Md Zesun Ahmed Mia

Curriculum Vitae

Curiosity drives me to seek new questions and create new knowledge. I believe progress in science comes from collaboration, open-mindedness, and the courage to explore beyond boundaries.

Education

08/2022–Present **Ph.D. in Electrical Engineering**, *Pennsylvania State University*, University Park, PA, USA, *CGPA: 4.00/4.00* | NeuroAI Lab | 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**, *NeuroAI Lab, Penn State*, State College, PA

- Leading research on hardware-algorithm co-design for efficient intelligence at scale, optimizing Transformer-based LLMs for near-memory and in-memory computing (CIM/PIM) architectures.
- Fabricated and characterized spintronic devices (e.g., hallbars, nanodots) using I-V and R-H loop measurements; utilized experimental data of emerging ferroelectric devices (e.g., FeFETs) to inform compact models and algorithm-hardware co-design.
- Architecting RMAAT (Astrocyte-inspired transformers), an adaptive memory compression and replay system reducing long-context sequence modeling overhead.
- Pioneered a hardware-aware training framework integrating hardware physics directly into system-level ML designs to co-optimize physical device constraints with ML accuracy.
- Developed adaptive Spike-Timing-Dependent Plasticity (STDP) learning rules for dynamic Spiking Neural Networks (SNNs) to enable semi-supervised lifelong learning in cybersecurity threat detection.

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*, Dhaka, Bangladesh

- Taught undergraduate courses: Solid State Devices, Digital Circuit Design, Semiconductor Device Physics.
- Developed lab modules and supervised projects on semiconductor devices and circuits.
- Supervised student projects on semiconductor devices and circuits.

02/2020–02/2021 **Lecturer**, *BUET*, Dhaka, Bangladesh, part-time

- Supervised labs (Digital Circuit Design).

Industrial Experience

05/2026–07/2026 **Machine Learning Engineer Intern**, *Micron Technology*, Richardson, TX

- Analyze ML workloads on Compute-in-Memory (CIM) architectures, quantifying energy, latency, and area trade-offs for next-generation AI accelerators.
- 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.

814-280-7244 • zesun.ahmed@psu.edu • zesun33.github.io
zesun-ahmed • zesun33

05/2025–07/2025 **Graduate Technical Intern**, *Intel Corporation*, Hillsboro, OR

- Developed predictive analysis frameworks using Artificial Intelligence and statistical modeling (JMP) to optimize recipe parameters and accelerate semiconductor process flows.
- Bridged domain gaps by integrating machine learning models with advanced material characterization to predict thin film deposition impact on technology node yields.
- Designed and executed Design of Experiments (DOE) evaluating first-of-its-kind process integration tools to establish robust process windows.

09/2020–07/2021 **R&D Engineer**, *SEMWAVES Ltd.*, London, UK, part-time

- Delivered a 50 kW hybrid renewable system (solar + smallscale hydro) for an offgrid Bangladeshi site, supporting reliable community power.
- Owned technical leadership and coordination (vendors, field teams, stakeholders); directed device selection, integration, QA/safety procedures, and optimization against load profiles and uptime targets.

Technical Skills

Core Expertise	Algorithm-hardware co-design (device-circuit-architecture-algorithm); Neuromorphic computing; AI hardware accelerators; DTCO & STCO; Transformer/LLM optimization; LLM serving analysis (TPOT/TTFT, PD/AFD disaggregation); in-memory computing (CIM/PIM); NVM for CIM architectures
ASIC/RTL & EDA	Logic synthesis, RTL design & verification (Verilog/SystemVerilog), ASIC design flow, FPGA prototyping; Cadence Virtuoso, Spectre, HSPICE, TCAD, Synopsys
Software & ML Sys.	Python, CUDA, C/C++, Shell/Bash; PyTorch, TensorFlow, JAX, ONNX, TensorRT; model compression (PTQ/QAT, pruning, distillation), hardware-aware ML, HPC, W&B, Matplotlib, Pandas
Device Physics	Spintronics, advanced NVM (FeFET, etc.), SRAM/DRAM design, compact modeling, Verilog-A, electrical characterization (DC/AC, IV/CV), NVM for CIM architectures
AI Dev.	Generative and agentic AI tools (Cursor, Copilot, etc.) for research, automation, and code development
Teaching & Comm	Scientific writing, peer review, research mentoring (15+ students), grant proposals, curriculum design, LaTeX

Research Interests

Neuromorphic Computing	Brain-inspired hardware, spiking neural networks (SNNs), event-driven sensing/compute, on-chip learning (STDP), crossbar synapses, oscillatory/phase-change/spintronic neurons, temporal coding, algorithm-device co-design, low-power edge intelligence
Machine Learning Hardware	DNN/GNN/Transformer accelerators, systolic and dataflow architectures, sparsity- and quantization-aware compute, mixed-signal MACs, compute-in-memory, near-memory compute, memory hierarchies, compilation and mapping, performance/energy modeling
Emerging Devices	Spintronics (STT/SOT-MRAM), FeFET/NCFET, RRAM/PCM memristors, 2D materials, analog programmability, endurance and variability
Semiconductor Process Integration	Device fabrication, FEOL/BEOL and BEOL-compatible memory, 3D monolithic integration, DTCO (design-technology co-optimization), compact modeling, variability and yield, process-device-circuit co-design
AI for Semiconductor	ML for process/device optimization, virtual metrology, yield prediction and defect classification, Bayesian optimization and surrogate modeling, inverse design, reinforcement learning for tool control

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. Pioneered a bioplausible transformer architecture leveraging neuron-astrocyte interactions to emulate self-attention mechanisms. Incorporated Hebbian and presynaptic plasticities with non-linearities and feedback dynamics, achieving superior accuracy and faster convergence on sentiment classification (IMDB), image classification (CIFAR-10), and language modeling (WikiText-2) tasks with improved perplexity compared to conventional transformers. Established the first scalable astromorphic framework for sequential tasks. See publication: [IEEE TCDS 2025].

TrilinearCIM: All-in-Memory Transformer Acceleration Lead Student Researcher, 2025–2026. Designed TrilinearCIM, a Double-Gate FeFET (DG-FeFET)-based Compute-in-Memory architecture that eliminates the core bottleneck of Transformer acceleration on CIM hardware: repeated runtime NVM reprogramming of dynamic attention operands (Q, K, V). Introduced a novel three-operand multiply-accumulate primitive ($Y = A \cdot B \cdot C$) enabled by DG-FeFET back-gate modulation, which maps static projection weights to the non-volatile top gate and applies dynamic token activations through the volatile back gate—executing the full attention dataflow in-memory without reprogramming the ferroelectric state. To our knowledge, the first architecture to perform complete Transformer attention computation exclusively in NVM cores without runtime reprogramming. Reduced global buffer capacity requirements by $\sim$3$\times$ by eliminating intermediate Q/K matrix storage. Evaluated on BERT-base (GLUE) and ViT-base (ImageNet, CIFAR-10/100): outperforms conventional FeFET CIM on 7 of 9 GLUE tasks, achieving up to 46.6% energy reduction and 20.4% latency improvement at 37.3% area overhead. See publication: [arXiv:2604.07628].

RMAAT: Bio-Inspired Long-Context Transformers Lead Student Researcher, 2024–2026. Developed RMAAT (Recurrent Memory Augmented Astromorphic Transformer), a recurrent transformer architecture for efficient long-context sequence processing. Integrated astrocyte-inspired memory compression via adaptive retention factors and linear-complexity attention from astrocyte short-term plasticity. Achieved competitive accuracy on Long Range Arena (LRA) benchmark with up to 1.73$\times$ training speedup and 4.4$\times$ memory reduction (3.4 GB vs. 15.0 GB) compared to standard recurrent transformers. Introduced AMRB (Astrocytic Memory Replay Backpropagation) training algorithm for memory-efficient recurrent learning. Accepted at ICLR 2026. See publication: [OpenReview].

Neuromorphic Cybersecurity with Lifelong Learning Lead Student Researcher, 2023–2025. Architected a Hierarchical Dynamic Spiking Neural Network (D-SNN) for Network Intrusion Detection Systems (NIDS) 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 scenarios, demonstrating superior adaptation to new attack types while mitigating catastrophic forgetting (5.3% improvement over static baseline). Demonstrated high operational sparsity ($\sim$0.0008 spike rate) suitable for neuromorphic hardware deployment. See publication: [arXiv], [ICONS 2025].

Spintronic Device-Based Memory Systems Lead Student Researcher, 2023–Present. Led end-to-end fabrication of spintronic memory arrays using e-beam lithography, sputtering deposition, and lift-off patterning. Performed comprehensive electrical characterization (I-V, R-H loops, retention, endurance) and extracted device parameters for compact model development. Advanced non-volatile memory technologies for neuromorphic computing and in-memory processing applications.

ECG Rats Processor Lead Student Researcher, 2023. Developed a custom digital signal processing (DSP) processor, "ECG Rats," for real-time ECG signal analysis. Implemented core modules for filtering, feature extraction, and arrhythmia detection. Achieved efficient processing with low latency and power consumption, suitable for wearable health monitoring applications.

Publications

[1] **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.

[2] **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). URL: https://arxiv.org/abs/2604.07628.

[3] **Md Zesun Ahmed Mia** and Abhronil Sengupta. "Leveraging Intrinsic Physics of Spintronic Devices for Neuromorphic Computing". In: *2026 IEEE 69th International Midwest Symposium on Circuits and Systems (MWSCAS)*. IEEE. 2026. URL: https://www.mwscas2026.org/.

[4] Arnob Saha, Bibhas Manna, Nikhil Kotikalapudi, **Md Zesun Ahmed Mia**, Rahul Kumar, Madhavan Swaminathan, and Abhronil Sengupta. "Bayesian Optimization of Crossbar-Based Compute-In-Memory System Design for Efficient DNN Inference". In:

IEEE Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS). 2026. URL: https://arxiv.org/abs/2605.08461.

[5] Anika Tabassum Meem, Muntasir Hossain Nadid, and **Md Zesun Ahmed Mia**. "Energy-aware spike budgeting for continual learning in spiking neural networks for neuromorphic vision". In: *Neuromorphic Computing and Engineering* 6.3 (2026), p. 034004. DOI: 10.1088/2634-4386/ae8627.

[6] **Md Zesun Ahmed Mia**, Malyaban Bal, and Abhronil Sengupta. "Delving deeper into astromorphic transformers". In: *IEEE Transactions on Cognitive and Developmental Systems* (2025).

[7] **Md Zesun Ahmed Mia**, Malyaban Bal, Sen Lu, George M. Nishibuchi, Suhas Chelian, Srini Vasan, 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.

[8] Arnob Saha, **Md Zesun Ahmed Mia**, Jiahui Duan, Kai Ni, and Abhronil Sengupta. "Toward Variation-Tolerant Ferroelectric Neural Computing: Special Session Paper". In: *2025 IEEE 68th International Midwest Symposium on Circuits and Systems (MWSCAS)*. IEEE. 2025, pp. 583–587.

[9] **Md Zesun Ahmed Mia** and Kazi Toukir Ahmed. "Ultra Low Cost, Low Power, High Speed Electronic Braille Device for Visually Impaired People". In: *2024 International Conference on Advances in Computing, Communication, Electrical, and Smart Systems (iCACCESS)*. IEEE. 2024, pp. 1–6.

[10] **Md Zesun Ahmed Mia**, Nazmus Saadat As-Saquib, Mehedi Hasan Himel, Mehedi Hossen Limon, and Samia Subrina. "Impact of Doping and Defects on Thermal Transport of Monolayer GaN Nanoribbons: A Molecular Dynamics Simulation Study". In: *2024 13th International Conference on Electrical and Computer Engineering (ICECE)*. IEEE. 2024, pp. 685–690.

[11] Tao Zhang, Mingjie Hu, **Md Zesun Ahmed Mia**, Hao Zhang, Wei Mao, Katsuyuki Fukutani, Hiroyuki Matsuzaki, Lingzhi Wen, Cong Wang, Hongbo Zhao, et al. "Self-sensitizable neuromorphic device based on adaptive hydrogen gradient". In: *Matter* 7.5 (2024), pp. 1799–1816.

[12] KM Ashraful Hoque Fahim, Md Jubair Hasan Khalid, **Md Zesun Ahmed Mia**, and Mirza Rasheduzzaman. "Study of 3-nm Cylindrical GAAFETs with Variations in High-k Dielectric Gate-oxide Materials". In: *2022 IEEE Symposium on Industrial Electronics & Applications (ISIEA)*. IEEE. 2022, pp. 1–5.

[13] Md Moinul Islam, Monjurul Haque, Saiful Islam, **Md Zesun Ahmed Mia**, and SMA Mohaiminur Rahman. "DCNN-LSTM based audio classification combining multiple feature engineering and data augmentation techniques". In: *Intelligent Computing & Optimization: Proceedings of the 4th International Conference on Intelligent Computing and Optimization 2021 (ICO2021) 3*. Springer. 2022, pp. 227–236.

[14] **Md Zesun Ahmed Mia**, Md Moinul Islam, Monjurul Haque, Saiful Islam, and SMA Mohaiminur Rahman. "Irfd: A feature engineering based ensemble classification for detecting electricity fraud in traditional meters". In: *2021 24th International Conference on Computer and Information Technology (ICCIT)*. IEEE. 2021, pp. 1–6.

Recognitions and Awards

- Melvin P. Bloom Memorial Outstanding Doctoral Research Award (2026)
- Harry G. Miller Fellowships in Engineering (2025)
- The Wormley Family Graduate Fellowship (2025)
- Arthur Waynick Graduate Scholarship (2024)
- Milton and Albertha Langdon Memorial Fellowship (2023)
- Melvin P. Bloom Memorial Fellowship (2022)
- Undergrad Dean's List (2016–2018)
- Honorable Mention, Notre Dame College, Dhaka (2014)

Professional Service and Affiliations

- Reviewer, Conference on Neural Information Processing Systems (NeurIPS) (2026)
- Reviewer, IOP Physica Scripta (2026)
- Reviewer, IEEE International Midwest Symposium on Circuits and Systems (MWSCAS) (2026)
- Reviewer, IEEE Transactions on Neural Networks and Learning Systems (TNNLS) (2025)
- Reviewer, Design Automation Conference (DAC) (2025)
- Reviewer, IEEE International Midwest Symposium on Circuits and Systems (MWSCAS) (2025)
- Member, Graduate Student Advisory Committee, Penn State (2024-2026)
- Student Member, Institute of Electrical and Electronics Engineers (IEEE) (2015-Present)
- Executive Member, IEEE Electron Devices Society (EDS) Bangladesh Section (2021-2022)

Outreach and Leadership

Mentorship	Mentored junior graduate students through neuroscience-inspired artificial intelligence research ideation and hands-on cleanroom fabrication workflows; supervised over 90 undergraduate students in senior capstone projects at Penn State, guiding end-to-end hardware/software integration and PCB design.
Research Leadership	Core contributor within the NeuroAI Lab for multi-disciplinary federally funded research (NSF EFRI, Army Basic Scientific Research), driving milestones across cleanroom fabrication, circuit design, and AI development.
Peer Review	Active technical reviewer for flagship journals and conferences including NeurIPS (2026), IOP Physica Scripta (2026), IEEE MWSCAS (2026), IEEE TNNLS (2025), Design Automation Conference (DAC) (2025), and IEEE MWSCAS (2025).
Academic Service	Serving on the Graduate Student Advisory Committee at Penn State (2024–2026). Previously served as a university Lecturer (ULAB, BUET), designing curricula and lab modules for foundational electronics and device physics.

References

Dr. Abhronil Sengupta	Associate Professor, Penn State University, Email: sengupta@psu.edu
Dr. Swaroop Ghosh	Professor, Penn State University, Email: szg212@psu.edu