GAHAN SARAIYA

+91 7359508173 ◇ gahansaraiya@gmail.com ◇ linkedin.com/in/gahan-saraiya ◇ github.com/gahan9

OBJECTIVE

With over 8+ years of experience as a Senior Software Engineer, currently working at AMD India Pvt Ltd. for managing AI/ML workload execution on HPC device (GPU). I possess a strong command over Python, HPC, UEFI Firmware, and DevOps, complemented by a thorough understanding of their workflows and architectures. My enthusiasm for open-source contribution and knowledge dissemination reflects my commitment to being a philanthropist in the realm of technology. I am eager to channel my expertise into the architectural design of optimized quantum machine learning algorithms tailored for multi-programming environments, as well as to explore the synergies between these cutting-edge technologies and DevOps practices.

SOFTWARE TOOLS

Languages	Python, C, C++, Shell Scripting, SQL, LaTeX, YAML
Frameworks	Tensorflow, Pytorch, TensorRT, Pytest, Edk2, PyCUDA, Pandas, UEFI, Selenium, Redfish
Architecture Design	Parallel Programming System Architecture, ROCm, HIP Programming
Tools and Services	GitHub Actions, Restful APIs, Teamcity, Terraform, Docker, Docker Compose, Jenkins, GiT
Cloud Technologies	GCP, AWS, Azure, Digital Ocean, JFrog Artifactory, TeamCity, JIRA, AWS Sage Maker
System Level IP	RAS, Compute Paritioning, Boot time optimization, CPER

EXPERIENCE

Senior System Software Designer at AMD India Pvt Ltd Mar 2024 - Present

- Conducting functional and performance benchmarking of HPC devices (GPUs) by developing and executing comprehensive test suites aligned with industry standards.
- Develop Load balanced optimal GPU performance criteria with thermal limits, power consumption, memory usage, and core utilization.
- Incorporating state-of-the-art industry standards design models, implementing and utilizing frameworks like TensorFlow, PyTorch, and GEMM operations.
- Provided precise steps for RAS (Reliability, Availability, and Serviceability) injection on MI2XX and MI3XX platforms.
- Designed precise steps for range of firmware capabilities, including partitioning, System Management Interface (SMI), HIP profiling, and parallel execution.
- Collaborating with a diverse team of engineers and researchers.

Senior Firmware Development Engineer at Intel Technology India Pvt Ltd Oct 2023 - Mar 2024

- Platform provisioning and infrastructure setup for mass automation of HPC products.
- Platform Orchestration Layer design of Universal Scalable Firmware (USF).
- Harnessing Azure OpenAI Generative AI adaptation for Firmware Development Productivity for assessment code review, guidance, basic security checks of code.
- GitHub Action and JIRA integration for managing Continuous deployment and test through CloudBees Jenkins with baremetal and docker nodes.
- Enhancing firmware-based boot-time performance for notebook devices.

Firmware Development Engineer at Intel Technology India Pvt Ltd Jun 2020 - Sept 2023

- Lead the next generation client platform and architecture design at USF POL alongside the domain ownership for Automation Framework, Platform Provisioning, Applied Generative AI & Prompt Engineering and Quantum Firmware Design.
- The end-to-end Software Development Life Cycle is managed, which includes licensing, legals, and IP planning, as well as applicable Code Security and Static Code Analysis.
- Jenkins configuration for Security tools - Snyk, Coverity
- Improving github code review hooks for python packaging and linting through bandit and flask as Jeninks pipeline
- automated testing of firmware functional validation of Firmware with TeamCity.
- Mentoring and knowledge development engagement activities.

Firmware Engineer Intern at Intel Technology India Pvt Ltd Jun 2019 - May 2020

- BIOS Firmware Development for Client products with the framework EDK2, UEFI Firmware, Involvement with assessing new debugging solutions and OpenBoard Package

PROJECTS

GPU Performance Analysis for Data Center Workloads 2024 - Present
Description: Solution design for in-band profiling through System Management Interface

- **Architected In-Band Profiling Solutions:** Designed and implemented robust in-band profiling mechanisms utilizing the System Management Interface (SMI) for comprehensive GPU telemetry.
- **Dynamic Workload Orchestration:** Developed systems for dynamic workload creation, intelligently adapting to real-time hardware resource availability (e.g., GPU utilization, memory) to optimize testing and analysis.
- **Comprehensive Hardware Metric Evaluation:** Engineered processes for capturing and analyzing critical hardware metrics, including GPU temperature, power consumption, and resource utilization, directly through SMI and in-band channels.
- **Granular Program Performance Analysis:** Implemented detailed program profiling techniques, including the capture and analysis of kernel call stack timings, to pinpoint performance bottlenecks and inefficiencies.
- **Advanced Analytical Metric Design:** Innovated a methodology to map multiple disparate key factors (e.g., thermal, power, utilization, kernel timings) into a unified, insightful data metric, significantly enhancing the depth and clarity of system performance analysis.

Prompt Engineering and Generative AI for firmware stack 2023 - 2024
Description: Solution design for accelerated silicon tape-in with accelerating firmware deliveries

- Work task summarization with Azure LLM Open AI services (GPT3.5 and GPT4)
- Code Review against standard static code analysis and common vulnerabilities finding
- Generation of Code Enhancement suggestion

Platform Provisioning with Pre-Production Reference Board 2023 - 2024
Description: Platform Provisioning with automating most common features for pre-production server processor such as:

- Remote BMC and BIOS Firmware Flashing
- OS Deployment from artifactory through Jenkins Pipeline
- Boot to EFI Shell and OS

- Configuration of Initial system setup with Jenkins pipeline steps with Baseboard management controller RedFish API

Universal Scalable Firmware 2022 - 2024
Description: Architecture design of Platform orchestration layer for Firmware Configuration

- Motivation of workgroup - Firmware Configuration – Past, Present, and Future
- Firmware Configuration

Firmware Automation Framework 2020 - Present
Description: Lead SDL for an automation foundation framework for high scaled volume of silicon devices with edk2 UEFI firmware;

- Seamless configuration of system without any manual intervention for BIOS configuration.
- SDL activities like legal compliance, security, licensing, static code analysis
- Continuous deployment and test with jenkins and tox configurations
- Release Artifact management to JFrog Artifactory using GitHub Actions
- https://github.com/intel/xml-cli

Multiview Video Summarization 2019 - 2020
Description: Summarizing video of multiple view angles of egocentric location in to one to generate effective video summary, minimizing overhead to explore multiple cam feed to discover accidents or anomalies.

Streaming Box 2017 - 2018
Description: An alternate streaming service to DTH and providing easy customized IPTV streaming service based on Kodi (formerly xbmc) Framework

- Backend Streaming Service, authentication and server security with Django Framework and REST API
- Web Portal to enable vendor customize Streaming Front-end UI for their users
- Cross compilation and Native compilation to enable various platform support as Android, Windows and MacOS

EDUCATION

M.Tech Computer Science Jul 2018 - Jun 2020
Nirma University, GPA: 8.49

B.E. in Computer Science and Engineering Aug 2013 - Apr 2017 Gujarat Technological University, CGPA: 8.08

12th Science, GSHSEB, Class XII Jun 2011 - Mar 2013
Science Stream, 74.92%

CERTIFICATIONS

2025 On-Device AI by Qualcomm & DeepLearning.ai
2025 Essential Terraform in AWS
2021 Product Assurance and Security Yellow Belt - Software
2020 Natural Language Processing with Classification and Vector Spaces Coursera S9DGGMCDMBGQ
2020 Improving Deep Neural Networks: Hyperparameter Tuning, Regularization and Optimization Coursera SMWQJKJYJW6G
2020 Structuring Machine Learning Projects - Coursera SEJML7H3TN5Z
2020 Neural Networks and Deep Learning - Coursera S2TETZHTLUSL
2020 Machine Learning - Coursera XVNXNPPZ3JMA

ACHIEVEMENTS

2023	Firmware Configuration Presentation at UEFI Developers Conference & Plugfest
2023	Tech Talk on Harnessing Multiprogramming across CPU, GPU and QPU (Quantum Processing Unit)
2021	Expert Lecture on CUDA Programming
2019-Present	4000+ reputation on Stackoverflow
2017-Present	Active volunteer on Peer Learning, Cultural and team building activity
2018	Rank #1 for Python challenges on hackerrank
2018	Rank #112 from 2 lacs+ participants in Techgig's National Coding Contest 2018

PUBLICATIONS

- HIP Parallel Programming with ROCm Lecture Series, YouTube 2025
- System Design for Parallel Programming Lecture Series, YouTube 2025
- Firmware Configuration – Past, Present, and Future. Presentation, UEFI DevCon 2023 2023