SNEHASIS ADDY

saddy@umass.edu | snehasisaddy.github.io | github.com/snehasisaddy | linkedin.com/in/snehasisaddy

Third-year PhD student in edge AI and machine learning, building on-device speech systems that cut the energy and latency of AI models.

RESEARCH EXPERIENCE

Research Assistant — Laboratory for Advanced System Software, UMass Amherst *September 2025 – Present | Amherst, MA*

HyVox: Hybrid Neuromorphic–von Neumann Speaker Identification (Under Submission)

- Built a hybrid neuromorphic–von Neumann system for always-on, open-set speaker identification: a neuromorphic chip embeds each audio window with sparse spiking, while an STM32 MCU scores speakers by cosine similarity, rejects unknowns, and enrolls new users and identifies them on-device without needing additional training.
- Designed an 832-neuron spiking neural network (SNN) that learns a time constant per neuron across fast, medium, and slow timescales, plus a 32-neuron embedding extractor trained on enrollment episodes to generalize to unseen speakers.
- Compressed the SNN 6.5x to fit the chip's synapse budget, then built chip-accurate spiking training (CAST), which trains through a differentiable replica of the chip's integer neurons and doubled top-1 accuracy on unseen speakers (31% → 61%).
- Reached 99.6% top-1 and 5.5% rejection EER with 2 members (94.9% and 13.2% with 20) at 1.63 mJ dynamic energy per 3-s window, with 42% lower latency than the fastest embedded ANN and 5.8x less total energy than the most efficient ANN.

Neuromorphic On-Device Spoken Language Understanding

- Building an end-to-end speech → phoneme → word → intent pipeline with no cloud speech recognition: the neuromorphic chip runs a feed-forward phoneme model, while the MCU holds context to assemble words and intents.
- This division of labor lets just 400 spiking neurons reach 60% span accuracy over a 2,000-word vocabulary.

Energy-Adaptive Wearable Intelligence (UbiComp 2027)

- Trained the baseline spiking models and built compute benchmarks for IMU-based activity recognition on SynSense Xylo, using FLOP profiling to map energy–accuracy trade-offs across operating points.
- Helped develop a single runtime-adaptive model that scales its computation to the energy available, widening the operating range 10–40x with up to 3.3x fewer spikes and <1.5% accuracy loss.

PUBLICATIONS

HyVox: Hybrid Neuromorphic–von Neumann Speaker Identification *(Under Submission)*
Snehasis Addy, Eduardo Calle Ortiz, Prashant Shenoy, VP Nguyen

Always-On Wearable Intelligence Through Energy-Adaptive Neuromorphic Computing *(To appear at UbiComp 2027)*
Eduardo Calle Ortiz, **Snehasis Addy**, Hui Guan, Deepak Ganesan, VP Nguyen

Flexible Polar Encoding for Information Reconciliation in QKD *(arXiv:2312.03100)*
Snehasis Addy, Sabyasachi Dutta, Somnath Panja, Kunal Dey, Reihaneh Safavi-Naini, Daniel Oblak

PROJECTS

Distributed Transaction Server — Advisor: Prof. Prashant Shenoy

- Built a fault-tolerant three-tier trading platform, keeping replicas consistent via Bully leader election and Paxos.

Flexible Polar Encoding for Information Reconciliation in QKD

- Designed a polar-code encoder for error reconciliation in quantum key distribution (QKD), deriving an O(N log N) reliability-sequence algorithm that works with any decoder.
- Converged ~47x faster than prior methods (~10s vs. ~468s at $N=2^{16}$) at comparable quality, with finite-key security proofs.

EDUCATION

UMass Amherst — Ph.D. in Computer Science Expected 09/28
Advisors: Prof. Prashant Shenoy, Prof. VP Nguyen | Amherst, USA

University of Calgary — M.S. in Physics May 2024
Calgary, Canada

IIT Dhanbad — B.Tech. in Electronics & Communication Engineering May 2021
Dhanbad, India

SKILLS

- **Research Areas:** Edge AI, On-Device Machine Learning, Ubiquitous Computing, HCI, Human-Centered AI
- **AI Model Efficiency:** Structured Pruning, Quantization-Aware Training (QAT), Model Compression, Chip-Accurate Fine-Tuning
- **Speech & Audio:** Speaker identification & verification, spoken language understanding, audio feature extraction, VoxCeleb
- **Embedded & Hardware:** STM32 MCUs, SynSense Xylo, neuromorphic chips, inference firmware, energy & latency profiling
- **Programming:** 3+ years experience with Python and MATLAB; C/C++, SQL, Linux, Docker, GCP, AWS
- **Machine Learning:** 3+ years experience with PyTorch and TensorFlow; NumPy, Pandas, scikit-learn