RESEARCH STATEMENT

Kommalapati Lahari
lahari1803ks@gmail.com
+91 7989410372

Previous Research Experience

May 2024 - July 2024

Research Intern IIT-Hyderabad (Driver behavior Analysis)

July 2024 - Feb 2025

InvGuard: Agentic AI for Smart Contract Security

Jan 2025 - Current

STMicroelectronics: Generative AI for (Design Space Optimization) and Verification

IIT Hyderabad – Research Internship (Driver Behavior Analysis)

- Built a real-time driver behavior analysis pipeline using PCA (82% variance) and K-Means clustering.
- Proposed Lateral Shift Ratio per 100, a novel metric for lane-keeping stability.
- Classified 166 diverging events into two behavioral clusters (aggressive vs non-aggressive).
- Performed ANOVA-based validation to ensure statistical significance.
- Presented and defended the work at the TRB Conference, Washington D.C.

InvGuard – Agentic AI for Smart Contract Security

- Designed InvGuard, a multi-agent AI system for autonomous vulnerability detection & repair.
- Developed modules: InvGetter, AttackModule, FixMaker, FixChecker for end-to-end invariant synthesis & patching.
- Achieved 96.2% invariant precision, mitigating 74/91 real-world exploits.
- Outperformed baselines by +11.5 pp invariant accuracy, +14.7 pp exploit coverage, +17.2 pp patch effectiveness.
- Fine-tuned on 600–700 curated samples, gaining insight into the high computational costs which come while dealing with SOTA LLMs.
- This experience reinforced my conviction that efficient, consumer-friendly AI models and hardware are essential for democratizing AI.
- Currently under review for FSE 2026.

STMicroelectronics: Generative AI for HLS Optimization and Verification

- At STMicroelectronics, I'm driving research at the intersection of agentic frameworks and hardware co-design. My primary responsibility involves designing an end-to-end pipeline for Design Space Optimization (DSO) and Microarchitecture Exploration using AI driven methodologies to enable intelligent decision-making across the hardware design stack.

- Within this pipeline, we adopted High-Level Synthesis (HLS) as the core modality for exploration allowing us to systematically evaluate design alternatives and extend the framework toward diverse optimization pipelines, such as kernel-level transformations.

- I conceptualized and implemented a generative AI workflow to optimize accelerator-level C++ code for synthesis, identifying and transforming key code segments to improve PPA metrics at block-level granularity. I integrated Open AI's SDK with proprietary accelerator code base, customizing its memory and context management for hardware constraints.

- Additionally, I established an automated verification pipeline for testing the above HLS Codes, auto-generating UVM testbenches, and achieving coverage closure through adaptive test generation. This work generalizes beyond a single HLS project to a scalable AI-assisted DSO pipeline, applicable across a wide range of algorithmic IPs.

- This project strengthened my conviction that efficiency research must unify algorithmic intelligence and hardware realization, forming the core of my Ph.D. vision

Current Research Interests & Original Ideas

- My current research focuses on developing efficient and sustainable AI accelerators by jointly optimizing algorithms and architectures. I see a paradigm shift from large-scale LLMs toward Edge Language Models (ELMs) systems optimized for batch size = 1 and memory-constrained deployment, enabling what I call deep edge intelligence through localized, adaptive inference.

- I am actively exploring transformers and post-transformer architectures, particularly state-space hybrids like the Mamba series, which offer linear-time inference with transformer-level accuracy. Recent results like Sonic-3, which achieves SOTA performance across the voice modality using state-space modeling, demonstrate the tangible potential of this paradigm.

- Building on this, my work involves Neural Architecture Search (NAS) and multi-dimensional model reduction through quantization, pruning, and sparsity to design architectures tailored for edge level latency and power constraints.

- The guiding lens for this exploration is the Roofline model, which highlights how edge systems are predominantly memory-bound rather than compute-bound. Instead of maximizing arithmetic throughput, my research targets memory access efficiency and digital in-memory computing paradigms that retain data locality while sustaining performance.

- Ultimately, my ongoing work envisions **a multi-context edge AI framework**, where compression-aware models power emerging applications in physical AI, robotics, and humanoid intelligence. I see **agentic AI frameworks** as the enabler of this shift where generative systems guide architectural exploration dynamically rather than through traditional static optimization. At its core, my research strives for **sustainability through efficiency**, aligning closely with Green-IC's vision of energy-conscious computation.

- As a starting point for my research, I would like to establish restrictive constraints and apply post-training optimization techniques to analyze existing models. I believe this analysis will generate high-quality research questions to guide my subsequent steps of my research.

Strengths and Weaknesses

Strengths

- My role at STMicroelectronics has given me a strong foundation and hands-on experience across the entire development pipeline from algorithm design to verification and hardware validation. This exposure has helped me understand the subtle challenges of translating algorithmic optimizations into efficient hardware implementations. It has also built a deep, layered understanding of the abstractions required to design high-performance AI accelerators.

- I bring a strong foundation in designing and implementing agentic system frameworks, which I view as a critical enabler for efficient and verifiable AI inference. These frameworks allow real-time adaptation and evaluation of models in operational settings key to optimizing not just accuracy, but also latency, throughput, and energy efficiency.

- My focus has always been on making inference efficient and sustainable, bridging adaptive AI behavior with hardware constraints. This understanding, combined with my ability to collaborate with peers specializing in circuit and architectural design, positions me to contribute effectively to Green IC's goal of developing ultra-low-power, high-performance systems for ubiquitous computing.

Weaknesses

- As a computer science graduate working at the intersection of algorithms and hardware, I recognize the need to deepen my theoretical grounding in advanced performance modeling (beyond basic roofline analysis), circuit-level low-power techniques, and formal energy efficiency optimization frameworks. While I have a working understanding of these areas, I seek to build the mathematical rigor and structured design methodology that come from formal, circuit-oriented training a gap I am proactively addressing.

- Sometimes my work stretches across too many areas from algorithms to hardware validation and I end up spreading myself a bit thin. I'm still learning to find a better balance, I believe keeping the breadth that fuels curiosity but also going deeper where it really counts.

Aspirations , personal motivation and End goals

- My journey into AI hardware research is driven by a deep conviction that democratizing AI is not only a technical pursuit but a societal necessity. The future trajectory of AI depends on how efficiently we can make intelligence accessible whether it remains confined to those with massive computational resources or becomes a universal utility available at the edge.

- This motivation crystallized during my undergraduate research on InvGuard, where fine-tuning large models on limited hardware exposed me firsthand to the prohibitive computational costs of today's AI workflows. That experience reinforced my belief that we need fundamentally more efficient, hardware-aware approaches not incremental improvements, but paradigm shifts that make AI both performant and sustainable.

- In the long term, I aspire to be a leading researcher in energy-efficient AI systems, advancing the co-design of algorithms and architectures that push the Pareto frontier of accuracy and power efficiency. My goal is to enable edge intelligence that operates locally and autonomously, minimizing dependence on large data centers and ensuring privacy, responsiveness, and massive power savings.

- Over the next 5-10 years, I aim to contribute to the evolution of AI hardware where efficient inference becomes the norm rather than the exception empowering AI that is scalable, sustainable, and beneficial to society at large. This vision aligns closely with Green IC's mission of creating ultra-low-power, high-performance integrated systems that define the future of ubiquitous computing.

Academic/Industrial references

Industrial references

- Director @STMicroelectronics : Nitin Kumar Chawla

Academic References:

- Dr.Raju Halder :
  https://www.iitp.ac.in/~halder/
  Contact No: +91 9608710115

- Dr. Somanath Tripathy :
  https://www.iitp.ac.in/people/faculty/2271:dr-somanath-tripathy
  Contact No: +91 8678088635

Positions applying and Funding Details

Which position(s) you are applying for?

Research Engineer with the motivation of prospective PhD candidate , in association with STMicroelectronics.

Provide funding details (self-funded, supported by our group, etc.) and Scholarships eligibility

I qualify for multiple competitive Ph.D. scholarships open to international candidates. Specifically:

- I meet the eligibility criteria for the NUS Research Scholarship under the NUS Graduate School and am applying as part of my Ph.D. application.
- I also intend to apply for the Singapore Millennium Foundation (SMF) Scholarship, which supports international students pursuing research in science and engineering at Singapore institutions.

I am currently in the process of completing these applications alongside my ECE-NUS Ph.D. submission. Additionally, I am being considered as a prospective candidate for industry co-support through my ongoing research engagement with STMicroelectronics, which may complement my academic funding pathway.

Which institutions you applied to and why?

I have been selectively exploring research groups whose work aligns with my focus on hardware-algorithm co-design and efficient AI inference. In particular, I've followed Prof. Dan Alistarh's group (IST Austria) on low-precision numeric formats, Prof. Priyanka Raina's Accelerate Lab (Stanford) on domain-specific accelerators, and Han Lab and Prof. Albert Gu's group (CMU) for efficient neural architectures.
While these groups offer complementary perspectives, the Green IC group at NUS stands out as my top choice for its strong emphasis on energy-efficient design, silicon validation, and circuit-algorithm synergy, which aligns directly with my research vision and industrial experience at STMicroelectronics.

The intended timeline and timing constraints (when you are available, for how long, etc.), in case you are selected ?

I am applying to the ECE-NUS Ph.D. program indicating you as my preferred supervisor, aiming for the August 2026 intake. In parallel, I would be available to join the Green IC group as a Research Engineer starting in December 2025 or January 2026, as per your recommendation.