

ASWINKUMAR RAMKUMAR

ramkumar4@wisc.edu [linkedin.com/in/aswinkumar99/](https://www.linkedin.com/in/aswinkumar99/) aswinkumar.me

Education

University of Wisconsin-Madison

December 2025

Master of Science in Electrical and Computer Engineering

Indian Institute of Technology Madras

2018 – 2022

Bachelor of Technology in Engineering Physics with Minor in Computing

Professional Experience

NVIDIA - Full-time

Bangalore, Karnataka

GPU Advocate

Jul 2022 – Aug 2024

- Developed a comprehensive [AI for Science](#) pipeline leveraging Modulus, Earth-2 and Omniverse platforms, integrating Physics Informed Neural Networks, Neural Operators & Ray Tracing for training, inference and visualization.
- Mentored over **20 research and enterprise teams** in 14+ hackathons, focusing on NVIDIA technology utilization, leading to significant code acceleration and optimizing for AI model convergence.
- Conducted 15+ workshops on **AI for Science, Computer Vision, LLMs and Accelerated Computing** as lead instructor, impacting learning at global supercomputing sites, including Alibaba, Infosys, IISc, NCI, NCHC, and CSIRO.
- Lead Mentor of research projects from **Siemens Research** and **NCRA** through the NVAITC research program, guiding them through advanced research methodologies, culminating in projects on track for publication in esteemed journals.

NVIDIA - Internships

Bangalore, Karnataka

AI HPC GPU Advocate Intern

Jun 2021 – Jul 2021

- Incorporated methods for scalable Deep learning training and addressed challenges concerning large batch training.
- Mentored research teams in the CINECA & C-DAC Hackathon to improve the throughput of applications by **up to 30x**.

AI GPU Advocate Intern

Oct 2020 – Dec 2020

- Architected a pipeline for real-time deep learning inferencing on video inputs utilising the DeepStream Framework. This pioneering work was deployed across multiple **Tier-1 Higher education and research institutions** globally.
- Accelerated throughput of AI inference pipeline on GPUs by **up to 28x** using the DeepStream framework.

Deep Learning Intern

Dec 2019 – Jan 2020

- Trained and enhanced state-of-the-art Deep Learning Networks for High-Performance Computing applications.
- The work was integrated into the AI for Science bootcamp under OpenACC and NVIDIA and was used as introductory training material for various premier research institutes such as **NASA, CINECA** and **IIT Bombay**.

Publications

MABViT (Modified Attention Block Enhances Vision Transformers)

Mar 2023 – Jul 2023

*Aswinkumar Ramkumar, Mahesh Ramesh, **Accepted** : Deployable AI workshop, AAAI Conference 2024*

- Engineered a novel transformer architecture to address the limitations of representation collapse in Pre-LN transformer models by implementing non-linearity within the attention block using Gated Linear Units applied to the Value tensor.
- Achieved superior performance on the Imagenet-1K dataset, surpassing state-of-the-art models in accuracy while utilizing only **half the number of parameters** compared with existing B/16 variants.

Coursework

- Advanced Robotics**
- Pattern Recognition and Machine Learning
- Computer Organisation
- Advanced Computer Architecture**
- Digital Systems Testing and Testable Design
- GPU Programming
- Parallel Scientific Computing
- Digital IC Design
- Virtual Reality Engineering

** - Currently enrolled

Selected Projects

Hardware Accelerated IDCT Algorithm | *B.Tech Project under Prof. Kamakoti Veezhinathan* **Aug 2021 – Jun 2022**

- Engineered a hardware synthesizable 64-point **Inverse Discrete Cosine Transform (IDCT) algorithm** in Bluespec Verilog, optimising it for various sizes (4, 8, 16, 32, 64 points) using a Butterfly architecture.
- Conducted rigorous testing and validation against a Python-based reference model, ensuring high precision and efficiency suitable for advanced computational applications such as in Audio and Video decoders.

Carry Save Multiplier Design | *Digital IC Design Course project* **Jul 2021 – Dec 2021**

- Designed and implemented a compact 8x8 carry save multiplier, leveraging CMOS technology for digital blocks like Inverters, NAND gates, and Full adders, ensuring efficient operation and verification.

Reinforcement Learning for Hexapod Robot | *Robotics Club, CFI (Centre for Innovation)* **Jul 2020 – Apr 2022**

- Spearheaded a team of 5 to develop a Hexapod robot and modelled the dynamics in the MuJoCo simulation environment with OpenAI gym to learn gaits using Reinforcement learning algorithms such as **PPO and DDPG**.
- Received special mentions at OpenHouse 2023, CFI and featured on National Television (NDTV).

VR Gloves with Thermal Interface | *Virtual Reality Engineering Course Project* **Jul 2021 – Dec 2021**

- Designed and prototyped VR gloves with an integrated thermal interface, enabling feedback in virtual environments, complemented by a custom battery pack for wireless operation and Bluetooth connectivity for software integration.
- Developed an interactive demo using the Unity game engine and Oculus SDK for Oculus Quest 2, demonstrating the gloves' capabilities in an immersive virtual environment.

Electronics Club Development Board | *Electronics club, CFI* **Dec 2020 – Mar 2021**

- Orchestrated a 12-member team in the end-to-end development of a **custom-built ESP32 SoC development board**, managing technical, financial, and logistical aspects for successful mass production and distribution.
- Crafted a comprehensive kit, distributing it to over 100 members of the community across India for an accessible \$25.

RISC-V RV32I Implementation in Verilog | *Computer Organisation Course Project* **Jul 2020 – Nov 2020**

- Designed and simulated a **RISC-V RV32I** processor in Verilog HDL, incorporating ALU operations, branch instructions, and IO interfacing, and validated through post-synthesis simulations using Yosys.

Technical Stack

Languages: Python, C++, C, CUDA, ROS, Julia, Fortran, MATLAB, Verilog, Bluespec, HTML

Frameworks: PyTorch, TensorFlow, JAX, Keras, OpenACC, OpenMP, MPI

NVIDIA Tools: Modulus (PINNs & Neural Operators), RAPIDS (Data Science & ML), NeMo (LLMs), DeepStream, TAO, TensorRT, Triton Inference Server (Deployment), USD Composer (Digital Twins), Robotics - Isaac Sim, Drive SDK

Software: Git, Docker, Apptainer, MuJoCo, Unity, LaTeX, LTSpice, Cura, Autodesk Eagle & Autodesk Fusion

Embedded Platforms: NVIDIA Jetson Nano, Arduino, Raspberry Pi, ESP32, ZedBoard Zynq-7000

Achievements

- Runner Up in "Yet Another Hackathon" 2019 - Developed an AI-based Real-time Sign Language Translator. **2019**
- Flipkart GRID 2.0 – Intelligent Picking - Top 43 in Round 2 of the challenge. (Team Blackbuck) **2020**
- Represented IIT Madras at the 7th Inter-IIT Technical Meet for PlutoX Hackathon - **only freshman selected**. **2018**
- Led a high school robotics team to secure a 1 Million INR grant for ATAL Tinkering & Innovation labs. **2016**
- Developed an Android app at age 14 with a database of algorithms for solving Rubik's cube in under 10 seconds. **2014**
- Elite World Records - Solved Rubik's Cube 236 times in 3 hours. **2013**
- Elite World Records - Solved 11 Mental Arithmetic Problems in 1 hour (Single Digit - 250 Rows). **2013**

Leadership Roles

Electronics Club Head, Centre for Innovation, IIT Madras **May 2020 – Apr 2021**

Electronics Club Coordinator, Centre for Innovation, IIT Madras **May 2019 – Apr 2020**

Publicity Coordinator, Shaastra, IIT Madras **May 2019 – Apr 2020**