Saurabh Dash

✉ *hello.saurabhdash@gmail.com*
⌂ *saurabhdash.com*
⬤ *saurabhdash*

Education

2019-July 2022 **Ph.D. (on leave of absence)**, *Electrical & Computer Engineering*, Georgia Institute of Technology, *GPA - 4/4*.
Awards: ECE PhD Fellowship

2014-2019 **Dual Degree (B.S. + M.S.)**, *Electronics & Electrical Communications Engineering + Minor, Computer Science,* Indian Institute of Technology, Kharagpur.
Awards: Order of Merit, IIT KGP for distinguished performance in the field of Technology

Select Publications

2026 *Tiny Aya: Bridging Scale and Multilingual Depth.* A. R. Salamanca, D. Abagyan, D. D'souza, A. Khairi, D. Mora, <u>S. Dash</u>, V. Aryabumi, S. Rajaee, M. Mofakhami, A. Sahu, T. Euyang, B. Prince, M. Smith, H. Lin, A. Locatelli, S. Hooker, T. Kocmi, A. Gomez, I. Zhang, P. Blunsom, N. Frosst, J. Pineau, B. Ermis, A. Üstün, J. Kreutzer, M. Fadaee. arXiv preprint [paper]

2025 *Aya Vision: Advancing the Frontier of Multilingual Multimodality.* <u>S. Dash</u>, Y. Nan, J. Dang, A. Ahmadian, S. Singh, M. Smith, B. Venkitesh, V. Shmyhlo, V. Aryabumi, W. Beller-Morales, J. Pekmez, J. Ozuzu, P. Richemond, A. Locatelli, N. Frosst, P. Blunsom, A. Gomez, I. Zhang, M. Fadaee, M. Govindassamy, S. Roy, M. Gallé, B. Ermis, A. Üstün, S. Hooker. Under review, International Conference on Machine Learning (ICML), 2026 [paper]

2025 *Command A: An Enterprise-Ready Large Language Model.* Team Cohere. arXiv preprint [paper]

2024 *How Does Quantization Affect Multilingual LLMs?* K. Marchisio, <u>S. Dash</u>, H. Chen, D. Aumiller, A. Üstün, S. Hooker, S. Ruder. Empirical Methods in Natural Language Processing (EMNLP), 2024 [paper]

2024 *Aya 23: Open Weight Releases to Further Multilingual Progress.* V. Aryabumi, J. Dang, D. Talupuru, <u>S. Dash</u>, D. Cairuz, H. Lin, B. Venkitesh, M. Smith, K. Marchisio, S. Ruder, A. Locatelli, J. Kreutzer, N. Frosst, P. Blunsom, M. Fadaee, A. Üstün, S. Hooker. arXiv preprint, 2024 [paper]

2024 *Outliers and Calibration Sets have Diminishing Effect on Quantization of Modern LLMs.* D. Paglieri, <u>S. Dash</u>, T. Rocktäschel, J. Parker-Holder. International Conference on Machine Learning (ICML) ES-FoMo II, 2024 [paper]

2023 *Intriguing Properties of Quantization at Scale.* A. Ahmadian*, <u>S. Dash</u>*, H. Chen*, B. Venkitesh, S. Gou, P. Blunsom, A. Üstün, S. Hooker. Neural Information Processing Systems (NeurIPS), 2023 [paper]

2023 *Associative Memory Augmented Asynchronous Spatiotemporal Representation Learning for Event-based Perception.* <u>S. Dash</u>*, U. Kamal*, S. Mukhopadhyay. International Conference on Learning Representations (ICLR), 2023 [paper]

2022 *Sequence Approximation using Feedforward Spiking Neural Network for Spatiotemporal Learning: Theory and Optimization Methods.* X.She, <u>S. Dash</u>, S. Mukhopadhyay. International Conference on Learning Representations (ICLR), 2022 [paper]

2022 *Temporal Point Process Modelling using Recurrent Graph Network.* <u>S. Dash</u>, X. She, S. Mukhopadhyay. International Joint Conference on Neural Networks (IJCNN), 2022. [paper]

2021 *A Heterogeneous Spiking Neural Network for Unsupervised Learning of Spatiotemporal Patterns.* X.She, <u>S. Dash</u>, S. Mukhopadhyay. Frontiers in Neuroscience, 2021 [paper]

2021 *Robust Processing-In-Memory with Multi-bit ReRAM using Hessian-driven Mixed-Precision Computation.* <u>S. Dash</u>, Y.Luo, A.Lu, S.Yu, S. Mukhopadhyay. IEEE Transactions on Computer Aided Design (TCAD), 2021 [paper]

2020 *Physics-Incorporated Convolutional Recurrent Neural Networks for Source Identification and Forecasting of Dynamical Systems.* P. Saha, <u>S. Dash</u>, S. Mukhopadhyay. Neural Networks [paper]

2020 *Hessian-Driven Unequal Protection of DNN Parameters for Robust Inference.* <u>S. Dash</u>, S. Mukhopadhyay. International Conference on Computer Aided Design (ICCAD), 2020 [paper]

Work Experiences

January 2026-Present **Member of Technical Staff, Cohere Inc., US**
- ○ Building Large Multimodal Models.
- ○ Building Inference Efficient (Latency and Memory Footprint) Large Language Models.

August 2022-January 2026 **Member of Technical Staff, Cohere Inc., CA**
- ○ Building Large Multimodal Models.
- ○ Building Inference Efficient (Latency and Memory Footprint) Large Language Models.

May 2021-August 2021 **Machine Learning Research Intern (MIND Team), Apple Inc., US**
- ○ Studied and implemented sample-adaptive augmentation strategies for improved generalization across various deep learning model families.
- ○ Obtained near SoTA (Adverserial AutoAugment) performance without the need to train a tightly coupled augmentation generator and classifier.

May 2018 - July 2018 **Statistical Signal Processing Intern, Texas Instruments, India**
- ○ Developed Kalman Filter based approach for tracking and correction of sampling clock jitter in high speed ADCs.
- ○ Proposed solution obtained suprious-free dynamic range (SFDR) of 80dB.

May 2017 - July 2017 **Visiting Student Researcher, Nanyang Technological University, Singapore**
Advisor: Dr.Arindam Basu
- ○ Explored a novel hardware amenable training rule to train binary and multi-bit synapses for neuromorphic classifiers to power the next generation of implantable Brain-Computer Interfaces.
- ○ Implemented a highly-efficient vectorized implementation to reduce the training time by $10\times$.

Invited Talks

July 2023 Intriguing Properties of Quantization at Scale: Memorial Sloan Kettering

October 2023 Intriguing Properties of Quantization at Scale: ML Collective

Research Experiences

Jan 2020 - April 2020 **Deep Generative Models: A study**
Advisors: Dr.Matthieu Bloch, GeorgiaTech | Academic Project [**Code, Report**]
- ○ Implemented and compared various aspects of deep generative modelling approaches - Generative-Adversarial Networks, Variational Autoencoders and Normalizing Flows.
- ○ Studied how recent variants like Wasserstein GAN(GP), β-VAE, Glow etc. mitigated issues faced by the original approaches.

Jan 2020 - April 2020 **Circuit Partitioning Using Graph Neural Networks**
Advisors: Dr.Sung-Kyu Lim, GeorgiaTech | Academic Project [**Code, Report**]
- ○ Implemented a deep-learning based fully differentiable approach to solve the problem of circuit partitioning using Graph Convolutional Networks.
- ○ Analytically derived custom loss gradients and extended pytorch autograd for sparse tensors to handle graphs with over 1M nodes.

July 2017 - April 2019 **Low Power On-Chip Learner for a Neuromorphic Classifier**
Masters' Dissertation | *Nominated by the panel for the best dissertation award.*
Advisors: Dr.Indrajit Chakrabarti, IIT Kharagpur and Dr.Arindam Basu, NTU Singapore
Developed and implemented network rewiring rule based sub-threshold analog circuits that can be incorporated on chip to train a neuromorphic classifier.

Reviewing Experiences

- ○ TMLR
- ○ EACL 2023
- ○ CVPR 2023
- ○ ICCV 2023
- ○ NeurIPS 2023

Other Achievements

- ○ Invited to **Tesla AI Day 2021** to discuss self-driving technology with the AI team.
- ○ **Contingent Head** of the first Indian Team to qualify for RoboCup 2017 held at Nagoya, Japan.

- Member of the **Bronze** winning team in FIRA MiroSot 2015 held at Daejeon, South Korea.
- Ranked among the **top 0.1% students** in IIT-Joint Entrance Examination among 485,000 students.
- **Kishore Vaigyanik Protsahan Yojna** (Young Scientist Award), 2013.
- Member of the **Gold** winning team in KPIT Sparkle 2015 - an inter collegiate innovation challenge.

Relevant Courses

- Statistical Machine Learning
- Math. Foundations of ML
- Probabilistic Graphical Models
- Convex Optimization
- PDEs for Image Processing
- Digital Signal Processing
- Information Theory
- Computational Neuroscience
- Real Analysis

Technical Competencies

Frameworks	PyTorch, JAX, TensorFlow
Languages	python, C, C++, Cuda, Triton, MATLAB

Service & Other

- Mentored a group of 20 students for a 7 day IEEE certified Robotics workshop leading to the design of a self-balancing robot.

Other Publications

2023 *Brain-Inspired Spatiotemporal Processing Algorithms for Efficient Event-Based Perception.* B. Chakraborty, U. Kamal, X. She, S. Dash, S. Mukhopadhyay. Design, Automation & Test in Europe (DATE), 2023 [paper]

2021 *Unsupervised Hebbian Learning for Point Set in StartCraft II.* B. Kang, H. Kumar, S. Dash, S. Mukhopadhyay. International Joint Conference on Neural Networks (IJCNN), 2022 [paper]

2021 *Reliable Edge Intelligence in Unreliable Environment.* M. Lee, X. She, B. Chakraborty, S. Dash, B. Mudassar, S. Mukhopadhyay. Design, Automation & Test in Europe (DATE), 2021 [paper]

2021 *Impact of HKMG and FDSOI FeFET drain current variation in processing-in-memory architectures.* N.E. Miller, Z. Wang, S. Dash, A.I. Khan, S. Mukhopadhyay. Journal of Materials Research, 2021 [paper]

2021 *Characterization of Drain Current Variations in FeFETs for PIM-based DNN Accelerators.* N.E. Miller, Z. Wang, S. Dash, A.I. Khan, S. Mukhopadhyay. IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS), 2021 [paper]

2020 *A Flexible Precision Multi-Format In-Memory Vector Matrix Multiplication Engine in 65nm CMOS with RF Machine Learning Support.* M. Mukherjee, Y. Long, J. Woo, D. Kim, N. Rahman, S. Dash, S. Mukhopadhyay. IEEE Solid-State Circuits Letters, 2020 [paper]

2017 *Low Power Implantable Spike Sorting Scheme based on Neuromorphic Classifier with Supervised Training Engine.* R. Pathak, S. Dash, M. Sharad. IEEE Computer Society Annual Symposium on VLSI (ISVLSI), 2017 [paper]