Comparative performance evaluation of routing algorithm and topology size for wireless network-on-chip

Asrani Lit¹, M. S. Rusli², M. N. Marsono³
¹Department of Electrical and Electronic Engineering, Universiti Malaysia Sarawak, Malaysia
², ³Division of Electronic and Computer Engineering, Universiti Teknologi Malaysia, Malaysia

ABSTRACT

Wireless Network-on-Chip or WiNoC is an alternative to traditional planar on-chip networks. On-chip wireless links are utilized to reduce latency between distant nodes due to its capability to communicate with far-away node within a single hop. This paper analyzes the impact of various routing schemes and the effect of WiNoC sizes on network traffic distributions compared to conventional mesh NoC. Radio hubs (4×4) are evenly placed on WiNoC to analyze global average delay, throughput, energy consumption and wireless utilization. For validation, three various network sizes (8×8, 16×16 and 32×32) of mesh NoC and WiNoC architectures are simulated on cycle-accurate Noxim simulator under numerous traffic load distributions. Simulation results show that WiNoC architecture with the 16×16 network size has better average speedup (∼1.2×) and improved network throughputs by 6.36% in non-uniform transpose traffic distribution. As the trade-off, WiNoC requires 63% higher energy consumption compared to the classical wired NoC mesh.

Keywords: Performance analysis, Routing algorithm, Traffic distribution, Wireless NoC

Copyright © 2019 Institute of Advanced Engineering and Science. All rights reserved.

1. INTRODUCTION

Technology scaling has allowed the integration of many-core design that range from hundreds to thousands of processing cores on a single integrated circuit (IC) [1-3]. Nevertheless, the trend of growing numbers of processing cores in Network-on-Chip (NoC) has caused higher latency and more power-hungry system architecture affected by the long distance multiple-hops communication [4-6]. To date, Wireless Network-on-Chip (WiNoC) is among the promising emerging technology to alleviate the aforementioned issues due to its some unique factors such as energy efficiency, high bandwidth delivery and low latency [4, 7-9]. Furthermore, these interconnects are also able to transmit data to distanced cores chip in single-hop with minimal energy consumption [10-13].

On-chip wireless interconnects was proposed as the viable alternative for wired communication in order to furnish an effective yet scalable WiNoC architecture [14]. There are numerous architectures of WiNoC that have been proposed, which can be categorized as pure wireless, 2D mesh-based, multiple-tier, small-world and irregular topology. Zhao et al. [15] proposed a multi-channel WiNoC called McWiNoC that uses conventional NoC architecture as a basis, which is very adaptable to the transmission range of its radio hubs. Meanwhile, a network-based processor array (NePA) that is a hybrid WiNoC that extends 2D mesh NoC with a single bidirectional link between two neighbouring nodes was proposed [16]. Authors in [17] have proposed an architecture known as WCube, a recursive wireless interconnects with a multi-tier structure that includes both of the wireless backbone and wired edges. This architecture can cope with scaling

Journal homepage: http://beei.org/index.php/EEI
limitation of the demand number of on-chip cores. Instead of long wire insertion to improve NoC performance [18], works in [19, 20] proposed a WiNoC architecture based on the small-world features. Wireless links are used to create one-hop shortcuts. These alternatives have proven to enhance the network performance in NoC. In [21], an irregular mesh-WiNoC topology that is established by wireless links has been proposed to provide high efficient and low cost distributed minimal table based routing strategy.

Several research challenges in the design of WiNoC communication architectures have been raised and in this research we are emphasis on investigating the impact of routing algorithm and topology size [2, 4, 22, 23]. The routing protocol is important as in general it will influence the most critical network metrics such as delay, throughput and power dissipation [24]. In addition, network topology designs equally vital as the ability if the network to efficiently disseminates packet depend on it underlying topology [25]. Therefore, the comprehensive understanding of efficient routing scheme and topology size in sentence traffic patterns aid in the development of optimal network topology and impact on performance, power and design cost in WiNoC architecture.

The main focus of this paper is the comparative performance evaluation of various routing schemes and traffic patterns with different sizes of network topology in common WiNoC architecture. In this work we analyzed the effect of increasing the network topology size and identify which architecture works best on certain routing policy and traffic model with regard to delay, throughput and power consumption. The results from this comparative evaluation enable and facilitate us to identify useful design trade-offs for optimal development of wireless on-chip network based design. This paper is structured as follows. Section 2 reviews the generic architecture of Wireless NoC. Section 3 discusses the popular routing algorithm for mesh network topology. The experimental setup is highlighted in Section 4. Section 5 spotlights the simulation results and discussion. Finally, the conclusion is given in Section 6 with recommendations for future work.

2. WIRELESS NOC ARCHITECTURE

The architecture formation in NoC is among the dominant factors that affect the overall performance and its network cost [25, 26]. NoC topology is constructed by the physical layout, links between IP cores and channels over the network. Therefore, the selection of topology will affect the hops count for a packet to traverse and also the link lengths between source and destination IP cores. Hence, this explains how topology impacts the network latency. Meanwhile, as data traversing process dissipates energy these hop counts indirectly affects network energy consumption [27-29].

Most conventional NoC architectures are based on multiple-hops wired interconnections among nodes. Unlike conventional NoC, WiNoC harnesses the wireless connectivity as a means of communication among nodes with reduced network latency, particularly for the far-away nodes. Figure 1 illustrates the generic example of 8×8 (64 nodes) mesh WiNoC architecture with 4×4 radio hubs that are distributed evenly over the network. As depicted, WiNoC can be visualized as a two-level network. The first level is a classical wired 8×8 mesh NoC topology, whereas the upper level is the wireless NoC constructed by a number of radio-hubs. In this example, each radio-hub has a concentration of 4 that is shared by 4 routers of the first level wired NoC network. NoC tiles are augmented with radio hub transceivers that allow single-hop communication between far-away tiles that otherwise would require multiple-hops in the classical wired network.

Figure 1. 8×8 WiNoC architecture with 4×4 radio hubs
In WiNoC, the role of radio hubs is in charge of single-hop packet transfer between remote nodes via wireless communication channel. The on-chip antenna and the transceiver are two essential modules in radio hub. The features of both elements are discussed accordingly in Section 2.1 and 2.2.

2.1. On-chip antennas

To be practical for the WiNoC architecture the on-chip antenna must be wideband, sufficiently small and highly efficient [30, 23]. Meanwhile, it has to furnish the best power gain at the minimal area overhead. A metal zig-zag antenna has been demonstrated to have these criteria [31]. Furthermore, this antenna also provides negligible effect of rotation on received signal strength, hence making it the most appropriate candidate for WiNoC application [15]. Figure 2 shows the detail of the zig-zag antenna structure.

![Zig-Zag Antenna](image)

Figure 2. The zig-zag antenna [8, 32]

2.2. Wireless transceiver architecture

The low power design of wireless transceiver is the pivotal factor to ensure the desired performance in the WiNoC system [33-36]. For that reason, low-power design considerations were taken into account for both architecture and circuit levels of the transceiver [37-39]. As depicted in Figure 3, a generic WiNoC OOK transceiver composed by a transmitter and a receiver that share the same antenna by means of a RF-switch [31, 40, 41]. As shown in figure, a token controller is present to assure that the wireless channel is not busy at the moment of a transmission [42]. If the channel is free incoming flit will be converted in a serial fashion by mean of the serializer. The main task of transmitters consists in adapting the data incoming from the electrical medium to the wireless medium by means of an antenna. In particular, a transmitter is constituted by a serial converter, which converts parallel streams of data (flits) in a serial fashion. An OOK modulator converts data in higher frequency signal that will be delivered to the antenna via a power amplifier (PA). The structure of the receiver is the opposite of the transmitter. Radio frequency signals will be converted in a baseband stream of data with a demodulator. A deserializer converts a serial stream in to a flit.

![OOK WiNoC Transceiver](image)

Figure 3. OOK WiNoC transceiver [31]

3. ROUTING ALGORITHM FOR MESH TOPOLOGY

The routing algorithm used in the underlying on-chip communication network is essential aspect that distinguishes various proposed NoC architectures [43]. In this section, we focus on well-known routing algorithm for mesh topology. With respect to mesh topology, it is easy to accomplish a shortest path deterministic routing by employing a simple variation of dimension order routing such as XY routing [44]. XY routing ensures deadlock and livelock freedom, but it provide no adaptiveness. This algorithm is a table-less routing technique whereby each packet is routed first in X direction and after it reaches the same X as the destination address, similarly moves along the perpendicular dimension.
A number of shortest path, deadlock-free, partially adaptive routing algorithms are based on restricting certain turns are West-first, North-last and Negative-first [45]. As their names imply, West-first requires that a packet is routed to the west direction first, if it is a productive direction, otherwise any shortest path can be taken. Similarly, North-last requires that if north is productive direction, to be taken last, while negative first requires that. All these schemes imply that certain turns are prohibited as shown in Figure 4. As can be seen in Figure 4(a), XY routing, by restricting four turns becomes completely deterministic. The partially adaptive routing algorithms restricts only two turns. West-first, by requiring that the west direction is taken first if required, allows all possible shortest path when a packet has to travel eastward, clearly favorings traffic toward the east direction. Likewise, North-last by demanding that traffic going toward the north takes the north direction last, allow all possible shortest paths for traffic going toward the south but again, only a single path for traffic going north. Clearly this algorithm favours traffic going from north to south. Finally, Negative-first requires that in case the packet destination is toward any negative axis, horizontal or vertical, along with any other direction, then the packet should be routed first toward that negative axis direction and afterward toward to other direction. All these schemes imply that certain turns are prohibited as shown in Figure 4(b)-(d).

The Odd-even model is one of the most popular partial adaptive wormhole routing algorithms in 2D mesh on-chip interconnection network [46]. Unlike the turn model which prohibits certain turn in all locations of the network, in the Odd-even model some turns are restricted only in even columns and some other turns are prohibited in odd columns. Odd even rules can be described by these dual rules. First, east turns cannot be taken in even columns as depicted in , and second, north turns cannot be taken in odd column as shown in Figure 4(e).

A routing algorithm called dynamically adaptive and deterministic (DyAD) combines the advantages of both deterministic and adaptive routing schemes is presented in [47]. This approach is based on the current network congestion since each router in the network continuously monitors its local network load and makes decision based on this information. When the network is not congested, the DyAD router works in a deterministic mode, thus enjoying the low routing latency enabled by deterministic routing. On the contrary when the network becomes congested, the DyAD router switches back to the adaptive routing mode, and thus avoids the congested links by exploiting other routing path.

![Figure 4. Possible turns in turn model routing algorithm](image)

(The solid lines indicate the allowable turns and the dash lines indicate unallowable turns)

## 4. EXPERIMENTAL SETUP

The evaluation between NoC and WiNoC architectures employed a cycle-accurate systemC based simulator called Noxim [48]. Mesh-based topology is chosen because of its natural layout easily map to an IC. In addition, due to its physical regularity this network is also scalable and adaptable with simple routing algorithm. In order to identify how the architectures reacts in different conditions, the simulation are carried out under various traffic scenarios namely hotspot, transpose and random. Simulations have been done for
three network scales namely 8×8 (64 nodes), 16×16 (256 nodes), and 32×32 (1024 nodes). Table 1 shows the simulation setup, while the traffic descriptions are shown in Table 2.

Table 1. Simulation setup

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network Sizes</td>
<td>8x8, 16x16, 32x32</td>
</tr>
<tr>
<td>Number of Radio Hub</td>
<td>4x4</td>
</tr>
<tr>
<td>Number of Channels</td>
<td>8</td>
</tr>
<tr>
<td>Simulation Time</td>
<td>100,000</td>
</tr>
<tr>
<td>Technology</td>
<td>65 nm</td>
</tr>
<tr>
<td>Clock Frequency</td>
<td>1 GHz</td>
</tr>
<tr>
<td>Switching Mechanism</td>
<td>Wormhole</td>
</tr>
<tr>
<td>Radio Access Control</td>
<td>Token Packet</td>
</tr>
<tr>
<td>Flit Size</td>
<td>32 bits</td>
</tr>
<tr>
<td>Routing Algorithm</td>
<td>XY, west-first, north-last,</td>
</tr>
<tr>
<td></td>
<td>negative-first, odd-even,</td>
</tr>
<tr>
<td></td>
<td>DyAD</td>
</tr>
<tr>
<td>Wireless Data Rate</td>
<td>16 Gbps</td>
</tr>
<tr>
<td>Wireless Communication</td>
<td>millimeter-wave</td>
</tr>
</tbody>
</table>

Table 2. Traffic patterns

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>Uniform distribution of traffic from source to destination where each node sends packets to others with the same probability.</td>
</tr>
<tr>
<td>Transpose</td>
<td>Bit-permutation traffic adopting transpose matrix.</td>
</tr>
<tr>
<td>Shuffle</td>
<td>Bit-permutation traffic from source to destination with shifted order address.</td>
</tr>
</tbody>
</table>

For different cases of traffic distribution, the experiments are carried out based on 8×8, 16×16, and 32×32 network sizes. The large size up to 32×32 (1024 cores) WiNoC is chosen to investigate the impact of radio hub allocation and placement for WiNoC architecture in handling large network. As previously mentioned, the comparative analysis between WiNoC and mesh NoC are made in terms of global average delay, network throughput and energy consumption.

5. RESULTS AND ANALYSIS

5.1. Effects of routing algorithm

Routing is the procedure that is employed to deliver the flits onward to the sentence directions over the network between from the source to its destination. The performance of NoC and WiNoC architecture have been evaluated with the popular different routing scheme namely XY, west-first, north-last, negative-first, odd-even and DyAD. The selection behind of these routing algorithm are because they are based on wormhole switching mechanism that provide deadlock and livelock freedom in the two-dimensional mesh topology [49].

To investigate the impact of routing algorithms in various topology sizes, three sizes (8×8, 16×16 and 32×32) of networks have been simulated under transpose traffic. The rationale behind the selection of this non-uniform traffic distribution is because of its practicality for the real world application [50]. Figure 5 shows how the performance of the NoC and WiNoC architecture with respect to network load for 8×8 network size. As can be observed, the performances of both architectures are almost identical for all routing schemes. However, from the context of routing algorithm, Odd-even is shown best algorithm since it can cater higher network load (saturation point at PIR 0.02).
As we increased the network size to 16x16 and 32x32, the performance of WiNoC is improved for all routing algorithm in comparison with the conventional mesh-NoC. This is shown in Figure 6 and 7 respectively. For instance, as can be seen in Figure 6(e) WiNoC has more capability in handling more network load in odd even routing algorithm which has the saturation point at 0.005 instead of 0.003 for NoC mesh. The reason of this is due to in the odd-even routing scheme, some turns are prohibited only in even column meanwhile some other turns are restricted in odd column. Hence, the degree of adaptiveness offered by this scheme is higher than others scheme.

Figure 6. Comparative performance evaluation between NoC and WiNoC architecture under different routing algorithm with 16x16 network size
Figure 7. Comparative performance evaluation between NoC and WiNoC architecture under different routing algorithm with 32x32 network size

5.2. Effects of Network Size
To investigate the effect of network size in both architectures, three various network sizes (64, 256 and 1024 nodes) have been simulated under numerous traffic distributions as described in Table 2. Odd-even routing scheme were adopted due to its advantage in handling more network load as has been explained in Section 5.1. This routing is utilized in the wired NoC layer until a wireless node with radio hub is found. Flits are then sent to its destination through the one-hop wireless channel.

5.2.1. Impact on Packet Injection Rate
Saturation throughput can be defined as the network throughput at which packet injection rate (PIR) begins to saturate. It is a common metric utilized to evaluate network performance [51, 52, 53]. At this point of network saturation throughput, the system is not effective in handling the network loads anymore. When PIR is set to 0.001, the rate of injection is 1 packet for every 1000 cycles. Since packets are set to 32 flits size, hence it resulted in 0.0032 flits/node/cycle, which corresponds to 0.001 packet/node/cycle of throughput. Figure 8 summarizes the comparison on the network latency of two simulated architectures namely mesh-NoC and WiNoC under different traffic load distributions as we vary the network dimensions.

Figure 8. Global average delay comparisons for various network sizes under different traffic load distributions
The latencies for both architectures are almost identical for all scenarios of traffic patterns and network sizes at low network commitment. However, as the network load increases, the latency of network begin to exponentially increase depending on traffic patterns and system sizes. In general, an architecture that has higher PIR demonstrates a better system because of its ability to manage higher network loads. For instance, in Figure 8(a), at transpose traffic WiNoC saturates at PIR 0.007 compared to NoC at PIR 0.009 has reveals that NoC has better performance in 8x8 system size. In contrast, Figure 8(b) for transpose distribution shows 16x16 WiNoC architecture is saturated at higher PIR, 0.006 compared to 0.004 for NoC which indicates that WiNoC has more desirable performance. Overall, from the network size perspective WiNoCs have disadvantages in 8x8 topology scale. In the contrary, as the topology scale become bigger (16x16 and 32x32) WiNoCs were shown to have more fitting performance due to the factor of radio hub that assist in single-hop long distance wireless communication.

5.2.2. Impact on network throughput

Throughput is another important metric indicator of the performance and quality of a network connection. It can be defined as how the network is able to process the requested packet injection rate and can be represented as flits/cycle. A high ratio of unsuccessful packet delivery will lead to lower throughput and degraded performance. Network throughput is affected by a number of factors such as network congestion and packet loss. Hence, the higher the throughput in the network reflects to more effective system.

Figure 9 shows the network throughput comparisons for various network sizes (8x8, 16x16 and 32x32) under different case of traffic scenarios. As illustrated in transpose traffic, bigger network WiNoC256 and WiNoC1024 have 6.36% and 1.82% respectively higher throughput in comparison with the conventional NoC architecture. The reason for this is the ability of the bigger system architecture that can cope with higher networks load as has been discussed in the impact of PIR in Section 5.2.1.

![Network throughput comparison](image)

Figure 9. Network throughput comparisons for various network sizes under different traffic load distributions

5.2.3. Impact on energy consumption

The reduced energy consumption leads to a better power characteristic in the system architecture. Figure 10 shows the energy comparisons between NoC and WiNoC for varied network sizes subject to different type of traffic patterns. As can be observed, the energy consumption has an incremental trend as the network increase in size. In addition, WiNoCs utilized more energy consumption for all cases networks sizes and traffic distributions.

This is due to the wireless communication activities introduced by on-chip transceivers in WiNoC composition. On the other hand, wireless radio hubs routers are more energy-hungry than conventional NoC routers. From the context of network size, as in transpose traffic for example when network topology grows bigger WiNoC consume lower energy consumption reduced from 87% (8x8), 63% (16x16) to 27% (32x32). This revealed that WiNoC benefited in energy saving when the network size is larger. The factor of this savings is due to the least wireless utilization of the radio hub in the bigger WiNoC as described in Section 5.2.4.
5.2.4. Impact on wireless utilization

The considered traffic settings (uniform random, transpose and shuffle) allow the network to work in different regions characterized by a different utilization of the wireless medium. In particular, wireless utilization can be defined as the ratio between the number of communications that use, totally or in part, the wireless medium and the total number of communications. As shown, the percentages of wireless utilization are inversely proportional with system size.

The high percentage (on average 86%) of wireless usage in 8×8 is because of the dense distributions of 16 radio hubs in 64 nodes WiNoC. In the contrary, for biggest network (32×32) from this experiment resulted in low percentage (on average 3%) of wireless usage due sparse distribution of the radio hubs. Optimally, the 16×16 network architecture give the most suitable concentration that use on average 27% wireless utilization. This is practical since the radio hubs were used for the purpose of long distance communication. Hence, this justifies the reason why 256 nodes give the best results in terms of performance as well as network throughput.

6. CONCLUSION

The objective of this research is to investigate the impacts of different routing strategy and the effects in varying the number of network sizes between the classical mesh NoC and WiNoC architecture. From experimental results, can be concluded that WiNoC architecture performs its best among other topology at 16×16 network size in nonuniform transpose traffic with better average speedup (~1.2×) and improved network throughput by 6.36%. However, it has the trade-off on higher energy consumption in comparison with conventional mesh NoC. For future work, we target to look into the implication of the number radio hub placements on several specific WiNoC architectures such as iWise, WCube and MeWiNoC.
ACKNOWLEDGEMENTS

The first author would like to thank his employer Universiti Malaysia Sarawak (UNIMAS) and Ministry of Education Malaysia for financing his PhD studies.

REFERENCES


BIOGRAPHIES OF AUTHORS

Asrani Lit received his B. Eng. and M. Eng. degree in Microelectronics & Computer System from Universiti Teknologi Malaysia. He is doctoral student in the Division of Electronic and Computer Engineering at the same university. His research interests focus on Network-on-Chip and on-chip interconnect architectures.

Mohd Shahrizal Rusli received his Ph.D in Electrical Engineering (2016), M. Eng. (Electronic and Telecommunication) (2010) and B. Eng. in Computer (2006) degree from Universiti Teknologi Malaysia (UTM), Johor. Between 2014 and 2015, he was attached to University of California Irvine, USA and University of Catania, Italy under the supervision of Prof. Dr. Nader Bagherzadeh and Assoc. Prof. Dr. Maurizio Palesi, respectively. He currently serves as senior lecturer at the Division of Electronic and Computer Engineering, UTM where his research interest and specialization are in the field of embedded processor system, deep learning application in embedded system, low-power and energy management in network-on-chip and wireless network-on-chip. He has published numerous journals, proceedings and a book chapter. He received several recognitions internationally and nationally such as best paper and national competition awards related to his research interest. He has been appointed as a reviewer for several journals. He is also an active researcher in VLSI and Embedded Computing Architecture Design (VeCAD) Research Group and has participated in several courses and seminars.

Muhammad Nadzir Marsono is an Associate Professor at Division of Electronic and Computer Engineering, Universiti Teknologi Malaysia. His research interests include system level integration, working in multi-processor System-on-Chip, Network-on-Chip, network algorithmics, and network processor architectures.