PhD Candidate in Computer Science
Stony Brook University (SBU), USA
Google Scholar
pramodh.rachuri.dev
srachuri@cs.stonybrook.edu

Sri Pramodh Rachuri

Educational Qualifications

PhD Candidate in Computer Science, **Stony Brook University**, New York, USA — 2020 – Present
Advisor – *Prof. Anshul Gandhi* **Topic** – *Sustainable Machine Learning at Edge* — *GPA – **4.0**/4.0*

B.Tech (Hons) in Electrical Engineering, **IIT Bhilai**, Chhattisgarh, India — 2016 – 2020
Advisor – *Prof. Arzad A. Kherani* **Topic** – *Multi-Access & Network Coding for Video Streaming* — GPA – **9.03**/10.00

Director's Gold Medal recipient for outstanding performance in academic and extracurricular activities.
R P Singh Memorial Award recipient for leadership and social welfare initiatives.

Work Experience

Research Intern, Nokia Bell Labs — June'24 – Sept'24

- Analyzed the characteristics and geographical challenges of green power generation, focusing on utilizing distributed green energy sources for shared GPU resources.
- Developed and integrated a local surplus green energy-aware scheduling system into an existing decentralized computation framework that operates on a collaborative cluster of independently owned edge nodes.
- Implemented system allows verifying the 'greenness' of the energy used for computation tasks using a blockchain-based ledger. These innovations have been filed as a patent application [1].

***Research Intern, AT&T Labs* (continuing collaboration)** — June'22 – Aug'22

- Analyzed anonymized production data traces from real-world radio equipment to compare radio resource consumption patterns across LTE macro/microcells and 5G macro/mmWave cells, identifying key differences relevant to vRAN pooling.
- Designed statistical techniques to guide the redistribution and optimal placement of vRANs. In simulations, we showed reductions up to *84%* in CapEx and *60%* in OpEx compared to existing deployment strategies in simulations. This work lead to a publication [1].

Research Intern, IIT Kharagpur — May'19 – July'19

- Used the non-stationary Multi-armed Bandit reinforcement learning method to design a solution for task distribution challenges in Fog computing. Demonstrated, through simulation, that the design scheme outperforms existing cloud and fog solutions, reducing overall latency by *59%*. Further details are available in the publication [8].

Research Projects

Sustainable Deep Learning Inference at the Edge — Aug'22 – Present

- Conducted power and energy consumption benchmarking of Deep Learning workloads on edge computing devices. Explored the effects of hardware and workload parameters, including CPU and GPU frequency, core count, batch size, and neural network model complexity. Simultaneous tuning these parameters resulted in a *42%* reduction in energy consumption without loss in accuracy when compared to DVFS (Dynamic Voltage and Frequency Scaling). More details in publication [4].
- Implemented a system and a online search algorithm for dynamically tuning the hardware parameters to optimize energy consumption and latency on edge devices as the workload arrival pattern change. Our algorithm outperformed existing baselines; lowered energy consumption by *31%* and tail latency by *14%*. More details in recent publication [2].
- Designing techniques to identify optimal model architectures for a given edge device or mobile phone to maximize performance metrics such as accuracy, throughput, latency, and energy efficiency. We aim to invent a method for manufacturers to create custom distilled models for each device to bring out the best possible performance.

Orchestrating Live Video Analytics on Disaggregated Systems at the Edge — Feb'21 – Jun'24

- Engineered an innovative microservice architecture for deploying independent components of a video analytics pipeline at the edge. Used Docker containers, and TensorRT on Nvidia Jetson Nano for the implementation. Experimentation showed that our deployment strategy incurred *60%* less overhead compared to a naive policy. More insights in the publication [6].
- Modularized the Video Analytics functions for scalability and orchestration using Kubernetes. Designed novel algorithms for scheduling and load-balancing, resulting in a *51%* increase in throughput and a *28%* increase in accuracy in comparison to a naive policy. In depth details are available in the publication [3].

***Optical Circuit Switching Policies for Disaggregated Infrastructure* (in collab. with TU Berlin)** — Aug'22 – Present

- Designed and implemented a realistic Optical Circuit Switching (OCS) emulation environment using Docker and Linux TC. It enables accurate modeling of hybrid OCS-OPS (Optical Packet Switching) datacenter networks and supports detailed, scalable experimental evaluation of network optimization strategies, including dynamic reconfiguration scenarios.
- Developed a middleware-based demand-aware OCS reconfiguration policy for disaggregated clusters running MapReduce. It continuously monitors network demand and reconfigures OCS paths to maximize bandwidth utilization while minimizing reconfig overhead. By balancing responsiveness and stability, it achieves up to 81% reduction in job completion times compared to baseline approaches. This work is currently under peer-review for publication.

Near Data Processing (NDP) Optimization for Apache Spark Jobs May'21 – Jan'22

- Collaborated on Open Infras Lab's customized version of Spark and Hadoop to implement intelligent push-down operations from Spark to Hadoop, optimizing job execution time, data transfer latency, and bandwidth utilization.
- Developed a Discrete Event Simulator using Simpy in Python to emulate large compute clusters. Simulations showed a *71%* reduction in query execution time. Empirical experiments on real hardware showed a *42%* reduction. More details in publication [5].

Accurate Modeling of Request Life Cycle in Two-Tier Storage Systems May'22 – Present

- Analyzing request arrival patterns in a two-tier storage system consisting of HDDs with SSD caching and studying the associated service, response, and wait times.
- Currently focused on modeling various distributions that closely mimic request behavior, aiming to enhance simulations and potentially inform future scheduling policies.

Multi-Access and Network Coding for Video Transmission (B.Tech Hons Thesis) Jan'19 – May'20

- Built an SD-WAN that can be used to serve any UDP stream with a high data rate and less delay jitter using multiple LTE dongles. In all experiments conducted, packet losses remained below *1%* while live streaming a 4K video content.
- Used Open Air Interface to build an LTE testbed using Software Defined Radios (SDRs) and used its simulator mode to make a scalable testbench for this project. More details in publications [7,9,10,11].

Technical Skills

- Languages – Python, Java, Go Lang, C, C++, MATLAB, Verilog, Shell Scripting.
- ML Frameworks – PyTorch, Keras, Scipy, SKLearn.
- Miscellaneous – Docker, Kubernetes, Android Studio, Git, LaTeX.

Academic and Extra Curricular Activities

- Serving as a PC member in TDIS 2025 workshop and as a Web Chair for ACM Sigmetrics 2025.
- Reviewer for IEEE IoT Journal 2022 and subreviewer for USENIX NSDI 2025; IEEE ICDCS 2024, IEEE MASCOTS 2024, USENIX ATC 2024; ACM Sigmetrics 2024, 2023; ACM/IFIP Middleware 2023.
- Taught guest lectures in the courses *CSE 544* (Probability and Statistics), *CSE 534* (Computer Networks) and *CSE 570* (Wireless Networks) at Stony Brook University.
- Invited as a speaker at CryptEng lab, TU Graz to present our publication on *Optimizing Near-Data Processing for Spark*.
- Received travel grants from SEC 2024, Sigmetrics 2023, ICDCS 2022 and MobiCom 2021.
- Student Representative in Grad Committee of CS Dept at Stony Brook University from Oct 2021 to May 2024.
- Serving as a Board Member of PanIIT USA since 2023.
- Selection committee member for Young Researcher Awards, IIT Bhilai in 2022.
- Founding Vice President (UG) of Alumni Association of IIT Bhilai from 2020 to 2023. Currently an Executive Member.
- Founding Secretary of Science and Technology Division in Council of Student Affairs, IIT Bhilai from 2017 to 2018.
- Student Representative in Undergrad Committee of EECS Dept at IIT Bhilai in 2018.
- Founded Electronics Club in 2016 and Founding coordinator of Institute's Innovation Cell (IIC) in 2018 at IIT Bhilai.

Publications

1. *Constellate:Establishing the Opportunity for Distributed Unit Pooling in Real-World 5G Radio Access Networks* Preprint
   S. P. Rachuri, A. Gandhi, G. Jung, S. P. Narayanan, A. Zelezniak
   IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS 2025)
2. *EcoEdgeInfer: Dynamically Optimizing Latency and Sustainability for Inference on Edge Devices* IEEEXplore
   S. P. Rachuri, N. Shaik, M. Choksi, A. Gandhi
   ACM/IEEE Symposium on Edge Computing (SEC 2024)
3. *OVIDA: Orchestrator for Video analytics on Disaggregated Architecture* IEEEXplore
   M. Singh, S. P. Rachuri, B. B. Cao, A. Sharma, V. Bhumireddy, A. Gandhi, F. Bronzino, S. Das, S. Jain
   ACM/IEEE Symposium on Edge Computing (SEC 2024)
4. *Evaluating the energy impact of device parameters for DNN inference on edge* ACM Digital Library
   A. Dutt*, S. P. Rachuri*, A. Lobo, N. Shaik, A. Gandhi, Z. Liu *Joint first-authorship
   International Green and Sustainable Computing Conference (IGSC 2023)
5. *Optimizing Near-Data Processing for Spark* IEEEXplore
   S. P. Rachuri, A. Gantasala, P. Emanuel, A. Gandhi, R. Foley, P. Puhov, T. Gkountouvas, H. Lei
   IEEE International Conference on Distributed Computing Systems (ICDCS 2022)
6. *Decentralized Modular Architecture for Live Video Analytics at the Edge* ACM Digital Library
   S. P. Rachuri, F. Bronzino, S. Jain
   ACM Workshop on Hot Topics in Video Analytics and Intelligent Edges (located in MobiCom 2021)

7. *Traffic Splitting for Delay Jitter Control in Multi-access Systems* Springer Link
   M. Sahu, S. P. Rachuri, A. A. Ansari, A. A. Kherani
   Springer Telecommunication Systems Journal - Volume: 80, Issue: 4
8. *Multi-Armed Bandit-based Decentralized Computation Offloading in Fog-Enabled IoT* IEEEXplore
   S. Misra, S. P. Rachuri, P. K. Deb, A. Mukherjee
   IEEE Internet of Things Journal - Volume: 8, Issue: 12
9. *On limiting Delay and Jitter characteristics at application-layer of Multi-connected Systems* IEEEXplore
   M. Sahu, S. P. Rachuri, A. A. Ansari, D. Tandur, A. A. Kherani
   IEEE 5G World Forum (5GWF 2020)
10. *An SD-WAN Controller for Delay Jitter Minimization in Coded Multi-access Systems* IEEEXplore
    A. A. Ansari, S. P. Rachuri, A. A. Kherani, D. Tandur
    IEEE International Workshop on 5G and Future Wireless Technology (located in IEEE ANTS 2019)
11. *Network-Coded SD-WAN in Multi-Access Systems for Delay Jitter Control* IEEEXplore
    S. P. Rachuri, A. A. Ansari, D. Tandur, A. A. Kherani, S. Chouksey
    International Conference on Contemporary Computing and Informatics (iC3I 2019)

Patents

1. *Certification of renewable energy* Patent Application
   S. P. Rachuri, N. Boskov, M. Khan, J. Kangas
   UK Patent Application No. GB2507011.1 (Filed, under review)