-50th, 2023.

- Gave a talk on "Neural Graphics: An Architectures Perspective" at ECE498SJP: Accelerator Architectures class at UIUC [2023].

- Gave a talk titled "Hardware Acceleration Opportunities in Neural Radiance Fields" at Intel [2022].

- Presented "Printed Machine Learning Classifiers" paper at MICRO-53rd, 2020.

- Gave a talk on "Hardware for Deep Learning" at ECE511: Advanced Computer Architecture class at UIUC [2021].

## SELECT ACADEMIC ACHIEVEMENTS AND AWARDS

🏆 **Dan Vivoli Endowed Fellowship (NVIDIA) for the 2024-2025 academic year.**

🏆 **IEEE Micro Top Picks - Honorable Mention, 2024.**

- Our paper "Space Microdatacenters" (MICRO-56th, 2023) ranks among the Top 24 papers published in computer architecture venues in 2023.

🏆 **ISCA-50th 25 year retrospective (1996-2020), 2023.**

- Selected for the retrospective of the years 1996 through 2020 on the 50th anniversary of ISCA for our paper "Printed Microprocessors" (ISCA-47th, 2020).

🏆 **Machine Learning and Systems Rising Star, 2023.**

🏆 **ECE-UIUC travel grant for attending ISCA-2023.**

🏆 **IEEE Micro Top Picks - Honorable Mention, 2021.**

- Our paper "Printed Machine Learning Classifiers" (MICRO-53rd, 2020) ranks among the Top 24 papers published in computer architecture venues in 2020.

🏆 **Prime Minister's Silver Medal**, NUST-CEME, 2018.

  – 2nd highest CGPA (3.95/4.00) in the Electrical Engineering Department at NUST-CEME.

## TECHNICAL SKILLS

- Programming languages: C++, Python, Verilog, SystemVerilog, CUDA, GLSL, Scala, SystemC, Tcl, Bash, Makefile

- Software: PyTorch, TensorFlow, TensorRT, Onnx-Runtime, Caffee, Keras, Synopsys-VCS, MATLAB, LATEX, Vim

- Frameworks: CUDA, Nvidia Nsight (Compute/Systems/Graphics), Vulkan, OpenGL, CHISEL, gem5.

## PROFESSIONAL SERVICES AND ACTIVITIES

- Workshop Organization:

  – Emerging Vision and Graphics Systems and Architectures (EVGA), co-located with ISCA 2024.

- Member, Computer Architecture Student Association (CASA) Jan 2021 - Present

- Reviewed papers for ISCA, MICRO, HPCA and ASPLOS.

## TEACHING AND MENTORING EXPERIENCE

- Mentored student projects in ECE 508 (Manycore Parallel Algorithms, Spring 2023) at UIUC.

- Frank Cai                                                              4th-year Undergraduate Student
  Project titled "Hardware Acceleration Opportunities for Speculative Decoding to Accelerate LLMs."
  Secured admission at UIUC's Graduate College.

- Ramakrishna Kanungo                                                          2nd year Masters Student
  Hardware Accelerator for Neural Radiance Field Inference. Secured industry job in load-store unit design.

- Abigail Wezelis                                                              2nd year Masters Student
  Design space exploration of data specific architectures for accelerating DNNs. Secured industry job in system design.

- Srijan Chakraborty                                                          2nd year Masters Student
  Synthesis and performance/area/power analysis of very large RTL designs.

## SELECT COURSES

📙 ECE498SJP: Accelerator Architectures (Sanjay Patel) - Spring 2023.

📙 CS534: Advanced Topics in Computer Architecture (Sarita Adve) - Spring 2022.

📙 ECE598JH: Advanced Memory and Storage Systems (Jian Huang) - Fall 2021.

📙 CS598DHK: 3D Vision (Derek Hoiem) - Fall 2021.

📙 CS498ME Architecture for Mobile and Edge Computing (Saugata Ghose) - Spring 2021.

📙 CS598LCE Languages and Compilers for Edge Computing (Vikram S. Adve) - Spring 2021.

📙 ECE598NSG: Deep Learning in Hardware (Naresh Shanbhag) - Fall 2020.

📙 ECE524/CS563: Advanced Computer Security (Adam Bates) - Spring 2020.

- CS598SVA: App-Cust Heterogeneous Systems (Sarita Adve) - Spring 2020.
- CS598JT: Energy Efficient Computer Architecture (Josep Torrellas) - Fall 2019.
- ECE511: Advanced Computer Architecture (Rakesh Kumar) - Spring 2019.
- CS533: Parallel Computer Architecture (Josep Torrellas) - Spring 2019.