

# A Performance Profiling Strategy for High-Performance Map Re-Projection of Coarse-Scale Spatial Raster Data

Babak Behzad<sup>1,2</sup>, Yan Liu<sup>1,3,4</sup>, Eric Shook<sup>1,3</sup>, Michael P. Finn<sup>5</sup>, David M. Mattli<sup>5</sup>, and Shaowen Wang<sup>1,2,3,4,\*</sup>

1. CyberInfrastructure and Geospatial Information Laboratory
2. Department of Computer Science
3. Department of Geography
4. National Center for Supercomputing Applications  
University of Illinois at Urbana-Champaign
5. Center of Excellence for Geospatial Information Science  
U.S. Geological Survey (USGS)

\* Corresponding author: [shaowen@illinois.edu](mailto:shaowen@illinois.edu)

## ABSTRACT

Computational performance profiling is often used to characterize the efficiency of high performance computing applications, and has been widely applied in many scientific and engineering domains. In cartography and geographic information science, however, little attention has been paid to this important method. The purpose of this research is to demonstrate an effective performance profiling strategy for detecting and eliminating performance bottlenecks of a parallel processing algorithm for high performance map re-projection of coarse-scale spatially explicit raster data. Map re-projection represents a fundamental cartographic transformation in geographic information analysis, and has become computationally intensive as large-, and multi-scale spatial data are treated. Therefore, this research takes the pRasterBlaster software package as our case study as it implements a suite of map re-projection techniques for spatial raster data such as coordinate translation, forward- and inverse-mapping, and resampling (Steinwand *et al.* 2005). Using the Message Passing Interface (MPI) (Snir *et al.* 1998), pRasterBlaster has been parallelized to exploit high performance and parallel computing resources for enabling the research and operation of the U.S. Geological Survey National Geospatial Program (Finn *et al.* 2012).

Our performance profiling strategy encompasses two interrelated components: 1) exploratory performance analysis, and 2) formulation of spatial computational domain (Wang and Armstrong, 2009). The performance analysis component uses profiling tools such as the Tuning and Analysis Utilities (TAU) toolkit (Shende and Malony 2006) and the Integrated Performance Monitoring (IPM) profiling infrastructure (Wright *et al.* 2009) to detect potential bottlenecks in the parallel processing algorithm of pRasterBlaster. The formulation of spatial computational domain complementarily focuses on gaining basic understanding of spatial domain representation for scalable parallel computation by eliminating the bottlenecks. For example, our performance analysis was able to detect a major performance bottleneck due to an imbalanced load imposed on a single processor. This bottleneck prevented pRasterBlaster from scaling beyond 384 processor cores. Consequently, an adjustment was made by regularizing the decomposition of spatial computational domain for balancing the distribution of loads among all available processors, which contributed to the elimination of the bottleneck.

A set of computational experiments was conducted on a supercomputer as part of the National Science Foundation Extreme Science and Engineering Discovery Environment (XSEDE) – a cutting edge cyberinfrastructure (Atkins *et al.* 2003). The results of our experiments have demonstrated that the strategy is able to efficiently use 4,096 processor cores for re-projecting a set of coarse-scale spatial raster data. The size of each dataset on average is multiple hundred megabytes. This computational performance improvement is significant compared to the scalability of the original parallel processing algorithm, and has enabled pRasterBlaster to process much larger datasets more rapidly as required by the USGS National Geospatial Program.

**KEYWORDS:** Computational performance profiling, high performance computing, map projection, parallel computing, spatial computational domain

#### **DISCLAIMER**

Any use of trade, product, or firm names in this paper is for descriptive purposes only and does not imply endorsement by the U.S. Government.

#### **ACKNOWLEDGEMENTS**

This work is supported in part by the National Science Foundation grants: BCS-0846655, OCI-1047916, and the computational resource award: SES070004 from the Extreme Science and Engineering Discovery Environment (XSEDE).

#### **REFERENCES**

- Atkins, D.E., K.K. Droegemeier, et al. 2003. Revolutionizing Science and Engineering Through Cyberinfrastructure: *Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*. Arlington, VA, National Science Foundation.
- Finn, M.P., Y. Liu, D.M. Mattli, Q. Guan, K.H. Yamamoto, E. Shook and B. Behzad 2012. pRasterBlaster: High-Performance Small-Scale Raster Map Projection Transformation Using the Extreme Science and Engineering Discovery Environment. Abstract presented at the *XXII International Society for Photogrammetry & Remote Sensing Congress*, Melbourne, Australia.
- Shende, S. and A. D. Malony 2006. The TAU Parallel Performance System. *International Journal of High Performance Computing Applications*, SAGE Publications, 20(2): 287-331.
- Snir, M., S.W. Otto, S. Huss-Lederman, D.W. Walker, and J. Dongarra. 1998 *MPI: The Complete Reference—The MPI Core*, 2nd Edition, Cambridge, Massachusetts: The MIT Press, Volume 1.
- Steinwand, D.R., M.P. Finn, J.R. Trent, E.L. Usery, and R.A. Buehler. 2005 Re-projecting Raster Data of Global Extent. Abstract presented at *Auto-Carto 2005, A Cartography and Geographic Information Society Research Symposium*, Las Vegas, NV.
- Wang, S. and M.P. Armstrong. 2009 A theoretical approach to the use of cyberinfrastructure in geographical analysis. *International Journal of Geographical Information Science*, 23 (2): 169-193.
- Wright, N.J., W. Pfeiffer and A. Snavely 2009 Characterizing Parallel Scaling of Scientific Applications using IPM. *The 10th Linux Clusters Institute International Conference on High-Performance Clustered Computing*. Boulder, CO. March 10-12.