

GOBINDA SAHA

2450 Sycamore Ln Apt. 36A, West Lafayette, Indiana, 47906

☎ 765-409-1042 ✉ gobinda.eee.buet@gmail.com  [Linkedin](#)  [GitHub](#)  [Website](#)

Objective

Seeking a full-time research/engineering role in AI/ML algorithm/software development. Bringing 6 years of research experience and 10 ML algorithm design publications, along with expertise in Python and PyTorch, I aim to advancing AI through continuous innovation in vision, language, and multimodal foundation models.

Education

Purdue University, West Lafayette, Indiana, USA

PhD in Electrical and Computer Engineering (Machine Learning)

Aug. 2017 – Dec. 2023 (expected)

Bangladesh University of Engineering and Technology (BUET), Bangladesh

Bachelor of Science (B.Sc.) & Master of Science (M.Sc.) in Electrical and Electronic Engineering

Jan. 2008 – Aug. 2015

Skills

Algorithm Design: ML Optimization, Continual Learning, Online & Few-shot Learning, Decentralized & Federated Learning

Applied Research: Computer Vision (Object Classification, Localization, Detection, Image Synthesis, Data Privacy), Reinforcement Learning (Model-free RL), Natural Language Processing/Language Models (Text Generation, Classification)

Programming Languages: Python, C, C++, Julia

Frameworks/Technologies: PyTorch, TensorFlow, Hugging Face (transformers, peft, trl, accelerate), AWS (EC2)

Research Experience

Nanoelectronics Research Laboratory & Center for Brain-Inspired Computing, Purdue University

Graduate Research Assistant | Advisor : Professor Kaushik Roy

Aug. 2017 – Present

Skills: Python, PyTorch, torchvision, scikit-learn, panda, OpenCV, gym, CUDA

- **Optimization Algorithms for Model Adaptation & Continual Learning (CL) without Forgetting**

- Invented Gradient Projection Memory (GPM)[3] based fast orthogonal gradient descent algorithm to achieve near zero forgetting on large-scale sequential object classification tasks
- Proposed adaptive Proximal Policy Optimization (PPO) method by introducing Scaled Gradient Projection (SGP)[1] and attained 12% more reward on continual reinforcement learning (Atari games) tasks than SoTA baselines
- Designed forgetting-free CL algorithm - SPACE[4] using PCA-driven network pruning and growth providing up to 5x energy efficiency during inference

- **Explainable Learning and Meta-Learning Algorithms for Fast Online Continual Learning**

- Invented memory efficient experience replay method - (EPR)[2] using saliency-based object localization and detection tools (Grad-CAM) to attain up to 5% accuracy gain over SoTA methods on image classification tasks
- Proposed episodic memory-free bi-level/meta optimization algorithm to obtain 4% higher accuracy with fast, few-shot online learning capability on computer vision benchmarks [8]

- **Communication Efficient and Data Private Decentralized Learning Algorithms**

- Collaborated in a team of 4 to design CoDeC[9] - a decentralized CL algorithm to achieve SoTA accuracy on image/bio-medical image classification tasks with up to 4.8x reduced data private communication among 16 agents
- Achieved 4% accuracy improvement by integrating knowledge distillation and Out-of-Distribution (OOD) detection in decentralized optimization setting with non-IID data distribution over 32 agents [10]

Internship Experience

GlobalFoundries, Santa Clara, CA, USA

ML Research Intern, Memory Solution Team

Jun. 2019 – Aug. 2019

Skills: Python, PyTorch, CUDA, HSPICE, Cadence, AWS (Amazon Elastic Compute Cloud)

- **Software Framework for Hardware-Algorithm Co-design for Deep Learning (DL) Inference Acceleration**

- Proposed a 10T bit-cell memory primitive for accelerating binary neural network inference as part of a team of 6
- Developed and published [5] Python-HSPICE based software framework to measure in-memory computing (IMC) array performance under nonidealities and process variations enabling accelerated chip design and testing for DL

Projects

Anomaly or Privacy-Driven Information Unlearning from Neural Network | Skill: *PyTorch* Oct. 2023

- Developed an unlearning algorithm to remove requested samples/classes from trained model and demonstrated 99% unlearning accuracy on ImageNet with a vision transformer model

Parameter-efficient Fine-tuning of Large Language Models (LLMs) | Skills: *transformers, peft, trl* Sept. 2023

- Performed supervised fine-tuning (SFT) on open pre-trained transformer (OPT) language model using low-rank adaptation (LoRA) to improve performance of LIMA question answering and FinGPT financial sentiment analysis tasks
- Implemented reinforcement learning from human feedback (RLHF) on OPT model with quantized LoRA and DeBERTa reward model to generate human-aligned responses on OpenOrca dataset

Semantic Image Retrieval with Multimodal Models | Skills: *transformers, panda* May 2022

- Designed semantic image retrieval system by contrastive pre-training of multimodal CLIP model on Flickr 8k dataset

Interior Point Linear Program (IPLP) Solver | Skill: *Julia* May 2020

- Implemented IPLP solver in Julia based on Mehrotra's Predictor-Corrector algorithm with Cholesky factorization

Hardware Accelerator Design for Neural Network Inference on FPGA | Skills: *Verilog, Quartus* Dec. 2018

- Built a hardware accelerator for MLP on Altera DE2-115 FPGA board to reduce inference time for MNIST classification

Photonic IMC Primitives Design for Spiking Neural Networks (SNNs) | Skills: *COMSOL, MATLAB* Jan. 2018

- Designed spiking neuron and synaptic arrays based on GST (phase change material) embedded optical micro-ring resonators to realize energy-efficient and fast photonic computing primitives [6,7]
- Developed a device-circuit-algorithm co-design framework to evaluate their performance as SNN inference accelerator

Relevant Coursework

- | | |
|--|---|
| <ul style="list-style-type: none">• Machine Learning• Computational Methods in Optimization• Natural Language Processing (Online Courses)• Multimodal Machine Learning (Online) | <ul style="list-style-type: none">• Data Structures and Algorithms• MOS VLSI Design• Advanced VLSI Design• System On Chip Design |
|--|---|

Teaching Experience

Bangladesh University of Engineering and Technology (BUET), Bangladesh

Faculty Member, Dept. of Electrical and Electronic Engineering

Aug. 2014 – Aug. 2017

- Supervised 200+ students in VLSI laboratory courses, fostering circuit/hardware design skills with Cadence and HSPICE tools, and guided 100 students in MATLAB numerical problem-solving assignments
- Part of a team of 9 to analyze causes of 2014 power blackout in Bangladesh & delivered 10 preventive recommendations
- Led a team of 4 to secure grant from World Bank in 2015 to foster computer-aided design research in Bangladesh

Selected Publications

1. [G. Saha](#), K.Roy, "Continual Learning with Scaled Gradient Projection", **AAAI**, vol. 37, no. 8, pp. 9677-9685, Jun. 2023.
2. [G. Saha](#), K.Roy, "Saliency Guided Experience Packing for Replay in Continual Learning", Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 5273-5283, 2023. (**Award Finalist - Top 1% paper**)
3. [G. Saha](#), I.Garg, K.Roy, "Gradient Projection Memory for Continual Learning", International Conference on Learning Representations (ICLR), 2021. (**Oral Presentation - Top 1% Papers**)
4. [G. Saha](#), I.Garg, A.Ankit, K.Roy, "SPACE: Structured Compression and Sharing of Representation Space for Continual Learning", IEEE Access, 9, 150480-15094, 2021.
5. [G. Saha](#), Z.Jiang, S.Parihar, C.Xi, J.Higman and M.Ahosan Ul Karim, "An Energy-Efficient and High Throughput in-Memory Computing Bit-Cell With Excellent Robustness Under Process Variations for Binary Neural Network", IEEE Access, 8, 91405-91414, 2020.
6. I. Chakraborty, [G. Saha](#), K.Roy, "Photonic In-Memory Computing Primitive for Spiking Neural Networks Using Phase-Change Materials", Phys. Rev. Applied, 11, 014063, Jan 2019.
7. I. Chakraborty, [G. Saha](#), A. Sengupta, K.Roy, "Toward fast neural computing using all-photonic phase change spiking neurons", Scientific Reports, 8, 12980, Aug 2018.
8. [G. Saha](#), K.Roy, "Amphibian: A Meta-Learner for Rehearsal-Free Fast Online Continual Learning", 2023. (Under review)
9. S. Choudhary, S. Aparna Aketi, [G. Saha](#), K.Roy, "CoDeC: Communication-Efficient Decentralized Continual Learning", arXiv:2303.15378, 2023. (Under review)
10. D. Ravikumar, [G. Saha](#), S. Aparna Aketi, K.Roy, "Homogenizing Non-IID datasets via In-Distribution Knowledge Distillation for Decentralized Learning", arXiv:2304.04326, 2023. (Under review)