

Academic Details

Examination/Degree	University	Year	GPA/%
Ph.D.:			
EECS, Computer Engineering	University of California, Irvine	2024-present	4.0/4.0
Master of Science:			1,5 (German Scale)
Comm. & Electronics Engineering	Technical University of Munich	July, 2021	First Class Distinction
Bachelor of Technology:			
Electronics & Communication Engineering	NIT Durgapur, India	May, 2017	9.14/10
Higher Secondary Examination	WBCHSE	May, 2013	88.8; Percentile : 99.97
10th Std. Examination	WBBSE	May, 2011	86.5

Publications & Patents

- **Characterizing State Space Model and Hybrid Language Model Performance with Long Context.** Saptarshi Mitra, Rachid Karami, Haocheng Xu, Sitao Huang, Hyoukjun Kwon. 2026 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS'26) [Code]
- **Lookup Tables for Ultra Low-Bit Operations.** Saad Ashfaq, Saptarshi Mitra, Alexander Hoffman, Darshan Ganji, Ahmed Hassanien, Sudhakar Sah, Ehsan Saboori, MohammadHossein AskariHemmat. (Patent Filed)
- **DeepliteRT: Computer Vision at the Edge .** Saad Ashfaq, Alexander Hoffman, Saptarshi Mitra, Sudhakar Sah, MohammadHossein AskariHemmat, Ehsan Saboori. The 34th British Machine Vision Conference (BMVC'23)
- **YOLOBench: Benchmarking Efficient Object Detectors on Embedded Systems.** Ivan Lazarevich, Matteo Grimaldi, Ravish Kumar, Saptarshi Mitra, Shahrukh Khan, Sudhakar Sah. Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCV'23)
- **DeepGEMM: Accelerated Ultra Low-Precision Inference on CPU Architectures using Lookup Tables.** Darshan C Ganji, Saad Ashfaq, Ehsan Saboori, Sudhakar Sah, Saptarshi Mitra, Mohammadhossein Askarihemmat, Alexander Hoffman, Ahmed Hassanien, Mathieu Leonardon. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPR'23)
- **Accelerating and Pruning CNNs for Semantic Segmentation on FPGA.** Pierpaolo Morì, Manoj-Rohit Vemparala, Nael Fasfous, Saptarshi Mitra, Sreetama Sarkar, Alexander Frickenstein, Lukas Frickenstein, Domenik Helms, Naveen Shankar Nagaraja, Walter Stechele, Claudio Passerone. Proceedings of the 59th ACM/IEEE Design Automation Conference, (DAC'22)
- **Accelerating Semantic Image Segmentation on FPGA.** Saptarshi Mitra. Technische Universität München'21
- **Arduino Based Foot Pressure Sensitive Smart Safety System for Industrial Robots.** Sayan Sarkar, Gautam Ghosh, Amitrakshar Mohanta, Atreye Ghosh, Saptarshi Mitra. Second IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT'17)
- **Implementation of Relay Hopper Model for Reliable Communication of IoT Devices in LTE Environment through D2D Link.** Anish Pradhan, Soumi Basu, Sreetama Sarkar, Saptarshi Mitra and Dr. Sanjay Dhar Roy. 10th International Conference on COMMunication Systems & NETworks (COMSNETS'18)
- **Performance of Different Power Control Schemes for a Hybrid LTE System with Channel Impairment.** Sreetama Sarkar, Saptarshi Mitra, Tamoghno Nath and Dr. Sanjay Dhar Roy. The 2017 International Electrical Engineering Congress (iEECON2017)

Experience

Microsoft

Research Intern, Applied Sciences Group

- Speculative tool execution for on-device agentic LLM inference

REDMOND, WASHINGTON
June '26 – Sep '26

Meta, Reality Labs

ASIC Intern Contextual AI

- On-device Contextual-AI for custom silicon in next-generation of XR-glasses.
- Hardware prototyping for Cascaded sensing to trigger high-power consuming event with lower-power consuming events.

REDMOND, WASHINGTON
June '25 – Sep '25

Deeplite	TORONTO, CANADA
Embedded AI Engineer in Research & Development	July '21 – May '24
- Runtime for Neutrino model optimizer, providing significant speedup through extreme Quantization - Automatic deployment of lightweight models on range of hardware platforms like embedded GPU, NPU, Raspberry Pi, MCU fostering hardware aware optimizing algorithms - Bringing up vendor specific SBC and OS to run compressed models in custom runtime	
Technical University of Munich	MUNICH, GERMANY
Student Research/Teaching Assistant	Oct '20 – Feb '21
- Automation of organization & grading of the Seminar course at the Chair of Design Automation - Tutoring and mentoring graduate level students in VHDL System Design Lab (EI7403)	
Intel	MUNICH, GERMANY
Internship & Working Student	March '19 – July '20
- During Internship, worked with Digital Design Verification team (Modem Technology Group) for development of a functional pre-silicon Verification IP (VIP) for a standardized AMBA debug interface. - As a part of Intel Architecture, Graphics and Software (IAGS) group, validated system-level debug framework and various features related to CPU run control	
Verizon	HYDERABAD, INDIA
Software Engineer	Jun '17 – July '18
Full stack developer (Angular & Spring boot) in Network & Technology Group	
TCS Innovation Labs	KOLKATA, INDIA
Research Intern	May '16 – July '16
Localization of mobile Robots with ROS	

Academic Projects

Accelerating Sparse Neural Networks for Semantic Image Segmentation on FPGA platforms	2020-2021
- Novel dilated convolution irrespective of dilation rate and fitting the sparse weights in the on-chip buffer - Efficient semantic image segmentation accelerator deploying state-of-the art model DeepLabV3+ - Hardware aware latency driven Genetic Algorithm based channel pruning for the custom accelerator	
Sustaining Reduction in Carbon Emissions Based on Policy Decisions post COVID-19 Crisis	2020
- Investigating the effects of stringent policy decisions imposed by various Governments on CO₂ emission - Data collection, cleaning and building a web-application that implements a comprehensive data analysis pipeline with ML models in back-end and UI for interactive visualization	
Radio-Controlled Multi-Functional Clock in Zync-7000 FPGA	2019-20
- Specification and RTL design of the Global Finite State Machine module of the clock - Integration with other modules like alarm clock, stopwatch and deployed on FPGA	
Hardware-Oriented Implementation of IDEA (encryption algorithm) in Spartan-3e FPGA	2018-19
- Used VHDL to implement and simulate the International Data Encryption Algorithm. - Implemented Resource Constrained Scheduling for optimum resource utilization	

Scholastic Achievement

Design Automation Conference (DAC) 2021 Young Fellow and best video award	Dec 2021
Deutschlandstipendium, German National Scholarship Program for academic excellence, TUM	2018-21
Scheme of Scholarship for College and University Student reg. of Government of India	2013
Got an award on the basis of Higher Secondary Examination held by West Bengal Council of Higher Secondary Education	
DST-INSPIRE Internship Science Camp	2011
Selected and Participated at JBNSTS, Kolkata	
Department of Science & Technology (DST) sponsored INSPIRE award	2010
For academic excellence from 6th to 10th standard	
'Best of Subject Category : Engineering' in India	2009
At IRIS science fair conducted by Intel corporation, CII and DST held at Ahmedabad	

Skills

C, Python, VHDL, OpenCL, TVM, TFLite, ONNX Runtime, Verilog, SystemVerilog, UVM, VCS, Arduino, MATLAB, LaTeX, Logisim, Dash (Framework with Flask and React), Angular, C++, Java, Spring Boot, Assembly Level Programming, GDB, Keil uVision, HTML, CSS

Extra Curricular Activities

- Semester Representative and Member of Quality Circle, MSCE, TUM** 2018-2021
- Volunteer at TEDx Jadavpur University** 2018
played an active role in inviting speakers and arranging the event
- Chief Events Coordinator of Prakriti - the Environmental club of NIT Durgapur** 2016-17
which was involved in activities like organizing plantation programs, observing earth hour, campus cleaning etc.
- Google Online Marketing Challenge (GOMC) 2016** 2016
Partnered with www.mayallyang.com on creating AdWords online marketing campaign for their expansion of Ecotourism . Ranked : Good
- DuckDuckGo and Mozilla(110n,e10s,webVR)** 2014-Present
Community member and Active Contributor
- Volunteer at Blood Donation Camps organized by State Blood Transfusion Council, West Bengal** 2014