Ubaid Bakhtiar, Ph.D. Candidate

Advisor: Bahar Asgari
University of Maryland, College Park

ubaidb@umd.edu
/ubaidb
/in/ubaid-bakhtiar/
/UbaidHunts
/UbaidBakhtiar
+1(301)768-5256

Resume

RESEARCH STATEMENT. General-purpose processors and specialized accelerators often fail to achieve the theoretical gains offered by sparsity due to underutilized hardware resources. To close the gap between peak and achievable performance, my research focuses on developing novel data scheduling, multi-tenancy, and dynamic reconfiguration techniques and integrate into sparse applications accelerators to mitigate the high resource underutilization and consequently, fully exploit the benefits of sparsity.

EDUCATION.

Ph.D. Electrical and Computer Engineering | University of Maryland-College Park | Fall, 2022- Present

BSc. Electrical Engineering | Lahore University of Management Sciences (LUMS), Pakistan | 2018-2022

PUBLICATIONS.

Chasoň: Supporting Cross HBM Channel Data Migration to Enable Efficient Sparse Algebraic Acceleration | Ubaid Bakhtiar, Amirmahdi Namjoo Bahar Asgari | MICRO 2025 | Acceptance Rate ~21%

A novel HBM-based sparse algebraic accelerator that improve the resource utilization by migrating the non-zero value across HBM channels. Chasoň achieves up to ~6× , ~20×, ~11× and ~3× speedup over state-of-the-art SpMV accelerator Serpens, Nvidia RTX 4090, Nvidia RTX 6000Ada and Intel Core-i9-11980HK respectively.

Acamar: A Dynamically Reconfigurable Scientific Computing Accelerator for Robust Convergence and Minimal Resource Underutilization | Ubaid Bakhtiar, Helya Hosseini, Bahar Asgari | MICRO 2024 | Acceptance Rate ~22.7%

Acamar is the first scientific computing accelerator that leverages partial dynamic reconfiguration of FPGA. It provides robust convergence across a wide range of workloads while enhancing resource utilization for sparse kernels—features that were lacking in the previous scientific computing accelerator designs.

Pipirima: Predicting Patterns in Sparsity to Accelerate Matrix Algebra | Ubaid Bakhtiar, Donghyeon Joo, Bahar Asgari | DAC 2025 | Acceptance Rate ~23%

Pipirima predicts the matrix sparsity pattern using lightweight predictor and leverage it to accelerate SpMM and SpMV kernels. It shows latency speed up of up to ~20× over state -of-the-art accelerators; ExTensor and Tensaurus.

Segin: Synergistically Enabling Fine-Grained Multi-Tenant and Resource Optimized SpMV | Helya Hosseini, Ubaid Bakhtiar, Donghyeon Joo, Bahar Asgari | IEEE-CAL 2025 | Acceptance Rate ~20%

Segin leverages a novel fine-grained multi-tenancy approach to allow multiple SpMV operations to be executed simultaneously on a single hardware with minimal modifications, enhancing resource utilization and improving throughput by 1.92×.

PROJECTS.

Lyra: Mapping Unstructured Sparse LLMs to FPGAs Using Bitmap-Based Dynamic Matrix Partitioning | Ubaid Bakhtiar, Sanjali Yadav, Donghyeon Joo, Ramyad Hadidi, Bahar Asgari | Under Review for HPCA 2026

An FPGA-based streaming accelerator designed specifically to efficiently support unstructured sparse LLMs while maintaining small power consumption. Lyra achieves a peak throughput of 99.8 GFLOP/s and delivers up to 4.6 GFLOP/s per watt across Llama2-13B, Llama2-7B, OPT-125M and OPT-2.7B models. Lyra exhibits 2× better resource utilization than DFX and offers up to 9× better performance-per-watt than Nvidia RTX Ada 6000 and 4080.

Lepus: Leveraging Sparsity to Upend Its Challenges | Helya Hosseini, Ubaid Bakhtiar, Donghyeon Joo, Bahar Asgari | Under Review for HPCA 2026

A lightweight SpMV accelerator that addresses hardware underutilization through multi-threading and adaptive resource management. Lepus improves resource utilization by up to 97.65% and achieves speedup of up to 10 × compared to state-of-the-art SpMV accelerators, Serpens while significantly reducing the number of costly reconfigurations.

RELEVANT EXPERIENCE.

Graduate Research Assistant | Computer Architecture and Systems Lab | University of Maryland | May, 2023-Present

- Research Advisor: Dr. Bahar Asgari
- Research Area: Computer Architecture and Domain Specific Designs
- Crafting domain-specific architecture designs to tackle computational challenges in sparse applications and devising methods to enhance their performance as well as simulation and prototyping on modern architectures, i.e. CPUs, GPUs, FPGAs

RELEVANT COURSEWORK.

Programming Languages and Computer Architecture | Domain Specific Architecture | Digital Computer Design | Compilers and Optimizations | Systems for Machine Learning

University of Maryland, College Park

Computer Architecture | Digital System Design | Embedded Systems | VLSI Design | Machine Learning

Lahore University of Management Sciences

TECHNICAL SKILLS.

Languages:	C, C++, Verilog, Python, OpenCL, MIPS/RISC Assembly, MATLAB
Hardware Platforms:	AMD Alveo Accelerator Cards, AMD ZYNQ SoC, RISC, CPU, GPU
Frameworks:	High-level synthesis, Xilinx Vitis, Xilinx Runtime TAPA, Rapidstream, Intel MKL, Nvidia cuSparse
Simulations:	In-house cycle-accurate simulator, CACTI, Synopsys DC, GPGPU-SIM, Nvidia Nsight Compute
Deployment:	Linux CLI, Docker, Jupyter, Git
Topics:	HW/SW Co-design, Hardware for generative AI, Scientific computing accelerator, AI/ML

AWARDS AND HONORS.

- Winner 2-Minute Video Contest, Student Young Fellow (DAC 2025)
- Student Travel Grant, ACM/IEEE 57th International Symposium on Microarchitecture (MICRO 2024)
- Summer Research Fellowship, University of Maryland-College Park (Summer 2023)
- Dean's Honor List Award, LUMS (2019-20)

PROFESSIONAL SERVICES.

- Student Volunteer @ SPICE: A Workshop Co-Located with International Symposium on Microarchitecture 2025
- Graduate Research Assistant, Computer Architecture & Systems Lab, University of Maryland-College Park
- Graduate Teaching Assistant, Advance Digital Computer Design, University of Maryland-College Park (Fall 2023)
- Graduate Teaching Assistant, Advance Digital Computer Design, University of Maryland-College Park (Fall 2022, Spring 2023)
- Teaching Assistant Feedback Control Systems, LUMS (Spring 2022)
- Teaching Assistant Electromagnetic Fields and Waves, LUMS (Fall 2021)
- Teaching Assistant Engineering Models, LUMS (Fall 2021)
- Teaching Assistant Circuits-II, LUMS (Spring 2021)