Extending the Roofline Model for Supercomputers

Boakye Dankwa, Fengguang Song
Department of Computer and Information Science
Indiana University Purdue University
Indianapolis, Indiana, USA

***Abstract*—The roofline performance model [1] is a simple 2-D visual tool that provides useful insight on kernel optimization on multi-core systems. In this project, an attempt is made to extend the roofline performance model to a 3-D roofline model for supercomputers, taking into consideration the peak floating point performance, the peak memory throughput and the peak communication throughput across nodes. A 3-D roofline performance model is constructed for 16 AMD Opteron Interlagos x86_64 nodes on BigRed II and used to bound the performance of the *SUMMA* parallel algorithm [2] on BigRed II.**

***Index Terms*—TODO; TODO; TODO;**

I. Introduction

Computationally intensive problems in industry and academia are usually solved using High Performance Computing (HPC) resources. These systems consist of dedicated high-end processors placed in close proximity and connected by high speed networks. Kernel optimization on such a system can be a daunting task. Therefore kernel performance models such as the roofline performance model proposed in [1], become very useful. The roofline model provides insightful kernel performance on multicore systems, however it cannot scale to a group of multi-core processors connected by a high speed network (i.e., a typical supercomputer configuration).

In this paper, we extend the roofline model to a group of multi-core processors connected by a high speed network by taking communication bottleneck into consideration.

TODO , [3], [4], [5]

II. Related Work

TODO

III. Proposed 3-D Roofline Model

Create 2-D model - (peak performance of multicore, memory bandwith using STREAM, operational intensity of kernels)

Create 3-D model - (peak performance of multicore, memory bandwith using STREAM, operational intensity of kernels, communication intensity)

A. Peak Network Throughput

TODO

B. Communication Intensity

TODO

IV. Results

TODO

V. Conclusions and Future Extensions

TODO

Acknowledgment

References

[1] S. Williams, A. Waterman, and D. Patterson, "Roofline: An insightful visual performance model for multicore architectures," *Commun. ACM*, vol. 52, no. 4, pp. 65–76, Apr. 2009. [Online]. Available: http://doi.acm.org/10.1145/1498765.1498785

[2] R. A. van de Geijn and J. Watts, "Summa: Scalable universal matrix multiplication algorithm," Austin, TX, USA, Tech. Rep., 1995.

[3] I. B. Peng, S. Markidis, E. Laure, G. Kestor, and R. Gioiosa, "Exploring application performance on emerging hybrid-memory supercomputers," in *2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, Dec 2016, pp. 473–480.

[4] M. Kong, L. N. Pouchet, and P. Sadayappan, "A roofline-based performance estimator for distributed matrix-multiply on intel cnc," in *2015 IEEE International Parallel and Distributed Processing Symposium Workshop*, May 2015, pp. 1241–1250.

[5] G. Ofenbeck, R. Steinmann, V. Caparros, D. G. Spampinato, and M. Pschel, "Applying the roofline model," in *2014 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS)*, March 2014, pp. 76–85.