

Performance Analysis in Mesh Network-on-Chip Topology by using Multilevel Network Partitioning

**¹A. Lit, M. Q. Edanan, H. H. A. Halim, ²F. Mahyan and T. H. Mahyan

Abstract

The increasing complexity of System-on-Chips (SoCs) has resulted in the bottlenecking of the system due to scalability problems in the bus system. This leads to the decrement of the performance of future SoCs with more complex circuitries inside them. Network-on-Chips (NoCs) was proposed as one of the solutions to overcome these issues especially regarding the communication between Intellectual Properties (IP) in a chip. The fundamentals in designing NoC include the selection of network topologies, and hence performance optimization is needed to ensure the full advantage of networking is taken. Therefore, multi-level Network Partitioning techniques are proposed to obtain the optimal design of networks based on its performance. The performance of a network is measured by its throughput, average queue size, waiting time and data loss. This technique is applied in a case study using MPEG-4 video application with four famous partitioning algorithms (Linear, Spectral, Tailor-Made and Kerningham-Lin). Experimental results show that second level of spectral partitioning gives the best performance compared to another network partitioning.

Keywords: Network-on-Chip, Performance Optimization, Multilevel, Network Partitioning.

I. INTRODUCTION

The development of multiprocessor SoCs (MPSoCs) demands more complexity on the system itself, as well as extremely demanding on the size of the system. However, the bus topology network does not scale well with size [2]. Hence, NoC was proposed to achieve good on-chip communications among SoC computational cores [3]. NoC aims to implement today's large scale networking technology into SoC. This includes implementing routers, links and the usage of data packets as the main components of data transmission. NoC platform is developed mainly using these elements in an ordered structure [4]. NoC networks tend to be robust and have almost no bottleneck issues. They are also very much scalable compared to SoC. While being complex and require a lot of wiring effort with increased size, NoC offers a more acceptable solution than what most MPSoCs have to offer.

2. MESH NOC TOPOLOGY

The mesh technology connects one node to another directly and indirectly to form a net-like connection between them. One node can connect to another through other intermediate nodes, and such actions are called hops. This makes this topology a multi-hop network. The advantage of this network is its flexibility upon failures. Instead of failing outright, a failure in the nodes may only cause performance drop due to packet losses the more hops are made. NoC commonly utilizes this type of topology. Figure 1 shows a 4x4 mesh NoC topology that consist of 16 functional cores.

**Corresponding author

¹ A. Lit is with the Universiti Malaysia Sarawak (UNIMAS), 94300 Kota Samarahan, Malaysia (Phone +6082-58-3319; Fax:+6082-58-3410; E-mail: lasrani@feng.unimas.my)

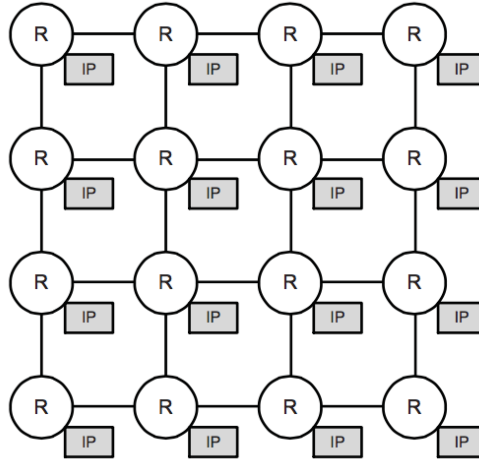


Figure 1 : 4x4 Mesh NoC Topology

3. M/M/1/B QUEUE PERFORMANCE

A simple M/M/1/B queue is exploited in this model. This model renders a discrete time Markov chain [14] analysis of queue where the time step is taken up equal to the time needed to send a packet. Poisson distribution traffic arrival process and the exponential distributed service time is presumed. The '1' entails there are one server and B implies the queue has a bounded buffer of size B. Figure 2 shows, the state transition diagram for the discrete-time Markov chain M/M/1/B queue. A homogeneous Markov Chain is taken as packet arrivals and departures are independent of the time index value [14]. Once a packet arrives, it is taken for granted that it could be served at the same time step with probability c .

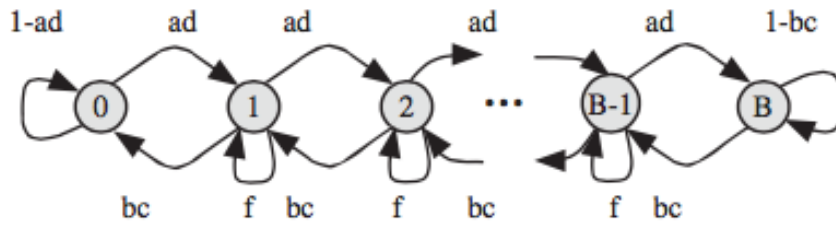


Figure 2: State Transition Diagram for discrete-time M/M/1/B

The queue performance is analyzed in term of throughput, average queue size, and packet delay. The output traffic or average of packets leaving the queue per time step is given by Equation (1):

$$N_a(out) = acs_o + \sum_{i=1}^{\infty} cs_i \quad (1)$$

The throughput of the queue in units of packets per time step is given by Equation (2):

$$Th = c(1 - bs_o) \quad (2)$$

The input traffic is measured in units of packets per time step is given by Equation (3):

$$N_a(in) = (1 \times a)(0 \times b) = a \quad (3)$$

The efficiency is defined as Equation (4):

$$\eta = \frac{c(1 - bs_o)}{a} \quad (4)$$

Data is lost when it is full and packets arrive but cannot leave. The average lost traffic is measured in units of packets per time step and is given by Equation (5):

$$N_a(lost) = s_B ad \quad (5)$$

The packet loss probability L is the ratio of lost traffic relative to the input traffic as shown in Equation (6):

$$L = \frac{Na(\text{lost})}{Na(\text{in})} = s_B d \quad (6)$$

The average queue size is given by Equation (7):

$$Q_a = \sum_{i=0}^B i s_i \quad (7)$$

Queue size is measured in units of packets given by Equation (8):

$$Q_a = \frac{\rho \times [1 - (B+1)\rho^B + B\rho^{B+1}]}{(1-\rho) \times (1-\rho^{B+1})} \quad (8)$$

The wait time is measured in units of packets per time step and given by Equation (9):

$$W = \frac{Q_a}{Th} \quad (9)$$

4. MULTI LEVEL NETWORK PARTITIONING TECHNIQUE

Network Partitioning (NP) is the process of separating a big network in to smaller partitions [9]. The goal is to minimize the cost of the communication over the partition boundaries. Because the number of core are only 12, the level of partitioning is chosen up to two level. In order to implement NP into the matrix, a graph partitioning algorithm is required. There are four NP techniques investigated; Linear, Spectral, Tailor-Made, and Kernighan-Lin partitioning. At Traffic Distribution Graph (TDG) stage, Figure 3 shows how's the MPEG-4 TDG were partitioned by employing four various NP algorithm. The dashed line indicate the boundary between the two partitions generated.

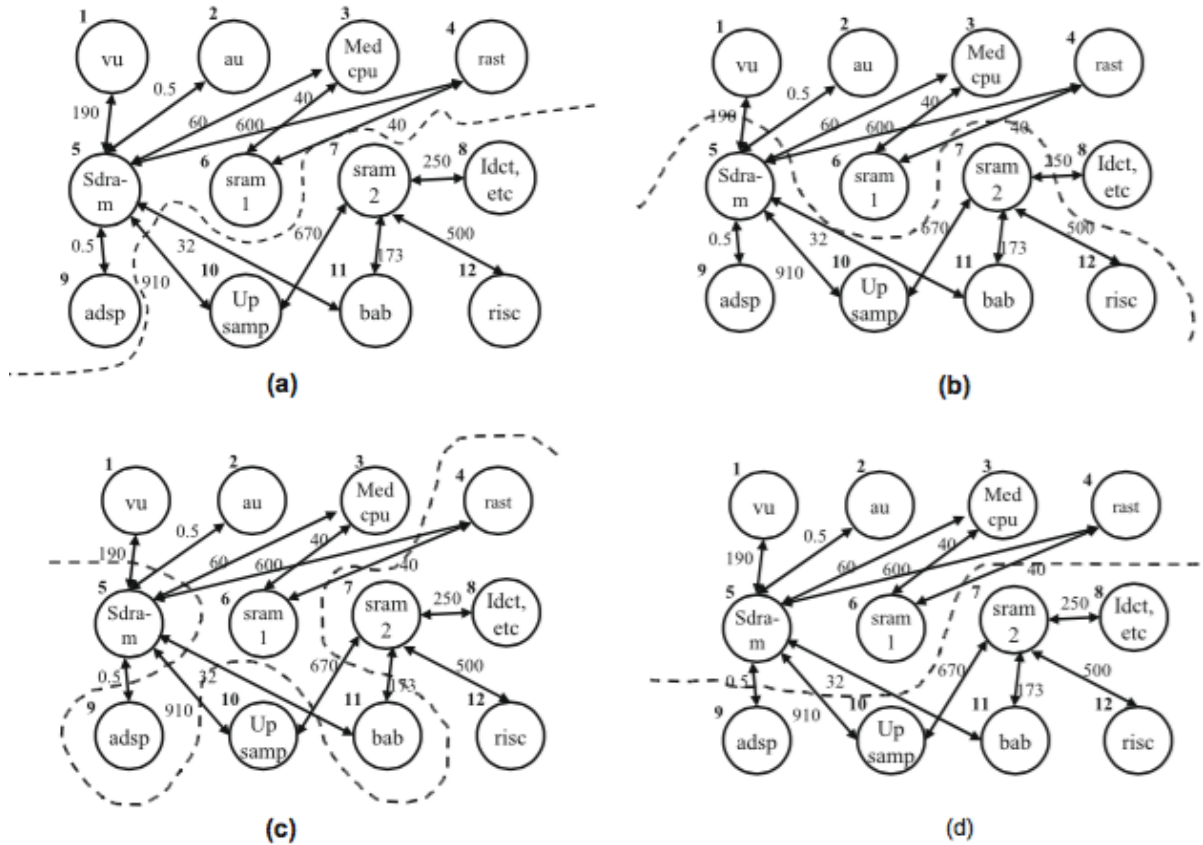


Figure 3: Network Partitioning in TDG (a) Tailor-Made, (b) Spectral, (c) Kernighan-Lin, and (d) Linear [7]

After the partitioning at TDG level, the IP is then mapped to mesh NoC topology. Figure 4 shows a mesh implementation of 16-nodes application before, first and second level of partitioning.

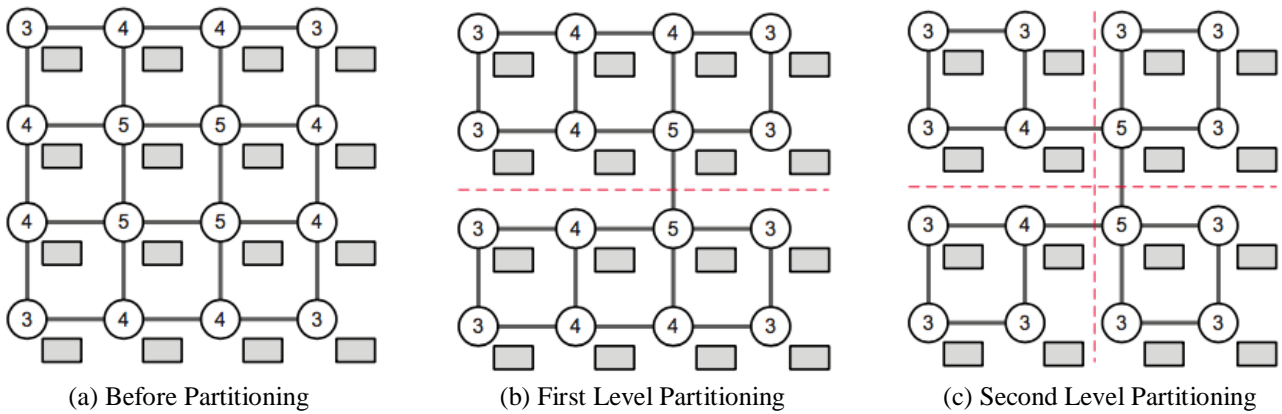


Figure 4 : Network Partitioning in Mesh NoC
 (a) Before Partitioning (b) First-Level Partitioning (c) Second-Level Partitioning

The particular divisions are disjointed by the dotted line. Meanwhile, the number in each node in the figure presents the number of ports of the router in this node. The figure distinctly shows that the greater partitioning results in moving out more global links which comply to cut down more area. The affect of multi level NP will be more significance when the number of nodes becomes bigger since the number of links withdrew by partitioning increases. Due to that, network partitioning is more profitable in scaling down the area cost as the number of cores in the application growths.

5. EXPERIMENTAL RESULTS

To analyze the performance of the proposed techniques, the experimental setup has been carried out by MATLAB software. As an example Figure 5 shows the simulation results for second-level of Spectral NP. The outputs produced are in term of throughput, average queue size, data loss and waiting time.

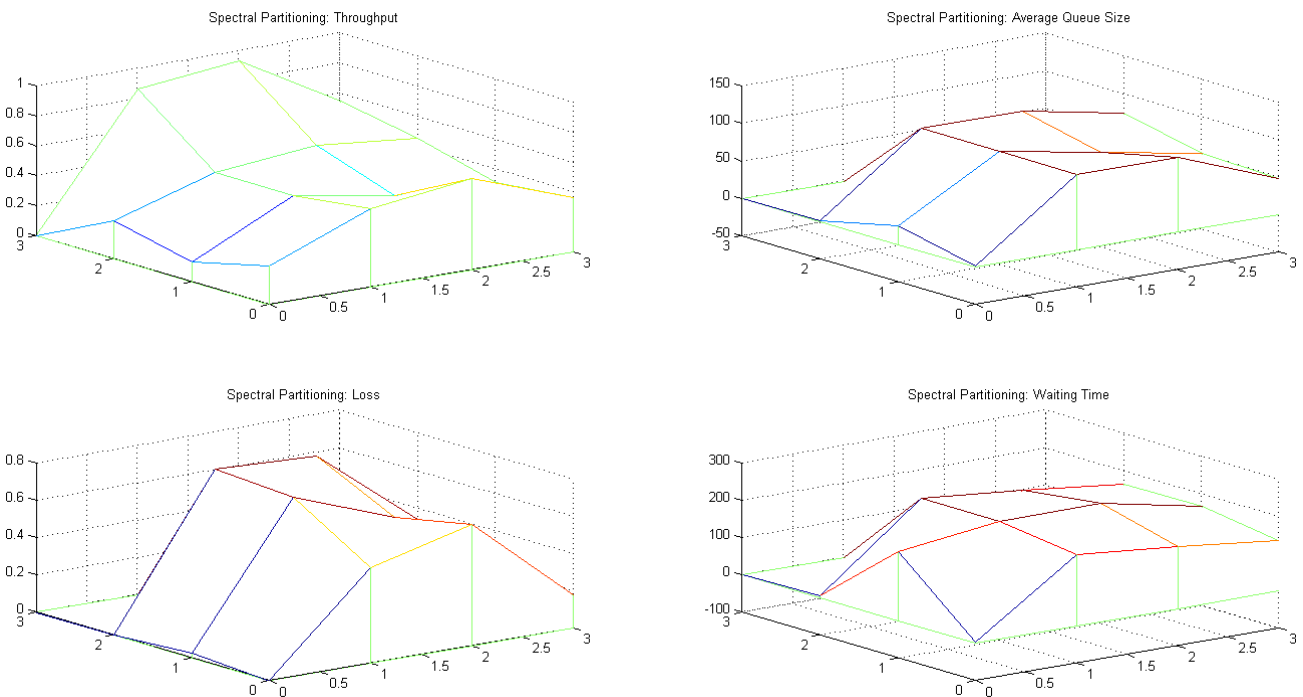


Figure 5 : Simulation result for second-level of Spectral Network Partitioning

The better performance are evaluated by high throughput, low queuing size, less data loss and waiting time. Figure 6 shows the comparison of NoC performance for different methods of NP and at the different level of its partitioning. As can be seen in Figure 6(a) the impact of second level of NP reduced the throughput of the NoC. Meanwhile, the queuing size and data loss are also decreased after the partitioning is made as shown in Figure 6(b) and Figure 6(c). However, in waiting time only

the spectral partitioning gives less waiting time compared to other NP techniques. As far as performance analysis is concerned, it can be concluded that spectral NP with the second level of partitioning gives the best-optimized performance.

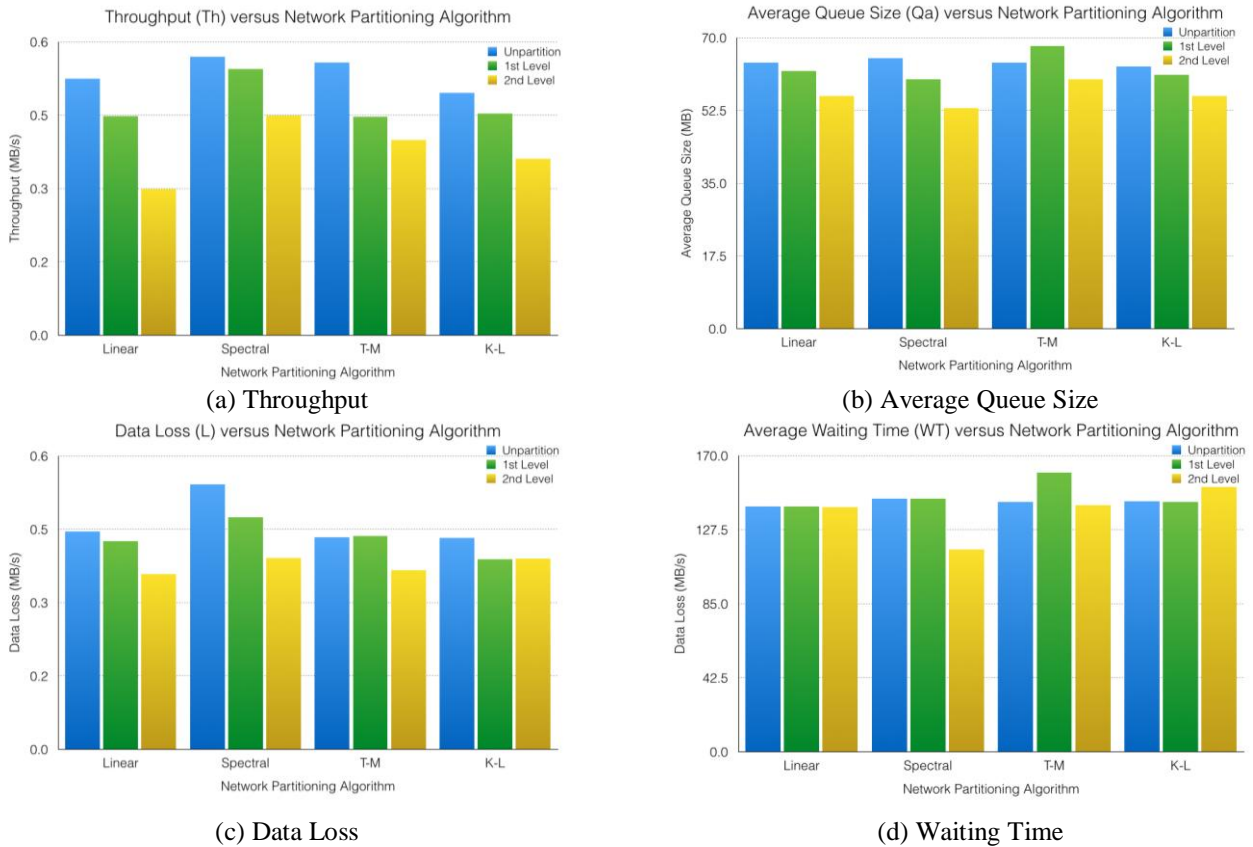


Figure 6 : Comparison of NoC Performance

6. CONCLUSION & FUTURE WORK

In this paper, the performance analysis of mesh NoC architecture is investigated by utilizing various multi-level NP techniques. The partitioning have been applied up to the second level in order to verify the performance of the network. These techniques showed various improvements and penalties as elaborated previously. From an experimental results, Spectral NP is shown as the best performing NP algorithm. There are many interesting possibilities for future research such as extending this modelling approach with a 3 dimensional mesh topology and investigate the impact of utilising other types of routing algorithm.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to Research and Innovation Management Centre (RIMC), Universiti Malaysia Sarawak (UNIMAS) for the financial support.

REFERENCES

- [1] Benini, L. and Micheli, G.D., 2006. Networks on Chips: Technology and Tools. 2nd ed. Morgan Kaufmann, United States of America
- [2] Larsson, E., 2005. Introduction to Advanced System-on-Chip Test Design and Optimization. Springer, Germany
- [3] Benini, L., 2006. Application specific NoC design. In *Proceeding of the IEEE Design, Automation and Test in Europe Conference (DATE '06)* vol.1 pp. 1-5.
- [4] Lit, A., 2011. *Network Partitioning and IP placement in Network-on-Chip*. M.Eng, Universiti Teknologi Malaysia, Malaysia.
- [5] Forouzan, B.A., 2007. Data Communications and Networking. 4th ed. McGraw Hill, New York.
- [6] Kiasari, A. E., 2013. Performance Analysis and Design Space Exploration of On-Chip Interconnection Networks. KTH Royal Institute of Technology, Stockholm, Sweden.
- [7] Elmiligi, H., Morgan, A.A., El-Kharashi, M.W., and Gebali, F., 2007. A topology-based design methodology for Network-on-Chip applications. In *Proceedings of the second International Design and Test Workshop (IDT '07)*, pp. 61-65
- [8] Mohar, B., 1997. Some Applications of Laplace Eigen Values of Graphs. *Algebraic Methods and Applications of NATO ASI Series C* pp. 227-275.
- [9] Fiedler, M., 1973. Algebraic Connectivity of Graphs. *Czechoslovak Mathematical Journal* pp. 298-305.
- [10] Kernighan, B. and Lins, S., 1970. An efficient Heuristic Procedure for Partitioning Graphs. *Bell Systems Technical Journal*, vol. 49, pp. 291-307.
- [11] Hosmark, R., 2002. Modeling and Prototyping of Network-on-Chip. Master, Sweden Jonkoping University, Sweden.
- [12] Morgan, A.A., Elmiligi, H., El-Kharashi, M.W., and Gebali, F., 2009. Area-Aware Topology Generation for Application-Specific Networks-on-Chip

Using Network Partitioning.

- [13] Zuo, L., Zhu, M. M., 2013. Toward Flexible and Fast Routing Strategies for Dynamic Network Provisioning.
- [14] Gebali, F. Computer Communication Networks Analysis and Design, Victoria : Northar Digital Design, 2004.