

Research engineer specializing in the co-design of machine learning algorithms and hardware accelerators. Expertise in device-circuit-ML integration and hands-on hardware prototyping.

Education

- 08/2022–
Present **Ph.D. in Electrical Engineering**, *Pennsylvania State University*, University Park, PA, USA, *CGPA: 4.00/4.00* | Advisor: Dr. Abhronil Sengupta | Expected Graduation: 05/2027
- 08/2022–
08/2024 **M.S. in Electrical Engineering**, *Pennsylvania State University*, University Park, PA, USA, *CGPA: 4.00/4.00* | Thesis: Neuromorphic Computing for Lifelong Learning
- 02/2015–
04/2019 **B.Sc. in Electrical and Electronic Engineering**, *Bangladesh University of Engineering & Technology (BUET)*, Dhaka, Bangladesh, *CGPA: 3.81/4.00*

Academic Research and Teaching Experience

- 08/2022–
Present **Graduate Research Assistant**, *Penn State*, University Park, PA
- Leading the design of a bit-serial ML accelerator for Transformer workloads, implementing CIM/PIM architectures with custom dataflow optimization for energy-efficient inference.
 - Pioneered device-circuit-ML co-design integrating spintronic NVM devices with hardware-aware training, demonstrating end-to-end flow from device fabrication to system-level deployment.
 - Developed brain-inspired ML systems including Astromorphic Transformer and RMAAT, with focus on hardware-aware model training and crossbar-aware quantization for CIM arrays.
- 08/2024–
05/2025 **Graduate Teaching Assistant**, *Penn State*, State College, PA
- Taught Cadence Virtuoso (schematic/layout), PDK usage, DRC/LVS, and analog/digital design flows; created hands-on lab modules and guided tool/debug workflows.
 - Supervised Capstone projects for 90+ undergraduate students across communications, electronics, and firmware: supported Raspberry Pi/Arduino development, software-defined radio experiments, PCB design, and system integration end-to-end.
- 02/2021–
08/2022 **Lecturer**, *University of Liberal Arts Bangladesh (ULAB)*, Dhaka, Bangladesh
- Taught undergraduate courses: Solid State Devices, Digital Circuit Design, Semiconductor Device Physics.

Technical Skills

- ML Systems & Accel.** Hardware-aware ML; ML accelerator/ASIC design; neural network accelerator architecture; AI inference engine design; GPU architecture fundamentals; long-context sequence modeling; device-circuit-ML co-design; Transformer/LLM acceleration; LLM serving (TPOT/TTFT, PD/AFD); Systolic Arrays; model compression (PTQ/QAT, pruning, distillation); mixed-precision (FP16/INT8); quantization error analysis; dataflow optimization; PIM
- AI Frameworks** PyTorch, TensorFlow, JAX, ONNX, TensorRT; model development and debugging; inference performance tuning; experiment tracking (W&B, Matplotlib); data tooling (Pandas, NumPy, Jupyter)
- Hardware & EDA** Cadence Virtuoso; Spectre; HSPICE; TCAD; COMSOL; ModelSim; Quartus; Vivado; Synopsys (Design Compiler, PrimeTime, VCS); ADC/DAC; mixed-signal design; CIM systems
- Digital/ASIC** RTL design; CMOS/FinFET circuit design; SRAM design; sense amplifier design; FPGA prototyping; synthesis; P&R; functional verification; DFT; static timing analysis; timing closure
- Programming** Python; CUDA; C/C++; MATLAB; Verilog; Shell/Bash; Linux/Unix; Git; Docker
- DTCO & STCO** Design-technology and system-technology co-optimization; cross-layer PPA; process-aware design; system-architecture co-design
- AI Dev.** Generative and agentic AI tools (Cursor, Copilot, etc.) for research, design, and code development
- Collaboration** Cross-functional collaboration; mentoring; technical writing; Microsoft Office; Google Workspace

Industrial Experience

- 05/2026–**Machine Learning Engineer Intern, Micron Technology, Richardson, TX**
 07/2026
- Model Non-Volatile Memory (NVM) device characteristics (technology-agnostic) to evaluate speedup and energy savings for memory-centric ML inference.
 - Characterize LLM serving performance (TPOT, TTFT) under Prefill-Decode and Attention-FFN disaggregation schemes (PD, AFD) for hyperscale inference deployment.
 - Investigate quantization error and CIM analog error mitigation techniques to preserve model accuracy under analog compute noise and limited ADC precision.
- 05/2025–**Graduate Technical Intern, Intel Corporation, Hillsboro, OR**
 07/2025
- Developed AI-driven models for process-circuit co-optimization in ML accelerator hardware, focusing on yield prediction and performance enhancement.
 - Applied machine learning to optimize thin film deposition parameters affecting AI hardware circuit characteristics and system performance.
 - Utilized statistical analysis (JMP) and AI models to predict process impact on circuit electrical parameters and device reliability for ML hardware.

Research Projects

- CIM-NVM Analysis for LLM Inference** Lead Student Researcher, 05/2026–07/2026. Analyzed ML workload performance on Compute-in-Memory architectures with Non-Volatile Memory at Micron Technology, evaluating TPOT, TTFT, energy, and latency benefits over digital accelerators. Investigated LLM serving optimization under Prefill-Decode and Attention-FFN disaggregation (PD, AFD) schemes. Studied quantization error and CIM analog error mitigation for LLM inference.
- Astromorphic Transformer** Lead Student Researcher, 2022–2025. Developed a bioplausible transformer architecture leveraging neuron-astrocyte interactions to emulate self-attention mechanisms. Incorporated Hebbian and presynaptic plasticities with non-linearities and feedback, achieving superior accuracy and faster convergence on sentiment classification (IMDB), image classification (CIFAR-10), and language modeling (WikiText-2) tasks. See publication: [IEEE TCDS 2025].
- Neuromorphic Cybersecurity with Lifelong Learning** Lead Student Researcher, 2023–2025. Developed a Hierarchical Dynamic Spiking Neural Network (D-SNN) for Network Intrusion Detection Systems combining static SNN detection (94.3% accuracy) with adaptive dynamic SNN classification using GWR-inspired structural plasticity and novel Adaptive STDP (Ad-STDP) learning. Achieved 85.3% overall accuracy on UNSW-NB15 benchmark in semi-supervised lifelong learning, demonstrating superior adaptation to new attack types while mitigating catastrophic forgetting (5.3% improvement over baseline). Demonstrated high operational sparsity suitable for neuromorphic hardware deployment. See publication: [arXiv], [ICONS 2025].

Select Publications

- Md Zesun Ahmed Mia**, Malyaban Bal, and Abhronil Sengupta. "RMAAT: Astrocyte-Inspired Memory Compression and Replay for Efficient Long-Context Transformers". In: *International Conference on Learning Representations (ICLR)*. 2026. URL: <https://openreview.net/forum?id=sTkJdbVxsI>.
- Md Zesun Ahmed Mia**, Jiahui Duan, Kai Ni, and Abhronil Sengupta. "Trilinear Compute-in-Memory Architecture for Energy-Efficient Transformer Acceleration". In: *arXiv preprint arXiv:2604.07628* (2026).
- Md Zesun Ahmed Mia**, Malyaban Bal, Sen Lu, George M. Nishibuchi, Suhas Chelian, Srini Vasani, and Abhronil Sengupta. "Neuromorphic Cybersecurity with Semi-Supervised Lifelong Learning". In: *Proceedings of the International Conference on Neuromorphic Systems (ICONS '25)*. IEEE Press, 2025, pp. 235–238. DOI: 10.1109/ICONS69015.2025.00043.
- Md Zesun Ahmed Mia**, Malyaban Bal, and Abhronil Sengupta. "Delving deeper into astromorphic transformers". In: *IEEE Transactions on Cognitive and Developmental Systems* (2025).

Recognitions

- Melvin P. Bloom Memorial Outstanding Doctoral Research Award (2026)
- The Wormley Family Graduate Fellowship, Harry G. Miller Fellowships in Engineering (2025)
- Arthur Waynick Graduate Scholarship (2024), Milton and Albertha Langdon Memorial Fellowship (2023), Melvin P. Bloom Memorial Fellowship (2022)

Professional Affiliations

- Reviewer, NeurIPS (2026), IOP Physica Scripta (2026), IEEE MWSCAS (2026)
- Reviewer, IEEE TNNLS (2025), Design Automation Conference (DAC) (2025), IEEE MWSCAS (2025)
- Student Member, IEEE (2015-Present), Executive Member, EDS, IEEE BD (2021-2022)