Kirkland WA 98034

SUDHEER CHUNDURI

https://www.linkedin.com/in/sudheer-chunduri/
sudheerc.atwork@gmail.com

EXECUTIVE SUMMARY

My background in research and software development was in the space of system-level performance engineering, specifically focusing on co-design, characterization and monitoring of various production HPC interconnects and workloads. Earlier, I worked on node-level and core-level performance tuning. I have developed various analysis tools for system monitoring and contributed to production software packages.

I am a passionate hands-on engineer interested in contributing to performance optimization across the stack with potential focus on scaling optimization for AI and HPC workloads in Cloud and on-premises systems.

EMPLOYMENT

Software Engineer	HPC Platform readiness and Performance team, GCP	November 2023 - Present

- Ensure the GCP compute and interconnect NPIs such as H3 (Intel SPR), C3D (AMD Genoa) vm families are HPC-ready by validating the functionality and performance through various regression tests using microbenchmarks and HPC applications.
- Identify the various system level bottlenecks pertaining to GCP systems for HPC performance, identify and evaluate the potential mitigation strategies. Also, model the performance for new NPIs such as Cloud RDMA.

Computer Scientist	Argonne National Laboratory	April 2022 - November 2023

- Conduct research and development activities on performance optimization and analysis of large-scale HPC systems with a specific focus on Exascale interconnects and distributed-memory programming (MPI).
- Contribute to the software development of major software projects such as MPICH (https://github.com/sudheer9/mpich; https://github.com/sudheer9/yaksa) and Darshan (https://github.com/sudheer9/darshan) and analysis methods for HPC interconnect performance evaluation.

Assistant Computer Scientist	Argonne National Laboratory	June 2017 – March 2022

- Analyzed approaches that reduced run-to-run variability and improved performance up to 12% for real applications. These helped reduce overall network congestion, these are deployed at NERSC and ALCF.
- Involved in the development of a benchmark that has presently become a standard for measuring congestion effects on HPC interconnects and used as a metric for future system procurements.
- Implemented and optimized the MPICH support for AMD GPUs. This code base will be utilized on the MPI installations on US's first Exascale system, Frontier, and other systems.
- Authored top-tier conference papers regularly as a primary author and contributor.

Postdoctoral Appointee	Argonne National Laboratory	June 2016 – June 2017

- Designed analytical models for workload performance evaluation on Intel's many-core hardware.
- A key contributor to the machine acceptance of a supercomputer that was featured in the top-10 supercomputers. The work was performed in close collaboration with Intel and Cray under tight deadlines.

Researcher	IBM Research India	June 2013 – May 2016

- Key contributor on the performance evaluation and optimization of production applications in the Oil & Gas domain on IBM Blue Gene systems and coordinated in a multi-lab collaboration as co-lead contributor.
- Optimized MPI performance on POWER8 processor with an order of magnitude better latency than the IBM Spectrum MPI implementation.

Assistant Professor	SSSIHL India	June 2011 – May 2013

- Introduced two novel courses and redesigned existing courses in HPC area such as parallel programming with GPU accelerators and programming for performance using the NSF XESDE supercomputer allocations.
- Innovated on the course delivery in collaboration with Industry experts and advised students in securing four Best poster and presentation awards at HiPC conference in 2012.

Ph.D., Computer Science   SSSIHL India   **Sep 2007 – May 2013**
- Topology and routing aware mapping tool for massively parallel processors, specifically relevant for Tori
- Topology aware implementation of Global Arrays data management for QMCPACK
- Optimal dynamic load balancing algorithm for large-scale codes involving identical computational tasks
- Optimizing assignment of threads to SPEs on IBM Cell BE for improved communication performance
- Advisor: Prof. Ashok Srinivasan, Florida State University

M.Tech., Computer Science   SSSIHL India   **June 2004 – May 2006**
- Thesis: Implemented a new file system in MINIX Operating system on IBM PowerPC embedded processor

TECHNICAL EXPERIENCE

- Experience in evaluation of HPC systems and especially interconnects through analytical, simulation, and experimental methods. The goal of the evaluation is system co-design, guiding architecture tradeoffs, workload characterization, and performance analysis tool development for Exascale systems.
- Defined and developed evaluations and optimizations for various Exascale interconnects such as HPE Slingshot and various other previous generation interconnects like Intel OPA and Cray Aries.
- Practical experience in performance optimization on various CPU and GPU processor architectures for on-node performance and interconnects for scale-out performance.

ADDITIONAL EXPERIENCE AND AWARDS

- Recognition with a spot bonus for the key role played in the Diorite NIC hardware issue (RX Wedge) that was a blocker for the GA of H3 and C3D vm families that use Diorite NICs.
- Impact Argonne Award for Extraordinary Efforts, Argonne National Laboratory, 2020 and 2021
- Certificate of Appreciation from the Director of IBM Research India for contribution towards an Oil & Gas customer project, 2016
- Mentored a post-doctoral candidate and hosted several Ph.D. students for internships
- Involved in various multi-institute collaborations spanning across labs, hardware vendors, and universities
- Served as a reviewer for various conferences, journals, and DoE specific reviewing activities

SELECTED PUBLICATIONS

Sudheer Chunduri, Kevin Harms, Taylor Groves, Peter Mendygral, Justs Zarins, Michele Weiland, and Yasaman Ghadar. Performance Evaluation of Adaptive Routing on Dragonfly-based Production Systems. In Proceedings of the 35th IEEE International Parallel & Distributed Processing Symposium, IPDPS 21, May 2021.

Sudheer Chunduri*, Taylor Groves*, Peter Mendygral*, Brian Austin, Jacob Balma, Krishna Kandalla, Kalyan Kumaran, Glenn Lockwood, Scott Parker, Steven Warren, Nathan Wichmann, and Nicholas Wright. GPCNeT: Designing a Benchmark Suite for Inducing and Measuring Contention in HPC Networks. In Proceedings of the International Conference for High Performance Computing, Networking, Storage, and Analysis, SC 19, Nov 2019.

Sudheer Chunduri, Elise Jennings, Kevin Harms, Christopher Knight, and Scott Parker. A generalized statistics-based model for predicting network-induced variability. In 10th International Workshop on Performance Modeling, Benchmarking and Simulation of High Performance Computer Systems (PMBS'19), SC'19, Nov 2019.

Sudheer Chunduri, Scott Parker, Pavan Balaji, Kevin Harms, and Kalyan Kumaran. Characterization of MPI Usage on a Production Supercomputer. In Proceedings of the International Conference for High Performance Computing, Networking, Storage, and Analysis, SC 18, pages 30:1–30:15, Piscataway, NJ, USA, 2018.

Sudheer Chunduri, Kevin Harms, Scott Parker, Vitali Morozov, Samuel Oshin, Naveen Cherukuri, and Kalyan Kumaran. Run-to-run Variability on Xeon Phi Based Cray XC Systems. In Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, SC'17, pages 52:1–52:13, New York, NY, USA, 2017. ACM.

SOFTWARE DEVELOPMENT SKILLS

C; Python; C++; MPI; OpenMP; CUDA; HIP; HPC performance and debugging tools;