An innovative routing scheme to reduce communication delay in DMesh networks

Xinyu Wang\textsuperscript{1a)}, Zhigang Yu\textsuperscript{2b)}, Kele Shen\textsuperscript{2c)}, and Moyuan Wang\textsuperscript{1d)}

\textsuperscript{1} College of Management Science and Engineering, Dongbei University of Finance and Economics, Dalian, China
\textsuperscript{2} Department of Computer Science and Technology, Tsinghua University, Beijing, China
\textsuperscript{a)} Distribute2008@163.com
\textsuperscript{b)} yuzg11@mails.tsinghua.edu.cn
\textsuperscript{c)} shenkele2006@yahoo.com.cn
\textsuperscript{d)} moyuanwang@aliyun.com

Abstract: Diameter and average distance are two important metrics in topology selection for on-chip networks. In order to lower down them in Mesh, DMesh is proposed by introducing diagonal links. Compared to Mesh, the port number in routers of DMesh almost doubles, thus high-radix routers are required. The ever increasing pin bandwidth enables design of high-radix routers to improve performance of the overall network. However, the original algorithm DXY always prefers to use the crossing links, and does not make efficient use of physical channels. Though quasi-minimal routing algorithm provides some adaptivity, it limits packets routing in the rectangular defined by source and destination nodes. In this paper, we present a novel deadlock-free and livelock-free routing algorithm for DMesh, and it provides much more flexibility for packets in order to make better use of different links. Simulation results validate the effectiveness of the novel routing scheme as compared to the existed algorithms.

Keywords: on-chip network, DMesh, routing algorithm, deadlock-free, livelock-free

Classification: Integrated circuits

References

1 Introduction

On-Chip Interconnection Network (OCIN) is a key technology for most modern embedded and smart systems by providing higher flexibility and better performance. Design of such networks is of great importance to the overall system, and further influences quality of the ultimate products [1].

Devising an OCIN architecture for specific applications includes three important aspects: flow control, topology development, and routing policy [2]. The flow control technique governs communication between routers, particularly, it determines when packets (or flits) can be forwarded through the switching component inside the router. Flit-level flow control is used to minimize the buffers in router, and hence reduces area footprint. In design of modern routers, input-queued buffers are usually adopted in NoC [3]. As the main storage part, buffers are commonly operated in first-in first-out (FIFO) manner [4]. In this paper, we only consider flit-level FIFO buffering routers with wormhole-based switching technique.

Topology defines the layout of nodes and physical links, and determines the minimal communicating distance between nodes. Mesh received much attention due to its attractive properties, such as regularity, small radix and low link complexity. However, the high diameter and average distance limit its further development in large scale OCINs. Advances in signaling technology has enabled the high-radix routers, which has inspired a lot of new topologies, such as clos [5], Dragonfly [6], Cluster Mesh [7], Spidergon Donut [8], Exchanged Crossed Cube [9], and DMesh [10, 11].

DMesh network introduced diagonal links to Mesh, which reduced the diameter and average distance, as well as inheriting many superior properties in Mesh. DMesh provides much more routing flexibility, and could alleviate traffic congestion in the network. A well-designed routing algorithm should fully exploit the adaptivity under the constraints of deadlock-freedom and livelock-freedom.
Load balancing and routing adaptivity are key aspects to achieve maximum throughput [12]. This paper aims to propose a new routing algorithm for DMesh to exploit the routing adaptivity and balance the load on different links, which ultimately yields to lower latency and higher throughput.

In the rest of the paper, we present related work in Section 2. Section 3 illustrates the idea behind the proposed routing scheme for DMesh networks, and also gives the livelock-freedom and deadlock-freedom proof. Section 4 gives the performance comparisons by comparing with DXY algorithm. Finally, Section 5 concludes the paper.

2 Related work

Among various topologies based on high-radix router, DMesh is migrated from Mesh with diagonal links incorporated, which holds many superior properties from Mesh. Figs. 1(a) and (b) shows the $8 \times 8$ Mesh and DMesh networks. For $n \times n$ networks, comparisons between Mesh and DMesh are shown in Table I. Compared to Mesh, average distance and diameter of DMesh are decreased by 20% and 50%, respectively.

<table>
<thead>
<tr>
<th>Topology</th>
<th>D</th>
<th>AD</th>
<th>BBW</th>
<th>NoL</th>
<th>RR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mesh</td>
<td>$2(n - 1)$</td>
<td>$\approx \frac{2n}{3}$</td>
<td>$2n$</td>
<td>$2n(n - 1)$</td>
<td>5</td>
</tr>
<tr>
<td>DMesh</td>
<td>$n - 1$</td>
<td>$\approx \frac{2n}{3}$</td>
<td>$6n$</td>
<td>$2(2n - 1)(n - 1)$</td>
<td>9</td>
</tr>
</tbody>
</table>

1D refers to Diameter;  
2AD refers to Average Distance;  
3BBW refers to Bisection BandWidth;  
4NoL refers to Number of Links;  
5RR refers to Router Radix.

Load balancing and routing adaptivity are key aspects to achieve maximum throughput [12]. This paper aims to propose a new routing algorithm for DMesh to exploit the routing adaptivity and balance the load on different links, which ultimately yields to lower latency and higher throughput.

In the rest of the paper, we present related work in Section 2. Section 3 illustrates the idea behind the proposed routing scheme for DMesh networks, and also gives the livelock-freedom and deadlock-freedom proof. Section 4 gives the performance comparisons by comparing with DXY algorithm. Finally, Section 5 concludes the paper.

![Fig. 1. An $8 \times 8$ network: (a) Mesh, (b) DMesh.](image)

2 Related work

Among various topologies based on high-radix router, DMesh is migrated from Mesh with diagonal links incorporated, which holds many superior properties from Mesh. Figs. 1(a) and (b) shows the $8 \times 8$ Mesh and DMesh networks. For $n \times n$ networks, comparisons between Mesh and DMesh are shown in Table I. Compared to Mesh, average distance and diameter of DMesh are decreased by 20% and 50%, respectively.

Given the topology, packets are directed to traverse the network from source to destination with the help of distributed routing algorithms. An excellent routing scheme should provide admissible paths in a manner that exploits adaptivity and flexibility to provide low latency and high throughput. Driven by XY routing in Mesh network, an intuitive dimensional order routing (called DXY) can be designed in which the diagonal channel is always selected if the both the offset along dimension X and that along dimension Y are nonzero. Clearly, the DXY algorithm is minimal with only one virtual channel to obtain deadlock-freedom.
Wang has verified that there is severe congestion on diagonal links with low utilization on vertical and horizontal links when minimal routing is adopted in DMesh network [11]. To solve this problem, derived from the concept of Network-based Processor Array (NePA), Wang proposed a distributed adaptive routing algorithm (quasi-minimal) by splitting the DMesh network into two subnetworks: E Subnet and W Subnet, as shown in Fig. 2. Though quasi-minimal provides much more flexibility and adaptiveness than DXY, when a packet comes to a node in the same row (or column) as the destination, it will be routed to destination with no adaptivity. In quasi-minimal routing, packets are always constrained in the rectangular defined by source and destination nodes, without taking into account of paths outside of the rectangular. Unless otherwise specified, we abbreviate quasi-minimal as QM in the following sections. In this paper, based on DMesh network, we presents a novel routing algorithm to further exploit routing adaptivity in order to improve performance of the overall chip.

![Fig. 2. Subnetworks of 8×8 DMesh: (a) WSubnet, (b) ESubnet.](image)

3 The random quasi-minimal routing (RQM) in DMesh network

In DMesh network, each router has 9 different ports, with 1 connected to local processing elements and 8 others connected to neighbor routers. In the rest of the paper, we use $X$ to represent the horizontal link, and $Y$ to represent the vertical link, and $C$ to represent the diagonal link along $y = x$, and $T$ to represent the diagonal link along $y = -x$. Fig. 3 shows the directions of the 8 inter-router links (ports).

![Fig. 3. Eight different directions of output ports in a router.](image)

### 3.1 Description of the proposed algorithm

The DXY routing is minimal, which always routes packets along the diagonal links if neither of the offsets along $X$ and $Y$ is zero. The quasi-minimal routing technique,
Algorithm 1 The proposed routing algorithm in WSubnet.

Input: Current node \((x_c, y_c)\) and destination node \((x_d, y_d)\);
Output: A selected output channel.

\[
\begin{align*}
&\text{off}_x = x_d - x_c, \text{off}_y = y_d - y_c; cset = \emptyset; \\
&\text{if } \text{off}_x \neq 0 \text{ and } \text{off}_y > 0 \\
&\quad \{ \\
&\quad \quad \text{cset} = \{T+, X-, Y+\}; \\
&\quad \quad \text{if } |\text{off}_x| - |\text{off}_y| >= 2cset = cset \cup \{C-\}; \\
&\quad \} \\
&\text{if } \text{off}_x \neq 0 \text{ and } \text{off}_y < 0 \\
&\quad \{ \\
&\quad \quad \text{cset} = \{C-, X-, Y-\}; \\
&\quad \quad \text{if } |\text{off}_x| - |\text{off}_y| >= 2cset = cset \cup \{T+\}; \\
&\quad \} \\
&\text{if } \text{off}_x \neq 0 \text{ and } \text{off}_y == 0 \\
&\quad \{ \\
&\quad \quad \text{cset} = \{X-\}; \\
&\quad \quad \text{if } |\text{off}_x| >= 2cset = cset \cup \{T+, C-\}; \\
&\quad \} \\
&\text{if } \text{off}_x == 0 \text{ and } \text{off}_y \neq 0 \\
&\quad \{ \\
&\quad \quad \text{if } \text{off}_y > 0 \text{cset} = \{Y+\}; \\
&\quad \quad \text{else } \text{cset} = \{Y-\}; \\
&\quad \} \\
&\text{if } \text{off}_x == 0 \text{ and } \text{off}_y == 0 \text{cset} = \{L\}; \\
&\text{return select(cset).}
\end{align*}
\]

Based on the NePA platform for quasi-minimal in [11], one DMesh network can be divided into two subnetworks. Fig. 2 shows the WSubnet and ESubnet for an 8 x 8 DMesh network. The WSubnet is responsible for transferring westward packets, while the ESubnet is responsible for eastward traffic. As for packets with offset along dimension \(X\) being 0, it can be injected into either WSubnet or ESubnet. When the source processing element starts packet transmission, it injects packets into the network via the internal link allocated to the sub-network; When a packet arrives at the destination router, it is ejected from the network via the eternal link.
Algorithm 1 shows the pseudo code of the proposed RQM routing scheme in WSubnet for DMesh. The algorithm for ESubnet can be derived in a similar way, and it was not shown in this paper in case of redundancy. Taking the coordinates of the current and destination nodes as inputs, it returns a selected channel as the output. First, compute the offsets along dimension $X$ and $Y$, and label them with $off_x$ and $off_y$. If both $off_x$ and $off_y$ are non-zero, the diagonal channel along minimal path, $X$ channel and $Y$ channel are candidate options; furthermore, when $|off_x| - |off_y|$ is $\geq 2$, the diagonal channel along non-minimal path is taken as an option. If $off_x$ is non-zero and $off_y$ is zero, the $X$ channel is selected in $cset$, furthermore, when $|off_x|$ is $\geq 2$, the two diagonal channels are also available. If $off_x$ is zero while $off_y$ is non-zero, the $Y$ channel is selected. If both of the two offsets are zero, packet is now in the destination router, and will be transmitted to its local processor via $L$ output port. Here, we use $L$ to represent the eternal channel of the router.

In quasi-minimal routing, packets are always constrained in the rectangular (REC($s$, $d$)) defined by source node $s$ ($x_s$, $y_s$) and destination node $d$ ($x_d$, $y_d$). That is, only routers residing in REC($s$, $d$) will service as the intermediate routers for transmitting the packet. Path diversity deceases with $|x_d - x_s|$ (or $|y_d - y_s|$) approaching to 0. For example, in $8 \times 8$ network, a packet starts at node (3, 1) and destines to node (1, 1). In QM routing, the packet has only one option, while in RQM routing algorithm, it has as many as seven different paths. Paths provided by the two algorithms are shown in Table II. At router (3, 1), QM provides only one output channel ($X-$), while RQM provides three optional output channels ($X-$, $T+$ and $C-$). Note that, the paths provided by RQM may vary in length. It is well accepted that short paths have higher priority when the corresponding links are non-congested. In the next subsection, we will show the priorities for selecting output channels.

<table>
<thead>
<tr>
<th>Alg.</th>
<th>Diversity</th>
<th>Path(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>QM</td>
<td>1</td>
<td>(3, 1) → (2, 1) → (1, 1)</td>
</tr>
</tbody>
</table>
| RQM  | 7         | (3, 1) → (2, 1) → (1, 1)  
|      |           | (3, 1) → (2, 2) → (1, 1)  
|      |           | (3, 1) → (2, 0) → (1, 1)  
|      |           | (3, 1) → (2, 2) → (2, 1) → (1, 1) |
|      |           | (3, 1) → (2, 2) → (1, 2) → (1, 1) |
|      |           | (3, 1) → (2, 0) → (2, 1) → (1, 1) |
|      |           | (3, 1) → (2, 0) → (1, 0) → (1, 1) |

3.2 Priority of the output channels
The proposed RQM routing algorithm is adaptive, which provides multiple paths for packets. At each intermediate router, the selection component returns one free output channel from channel set ($cset$) according to the predefined rules. If none of the channels is free, the packet will wait until one channel is released.
If more than one channel is free, the free channels from \( cset \) constitute the available channel set (\( aset \)). As shown in Fig. 4, the minimal channels usually hold a higher priority. Here, channels in \( aset \) are divided into two sets: the minimal channels along the minimal path (\( MCS \)), and the quasi-minimal channels that immediately lead packets to travel on non-minimal routes (\( QMCS \)). The minimal channels are given higher priority since they could direct packets to destination more quickly. When \( MCS \) is empty, packets could bypass the congested area and proceed to its destination by use of channels from \( MCS \). It may happen that more than one packet compete for the same output channel, the round robin scheme is adopted in the allocator design for routers.

### 3.3 Deadlock and livelock freedom

The proposed algorithm allows packets to traverse the network along non-minimal paths, therefore, proof of the livelock freedom should be presented. Deadlock and livelock freedom are two fundamental requirements in design of routing schemes. Here in this subsection, we will show that the proposed algorithm can fulfill the rules.

**Lemma 1.** The RQM routing algorithm for DMesh is livelock free.

**Proof.** As known, livelock occurs when there exists packets indefinitely wandering in the network. Clearly, mis-routing is allowed in RQM routing algorithm, however, for any packet, it will ultimately reach its destination node in the finite time. Now we analyze the output channels provided by the RQM algorithm. Clearly, channel set provided by RQM routing (\( cset_{RQM} \)) is a super set of that provided by quasi-minimal \( cset_{QM} \). Here, we divide the channel set into two disjoint subsets:

\[
cset_1 = cset_{QM}; \quad cset_2 = cset_{RQM} - cset_{QM}.
\]

Take a packet at node \((x_c, y_c)\) for example, if the packet travel along channels belonging to \( cset_1 \), it goes one step further towards the destination, and will traverse no more than \(|x_d - x_c| + |y_d - y_c|\) hops in the network. If the packet travel along
channels belonging to \( cset_2 \), it must go a step further along dimension \( X \). Note that, when the packet arrives at the column of destination node, it will route along the minimal path. To sum up, packet routed by RQM algorithm will never wander permanently in the network. Since packet will always reach its destination router in some definite time, the RQM routing algorithm for DMesh is livelock free.

Deadlock is such a situation that occurs when a sequence of packets are waiting for one another to release resources, and hence are blocked indefinitely. In NoC, channel resources are shared and competed among different packets, and deadlock will never occur if there exists no cyclic channel dependency in the network [13, 14]. Since the proposed algorithm is based on NePA platform [11], the DMesh network is split into two acyclic subnetworks, and no cyclic channel dependency occurs as long as packets in different subnetworks use separate channels. Therefore, the proposed RQM routing algorithm for DMesh is deadlock free.

Though the proposed RQM is non-minimal, its property of finite hop counts makes it free from livelock. Due to the select function and its adaptivity, its performance advantage will exhibit more adequately, details about comparisons and discussions between RQM, DXY and quasi-minimal routing algorithms are described in the next section.

4 Evaluation and discussion

4.1 Experimental setup

Measurements are implemented in a cycle-accurate C++ network simulator. Simulations are performed on a 8 × 8 2D DMesh, and all network channels are 64 bits wide with a delay of one cycle. Each router is modeled input-queued with four pipeline stages including buffer write, routing computation, switch allocation, and switch traversal. Each input buffer has a total capacity of 4 flits. Wormhole flow control is applied throughout all simulations. Each packet contains 16 flits, with 1 head flit, 14 data flits, and 1 tail flit. In all simulations, 20,000 cycles are used to warm up the network, and another 80,000 cycles are used to capture the latency and throughput of the network.

4.2 Traffic pattern and evaluation metrics

Permutation traffic patterns have been used in the past to stress and evaluate routing algorithms [15]. We select bit-reversal, bit-complement and shuffle for comparisons. Details about the three permutation traffic patterns can be found in chapter three in [16]. Also, taking bit-reversal as the baseline traffic, hotspot traffic is given by taking \((n/2, n/2)\) and \((n-1-n/2, n-1-n/2)\) as the hotspot nodes. Given a hotspot percentage \( h \), a newly generated packet is directed to each hotspot node with an additional \( h \) percent probability. In this paper, we set \( h \) to 5%.

To verify the effectiveness of the proposed RQM, we evaluate its performance by comparing with DXY and quasi-minimal (abbreviated as QM) under different traffic patterns. Throughput (flits/node/cycle) and latency (cycles) are two important performance metrics for evaluating routing algorithms in OCINs. In the following subsections, packet injection rate (pir) is defined as the rate at which packets are injected into the network for each router, so it is measured in flits/node/cycle.
4.3 Results discussion

In the RQM algorithm, the select function gives high priority to the minimal output channels when there are more than one optional free channel. On the one hand, the select function could make efficient use of diagonal channels to cut down the traveling hop counts under low load rate. On the other hand, when load rate is high, the resource competition is fierce, so the flexibility provided by RQM could efficiently balance channel utilization among regular links and diagonal links. This balancing could directly be transferred to reduced average delay and enhanced throughput in the whole network.

Fig. 5 shows the latency-pir curves for the bit-reversal traffic pattern in $8 \times 8$ DMesh. Under this traffic, the RQM shows best performance, DXY gives the worst performance, and QM falls midway between the two. The saturation point of them are 0.41, 0.19 and 0.23 (flits/node/cycle), respectively. The performance improvement of RQM over DXY and QM are about 116% and 78% in the respect of saturation point. As for throughput, the peak throughput of RQM, QM and DXY are 0.409, 0.227 and 0.194. Both QM and RQM beat DXY perfectly by providing adaptivity and flexibility for packets. Besides, the proposed RQM offers better performance than QM because it makes more balance traffic between regular channels and diagonal channels as well as that it provides much more adaptivity by allowing packets to route outside the rectangular between source and destination.

Fig. 5. Performance comparisons of different algorithms under bit reversal traffic: (a) latency-pir curve, (b) throughput-pir curve.

Fig. 6. Performance comparisons of different algorithms under bit complement traffic: (a) latency-pir curve, (b) throughput-pir curve.
Fig. 6 shows performance comparisons under bit complement traffic. This type of traffic is used in several simulation scenarios. The RQM achieves the optimal performance with saturation point at 0.25, QM gives the suboptimal performance with saturation throughput 0.20, while DXY performs worst due to its deterministic property and congestion on the diagonal links. For a pir value of 0.19, the average latency of RQM, QM and DXY are 48.11, 61.15 and 93.78, latency reduction of RQM and QM over DXY are about 48% and 34%. Fig. 7 shows the latency-injection rate curves for each algorithm on the shuffle traffic pattern, which exhibits similar trend as in bit complement traffic.

Fig. 7. Performance comparisons of different algorithms under shuffle traffic: (a) latency-pir curve, (b) throughput-pir curve.

Fig. 8. Performance comparisons of different algorithms under hotspot traffic: (a) latency-pir curve, (b) throughput-pir curve.

Fig. 6 shows performance comparisons under bit complement traffic. This type of traffic is used in several simulation scenarios. The RQM achieves the optimal performance with saturation point at 0.25, QM gives the suboptimal performance with saturation throughput 0.20, while DXY performs worst due to its deterministic property and congestion on the diagonal links. For a pir value of 0.19, the average latency of RQM, QM and DXY are 48.11, 61.15 and 93.78, latency reduction of RQM and QM over DXY are about 48% and 34%. Fig. 7 shows the latency-injection rate curves for each algorithm on the shuffle traffic pattern, which exhibits similar trend as in bit complement traffic.

Hotspot is considered to be a more realistic model since processes usually communicate frequently with only some nodes (e.g., memory resources, I/O resources). If we consider the hotspot traffic scenario (Fig. 8), it is observed that a network adopting DXY performs poorly due to its determinism in distributing packets. Here the algorithm RQM offers the best performance because it efficiently balances traffic among the network. The algorithm QM falls midway between the other two. The adaptivity for packets will directly translate to better performance behavior in the area of congestion, that’s the reason why QM performs better than DXY.

The ideal case of channel utilization is each channel have equal fraction of total passing traffic. If some channels have more traffic than the others, they may become the bottleneck and directly restricts the performance of the entire system. However, for the randomness in packet generating process and limitation from routing
algorithm, the ideal case will not come true in the real NoC. All that we can do is providing more adaptivity when designing the routing algorithms. The adaptivity in RQM could beautifully distribute traffic more evenly than QM and DXY in the network, which could directly be transferred to higher performance. From the simulation results presented in this subsection, RQM brings improvements over both DXY and QM under different traffic patterns.

5 Conclusion

To exploit advances in technology, high radix routers are adopted in order to convert pin bandwidth to reduced latency in networks. This paper introduces the DMesh network, and analyzes its superior properties compared to Mesh. In order to exploit the routing adaptivity and flexibility, this work presents a novel random quasi-minimal routing algorithm, which allows packets to route outside the rectangular defined by source node and destination node. Moreover, we give a theoretical justification for RQM in the aspect of deadlock-freedom and livelock-freedom. Simulation results show that the proposed RQM can provide much more adaptivity, and brings improvement when compared to the DXY and quasi-minimal routing algorithms.

Acknowledgments

This research is now supported by National Natural Science Foundation of China (No 61402086).