

Kumudha Narasimhan

✉ kumudhakn@gmail.com

📍 United Kingdom

🔗 kumudhan.github.io/

🌐 in/kumudha-narasimhan

SUMMARY

Technical and people leader with 10 years of domain experience in AI and HPC performance optimization across diverse architectures. Skilled in defining and executing software technology roadmaps, driving productization of research into production-ready solutions, and ensuring quality and integration maturity. Proven ability to engage with customers, open-source communities, and industry partners to gather requirements, influence adoption, and deliver high-impact solutions. Experienced in mentoring and managing engineering teams while aligning technical execution with business strategy.

SKILLS

Performance optimization

oneDNN, oneMKL, ONNX Runtime and using cuBLAS, cuDNN, CUTLASS, llama.cpp

Research projects Managed

EU projects - AERO (open cloud acceleration), SYCLOPS (AI acceleration)

Programming Paradigms

SYCL, CUDA, C++, Python

Leadership & Process

Agile methodologies, Technology roadmapping, OKR alignment, process audit, mentoring, requirements prioritization, competitive analysis

EXPERIENCE

Jan 2023 – Present
Edinburgh, UK

Software Engineering Manager

Intel Subsidiary - Codeplay Software

- Defined and executed multi-quarter roadmaps for AI software stacks, including llama.cpp optimizations, SYCL backend for CUTLASS, and portable backends for oneDNN/oneMath, ensuring alignment with Intel's product strategy.
- Collaborated with customers, partners, and EU project (AERO, SYCLOPS) members to gather requirements, track deliverables, and ensure timely delivery.
- Provide people-focused leadership to 18+ engineers across 4 Agile teams, supporting onboarding, mentoring, performance reviews, and HR transitions while fostering an inclusive, high-performance culture.
- Co-chaired the Math SIG (now under the UXL Linux Foundation), engaging open-source users of Intel's oneMath, presenting release updates, gathering feedback to shape internal roadmaps, and driving cross-vendor design discussions for oneMath.
- Partnered with cross-functional stakeholders (HR, Agile coaches, infrastructure, leadership) to improve organizational processes, track OKRs, and guide initiatives.

Nov 2020 – Dec 2022
Edinburgh, UK

Senior Staff Engineer

Codeplay Software

- Led SYCL integration in ONNX Runtime, ensuring optimized performance on heterogeneous hardware and enabling broader adoption of SYCL in other platforms.
- Enhanced engineering maturity by defining CI pipelines, reproducibility standards, and merge review protocols, contributing to ISO audit readiness.
- Represented company in external communities by co-authoring peer-reviewed publications, presenting at conferences (PMAM, TACO, P3HPC), and delivering customer demos.
- Contributed to error analysis and validation in performance portability experiments, ensuring reproducibility and correctness across architectures.

Sep 2019 – Nov 2020
Edinburgh, UK

Staff Software Engineer

Codeplay Software

- Developed GPU backends (cuBLAS, cuDNN) for oneMKL and oneDNN, increasing portability and adoption across hardware platforms.
- Collaborated with compiler teams and open-source projects to enable SYCL as a backend for ONNX Runtime, strengthening ecosystem engagement.
- Optimized neural network inference on custom hardware for real-time performance as part of a customer-facing project.
- Presented technical tutorials at major events (IWOCL, Intel Dev Summit), evangelizing SYCL and performance portability to the wider community.

Jul 2018 – Jul 2019
Bengaluru, India

Senior Software Engineer

Samsung Research India

- Analyzed *custom profilers and debuggers* for Samsung Tizen native and web applications, enhancing their functionality based on IoT team requests.
- Investigated and resolved issues reported in the Tizen Studio IDE to improve developer experience.
- Coached colleagues for the professional-level internal competitive coding exam, focusing on data structures and algorithms.

Jun 2012 – Jul 2015
Bengaluru, India

Technology Analyst at Investment Banking Division

Goldman Sachs

- Building and standardizing frameworks in C# .NET and Java for all applications developed within the division with *focus on scalability and performance*.
- Inherited several C# projects and led their sustenance, ongoing development, and enhancements, including re-architecting the resource discovery system, entitlement system, and provisioning system.
- Communicated effectively with internal and business stakeholders, inducted and mentored new joiners, and resolved multiple critical issues within deadlines while managing expectations.

EDUCATION

2018
Bangalore, India

M.Sc. (by research), High Performance Computing & Compilers

Indian Institute of Science (Advisor: Prof. Uday Reddy)

Thesis: Optimizing dense matrix computations with PolyMage

2012
Bangalore, India

Bachelor of Engineering, Computer Science and Engineering

M S Ramaiah Institute of Technology

Selected Publications

- Programming Model Extensions for General-Purpose Processing-In-Memory. [ISC 2024]
- A Performance Analysis of Leading Many-Core Technologies for Cellular Automata Execution [Europar 2023]
- A practical tile size selection model for affine loop nests. [ICS 2021]
- Towards Cross-Platform Performance Portability of DNN Models using SYCL. [P3HPC@SC 2020]

Full list of Publications - <https://dblp.org/pid/208/1873.html> ↗

List of presentation / talks - <https://kumudhan.github.io/#publications> ↗