Vivek L. Kale, PhD

Phone: +01 217-369-7996 *Email*: vivek.lkale@gmail.com *Web*: http://vlkale.github.io
U.S. Citizen

Education

Bachelor of Science, Computer Science
University of Illinois at Urbana-Champaign, May 2007

Doctor of Philosophy, Computer Science
University of Illinois at Urbana-Champaign, May 2015
Advisor: William D. Gropp

Honors

- SC 2017 Early Career Program Invitee and Participant
- 2015 Heidelberg Laureate Forum Invitee and Participant.
- Fellow at Lawrence Livermore National Laboratory through Lawrence Scholar Program.

Courses Taken

Designing and Building Scientific Applications, Parallel Software Patterns, Program Optimization, Parallel Computer Architecture, Algorithms, Artificial Intelligence, Software Engineering

Technical Skills

Languages: C, C++, python, Fortran, bash, csh, CUDA

Tools: LaTeX, gnuplot, emacs, autoconf, cmake, svn

Libraries for Parallelism: POSIX threads (Pthreads), MPI (mpich3), OpenMP (gomp, llvm), OpenACC (pgi)

Performance Profiling Tools and Libraries: hpcToolkit, PAPI, nvprof, gprof

Experience

Sandia National Laboratories Senior Member of Technical Staff August 2022 - present

- Developing and testing features in the US DoE's LLVM's OpenMP implementation.
- Contributing to OpenMP 6.0 Specification, specifically on topics of affinity, loop transformations, accelerators and tasking.
- Prototyping tunable locality-aware loop scheduling strategy features for OpenMP, and generally user-defined loop schedules, for LLVM's OpenMP implementation.
- Owner of Kokkos Software Ecosystem's Kokkos Tools, which provides profiling and debugging capabilities for Kokkos programs (for performance portable parallel programs) as well as sophisticated auto-tuning and performance analysis capabilities.

- Contributor to the DOE ASCR Xstack project on automated test generation for parallel programs via LLVM. Developing a source-to-source translator via the ROSE compiler plugin for the LLVM's clangASTRewriter to translate a Kokkos program to a Kokkos Model (simplified version of Kokkos) program for analysis by LLVM's Klee symbolic execution library.

Brookhaven National Laboratory Computational Scientist May 2019 - August 2022

- Contributed to developing an LLVM OpenMP implementation, specifically the OpenMP implementation's compiler and its runtime, targetted for Department of Energy's upcoming Exascale Supercomputer platforms.
- Designed and implemented OpenMP task-to-multiGPU scheduling strategies to improve within-node load balancing of applications running on supercomputers having multiple GPUs per node.
- Developed tunable locality-aware loop scheduling strategies, and generally user-defined loop schedules, in LLVM's OpenMP implementation, in the context of MPI+OpenMP applications runnning on supercomputers having multicore processors and GPUs.
- Contributed to the OpenMP Language Committee to support OpenMP parallelization on multiple GPUs of a node for C, C++ and Fortran, and for user-defined schedules in OpenMP.
- Developed benchmarks and evaluating OpenMP implementations, e.g., LLVM's OpenMP, NVIDIA's OpenMP, on Exascale Supercomputers.
- Led hackathons (including virtual) for using OpenMP on Department of Energy's Exascale Supercomputers.
- Served as Technical Project Manager for US DoE Exascale Computing Project's SOLLVE project to develop LLVM's OpenMP.
- Represented Brookhaven National Laboratory in the OpenMP Architecture Review Board.

Charmworks, Inc. Software Developer June 2018 - April 2019

- Collaborated with Lawrence Livermore National Lab on a proposal for a synergistic loop scheduling and load balancing strategy.
- Worked on making User-defined Loop Scheduling portable across different parallel programming library, done with Oak Ridge National Lab through DoE Exascale Computing Program.
- Added examples of loop scheduling in OpenMP in the Examples section of OpenMP Specification.
- Worked on a NSF startup SBIR proposal for loop scheduling for desktop computers.
- Collaborated on developing a proposal to add an OpenMP User-defined Schedule to the OpenMP specification based on an OpenMPCon 2017 paper, presenting a proposal at the OpenMP F2F in Santa Clara and the upcoming F2F in Toronto.
- Worked on papers for User-defined Loop Scheduling for publication.
- Assisted with slides for pitch and marketing of Charm++ software, and providing feedback for tutorials on Charm++.
- Integrated a shared memory library for sophisticated loop scheduling strategies, including some based on my dissertation, into the current version of Charm++.

University of Southern California / ISI Computer Scientist December 2016 - June 2018

- Worked with postdoc from LLNL on a proposal to study techniques that combine loop scheduling and load balancing to improve performance of scientific applications.
- Worked with OpenMP Language Committee to support user-defined loop schedules in OpenMP.

- Translated an x-ray tomography code written in Matlab code to C code and then parallelizing it to run on a supercomputer having nodes with GPGPUs.
- Made modifications to LLVM compiler to support new OpenMP loop schedules.
- Ensured external network infrastructure to support transfer of application code's input data files were adequate for an application code's efficient execution using the Globus Toolkit.
- Worked in team to manage computational performance aspects of running an application program involving Fast Fourier Transformation and image reconstruction algorithms.

Charmworks, Inc. Developer Jan. 2016 - Nov. 2016

- Implemented mixed static/dynamic loop scheduling strategies within Charm++'s thread scheduling library.
- Helped to improve portability of Charm++ to a variety of platforms.
- Assisted with business aspects of a high-tech startup.

University of Illinois Postdoctoral Associate Jul. 2015 – Dec. 2015

- Developed library that allows application programmers to use strategies from dissertation.
- Adapted a plasma physics application code to work on a GPGPU processor and Intel Xeon Phi.
- Incorporated over-decomposition and locality-aware scheduling into strategies from dissertation.

Lawrence Livermore Nat'l Lab Lawrence Scholar Feb. 2012 – Jun. 2014

- Measured MPI communication delays for micro-benchmarks codes run on supercomputers and worked to find tools to measure dequeue overheads of OpenMP loop schedulers.
- Created a software system for automated performance optimization and application programmer usability of low-overhead hybrid scheduling strategies.
- Developed a ROSE-based custom compiler for automatically transforming MPI+OpenMP applications to use low-overhead scheduling techniques and runtime.
- Assessed further opportunities for performance improvement of low-overhead schedulers, including improvement of spatial locality of low-overhead schedulers.

Lawrence Livermore Nat'l Lab Scholar Jun. 2011 - Sep. 2011

- Experimented with different OpenMP parameters of implemented MPI+OpenMP application code to understand performance optimizations on LLNL supercomputers.
- Developed software design for low-overhead loop scheduling library based on libgomp software design.

Lawrence Berkeley Nat'l Lab Summer Scholar Aug. 2010 - Sep. 2010

- Analyzed results for the performance tests developed on NERSC machines.
- Compared with collectives in reference to MPI (mpich2) runtime system.

Lawrence Livermore Nat'l Lab Scholar May. 2010 - Aug. 2010

- Modified libgomp runtime system in order to integrate low-overhead schedulers within it.
- Developed an algorithm multi-stage low-overhead loop scheduler with each stage associated with a level in the memory hierarchy, allowing for MPI-shared memory extensions to be used in conjunction with the low-overhead loop scheduling strategies.

Goldman-Sachs Summer Analyst Jun. '09 – Sep. '09

- Wrote code for testing trading system infrastructure functions under extreme market conditions.

- Analyzed performance bottlenecks of system infrastructure functions.

Proteus Technologies, LLC Software Developer Aug. 2007 – Apr. 2008

- Primarily responsible for developing, testing and documenting a service-oriented software application for health and status monitoring of large-scale parallel and distributed networked systems.
- Developed company standards for software development (System Requirements Specifications, Design Documentation).
- Designed and implemented algorithms for cost optimization applications. Used dynamic programming, discrete optimization heuristics, and APIs.

List of Publications

Papers

1. Mathialakan Thavappiragasasm and Vivek Kale. ***CPU-GPU Performance Tuning for Improving Performance of Modern Scientific Applications on Exascale Supercomputers***. IEEE HiPC 2023. Goa, India. December 18-21, 2023.
2. Shravan Kale, Kevin Huck, David Boehme, Vanessa Surjadidjaja and Vivek Kale. ***Performance Analysis and Auto-tuning Tools for Performance Portable Parallel Programs***. 2023 ACM/IEEE International Conference for High Performance Computing Networking, Storage, and Analysis. Denver, CO, USA. November 12-17, 2023.
3. Vivek Kale, Vanessa Surjadidjaja, Christian Trott and James Brandt. ***Data Order Reduction for Performance Monitoring of Supercomputers via the Kokkos Tools Sampler Utility*** LDMSCon 2023. Boston, MA, USA. June 13-15, 2023.
4. Vivek Kale and Shyamali Mukherjee. ***Tools to Rapidly Develop Sophisticated HPC Software Libraries*** SIAM Computational Science and Engineering Conference 2023. Amsterdam, Netherlands. March 2, 2023.
5. Mathialakan Thavappiragasam and Vivek Kale. ***OpenMP's Asynchronous Offloading for All-pairs Shortest Path Graph Algorithms on GPUs*** HiPar 2022 Workshop at The 2022 ACM/IEEE International Conference for High Performance Computing Networking, Storage, and Analysis. November 16, 2022. Dallas, Texas, USA.
6. Mathialakan Thavappiragasam, Vivek Kale, Oscar Hernandez and Ada Sedova. ***Addressing Load Imbalance in Bioinformatics and Biomedical Applications: Efficient Scheduling across Multiple GPUs*** In Proceedings of 12th International Workshop on High Performance Bioinformatics and Biomedicine. December 9, 2021. Houston, Texas, USA.
7. Barbara Chapman, Buu Pham, Charlene Yang, Christopher Daley, Colleen Bertoni, Dhruva Kulkarni, Dhruva Kulkarni, Dossay Oryspayev, Ed D'Azevedo, Helen He, Johannes Doerfert, Keren Zhou, Kiran Ravikumar, Mark Gordon, Mauro Del Ben, Meifeng Lin, Melisa Alkan, Michael Kruse, Oscar Hernandez, P. K. Yeung, Paul Lin, Peng Xu, Swaroop Pophale, Tosaporn Sattasathuchana, Vivek Kale, William Huhn and Dhruva Kulkarni. ***Outcomes of OpenMP Hackathon: OpenMP Application Experiences with the Offloading Model: Part 1*** In Proceedings of 17th International Workshop on OpenMP, IWOMP 2021, Bristol, UK, September 14–16, 2021.
8. Barbara Chapman, Buu Pham, Charlene Yang, Christopher Daley, Colleen Bertoni, Dhruva Kulkarni, Dhruva Kulkarni, Dossay Oryspayev, Ed D'Azevedo, Helen He, Johannes Doerfert, Keren Zhou, Kiran Ravikumar, Mark Gordon, Mauro Del Ben, Meifeng Lin, Melisa Alkan, Michael Kruse, Oscar Hernandez, P. K. Yeung, Paul Lin, Peng Xu, Swaroop Pophale, Tosaporn Sattasathuchana, Vivek Kale, William Huhn and Dhruva Kulkarni. ***Outcomes of OpenMP Hackathon: OpenMP Application Experiences with the Offloading Model: Part***

2 In Proceedings of 17th International Workshop on OpenMP, IWOMP 2021, Bristol, UK, September 14–16, 2021.

9. Seonmyeong Bak, Colleen Bertoni, Swen Boehm, Reuben Budiardja, Barbara M. Chapman, Johannes Doerfert, Markus Eisenbach, Hal Finkel, Oscar Hernandez, Joseph Huber, Shintaro Iwasaki, Vivek Kale, Paul R.C. Kent, JaeHyuk Kwack, Meifeng Lin, Piotr Luszczek, Ye Luo, Buu Pham and P.K. Yeung. ***OpenMP Application Experiences: Porting to Accelerated Nodes***. In Journal of Parallel Computing. October 23rd, 2021.
10. Vivek Kale, Wenbin Lu, Anthony Curtis, Abid Malik, Barbara Chapman and Oscar Hernandez. **Toward Supporting MultiGPU targets via taskloop and User-defined Schedules**. Proceedings of the 2020 International Workshop of OpenMP. September 23-25, 2020. Austin, USA. (virtual)
11. Jonas H Müller Korndörfer, Florina M. Ciorba, Akan Yilmaz, Christian Iwainsky, Johannes Doerfert, Hal Finkel, Vivek Kale, Michael Klemm. ***A Runtime Approach for Dynamic Load Balancing of OpenMP Parallel Loops in LLVM***. The International Conference for High Performance Computing Networking, Storage, and Analysis. November 19, 2019. Denver, Colorado, USA.
12. Vivek Kale, Christian Iwainsky, Michael Klemm, Jonas H. Muller Korndorfer, Florina M. Ciorba. ***Toward a Standard Interface for User-Defined Scheduling in OpenMP***. International Workshop on OpenMP. September 23, 2019. Auckland, New Zealand.
13. Vivek Kale and William D. Gropp. ***Composing Low-Overhead Scheduling Strategies for Improving Performance of Scientific Applications***. IWOMP 2015. October 2015. Aachen, Germany.
14. Simplice Donfack, Vivek Kale, Laura Grigori and William D. Gropp. ***Hybrid Static/Dynamic Scheduling for Already Optimized Dense Matrix Factorizations***. IPDPS 2012. May 2012. Shanghai, China.
15. Vivek Kale, Abhinav Bhatele and William D. Gropp. ***Weighted Locality-sensitive Scheduling for Noise Mitigation on Multicore Clusters***. HiPC 2011. December 2011. Bangalore, India.
16. Vivek Kale and William D. Gropp. ***Load Balancing for Regular Meshes on a Cluster of SMPs with MPI***. EuroMPI 2010. September 2010. Stuttgart, Germany. *(Selected as a Best Paper)*
17. Torsten Hoefler, James Dinan, Darius Buntinas, Pavan Balaji, Brian Barrett, Ron Brightwell, William Gropp, Vivek Kale and Rajeev Thakur. ***MPI+MPI: A New Hybrid Approach to Parallel Programming with MPI Plus Shared Memory***. EuroMPI 2012. September 2012. Madrid, Spain.
18. Amanda Randles, Vivek Kale, Jeff Hammond, William D. Gropp and Efthimios Kaxiras. ***Performance Analysis of the Lattice Boltzmann Model Beyond Navier-Stokes***. IPDPS 2013. May 2013. Boston, USA.
19. Vivek Kale. ***Towards Using and Improving the NAS Parallel Benchmarks: A Parallel Patterns Approach***. ParaPLoP 2010. April 2010. Carefree, USA.
20. Vivek Kale and Edgar Solomonik. ***Parallel Sorting Pattern***. ParaPLoP 2010. April 2010. Carefree, USA.
21. Vivek Kale. ***The Correlation between Parallel Patterns and the NAS Parallel Benchmarks***. ICSE 2010. May 2010. Johannesberg, South Africa.

Extended Abstracts

1. Vivek Kale and Martin Kong. ***Enhancing Support in OpenMP to Improve Data Locality in Application Programs Using Task Scheduling***. OpenMPCon 2018. September 2018. Barcelona, Spain.

2. Vivek Kale and William D. Gropp. ***A User-defined Schedule for OpenMP***. Extended Abstract. OpenMPCon 2017. September 2017. New York, USA.
3. Vivek Kale. ***A Pattern Language for Dynamic Scheduling***. ParaPLoP 2011. May 2011. Carefree, USA.

Posters

1. Raul Torres, Vivek Kale, Abid Malik, Tom Scogland, Roger Ferrer and Barbara M. Chapman. ***Support in OpenMP for Multi-GPU Parallelism***. The International Conference for High Performance Computing Networking, Storage, and Analysis. November 19, 2021. St. Louis, Missouri, USA.
2. Vivek Kale and Oscar Hernandez. ***Performance Portability of User-defined Loop Schedules***. DOE PPP 2019. April 2019. Denver, USA.
3. Vivek Kale, Harshitha Menon and Karthik Senthil. ***Adaptive Loop Scheduling with Charm++ to Improve Performance of Scientific Applications***. SC 2017. November 2017. Denver, USA. *(Selected as a Candidate for Best Poster)*
4. Vivek Kale, Simplice Donfack, Laura Grigori and William D. Gropp. ***Balancing the Trade-off Between Load Balancing and Locality to Improve Performance of Scientific Applications***. SC 2014. November 2014. New Orleans, USA.
5. Vivek Kale, Amanda Randles and William D. Gropp. ***Locality-Optimized Mixed Static/Dynamic Scheduling for Load Balancing on SMPs***. EuroMPI/ASIA 2014. September 2014. Kyoto, Japan.

Membership of Organizations

- Society for Industrial and Applied Mathematics
- Association for Computing Machinery
- Institute of Electrical and Electronics Engineers
- Association for Computing Machinery's SIGHPC

Teaching Experience

1. Teaching Assistant for Programming Studio (CS 242) at University of Illinois at Urbana-Champaign for Spring 2009.
2. Teaching Assistant for Programming Studio (CS 242) at University of Illinois at Urbana-Champaign for Fall 2009.
3. Teaching Assistant for Programming Studio (CS 242) at University of Illinois at Urbana-Champaign for Spring 2010.

Services

- Reviewer for 2017 Elsevier Parallel Computing Journal
- Member of Organization Committee for SC 2018's Early Career Program
- Member of Selection Committee for SC 2018's Experiencing HPC for Undergraduates Program