Performance Evaluation of NoC Architectures for Parallel Workloads

Henrique C. Freitas[†], Marco A. Z. Alves, Lucas M. Schnorr, Philippe O. A. Navaux
Informatics Institute, Universidade Federal do Rio Grande do Sul, RS, Brazil
{hcfreitas, mazalves, lmschnorr, navaux}@inf.ufrgs.br

Abstract

Network-on-Chip is the state-of-the-art approach to interconnect many processing cores in the next generation of general-purpose processors. In this context, the problem is to choose NoC architectures capable of achieving high performance for parallel programs. Therefore, the main goal of this paper is to evaluate the performance of three NoC architectures using well-known parallel workloads.

1. Methodology and Results

Due to the many-core processors, parallel programs can be an important alternative to explore the high number of cores. For this reason, this paper describes an evaluation methodology focusing on parallel programs and NoC performance results, as follows: i) Parallel workloads from NAS (Numerical Aerodynamic Simulation) [1] version 3.3 based on MPI (Message Passing Interface) were executed on a real eight-core machine. ii) The Triva prototype [2] was used to provide an alternative way to analyze parallel applications. The Triva benefits are the visualization of communication patterns generated by parallel applications in different time intervals, and the possibility to analyze these patterns together with the network topology involved in the execution. iii) An analyzer tool was designed in Python for reading trace files and processing performance results. The analyzer is divided into two main parts: performance analytical models based on NoC architectures, and transmission time analysis considering time to send packets related to transmission start time. Architecture characteristics to develop the analyzer were obtained from SoCIN – System-on-Chip Interconnection Network [3] and MCNoC – Multi-Cluster NoC [4].

The total transmission time shown in Figure 1 considers the influence of all collective communication patterns. It is important to notice that the main impact on these results is from one-to-one pattern. This impact is related to the higher number and larger size of messages than other patterns. In addition, the number of hops and the influence of routers add more cost to transmit messages. Therefore, MCNoC decreases transmission time relative to 2x4 mesh and torus up to 21.2% and 19% for BT, 11.6% for CG, 49.5% and 45.8% for EP, 23.5% and 21.3% for FT, 26.1% and 25% for IS, 15.8% for LU, 16.2% for MG, and 21% and 18.8% for SP, respectively.
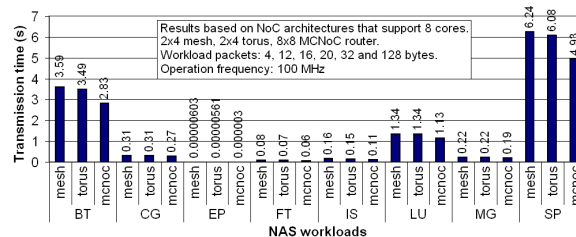

Figure 1. Total transmission time

To conclude, MCNoC is capable of configuring topologies through programmable routers and it can achieve a higher performance than a traditional NoC based on mesh or torus topologies. MCNoC has a lower impact of routers since they map communication patterns decreasing the number of hops.

References

[1] M. Frumkin, H. Jin and J. Yan, "Implementation of NAS Parallel Benchmarks in High Performance Fortran", In NAS Technical Report NAS-98-009, NASA Ames Research Center, 1998.

[2] L. M. Schnorr, G. Huard and P. O. A. Navaux, "3D Approach to the Visualization of Parallel Applications and Grid Monitoring Information", IEEE/ACM GRID, Tsukuba, Japan, 2008.

[3] C. A. Zeferino and A. A. Susin, "SoCIN: A Parametric and Scalable Network-on-Chip", IEEE SBCCI, São Paulo, Brazil, pp. 169-174, 2003.

[4] H. C. Freitas, T. G. S. Santos and P. O. A. Navaux, "NoC Architecture Design for Multi-Cluster Chips", IEEE FPL, Heidelberg, Germany, pp. 53-58, 2008.