

Overlaid Mesh Topology Design and Deadlock Free Routing in Wireless Network-on-Chip

Dan Zhao and Ruizhe Wu
 The Center for Advanced Computer Studies
 University of Louisiana at Lafayette
 Lafayette, LA 70503
 {dzhao,rxw2563}@cacs.louisiana.edu

Abstract—To bridge the widening gap between computation requirements of terascale application and communication efficiency faced by many-core processor chips, wireless Network-on-Chip (WiNoC) has been proposed by using the recently developed CMOS ultra wideband interconnection. In this research, we propose an unequal RF nodes overlaid mesh topology design to improve the on-chip communication performance. A network capacity model is developed for fast searching of optimal topology configuration. A high-efficient, low-cost zone-aided routing scheme is designed to facilitate deadlock freedom. The simulation study demonstrates topology modeling effectiveness, routing efficiency, and promising network performance of the overlaid mesh WiNoC over a regular 2D mesh baseline.

Keywords—wireless network-on-chip; overlaid mesh topology; zone aided routing; octagon turn model; deadlock avoidance.

I. INTRODUCTION

With silicon technology scaling, multi-processor chips (CMPs) are moving towards many-core structures to achieve energy-efficient performance. Network-on-chips (NoCs) are in replace of conventional shared-bus architectures to provide scalable and energy-efficient communication for CMPs using integrated switching network. In the meantime, RF/wireless interconnect technology [1]–[3] has emerged recently to address future global routing needs and surpass the fundamental limitation of hard-wired electronic interconnects. Among them, the UWB interconnect (UWB-I) by using the carrierless impulse-radio ultra wideband (UWB) technology [4], [5] brings in new opportunity for low-power, high bandwidth, short-range communication. The high flexibility and free-of-wiring make UWB-I an attractive solution for the on-chip inter-core communication.

The UWB-I is based on transverse electromagnetic wave propagation by using on-chip antennas. Since its signal has very short pulse duration, high data rate with constant signal-to-noise ratio is achieved by increasing the bandwidth. It consumes low power (e.g., a few milliwatts), attributed to its very low duty cycle (typically $< 0.1\%$). With carrier free implementation, the RF circuits can be very simple. The pulse

This research was supported in part by the National Science Foundation under grant number CNS-0821702 and CCF-0845983.

position modulation is typically used to modulate a sequence of very sharp Gaussian monocycle pulses. The received signal and the template signal is automatically correlated during the demodulation. Various CMOS-integrated high transmission gain printed antennas such as linear dipole, meander dipole, folded dipole, zigzag dipole and loop antenna [6], have been proposed for on-chip implementation.

Following ITRS projection, it is possible to build RF circuits operating at $\sim 20GHz$, achieving a data rate of $\sim 20Gbps$ /band (with $1bps/Hz$ bandwidth efficiency) in $32nm$ CMOS technology. With multiple bands, the aggregate data rate can be further improved for tera-scale computing. With RF technology scaling, the required RF circuitry and antenna sizes will scale down, tremendously reducing implementation cost and increasing on-chip interconnection flexibility. Moreover, the energy consumption per bit is expected to scale down too. Table I [7] summarizes the UWB interconnect scalability which facilitates the fundamental architectural shift to many-core and tera-scale computing.

Table I
UWB INTERCONNECT SCALING

| Technology (nm) | 90 | 65 | 45 | 32 | 22 |
|---|-------|-------|-------|-------|------|
| Cut-off freq. (GHz) | 105 | 170 | 280 | 400 | 550 |
| Data rate per band (Gbps) | 5.25 | 8.5 | 14 | 20 | 27.5 |
| Dipole antenna length (mm) | 8.28 | 5.12 | 3.11 | 2.17 | 1.58 |
| Meander type dipole antenna area (mm ²) | 0.738 | 0.459 | 0.279 | 0.194 | 0.14 |
| Power (mW) | 33 | 40 | 44 | 54 | 58 |
| Energy per bit (pJ) | 6 | 4.7 | 3.1 | 2.7 | 2.1 |

With the high bandwidth, low power and ultra-short range communication provided by UWB-I, the wireless radios are deployed on chip in replace of wires to establish a wireless Network-on-Chip (WiNoC). A WiNoC consists of a number of RF nodes, i.e., wireless routers, each associated with a processor tile. The RF node has a predetermined transmission range. The processor tiles access the network via RF nodes, and their packets are delivered to destinations through multi-hops across the network. Multiple channels are assigned among the nodes to ensure transmission paral-

lelism and reduce channel contention. A node may receive packets from its neighbors fall within its transmission range along dedicated channels, and thus are connected by wireless links. A collection of RF nodes connected by high BW wireless links, forming WiNoC topology. RF node architecture mainly implements routing, channel arbitration, virtual output queuing and congestion control mechanisms for high-speed cost-efficient on-chip communication.

In this work, we design an overlaid mesh wireless NoC, where unequal RF nodes are dispersed on-chip as wireless routers to forward data (Sec. II-A). By configuring the RF nodes placement, various topologies may be formed. An effective network capacity modeling scheme is designed for fast allocation of an optimal topology from a large searching space (Sec. II-B). Benefited from long link transmission, we further develop a simple and efficient logic-based zone-aided routing scheme for distributed and deadlock-free routing (Sec. III-A, Sec. III-B). Routing efficiency is further improved by an enhancement technique which greatly improves transmission concurrency by evenly distributed traffic density. The thus induced deadlock problem is resolved by a simple buffer ordering scheme. (Sec. III-C).

II. OVERLAID MESH TOPOLOGY DESIGN

A. Architectural Topology Design

To improve network performance of WiNoC where the RF nodes are placed in a regular mesh, *unequal RF nodes*, i.e., wireless routers, are deployed in a way that *small RF nodes* (with shorter transmission range T and lower link bandwidth) form a mesh while *big RF nodes* (distributed at distance of nT with longer transmission range of $\sqrt{2}nT$ and higher link bandwidth) are overlaid to form a fully connected mesh as shown in Fig. 1. As thus, we introduce two types of meshes, a *base mesh* which is simply a regular 2D mesh formed by both small and big nodes, and a *full mesh* formed only by big nodes where the big nodes within a grid are fully connected to each other. Further, the big node has starry ends constituting of unidirectional direct links to the nearby small nodes which fall within its transmission range.

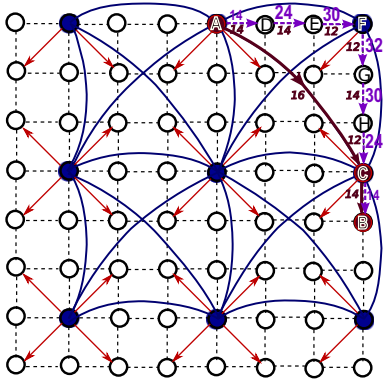


Figure 1. Illustration of an overlaid mesh WiNoC.

The overall topology is a flat structure with both short and long wireless links formed by small and big RF nodes. It looks like a full mesh overlaid on a base mesh, thus we simply name it an *overlaid mesh* topology. It's worth mentioning that more channel bandwidth is allocated to the big nodes in order to alleviate denser traffic congestions at the big nodes (due to higher node radix).

An overlaid mesh potentially improves WiNoC performance from two aspects, *reduced hop count* with long range wireless links and *reduced traffic congestion* due to efficient traffic distribution. On one hand, routing cost can be reduced due to significant reduction in the hop count by using long links as much as possible in overlaid mesh. The direct links from the big nodes to its nearby small nodes further reduce the hop count. For example, as shown in Fig. 1, only two hops are needed to deliver a packet from A to B via C while it requires a total of 7 hops to route the packet along a regular 2D mesh which results in 71% hop count reduction. On the other hand, using long links may relieve the traffic congestion problem in the network specially for steaming data flows. For instance, considering all-to-all communication in a regular 2D mesh and an overlaid mesh, respectively. Without loss of generality, we compare the traffic density on selected links after applying shortest path routing for all-to-all communication. As we can see in a regular 2D mesh, the traffic hot spots may be formed at nodes E, F, G and H while the traffic is more evenly distributed in the overlaid mesh.

B. Overlaid Mesh Topology Configuration and Optimization

In an overlaid mesh, big RF nodes are placed in a way to tradeoff between routing path cost and network congestion. For a $N \times N$ WiNoC with big nodes deployed at distance of nT , we may generate several different topologies by changing the big nodes placement distance. When increasing n (i.e., big nodes are separated farther), a packet may be delivered to the destination with less hops by using longer links formed by big nodes. However, as farther separated big nodes have higher radix (due to more starry ends), the traffic would be more congested at big nodes, thus increasing the end-to-end delay. In the meantime, with less number of big nodes in the network, the number of channels needed may be reduced. As a result, per-link BW may be improved¹. It becomes essential to study the impact of big nodes placement and the corresponding number of big nodes on topology formation, and consequently the impact on network performance.

Various configurations of overlaid mesh topology can be formed with varying big nodes placement for a given network scale. It is important to find out an optimal or near-optimal configuration which results in the best possible

¹With a single band implementation, the overall bandwidth will be split among the number of channels assigned in the system.

network performance at given network scale. Such a topology design problem is basically an big nodes placement problem. As big nodes may likely form traffic hot spots in the network, they will be distributed at the center of their neighboring small nodes where their starry ends may reach out as equally as possible. Now the question is: which placement may result in the best possible network capacity? We derive an efficiency network capacity modeling scheme to fast approach an optimal topology configuration without running comprehensive WiNoC network simulation.

Network Capacity Modeling: Considering a wireless link l_{AB} with node radix of n_A (or n_B) for node A (or B), the possibility of transmitting a packet between A and B is $P_{AB} = \frac{1}{n_A \cdot n_B}$. Assume that m paths are routed through it from n_1 sources $S = \{S_1, S_2, \dots, S_i, \dots, S_{n_1}\}, i \in n_1$ to n_2 destinations $D = \{D_1, D_2, \dots, D_j, \dots, D_{n_2}\}, j \in n_2$. Considering the worst case traffic contention in the collision domain of nodes A and B where all neighbors are involved in potential data transmission and only one transmission is allowed per frame time to avoid collision, the expected transmission time on link l_{AB} for m paths passing through it would be $ETT(m) = m \cdot n_A \cdot n_B \cdot FrameTime$. Thus the possibility of transmitting a packet for all m paths passing through link l_{AB} is:

$$P_{AB}(m) = \frac{1}{m \cdot n_A \cdot n_B} \quad (1)$$

A link capacity, i.e., the maximum packet carrying capacity of a link, is strictly coupled with the effective bandwidth of the wireless link by considering channel contention overhead. It is determined by $Cap(l_{AB}) = P_{AB} \cdot BW_{AB} = \frac{BW_{AB}}{n_A \cdot n_B}$. When excessive packets flow in a path $S_i \rightarrow D_i$, at least one packet will be queued at the bottleneck wireless router and keeps the bottleneck router saturated. Thus pushing more packets into the path cannot improve the flow's throughput. We may find the bottleneck capacity of a link along path $S_i \rightarrow D_i$:

$$Link_Cap(S_i \rightarrow D_i) = \min\left\{\frac{BW_{A_i B_i}}{m_j \cdot n_{A_i} \cdot n_{B_i}}\right\} \quad (2)$$

where $i \in$ any link $l_{A_i B_i}$ on path $S_i \rightarrow D_i$ and $j \in$ all paths passing through $l_{A_i B_i}$. Therefore, the minimum capacity of path $S_i \rightarrow D_i$ is determined by:

$$Path_Cap(S_i \rightarrow D_i) = HopCount_{S_i \rightarrow D_i} \times Link_Cap(S_i \rightarrow D_i) \quad (3)$$

The overall network capacity can be simply calculated as the sum of all path capacity.

$$Net_Cap(N \times N) = \sum_{j=1 \dots N}^{i=1 \dots N} Path_Cap(S_i \rightarrow D_j) \quad (4)$$

Based on the above modeling, a simple and fast simulator is developed to estimate the overall network capacity under

different topology configuration. Under certain traffic pattern such as uniform, the estimator will record the total number of traffic flows passing through each link and calculate the overall network capacity. The one which deliver the maximum capacity is chosen to be the optimized topology design under given network scale.

III. DEADLOCK-FREE OVERLAID ROUTING

With the great freedom and large bandwidth offered by this emerging WiNoC technology, a deadlock-free logic-based routing scheme is developed for low latency on-chip application. Aiming at low-cost and high-efficient implementation, we propose a *zone-aided routing* scheme, where the whole chip plane is virtually divided into several zones, each associated with one big node serving as the zone header and a group of small nodes. To take the advantage of overlaid mesh topology, routing is performed in a way to use long links as much as possible to shorten routing path lengths. The proposed routing scheme mainly includes three steps, virtual zone division, basic routing scheme, and enhancement techniques.

A. Virtual Zone Division

In order to efficiently utilize the long links in the full mesh, a small source node will first deliver its packet to the closest big node. In other words, it will be shortest-XY routed to a neighboring big node. All the small nodes which will forward their packets to the same big node will be grouped into one *virtual zone* with the big node as the *header* of this zone. The whole network can thus be divided into several virtual zones where the zone headers are located at the center of these zones. Note that some zones at the edge of the network are just in partial size. The zone division for a 8×8 WiNoC is illustrated in Fig. 2.

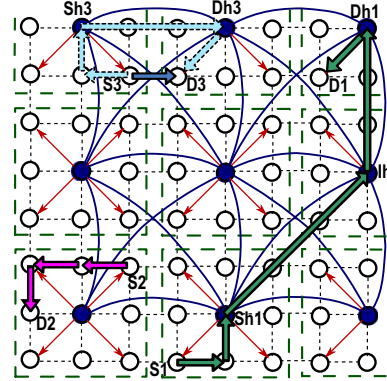


Figure 2. Illustration of zone aided routing.

Assume the XY-coordinate of the left bottom node $(1, 1)$ and the right top node (N, N) in a $N \times N$ network with virtual zone size of $n \times n$. For any small RF node S with its XY-coordinates (x_S, y_S) , a fast process is developed to determine its zone $Z(X_R, Y_R)$ and zone header $H(x_B, y_B)$:

$X_R = \lceil \frac{x_s}{n} \rceil$; $Y_R = \lceil \frac{y_s}{n} \rceil$; $x_B = \lceil \frac{x}{2} \rceil + (X_R - 1) \cdot n$; $y_B = \lceil \frac{y}{2} \rceil + (Y_R - 1) \cdot n$. For example, node $A(7, 6)$ in Fig. 2 is located in zone $Z(\text{III}, \text{II})$ with its header $H(8, 5)$.

B. Basic Routing Scheme

Intuitively, the shortest-path routing can be performed on the overlaid mesh. However, the shortest-path routing involves high implementation cost in order to maintain the forwarding table at routers. Further, the shortest-path routes can easily form cyclic dependency, inducing deadlocks which are not easy to get resolved. The proposed zone-aided routing scheme facilitates simple and efficient logic-based implementation and at the same time shortens the routing paths to the maximum possible by taking the advantage of long links in the overlaid mesh. Most importantly, it ensures deadlock freedom.

Without loss of generality, the basic routing scheme makes the packet forwarding decision at the routers based on the source and destination nodes' position in the zones. When we transmit packets between a source and destination pair, the source will first check where the destination is located based on the destination's address contained in the packet header.

- If the source and destination nodes fall within the same zone, the packet is forwarded to the destination using XY-routing along the short links of the base mesh, such as the path $S_2 \rightarrow D_2$ with 3 hops in Fig. 2.
- If they belong to different zones, the packet is first XY-routed to the source zone header (if the source is a small node). The packet is then routed along the full mesh with long links using turn-restricted shortest-path routing to reach destination zone header. More specifically, the tilted long links are used to forward the packet until reaching the header located on the same row/column of the destination zone header. The packet is continuously forwarded along the horizontal/vertical long links towards the destination zone header. Finally, the packet directly hops to its destination small node from the destination zone header. For example, a packet generated at source S_1 is first XY-routed to the source header S_{h_1} and then routed to an intermediate header I_h located on the same column of destination header D_{h_1} . The packet is then vertically routed up to D_{h_1} and reaches the destination D_1 via the direct link from D_{h_1} , resulting in a total of 5 hops.

1) *Octagon Turn Model Design*: To ensure deadlock free routing in full mesh, we propose a new octagon turn model as illustrated in Fig. 3. The model involves two abstract cycles, a clockwise cycle and a counter clockwise cycle, each formed by eight turns. It prohibits certain turns in each cycle to break all the cyclic dependencies. The routing employing the remaining turns prevents circular waiting on network resources, such as buffers or channels, and is thus deadlock free.

Essentially, a turn involves a 135 degree change of traveling direction. The eight types of turns defined in the clockwise cycle are $W \rightarrow SE$, $NW \rightarrow S$, $N \rightarrow SW$, $NE \rightarrow W$, $E \rightarrow NW$, $SE \rightarrow N$, $S \rightarrow NE$, and $SW \rightarrow E$. Similarly, the other eight types of turns defined in the counter clockwise cycle are $E \rightarrow SW$, $NE \rightarrow S$, $N \rightarrow SE$, $NW \rightarrow E$, $W \rightarrow NE$, $SW \rightarrow N$, $S \rightarrow NW$, and $SE \rightarrow W$. As thus, two rules are restrictively followed by the octagon turn model.

Rule 1. Any packet is not allowed to make the four turns i.e., $W \rightarrow SE$, $N \rightarrow SW$, $E \rightarrow NW$, and $S \rightarrow NE$ at a node as in the clockwise abstract cycle.

Rule 2. Any packet is not allowed to make the four turns i.e., $NE \rightarrow S$, $NW \rightarrow E$, $SW \rightarrow N$, and $SE \rightarrow W$ at a node as in the counter clockwise abstract cycle.

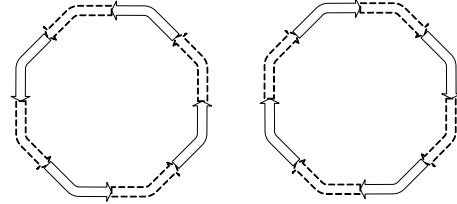


Figure 3. The octagon turn model.

2) *Turn-Restricted Shortest-Path Routing*: The turn-restricted shortest-path routing is performed in full mesh to deliver packets between big nodes or the zone headers. As designed in the above basic routing scheme, all 45-degree and 90-degree turns are prohibited when routing in full mesh besides the prohibited turns designated in the octagon turn model. The 0-degree and 180-degree turns are incorporated without introducing cycles. The routing is *minimal* as it routes a packet via a shortest path between a source-destination pair. The turn-restricted shortest-path routing is deadlock free as it follows the rules of octagon turn model and prohibits all 45-degree and 90-degree turns in full mesh.

While the XY-routing on the base mesh and the turn-restricted shortest-path routing on the full mesh are deadlock free respectively, it can be derived that **the basic zone-aided routing scheme in overlaid mesh is deadlock free** by contradiction. Assume a set of packets form a resource usage dependency cycle. The formed waiting path must interweave all types of routing paths. According to zone-aided routing, short links for routing within the zone and long links for traversing the zones. So any packet may enter and traverse the full mesh at most once. In order to form a cycle, the waiting path must include a deadlocked path segment where a packet gets off the full mesh and is forwarded to the small destination node. However, when a packet leaves the full mesh, it is consumed immediately by the destination small node via a direct link from the destination's zone header. So no circular waits will be formed.

C. Routing Efficiency Enhancement

Applying the basic routing scheme reduces routing cost in terms of hop count by using long links as much as possible, which however may cause severe traffic congestion at big nodes. In order to achieve more even traffic distribution in a network so as to alleviate traffic congestion at big nodes, we propose a routing enhancement technique. The basic idea is that if any pair of source and destination are not located in the same zone while their Manhattan distance $|y_D - y_S| + |x_D - x_S|$ falls within a threshold distance, XY-routing is performed instead of the turn-restricted shortest-path routing to deliver packets between them. The key is the threshold distance setting.

With the primary purpose to even out the traffic density with proper threshold setting, routing enhancement may reduce hop count if a source-destination pair is located near the border of two adjacent zones. For example in Fig. 2, a packet is sent from source S_3 to destination D_3 which are in two adjacent zones. Following the basic scheme, a packet at S_3 will first be forwarded to S_{h_3} and then is connected to D_{h_3} before reaching D_3 , which takes 4 hops. It only requires 1 hop by applying XY-routing between S_3 and D_3 , a 75% reduction in path length.

1) *The Setting of Threshold Distance:* The routing efficiency with enhancement varies with different setting of threshold. For example in Fig. 4, when the threshold is set at 7, averagely 29.7% more traffic is distributed at a big node than a small node. When the threshold arises to 15, the traffic is quite evenly distributed in the whole network. In general, the traffic density would be evened out at higher threshold. However, a larger threshold may lead to longer routing paths without taking the benefit from long link transmission. It becomes essential to study the impact of threshold distance setting on the network performance and strike a balance between routing cost reduction and traffic congestion alleviation.

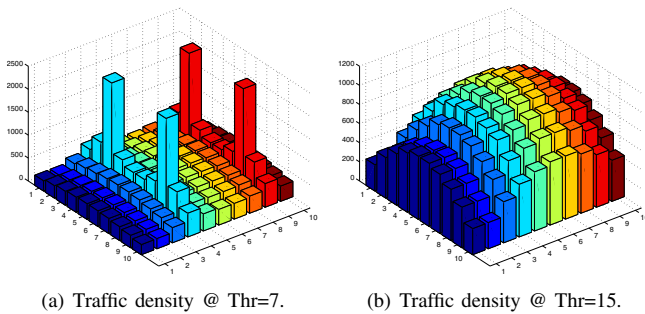


Figure 4. Traffic density under various thresholds.

To quickly latch on the best threshold setting under certain topology configuration with big nodes separation distance nT and network scale $N \times N$, we determine the upper and lower bonds of its searching range. Roughly, the upper

bound is set at $2(N - 1 - \frac{n}{2})$, where about 96% of all-to-all traffic will perform XY-routing on the base mesh. Thus the overlaid mesh seldom plays a role in routing. The threshold is lower bonded by the point where the XY-route between a source-destination pair is not a shortest path between them. Roughly, when the threshold is set less than n , the packets will simply reach the adjacent zone. Thus, routing with long links may result in more hop count. In a nutshell, the threshold searching space is bounded by

$$n < Thr < 2(N - 1) - n \quad (5)$$

2) *Deadlock Avoidance:* Improving routing efficiency by enhancement however may cause deadlock in routing due to the fact that the introduced XY-routes cross the borders of multiple zones. For example as shown in Fig. 5, data streams are delivered along five routing paths. Among them, four are enhanced XY-routes. Node A wants to send packets p_1 to H , but H 's dedicated virtual output queue² (VQ) is occupied by packets p_2 from G and A has to wait. Similarly, nodes H , E , D and C are waiting to send their packets p_2 , p_3 , p_4 and p_5 as their dedicated VQs at nodes F , D , C and B are blocked by p_3 , p_4 , p_5 and p_1 respectively. Thus the set of packets p_1, p_2, \dots, p_5 generate a cyclic dependency and wait on each other to release the buffers. A deadlock occurs.

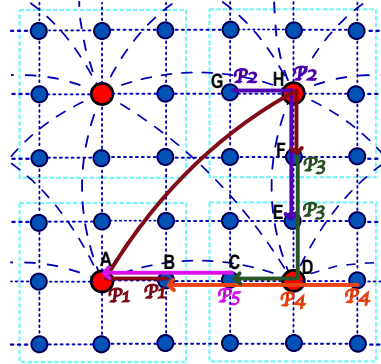


Figure 5. Illustration of deadlock avoidance.

Deadlock can be avoided by a simple buffer ordering scheme. Each VQ will maintain two units of buffer which are ordered into two numbered buffer classes. The 1st class buffer is used to store the packets delivered along the basic zone-aided routing paths while the 2nd class buffer is reserved for storage of packets set along enhanced XY-routes. With buffer ordering, the circular waits on the waiting path created in Fig. 5 will be broken in a way that p_5 can be forwarded to B and stored in 2nd class buffer as p_1 is stored in the 1st class buffer of B . Similarly, p_2 in 2nd class buffer won't block p_1 at H . Consequently, D and F can send out p_4 and p_3 respectively. As a result, deadlock freedom is achieved.

²A dynamic virtual output queuing strategy [8] is employed here for buffering.

IV. PERFORMANCE EVALUATION

A. Simulation Setup

A simulator is developed to evaluate the performance of the proposed WiNoC platform under various network configurations, traffic patterns and network scales. A WiNoC with omnidirectional radio range is built to cover the communication among the processor tiles. Unequal RF nodes (i.e., small/big nodes with short/long transmission range and low/high link BW) are properly distributed to construct an overlaid mesh topology. At a given network scale, an overlaid mesh is configured by varying big nodes placement. A topology generator is further developed to automatically generate a variety of topology configurations under different network scales. The overlaid routing scheme (with both basic and enhanced zone-aided routing as discussed in Sec. III) is developed for efficient and cost-effective routing in the overlaid mesh WiNoC. Multi-channeling which is not the focus of this paper, is facilitated with different RF nodes transmitting in parallel on distinct channels. A virtual output queuing strategy is used for cost efficient buffering. A backpressure based flow throttling scheme is implemented for congestion control.

The overlaid mesh WiNoC network performance is evaluated in terms of end-to-end delay and network throughput at various traffic injection rate. The *network throughput* is the average rate of successful message delivery over the network. The *end-to-end delay* is defined as the time needed to deliver a packet successfully from a source to a destination node. The *injection rate* is set as a fraction of the total number of packets injected in a certain traffic pattern. At each injection rate setting, a total of 8000 packets are generated and injected into the network. For synthesis traffic simulation, the wireless bandwidth of small node is set at 1Gbps while the big node's bandwidth quadruples.

B. Topology Configuration Performance Impact

We will study the performance impact of topology configuration granting to big nodes placement. Fig. 6 verifies how accurate of our network capacity modeling scheme proposed in Sec. II-B and how effective of using it for fast and efficient topology configuration. We compare the network capacity obtained under the fast estimator with the network throughput under the complex WiNoC simulator as described in Sec. IV-A, where the performance trend under various topology configuration in terms of different RF nodes placement is studied for a 10×10 WiNoC. As we can see, the best performance is achieved when the big nodes separation distance is set at $6T$, forming a 2×2 full mesh. More important, the estimated network capacity follows exactly the same trend as the simulation result and reaches its peak performance at $6T$.

We further compare the selections of best-performance topology configuration obtained under both the estimation

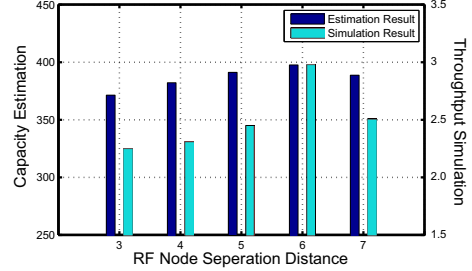


Figure 6. The comparison of estimated and simulated throughput.

and simulation under various network scales. As we can see in Fig. 7, the estimator's topology configuration choice matches well with the choice of the simulator. For example, a 3×3 RF nodes placement where its separation distance is $3T$, is chosen as the optimal configuration for the 8×8 overlaid mesh WiNoC. We argue that the network capacity estimator can be employed for fast searching of an optimal topology configuration at given network scale in favor of its estimation accuracy and searching speed without comprehensive network simulation.

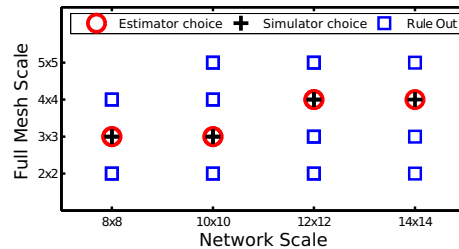


Figure 7. Selection of optimal topology configuration.

C. Routing Efficiency and Routing Cost

We evaluate the routing performance in terms of hop count. As discussed in Sec. III, the overlaid routing improves routing efficiency and reduces routing cost with a logic-based routing scheme instead of a table-based scheme, which however doesn't guarantee shortest path. We compare the average hop count with the shortest-path routing. As shown in Fig. 8, the basic routing has a hop count very close to the shortest-path routing. For example, under 8×8 , 10×10 and 12×12 WiNoC, the hop count is increased by 1%, 4.8% and 9% respectively. The hop count is further increased by 0.2%, 2.8% and 14% respectively by applying the enhancement. Even that, the overlaid mesh demonstrates its advantage in hop count reduction (by using long links) over the regular 2D mesh (i.e., baseline). As we can see, the reduction may reach 17% in average and as high as 20.8%.

We further study the average hop count changing with the threshold under different network scales and the results are given in Fig. 9. The impact of threshold setting on the hop count are in two folds. On one hand, using enhanced XY-routes may reduce hop count for those nodes located near the border of two adjacent zones. For a small threshold setting, such reduction in hop count is more obvious. For

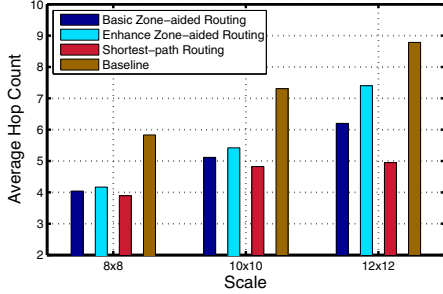


Figure 8. Hop count comparison with shortest-path and baseline.

example, the average hop count of the shortest-path routing under 10×10 WiNoC is 4.98 as in Fig 8. The hop count of the enhanced routing at threshold=7 is increased by 4.8% when compared with shortest-path routing. However, the increment may reach up to 20% when threshold doubles (=15). On the other hand, increasing the threshold means more packets will transit over the enhanced XY-routes instead of using long links to shorten the routing path which finally leads to the increase of hop count. For example, when the threshold changes from 7 to 11 under the 8×8 WiNoC, the hop count is increased by 11%. However such hop count increment will gradually slow down when the threshold increases further. For instance, when the threshold changes to 15, the hop count is increased just by 2%.

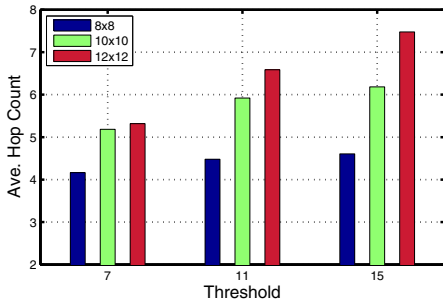


Figure 9. Hop count under different threshold.

D. WiNoC Network Performance

Threshold Performance Impact: We study the impact of threshold setting on the network performance of an example 12×12 overlaid mesh WiNoC where the optimal configured topology with 3×3 full mesh (RF nodes separation distance=5) is used for experiment under uniform traffic. As we can see from figures 10 and 11, a better performance is achieved at a higher threshold by taking the advantage of congestion alleviation at big nodes. For example, the best performance is obtained when threshold is 15. Although higher threshold increases hop count in general, more evenly distributed traffic (e.g., the traffic density is dropped by 73.6% and 31.3% respectively at threshold of 7 and 15 when comparing to the density at threshold=2) may greatly improve the transmission concurrency while relieving traffic congestion, resulting in better end-to-end performance.

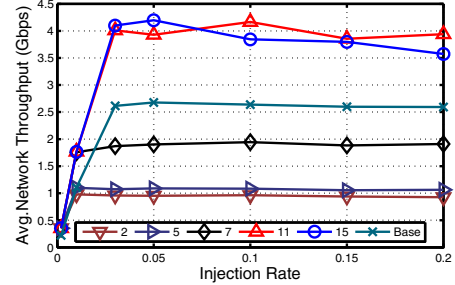


Figure 10. Network throughput under various threshold setting.

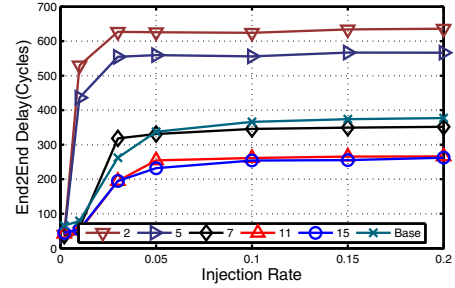


Figure 11. End-to-end delay under various threshold setting.

When comparing the performance with the baseline, we can see that the overlaid mesh outperforms the 2D mesh. Its end-to-end delay is reduced by 25% at the best threshold setting (=15) while the network throughput is improved by 40%. Such performance improvement was attributed to three key factors: improved transmission concurrency, traffic redistribution and reduced hop count. We have discussed hop count reduction in Sec. IV-C, we will briefly discuss the other two factors. The transmission concurrency is improved by the increased wireless connectivity in overlaid mesh and consequently more effective bandwidth distribution with multi-channeling. In the meantime, the overlaid mesh shows a more evenly distributed traffic when comparing the traffic density at threshold=15 with baseline, thus further alleviating traffic congestion.

Traffic Pattern Performance Impact: The network performance under two synthetic traffics are studied: uniform and transpose. Under *uniform traffic*, a.k.a., all-to-all, each RF node uniformly injects packets into the network with randomly generated destinations. A node at (x, y) sends its packets to the node at (y, x) generates the *transpose traffic*. As we can see from figures 12 and 13, overlaid mesh demonstrated its performance advantage over long range communication oriented applications. For example, there are sufficient long traverse range traffics under uniform traffic. The overlaid mesh may thus reduce the delay with the use of long links and at the same time evenly distributing the traffic with enhancement technique. The reduction in delay may reach as high as 33% over the baseline while the throughput is improved by 51%. Meanwhile in simplex transpose traffic, most traffics cannot get benefit from long link transmission, as only the traffic between the top-left and bottom-right

nodes will traverse through the tilted long links. Even that, the overlaid mesh can still improve the throughput by 11% and reduce the delay by 9%.

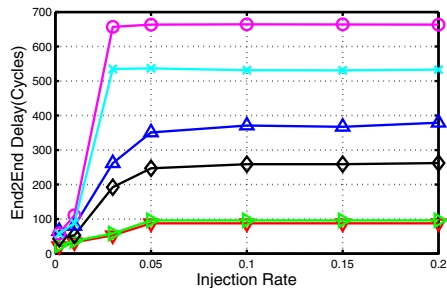


Figure 12. End-to-end delay under different traffic patterns.

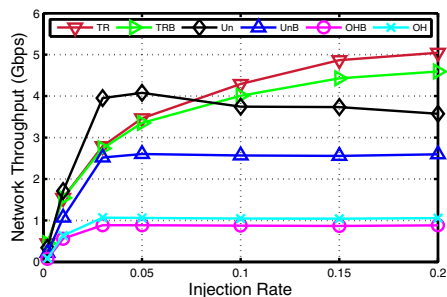


Figure 13. Network throughput under different traffic patterns.

Scalability Study: Moreover, the scalability of overlaid mesh WiNoC is investigated with uniform traffic under three different network scales: 8×8 , 10×10 and 12×12 . As we can see from figures 14 and 15, the network throughput scales up while the end-to-end delay scales down with the increase of the network size. For instance, when scaling from 10×10 to 12×12 WiNoC, the delay is increased by 33%. It's mainly due to the hop count increment by 25%. The throughput reflects the performance changing with the communication concurrency level at various network scale. Roughly, the maximum achievable packet transmission concurrency scales with the network scale ratio.

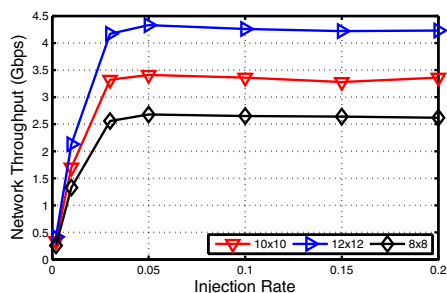


Figure 14. Network Throughput at different network scales.

V. CONCLUSION

We have proposed an overlaid mesh WiNoC platform to improve the on-chip communication of future many-core

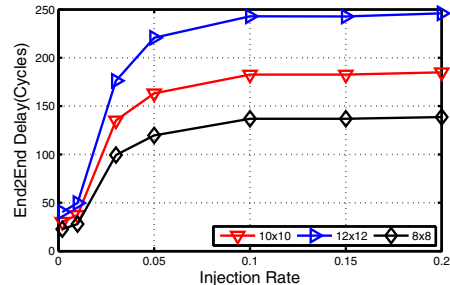


Figure 15. End-to-end delay at different network scales.

chips. The WiNoC topology was designed in a way to achieve a performance optimized configuration by proper big nodes placement. An effective and efficient topology configuration model has been developed for fast searching through the design space without comprehensive network simulation. A high-efficient, low-cost zone-aided routing scheme has been designed to facilitate deadlock freedom while ensuring routing efficiency. The simulation study has demonstrated the promising network performance of the overlaid mesh WiNoC over a regular mesh.

REFERENCES

- [1] M.-C. F. Chang, V. Roychowdhury, L. Zhang, H. Shin, and Y. Qian, "RF/wireless interconnect for inter- and intra-chip communications," *Proc. of The IEEE*, vol. 89, no. 4, pp. 456–466, April 2001.
- [2] B. A. Floyd, C.-M. Hung, and K. K. O, "Intra-chip wireless interconnect for clock distribution implemented with integrated antennas, receivers, and transmitters," *IEEE Journal of Solid-State Circuits*, vol. 37, no. 5, pp. 543–552, May 2002.
- [3] K. Kimoto and T. Kikkawa, "Transmission characteristics of gaussian monocycle pulses for inter-chip wireless interconnections using integrated antennas," *Japanese Journal of Applied Physics*, vol. 44, no. 4B, pp. 2761–2765, 2005.
- [4] T. Kikkawa, P. K. Saha, N. Sasaki, and K. Kimoto, "Gaussian monocycle pulse transmitter using $0.18\mu\text{m}$ cmos technology with on-chip integrated antennas for inter-chip uwb communication," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 5, pp. 1303–1312, May 2008.
- [5] W. M. N. Sasaki, K. Kimoto and T. Kikkawa, "A single-chip ultra-wideband receiver with silicon integrated antennas for inter-chip wireless interconnection," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 2, pp. 382–393, February 2009.
- [6] K.K. O et al, "The feasibility of on-chip interconnection using antennas," in *Proc. of ICCAD*, 2005, pp. 979–984.
- [7] D. Zhao, Y. Wang, J. Li, and T. Kikkawa, "Design of multi-channel wireless noc to improve on-chip communication capacity," in *Fifth ACM/IEEE International Symposium on Networks-on-Chip*, 2011, pp. 177–184.
- [8] R. Wu, Y. Wang, and D. Zhao, "A low-cost deadlock-free design of minimal-table rerouted xy-routing for irregular wireless nocs," in *Fourth ACM/IEEE International Symposium on Networks-on-Chip*, 2010, pp. 199–206.