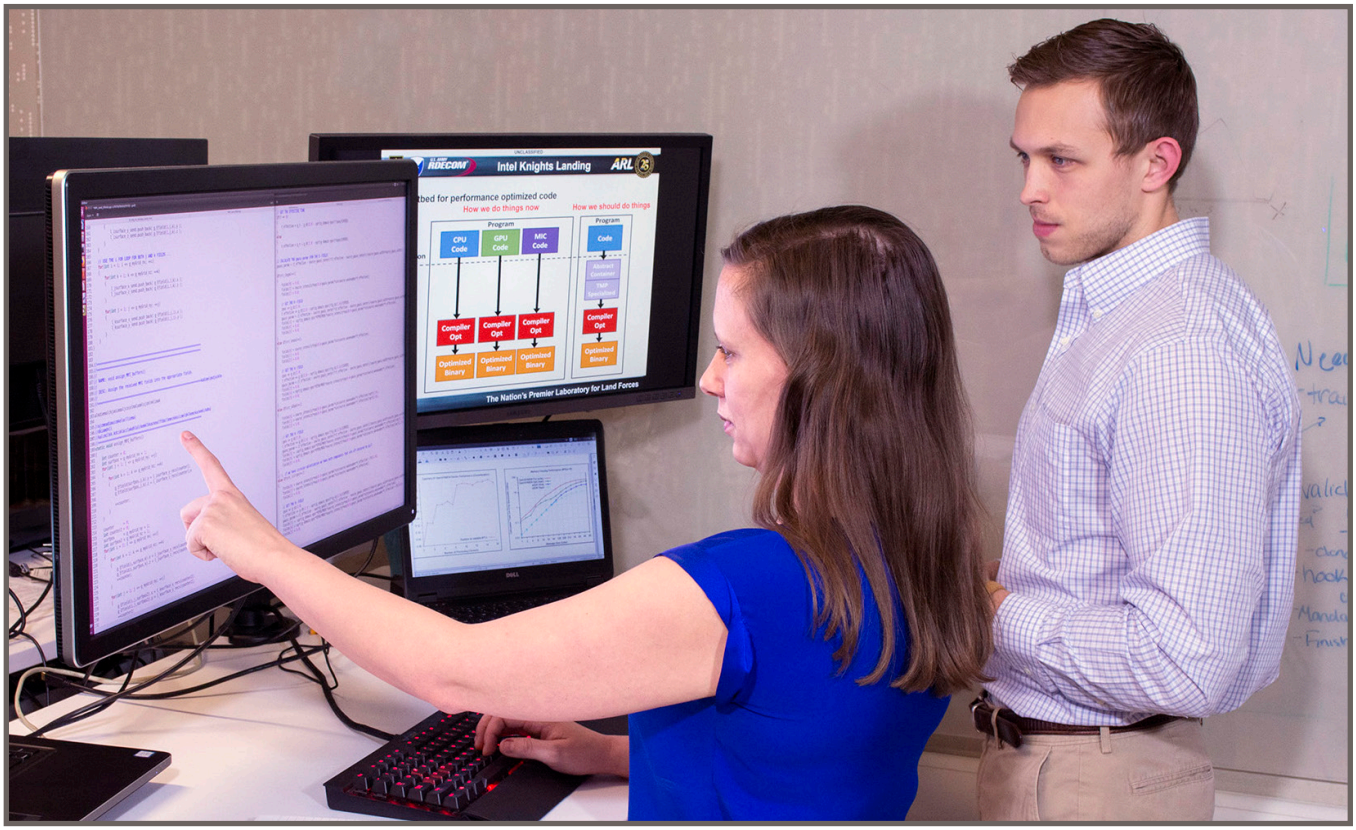
AUTO-TUNING BENCHMARK FOR HPC ACCELERATORS



Project Description

Our objective is to develop a benchmarking suite that utilizes parameterized kernel auto-tuning techniques to provide quick and accurate performance predictions on Army-relevant applications and platforms. We investigate and implement novel approaches to auto-tune computational kernels on heterogeneous platforms to determine optimal source-level parameterization for each architecture. Our goals is to gain deeper understanding of the connection between the optimal parameters and processor architecture. A major challenge is the widely diverse processing core types available. Current codes lack portability across heterogeneous computing platforms. Since we have little flexibility to change the available processors, it becomes vital to optimally use the processors available and to predict performance before selection.

Relevance of Work to DOD

Current high performance computing platforms employ a diverse range of resources to achieve maximum performance: shared and distributed memory, accelerators, many cores, multi-core integrated chips, FPGAs, etc. and we know the future portends even greater heterogeneity. To avoid having to recode applications software, we need to write portable software and be able to predict how well a code and its algorithms will perform on a given platform. This will enable the best use of available computing resources and fastest solution for all DOD problems.

Computational Approach

We are conducting research based on the 13 computational dwarfs concept developed at the University of California Berkeley. We are leveraging the SHOC suite from Oak Ridge National Laboratory, and OpenDwarfs

from Virginia Tech to guide kernel selection and implementation. We have developed an auto-tuning method based on parameterized kernels to automatically analyze and determine the best kernels for a given architecture. Our approach of predicting performance is similar to that of Oak Ridge National Laboratory except we add auto-tuning methods to obtain optimal benchmarks consistently while maintaining a common code base. Our methods are implemented on state-of-the-art heterogeneous computing systems at ARL.

Results

We have a rapidly maturing portable benchmarking suite comprised of computation kernels most applicable to Army applications that utilizes just-in-time (JIT) compile techniques and auto-tuning methods to determine optimal parameters for each computing architecture. One of the major advancements that this suite provides is a kernel generator method to automatically generate kernels to explore different combinations of parameters for a given kernel on a given architecture. We have complete results for two kernels fully utilizing JIT compile with auto-tuning methodologies. We are integrating three additional kernels into the suite. Results to date show that auto-tuning provides a measurable favorable impact on performance for each kernel, especially when compared between architectures. When compared to the current alternatives, our work provides a more complete, deeper understanding of the impact of parameters on software performance.

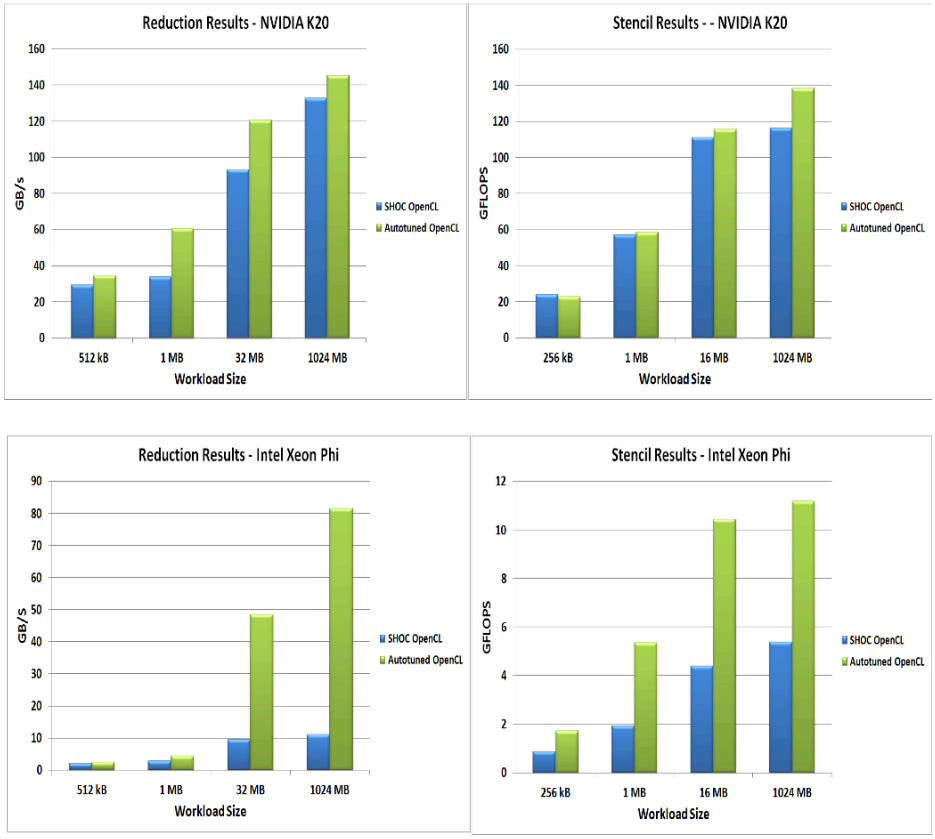


Figure 1:
Results for the reduction kernel (on the left) and the stencil kernel (on the right) using different benchmark implementations using a range of workloads on Intel MIC and NVIDIA K20 platforms

Future

Our path forward is the integration of more complex kernels into the benchmark suite, and extension of the benchmarking suite to other application types. In addition, we will explore ways to reuse the kernel generator to produce optimal kernels for a variety of different architectures. This will lead to more "tuned" kernels for each architecture without having to maintain separate code bases. The kernel generator will be the only element that needs to be maintained and it will be able to produce a variety of individual kernels for each architecture. We are also exploring transition opportunities to CERDEC and other organizations.

Contact Information

Jamie Infantolino
Army Research Laboratory
Computational and Information Sciences Directorate
jamie.k.infantolino.civ@mail.mil 410.278.7121

Co-Investigators

James A. Ross, Song J. Park, Dale R. Shires & Thomas M. Kendall
Army Research Laboratory
Computational and Information Sciences Directorate

David A. Richie
Brown Deer Technology