A Path-Wise Scheme for Simpler Mesh-of-Tree Model in Network-on-Chip Designs

Yu Lu
College of Science and Engineering
James Cook University
Cairns, QLD 4870, Australia
E-mail: yulu@gmail.com

***Abstract*—The needs of the chip are growing on a daily basis as extremely large size integrated circuits evolve fast. As a result, the network-on-chip (NoC) was suggested and successfully developed to address the issue of communication on-chip. Current research focuses on NoC topology and routing algorithms that have a substantial influence on network performance. Because low power is a priority, the Mesh-of-Tree (MoT) architecture is preferable. However, certain MoT routers may perform extensive communication activities, which means they consume more power and are more likely to become hotspots. To improve power consumption, we presented a Simpler Mesh-of-Tree (SMoT) architecture based on MoT in this study. We also suggested a path-wise routing scheme for the Internet of Things, which takes into account both network traffic and the shortest path between two points. By altering the direction of data forwarding to reduce network latency, our methodology spreads communication over the whole network. When compared to MoT, our suggested strategy can lower network power consumption by 5.39% -23.3% while also reducing network latency by 4.23%–27.28%.**

Keywords-Noc; Low Power; Routing Scheme

I. INTRODUCTION

Because of its increased efficiency and the ongoing expansion in the number of computational and storage blocks integrated in a single chip, NoC has become the trend of high-performance microprocessors. Many studies are now focusing on NoC design, including topology and routing strategy. Mesh, cmesh, torus, and MoT [1] are some of the topologies that have been proposed. Mesh has gotten the most attention because of its lower node degree. However, due of its larger diameter, it has a higher power consumption.

Root, stem, and leaf routers are used in the MoT architecture, which has a smaller diameter and lower node degree. Root routers may perform intensive communication activities on the Internet of Things, which means they use more power and are more likely to become hotspots. We proposed the SMoT topology in this study, which would diminish this phenomenon while reducing power usage. We also suggested a path-wise routing method for SMoT, which spreads communication around the network by altering packet forwarding direction to reduce network latency.

II. RELATED WORK

Several recent studies have aimed to create a network with low latency and low power consumption. Node-Router Decoupling was suggested by Chen et al., who used power-gating techniques to reduce power consumption [2]. Ghosal, Prasun, and colleagues proposed a Level-2 Mesh topology with two tiers of linkages, one for long-distance communication and the other for short-distance communication [3]. When compared to 3D FMT, Viswanathan designed a revolutionary 3D structure with a 75 percent reduction in the number of vertical links [4].

To minimize network latency, Koibuchi et al. created random shortcuts by supplementing conventional topologies with random links [5]. By delivering packets through less crowded routers, Ebrahimi et al. presented an adaptive routing algorithm [6]. To decrease power dissipation, Ezz-Eldin et al. introduced a low leakage power switch using power supply gating and adaptive virtual channel technique [7].

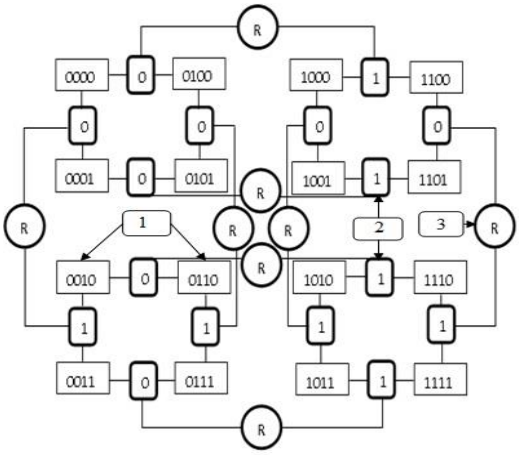
III. MOTIVATION

One of the most fundamental restrictions in NoC designs is power consumption. The linked network of an MIT Raw processor with 16 PEs consumes 36 percent of the total power consumption. The Tera-scale chip was created by Intel, and its associated power usage accounted for 40% of the total power consumption. As a result, NoC has a high-power efficiency. Meanwhile, as process technology advances, communication between cores will become a constraint in further improving performance. Take the Intel Single-Chip Cloud Computer for example (SCC). Each SCC chip has 48 Pentium processors that were coupled by a 46 mesh. It takes at most 10 hops to send a packet from the source to its target node.

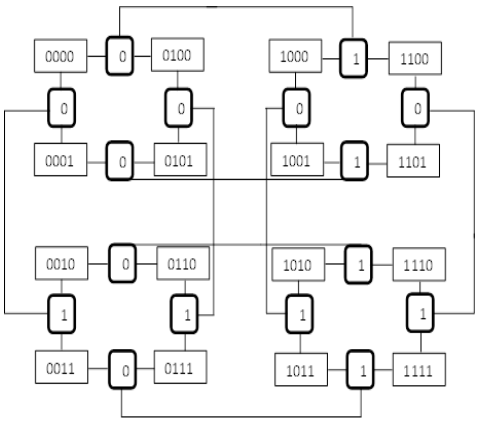
As a result, we presented a more power efficient SMoT architecture in this study. We also proposed a path-wise routing method for SMoT to address the requirement for reduced network latency. Our approach allowed us to reduce the cost of communication, including electricity and latency.

IV. ARCHITECTURE OF THE SIMPLER MESH-OF-TREE

There are three types of routers in the MoT topology: root, stem, and leaf. Row or column routers are the stem routers. Two cores, one row router, and one column router are connected by each leaf router. Until the final level stem routers join the root routers, the stem routers are structured in a binarytree-like design.



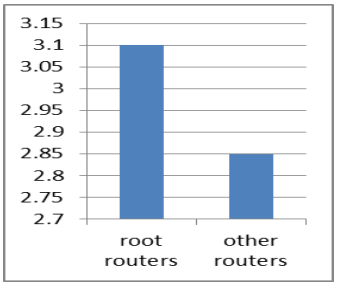
(a)



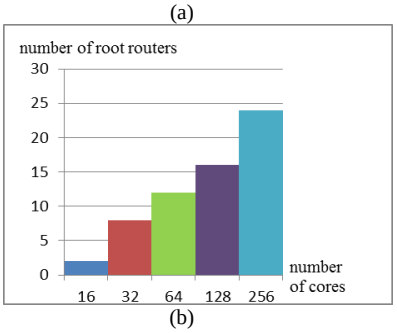
(b)

Figure 1. (a): 32-core MoT (1:leaf, 2:stem, 3:root); (b): 32-core SMoT

Based on the original design depicted in Figure 1(a), we proposed the SMoT, which uses connections instead of root routers to minimize power consumption and latency. For starters, when a network's traffic increases, root routers can quickly become hotspots. The average link use of root and other routers is shown in Figure 2(a). Because other routers have more ports than root routers, packets that arrive at a particular router might pick an appropriate port to be received or sent, or just wait for a while in the buffers. Furthermore, if a significant number of cores are merged, the number of root routers will increase. Figure 2(b) depicts the root routers' growth as the number of cores grows.



(a)



(b)

Figure 2. (a) The average link use of root and other routers; (b) Depicts the root routers' growth

V. PATH-WISE SCHEME

For SMoT, we describe a path-wise routing strategy. When a packet arrives at a leaf router, calculate the congestion coefficient for each neighboring router, which shows the traffic surrounding this router, and select the next hop from the shortest path with the lowest congestion coefficient. This method can minimize network congestion while also providing the shortest path.

In our approach, different routing strategies are applied to the leaf and stem routers, and a few fundamental parameters are required. The number of row routers that a packet passes through on its way from the current node to the destination is represented by Δx, while the number of column routers that a packet passes through on its way from the current node to the destination is represented by Δy. In a M*N MoT structure (M and N denote the number of rows and columns respectively), we also employ a simpler addressing method.

Core_id	Col_ADDR	Row_ADDR
$\log_2 C$	$\log_2 N$	$\log_2 M$

(a)

Col_ADDR	Row_ADDR
$\log_2 N$	$\log_2 M$

(b)

Figure 3. (a) The address for a core; (b) The address for a leaf router

The following is the procedure of the proposed algorithm for leaf routers:

Step1: Packet input.

Step2: Compute the optimal port for packet output.

Step2.1: Calculate the value of Δx and Δy:

$$\Delta x = Des\_addr_{row\_addr} - Src\_addr_{row\_addr} \quad (1)$$

$$\Delta y = Des\_addr_{col\_addr} - Src\_addr_{col\_addr} \quad (2)$$

Step2.2: If Δx≠0 and Δy≠0, which indicates the source and destination nodes are in separate rows and columns, choose one neighbor router as the output:

Step2.2.1: Compute the feasibility P1 and P2 that a packet is sent to the router in the next row or column:

$$P_1 = \frac{\Delta x + P_{algorithm}}{\Delta x + \Delta y + 2P_{algorithm}} \quad (3)$$

$$P_2 = \frac{\Delta y + P_{algorithm}}{\Delta x + \Delta y + 2P_{algorithm}} \quad (4)$$

Where Palgorithm denotes the total number of pathways between a packet's source and destination node, in our proposed SmoT model, Palgorithm is set to 2.

Step2.2.2: Calculate the congestion coefficient for neighbor routers other than the one where the packet was sent, using the equation below:

$$Con_i = P_i \times \frac{Buf\_SIZE_{out1} + Buf\_SIZE_{out2}}{2} \quad (5)$$

Where Coni is the produced congestion coefficient, Pi denotes the probability that a packet is sent to the row router or column router, i=1, 2; Buf_SIZEout1 and Buf_SIZEout2 reflect the number of available buffers in the current output port as well as the number of available buffers in the next available output port.

Step2.2.3: Choose a router with the optimal Coni as the output stop for the packet, along with its output port.

Step2.3: If Δx=0 and Δy≠0, which indicates the source and destination nodes are in identical rows, identify the relevant output port for the row router as the next stop.

Step2.4: If Δx≠0 and Δy=0, which indicates the source and destination nodes are in identical columns, identify the relevant output port for the row router as the next stop.

Step2.5: If Δx=0 and Δy=0, which indicates the source and destination nodes belong to a single router, if Core_id=0, forward the incoming packet to L_Core; if Core_id=1, forward the incoming packet to R_Core.

Step2.6: When two or more packets fight for the same output port, the packet with the larger number of hops wins.

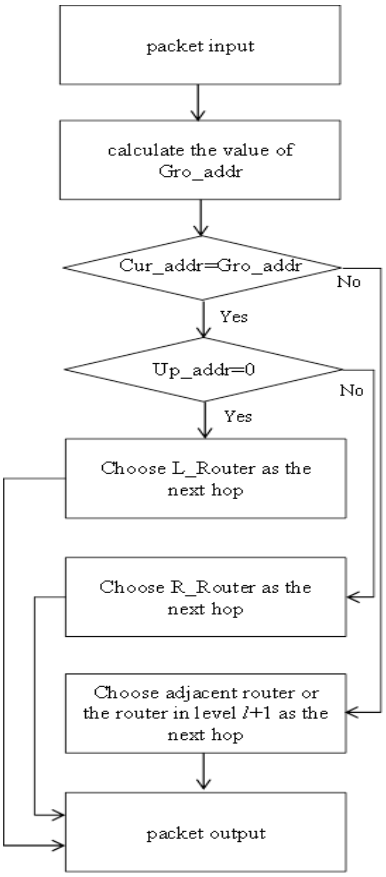


Figure 4. The flow diagram of routing algorithm for leaf routers

Step3: Forward the packet via the determined output port.

Meanwhile, the algorithm for l-th stem routers is presented in Figure 4. Specifically, when a packet comes in, compute the target router in level l. if the present router's location is not the same as the address of the target router, select (l+1)-th router or l-th neighbor one as the next stop, with the condition of current router's level. If not, calculate the address of the target router in level l-1, forward the packet to L_Router or R_Router based on the target router.

VI. EXPERIMENT AND EVALUATION

A. Experimental setup

In this section, we choose two metrics to evaluate the performance of SMoT, i.e., the power consumption and network latency. These two metrics are well adopted in previous works.

In the evaluation of performance, power consumption is a significant factor. It will be meaningful if we can lower it while obtain comparable performance. Overall, the power comsumption is computed as follows:

$$Network\_power = P_{dynamic} + P_{static}$$

The time it takes for a packet to travel from its source to its destination is known as network latency in NoC. One of the primary performance concerns we utilize here is the average latency.

$$Network\_latency = Total\_Cycles / Total\_Flits$$

B. The experimental environment

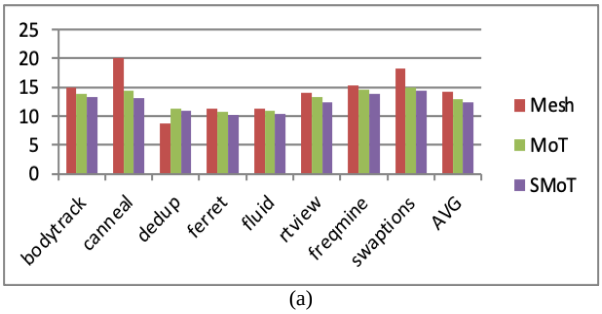
We conduct experiments with the Gem5 simulator, which is a public well-known platform. We choose the PARSEC dataset for evaluation following the the common practice in prior stuides. To be concrete, the details of experimental setup is depicted in Table 1.

TABLE I. ENVIRONMENTAL SETUP

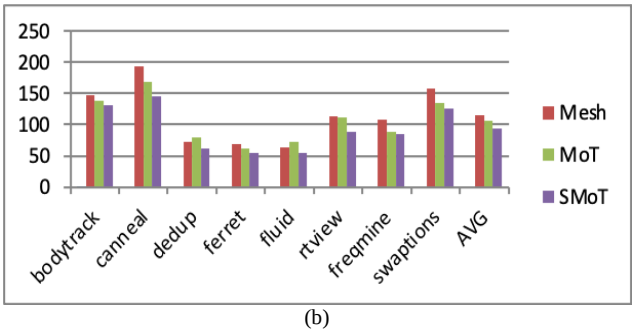
Processor configuration			
CPU_TYPE	Timing	Timing	
CPU_NUMS	16	64	
NoC configuration			
TOPOLOGY	Mesh	MoT	SMoT
SCHEDULING_POLICY	xy routing scheme	deterministic routing scheme	path-wise routing scheme
BENCHMARK	PARSEC	PARSEC	PARSEC

C. Experimental results

The network power comparison is presented in Figures 5(a) and 5(b). As is observed, our proposed SMoT has apparently gained lower power consumption. Specifically, it reduces the power consumption by 5.39%-23.3% compared to the original MoT. We argue this is mainly due to two fundamental factors. On one hand, leveraging short paths between the subnets is able to decrease the number of hotspots, since most routers hold more ports compared to the root ones, packets would obtain more choices for paths. On the other hand, the power consumption is directly reduced with less routers.



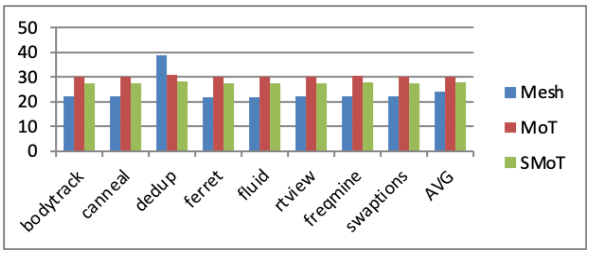
(a)



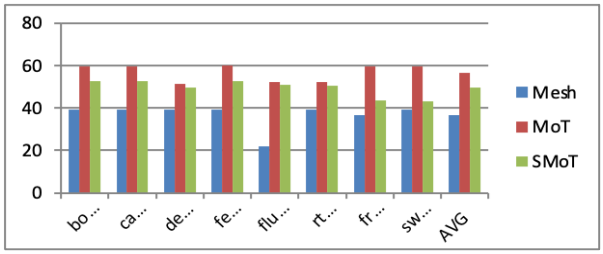
(b)

Figure 5. (a): Network power for 16 cores; (b): Network power for 64 cores

Meanwhile, the results of network latency is shown in Figures 6(a) and 6(b). It is noticed that the network latency of the proposed SMoT was decreased by 4.23%–27.28% than a typical MoT. We believe the reason is that the path-wise scheme improves the network efficiency and further decreases the latency caused by large number of hop counts.



(a)



(b)

Figure 6. (a): Network latency for 16 cores, (b): Network latency for 64 cores

Overall, the above experimental results verify that the proposed algorithm obtains less power consumption and lower network latency while maintaining better performance compared to previous studies. On one hand, leveraging short paths between the subnets is able to decrease the number of hotspots, since most routers hold more ports compared to the root ones, packets would obtain more choices for paths. On the other hand, the power consumption is directly reduced with less routers.

VII. CONCLUSIONS

The architecture design plays an important role in the performance of NoC applications. In this work, we propose a SMoT architecture along with a path-wise scheme which takes into account both network topology and the congestion coefficient. Experimental results demonstrate that our proposed model obtains better performance while reducing the power consumption and network latency effectively and efficiently.

REFERENCES

[1] Manna, K., Chattopadhaya, S. Sengupta, I., An efficient routing technique for mesh-of-tree-based NoC and its performance comparison. International Journal of High Performance Systems Architecture, v4, n1, pp. 25-37, 2012.

[2] Lizhong Chen et al., NoRD: Node-Router Decoupling for Effective Power-gating of On-Chip Routers. Proc.

Of IEEE/ACM 45th International Symposium on Microarchitecture, MICRO 2012, pp. 270-281, 2012.

[3] Ghosal, Prasun, Das, Tuhin Subhra, FL2STAR. A novel topology for on-chip routing in NoC with fault tolerance and deadlock prevention. 2013 IEEE International Conference on Electronics, Computing and Communication Technologies, CONECCT2013, 2013.

[4] Viswanathan et al., An optimised 3D topology for on-chip communications. International Journal of Parallel, Emergent and Distributed Systems, v29, n4, pp. 346-362, 2014.

[5] Michihiro Koibuchi et al., A Case for Random Shortcut Topologies for HPC Interconnects. 2012 39th International Symposium on Computer Architecture, ISCA 2012, pp. 177-188, 2012.

[6] Ebrahimi et al, LEAR-A low-weight and highly adaptive routing method for distributing congestions in on-Chip Networks. 20th Euromicro International Conference on Parallel, Distributed and Network-Based Processing, PDP 2012, pp. 520-524, 2012.

[7] Ezz-Eldin, Rabab, Magdy A. El-Moursy, and Amr M. Refaat. "Low leakage power NoC switch using AVC." Circuits and Systems (ISCAS), 2012 IEEE International Symposium on. IEEE, 2012.