Somesh Singh

Post-doctoral Researcher
Team ROMA, INRIA
LIP, ENS de Lyon
69007 Lyon
France

Email: somesh.singh1992@gmail.com
somesh.singh@ens-lyon.fr
Webpage: https://ssomesh.github.io
Github: https://github.com/ssomesh/

Research Interests

High-Performance Computing; Parallel Computing; High-Performance Graph Analytics; Sparse Tensor Computations.

Area of Research

My research interests span the broad areas of high-performance computing and parallel computing. My current focus is on making the processing of *irregular* workloads on parallel platforms *performant*, and bridging the gap between programmability and efficiency on such platforms.

Graduate Courses

Mathematical Concepts for Computer Science, Advanced Data Structures and Algorithms, Computer Architecture, High-Performance Parallel Computing, Program Analysis, Modern Compilers, Indexing and Searching in Large Datasets, Probability and Computing, Pattern Recognition and Machine Learning, Digital Design Verification, CAD for VLSI Systems.

Programming Languages

- Fluent: C/C++, CUDA, OpenMP
- Familiar: OpenCL, Python, MATLAB, LLVM

Publications

- **Somesh Singh**, Tejas Shah and Rupesh Nasre, "ParTBC: Faster Estimation of Top-k Betweenness Centrality Vertices on GPU", *ACM Transactions on Design Automation of Electronic Systems (TODAES)*, vol. 27, no. 2, pp. 12:1–12:25, 2021. https://doi.org/10.1145/3486613.

- **Somesh Singh** and Rupesh Nasre, "Graffix: Efficient Graph Processing with a Tinge of GPU-Specific Approximations", *49th International Conference on Parallel Processing (ICPP 2020)*, pp. 23:1–23:11. https://doi.org/10.1145/3404397.3404406.

- **Somesh Singh** and Rupesh Nasre, "Optimizing Graph Processing on GPUs using Approximate Computing: Poster", *24th ACM SIGPLAN Annual Symposium on Principles and Practice of Parallel Programming (PPoPP 2019)*, pp. 395–396. https://doi.org/10.1145/3293883.3295736.

- **Somesh Singh** and Rupesh Nasre, "Scalable and Performant Graph Processing on GPUs using Approximate Computing", *IEEE Transactions on Multi-Scale Computing Systems (TMSCS)*, vol. 4, no. 3, pp. 190–203, 2018. https://doi.org/10.1109/TMSCS.2018.2795543.

- R. De Maria, J. Andersson, V.K.B. Olsen, L. Field, M. Giovannozzi, P.D. Hermes, N. Høimyr, S. Kostoglou, G. Iadarola, E. McIntosh, A. Mereghetti, J. Molson, D. Pellegrini, T. Persson, M. Schwinzerl, E.H. Maclean, K.N. Sjobak, I. Zacharov and **S. Singh**, "SixTrack V and runtime environment", *International Journal of Modern Physics A (IJMPA)*, vol. 34, no. 36, 1942035, 2019. https://doi.org/10.1142/S0217751X19420351. (**Invited paper**)

- R. De Maria, J. Andersson, V.K.B. Olsen, L. Field, M. Giovannozzi, P.D. Hermes, N. Høimyr, S. Kostoglou, G. Iadarola, E. McIntosh, A. Mereghetti, J. Molson, D. Pellegrini, T. Persson, M. Schwinzerl, E.H. Maclean, K.N. Sjobak, I. Zacharov and **S. Singh**, "SixTrack Project: Status, Runtime Environment and New Developments", *13th International Computational Accelerator Physics Conference (ICAP 2018)*, pp. 172–178. https://doi.org/10.18429/JACoW-ICAP2018-TUPAF02.

PROFESSIONAL EXPERIENCE

- Post-doctoral researcher with INRIA at LIP, ENS de Lyon. *November 2021 – Present*
- Post-doctoral researcher with CNRS at LIP, ENS de Lyon. *September 2021 – October 2021*
- Research Intern *September 2020 – December 2020*
  - Microsoft Research, Bengaluru, India.
  - Worked on parallelizing approximate nearest neighbor search (ANN) algorithm on GPU. We achieved a throughput of 30K queries per second, providing high 90s recall for ANN on billion-scale data.
  - Technologies involved: C/C++, CUDA
- Research Intern *June 2020 – August 2020*
  - Intel Research, Bengaluru, India.
  - Worked on optimizing parallel graph analytics for shared memory systems with Optane Persistent Memory Modules (PMMs). Developed strategies for data placement and graph partitioning on tiered memory systems for improving the performance of graph applications.
  - Technologies involved: C/C++, OpenMP

ACCOMPLISHMENTS AND AWARDS

- Google Summer of Code 2018 participant with CERN-HSF.
  - Developed a standalone optimized parallel implementation of (a part of) SixTrackLib, a particle-tracking library.
  - Technologies involved: C/C++, OpenCL 1.2
- Google Summer of Code 2017 participant with CERN-HSF.
  - Developed SALLOC, an arena based memory allocator for SIMT architectures, with support for a thread-safe C++ STL-style *vector* container.
  - Technologies involved: C/C++, CUDA
- Secured 4th place in HiPC 2016 Student Parallel Programming Challenge (Intel Xeon-Phi track) (Team of 2).
  - Implemented an efficient scheme for labeling connected clusters in a 3-dimensional grid using the Union-Find data structure. All points in a cluster were to be assigned the same label.
  - Technologies involved: C++, OpenMP.
- Secured 4th place in HiPC 2015 Student Parallel Programming Challenge (Intel Xeon-Phi track) (Team of 3).
  - Implemented an efficient parallel version of the KMeans++ algorithm for assigning membership to each data point in a high dimensional unlabeled data set, to maximize the Dunn-index.
  - Technologies involved: C++, OpenMP.
- Awarded ACM SIGPLAN PAC grant for attending PPoPP 2019.
- Awarded the STAR TA award for contributions as a Teaching Assistant to the course "GPU Programming" for the period July – November 2017.

SERVICES

- Reviewer: IEEE TPDS (2021), ParCo (2021), CCPE (2021), INAE Letters (2018), IEEE ESL (2017).
- Extended Review Committee: ECOOP 2022.
- Artifact Evaluation Committee: PLDI 2022, ESOP 2022, ECOOP 2022, ECOOP 2021, PPoPP 2021, ECOOP 2020, PPoPP 2018.
- Student Volunteer (SV) for SPLASH/ECOOP 2020.
- Organizer for CUDA Workshop during Exebit 2018 at the Indian Institute of Technology Madras.

Talks

- [Virtual] Talk at the "Parallel Processing in Data Science Applications" mini-symposium within SIAM PP'22.   February 2022
- [Virtual] Talk at FORTH-ICS, Greece.   July 2021
- [Virtual] Talk at Queen's University Belfast, Ireland.   May 2021
- [Virtual] Talk at Lawrence Berkeley National Laboratory, USA.   March 2021

Projects and Previous Internships

Course Projects

- Supergraph Containment Search (`Team of 2`).   October – November 2016
  - Implemented an efficient supergraph containment search technique using the *filtering* and *verification* framework, in C++.
  - Optimized the online processing time required for finding the (small) graphs, in the database, that are present in the (large) query graph. Our team **won** the contest for minimizing the querying time over 200 query graphs for a database containing 70K graphs during the *verification* phase.

- Five stage RISC pipeline.   October – November 2015
  - Implemented a five stage pipeline for a RISC processor with operand-forwarding using Bluespec.

- Domain Specific Language For Circuit Design (`Team of 2`).   March – April 2015
  - Implemented an internal DSL, in Python, that allows specifying a boolean expression in the Disjunctive Normal Form (DNF), and supports generating a netlist, comprising AND, OR, NOT logic gates, for the minimal form of the boolean expression.

Other Projects (Undergraduate)

- Object Tracking in Video Using Parallel Computing   March – April 2014
  - Implemented a sum of absolute differences (SAD) based parallel block-matching algorithm for tracking the object of interest in a video, in CUDA.

- Online Gaming   April 2013
  - Designed an interactive single player online game, using HTML5 and JavaScript, that can be played on an internet browser.

Internships (Undergraduate)

- RTOS based Embedded Software Design and Verification of Serial Communication
                                                                           June – July 2013
  - Intern at Larsen and Toubro SIPL, Bengaluru, India.
  - Implemented a device driver for an external UART device for the RTOS, VxWorks; established communication between host-PC and target PowerPC board using a serial protocol.

- Autonomous Mobile Robots – A Study   May - June 2012
  - Intern at the Indian Institute of Technology Delhi, India.
  - Programmed a mobile robot (iRobot) to move autonomously in an unstructured environment, using 'kinect' for visual feedback, using the Player/Stage software.

MENTORING

- Mentor at C-DAC HPC Hackathon 2021.
- Mentor for a masters project   *2017-18*
  - Objective: Faster estimation of top-k betweenness centrality vertices in a graph, aided by approximate computing.
  - Technologies involved: Java
- Mentored two undergraduate students. They worked on:
  - Graph-based Image Segmentation   *December 2016*
    * Modeled an image as a weighted graph and performed image segmentation using various graph algorithmic techniques on the underlying graph.
    * Technologies involved: C++, OpenCV
  - Image Segmentation and Object Tracking on GPU   *May - June 2015*
    * Implemented a parallel seed-based region growing algorithm for image segmentation.
    * Technologies involved: C/C++, CUDA, OpenCV

EDUCATION

Doctor of Philosophy (Ph.D.) + Master of Science   *July 2014 - June 2021*
Indian Institute of Technology Madras
- Thesis: "Scalable and Performant Graph Processing on GPU using Approximate Computing"
- Adviser: Dr. Rupesh Nasre

Bachelor of Technology in Computer Science and Engineering   *July 2010 - May 2014*
National Institute of Technology Uttarakhand