

Sunil Kumar | Google Ph.D. Fellow

✉ sunilk@iiitd.ac.in | [in linkedin.com/in/justasunil](https://www.linkedin.com/in/justasunil) | [github justasunil.github.io](https://github.com/justasunil)

Professional Summary

6+ years of research experience in **high-performance computing (HPC)**, **performance engineering**, **workload characterization**, and **power-performance optimization** on multicore systems. Experienced in **PMU/PMC-based profiling and runtime-level resource management** to improve system throughput, fairness, and energy efficiency. Strong knowledge of computer architecture, including cache hierarchies, memory controllers, interconnects, DRAM bandwidth, and SIMD/matrix extensions (AVX/AMX). Extensive hands-on experience with Linux-based power management, including **DVFS**, **uncore frequency scaling (UFS)**, **power capping (PCAP)**, **thread tuning**, **SMT control**, and **processor power-state management**.

Tools and Technologies

Programming and Development: C/C++, Python, Java, Bash/Shell Scripting, CUDA, Fortran, Git, GNU Make, CMake

Performance Profiling and Telemetry: Linux Perf, Intel VTune, LIKWID, Intel/AMD MSR Counters, NVML, NCCL

Education

Indraprastha Institute of Information Technology, Delhi, India
Ph.D. in Computer Science and Engineering

July 2021 – Present
CourseWork CGPA: 8.69

Indraprastha Institute of Information Technology, Delhi, India
B.Tech. in Computer Science and Engineering

August 2017 – April 2021
CGPA: 7.0

Research Experience

Google Ph.D. Fellow @HiPeC Lab, IIIT Delhi, India

July 2021 – Present

- Developed a runtime-level power-management framework for multicore CPUs to optimize system throughput, power efficiency, and QoS under strict power budgets.
- Designed workload profiling and phase-aware DVFS/UFS algorithms to optimize core and uncore frequencies based on application characteristics.
- Developed dynamic power-cap allocation algorithms to improve throughput and power utilization across co-running applications while respecting system power constraints.
- Evaluated ten HPC applications on a 144-core, quad-socket Intel Xeon system, achieving **26.3% higher throughput and 27.7% better power efficiency** than the baseline, outperforming SLURM power scheduling by 16.4% and 18.6%, respectively.

Research Assistant @HiPeC Lab, IIIT Delhi, India

May 2021 – June 2021

- Developed a multi-node performance profiler for PMU-counter analysis of MPI+OpenMP hybrid applications.
- Analyzed computation and MPI communication overheads across an 8-node Chameleon Cloud cluster.
- Profiled exascale proxy applications to characterize MPI-rank load imbalance and runtime slack.

B.Tech. Thesis Researcher @IIIT-Delhi, India

June 2020 – April 2021

In collaboration with Lawrence Berkeley National Laboratory (LBNL), USA

- Developed an energy-efficient runtime for OpenMP and HCLib task-based applications, achieving **19.4% geometric energy savings with only 3.6% geometric performance slowdown** on a 40-core Intel Xeon Haswell server. Published at **SC'21**.

Awards and Achievements

Selected for Google DeepMind APAC Research Symposium	<i>2026</i>
Euro-Par Travel Scholarship	<i>2025</i>
HiPC SRS Travel Support	<i>2024</i>
HiPC SRS Travel Support	<i>2022</i>
Google Ph.D. Fellowship (one of six awardees globally in the systems and networking domain) [Link]	<i>2022</i>
Best Artifact Reproducibility Finalist (one of five finalists for an SC conference paper)	<i>2021</i>
SIGHPC Travel Grant (one of seven awardees globally for an SC conference paper) [Link]	<i>2021</i>
Outstanding B.Tech Project Award	<i>2021</i>

Peer Reviewed Publications

Sunil Kumar and Vivek Kumar, Reward-Driven Power Scheduling under Power Capping with DVFS-UFS Co-Optimization, in *Proceedings of the IEEE HiPC*, Bangalore, India, December 2026. [To appear]

Sunil Kumar, Vivek Kumar, and Sridutt Bhalachandra, Power Scheduling for Maximizing Throughput and Fairness in Co-running Applications, in *ACM Transactions on Architecture and Code Optimization (TACO)*, May 2026. [Link]

Sunil Kumar and Vivek Kumar, KarmaPM: Reward-Driven Power Manager, in *Proceedings of the European Conference on Parallel Processing (Euro-Par'25)*, Dresden, Germany, August 2025. [Link]

Sunil Kumar, Akshat Gupta, Vivek Kumar, and Sridutt Bhalachandra, Cuttlefish: Library for Achieving Energy Efficiency in Multicore Parallel Programs, in *Proceedings of the Supercomputing Conference (SC'21)*, St. Louis, MO, USA, November 2021. [Link]

Patents

Reward-based Bidirectional Power Scheduling on Multiprocessor Systems for Maximizing Throughput and Fairness

Filed with IIIT Delhi at the *Indian Patent Office*. Based on the paper KarmaPM.
Filing date – October 07, 2025

Student Research/Doctoral Symposium Publications

Sunil Kumar, "Power Scheduling on Multicore Multiprocessor Systems for Maximizing Throughput and Fairness," in *PhD Symposium, Euro-Par'25*, Dresden, Germany, August 2025.

Sunil Kumar, Vivek Kumar, and Sridutt Bhalachandra, "RPM: Reward Power Manager for Power Distribution over a Cluster," in *Student Research Symposium, HiPC'24*. [Link]

Sunil Kumar, Vivek Kumar, and Sridutt Bhalachandra, "Energy Efficiency under Limited Power Budget," in *Student Research Symposium, HiPC'22*. [Link]

Teaching Experience

GPU Programming Specialization *TAship Winter, Monsoon 2025*
Managed the system related issues occurred in CUDA programming and testing

Operating Systems *TAship Monsoon 2024*
Supported the course on kernel-level threading and memory management in C language

Parallel Runtimes for Modern Processors *TAship Winter 2024*
Conducted tutorial on GDB debugger and memory leakage detection using valgrind tool

Operating Systems *TAship Monsoon 2022*
Conducted weekly tutorials on system calls, GNU Makefile, Bash scripting, VM installations

Foundations of Parallel Programming *TAship Winter 2022*
Supported the course on developing runtimes for parallel programs and testing scalability on NUMA architecture

Advanced Programming *TAship Monsoon 2021*
Supported the course on OOPs concept in Java and android game development using javafx library

Invited Talks and Poster Presentations

Doctoral Symposium Lightning Talk *Euro-Par, August 2025*

Research Meetup Talk featuring Prof. Vivek Sarkar, Georgia Tech, USA *IIIT Delhi, July 2025*

SRS Lightning Talk *HiPC, December 2024*

Research Meetup Talk featuring Dr. Suraj Kumar, INRIA, France *IIIT Delhi, April 2024*

RIISE Poster Presentation *IIIT Delhi, May 2023*

RIISE Talk *IIIT Delhi, March 2022*

Invited Talk in Foundation of Parallel Programming Course *IIIT Delhi, March 2022*

Invited Alumni Talk at Summer School *IIIT Delhi, May 2021*

Guiding students on how to start research career in graduation

Invited Talk in Foundation of Parallel Programming Course *IIIT Delhi, March 2021*

Co-supervised B.Tech./M.Tech. Thesis Research Projects

- Thermal-Aware Speed Control of Cooling Fans in Multicore Servers** *M.Tech., 2026*
Optimizing cooling costs by operating individual fans at variable speeds instead of a uniform speed
- Power-Aware Execution of LLM Inference on CPU Accelerators** *B.Tech., 2026*
Dynamically allocating power between compute and memory resources based on their requirements during the prefill and decode phases
- Energy-Efficient SIMD Phases in LLM Inference on CPU Accelerators** *B.Tech., 2025*
Mapping SIMD layers to specific instruction sets (SSE, AVX, AVX2, AVX-512, and AMX) based on their power and performance characteristics
- Analysis of Thermal Variations across Diverse HPC and ML Workloads** *B.Tech., 2025*
Evaluated thermal variations across diverse HPC and ML workloads on a quad-socket Intel Xeon 3rd-generation server with 144 cores
- Profiling Thermal Sensors on Motherboards Using IPMI Tools** *B.Tech., 2025*
Analyzed platform-level power consumption, including compute and cooling power, using off-chip thermal sensors and linux IPMI tools
- Maximizing Performance under Power Capping Using Bayesian Optimization** *B.Tech., 2024*
Developed an ML-based Bayesian optimization approach to maximize application performance under a fixed power budget
- Energy Profiling of SIMD Programs Using OpenBLAS and Intel MKL** *B.Tech., 2024*
Performed comparative power and performance analysis of SIMD programs using OpenBLAS and Intel MKL
- Energy and Performance Profiling of GPUs Using NVML and CUPTI Libraries** *B.Tech., 2022*
Evaluated GPU energy and performance profiling techniques on NVIDIA G100 and V100 GPUs