



# LOFT: A low-overhead fault-tolerant routing scheme for 3D NoCs



Jun Zhou<sup>a,b,\*</sup>, Huawei Li<sup>a</sup>, Tiancheng Wang<sup>a,b</sup>, Xiaowei Li<sup>a</sup>

<sup>a</sup> State Key Laboratory of Computer Architecture, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, PR China

<sup>b</sup> University of Chinese Academy of Sciences, Beijing, PR China

## ARTICLE INFO

### Article history:

Received 28 October 2014

Received in revised form

2 August 2015

Accepted 25 August 2015

Available online 2 September 2015

### Keywords:

Networks-on-chip

3D Mesh

Permanent fault

Fault-tolerance

Routing scheme

## ABSTRACT

As one of the main trends of communication technology for 3D integrated circuits, the 3D networks-on-chip (NoCs) have drawn high concern from the academia. The links are main components of the NoCs. For the permanent link faults, the fault-tolerant routing scheme has been regarded as an effective mechanism to ensure the performance of the 2D NoCs. In this paper, we propose a low-overhead fault-tolerant routing scheme called LOFT for 3D Mesh NoCs without requiring any virtual channels (VCs). LOFT is a deadlock-free scheme by adopting a logic-based routing named LBDR<sup>2</sup> guided by a turn model Complete-OE. The experimental results show that LOFT possesses better performance, improved reliability and lower overhead compared with the state-of-the-art reliable routing schemes.

© 2015 Elsevier B.V. All rights reserved.

## 1. Introduction

3D Integration Technology [1] is developed based on the traditional 2D integrated circuits (ICs). This technology piles layers of wafer and die together vertically, which can shorten the length of lines in the chip to reduce the systemic communication latency and overhead greatly. The network-on-chip (NoC) is a mainstream and effective communication technique for 3D ICs. Fig. 1 shows a typical  $4 \times 2 \times 3$  3D Mesh NoC. There are 3 layers, 24 processing elements (PEs) in this network. Each PE communicates with its own router/switch node (“node” for short below) through the network interfaces (NIs), and different nodes are connected to adjacent ones with the horizontal or vertical links.

The faults in the bidirectional links between different nodes of 3D Mesh NoCs are concerned in this paper. Generally, the faults include temporary faults and permanent faults. In this paper, because the latter have greater influence on the performance of the networks [2], we mainly focus on the permanent faults, which are not repairable.

Faults of links in the 2D NoCs have been explored much so far, but few were investigated in the 3D scenario. In general, there are four major fault-tolerant schemes for the permanent link faults in NoCs:

1. The first scheme is using the redundancy technology [2]. This scheme needs to add redundant links to replace the invalid ones to achieve fault-tolerance, incurring large overhead on area and power consumption, which will be more significant with the increase of the chip scale.
2. The second scheme is controlling the data to bypass the faulty links with the peripheral logical circuits of them, to route the packets through the valid links [3]. The alteration to the physical structure of the circuits will also result in a high systemic cost.
3. The third scheme is adopting some techniques, such as time division multiplexer (TDM) [4], to be implemented for the fine-grained micro-architecture of links in NoCs. Note that, even if the faults occur, the single links are usually not damaged completely. This scheme is effective to save the strictly limited resources to fulfill the tasks of the NoCs. However, this scheme will also induce the increment of the overall overhead for the added control logics.
4. The fourth scheme is implementing the fault-tolerant routing scheme [5–11]. This scheme is a low-cost and lightweight, and especially applicable for the resource-restricted chip multiprocessors (CMPs).

On the other hand, the turn models and virtual channels (VCs) [12] are the two main techniques to guarantee the deadlock-freeness of the networks. This is because deadlock is prone to be caused by the abstract dependency cycles of communications (“abstract cycle” for short below) formed with one or several different data streams under the wormhole flow control in the channel dependency graphs (CDGs) [12,13]. However, the usage of VCs will induce extra buffer area and complex control logics.

\* Corresponding author at: State Key Laboratory of Computer Architecture, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, PR China.

E-mail addresses: [zhoujun@ict.ac.cn](mailto:zhoujun@ict.ac.cn) (J. Zhou), [lihuawei@ict.ac.cn](mailto:lihuawei@ict.ac.cn) (H. Li), [wangtiancheng@ict.ac.cn](mailto:wangtiancheng@ict.ac.cn) (T. Wang), [lxw@ict.ac.cn](mailto:lxw@ict.ac.cn) (X. Li).

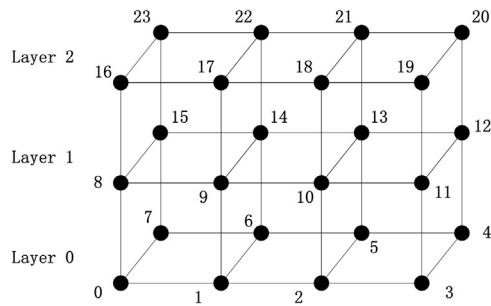


Fig. 1. A typical  $4 \times 2 \times 3$  3D Mesh NoC.

Therefore, the turn model, which can guide the implementation of the (fault-tolerant) routing scheme, is especially suitable for the overhead-sensitive circuits to avoid the deadlock.

In this paper, we propose a low-overhead fault-tolerant routing scheme named LOFT for 3D Mesh NoCs. Our main contributions are as follows:

1. We present a flexible logic-based 3D routing algorithm extended from a low-cost 2D routing with the consideration of the fault-tolerance, 3D scenario and other factors (see Section 3.2 in detail) to outperform the common table-based ones in terms of the communication latency and area overhead.
2. A new 3D turn model for guiding the routing is proposed to achieve the deadlock-freeness of the networks. The new model is an extension from a classical 2D turn model and its variants to obtain good turn fairness and traffic flow balance in 3D Mesh NoCs.

LOFT supports both single-fault and multi-fault models without the distinction between the horizontal and vertical faulty links. Compared with the other state-of-the-art schemes, LOFT has better performance, improved reliability and lower overhead as indicated through the results of the simulation and synthesis.

The rest of this paper is organized as follows: Section 2 introduces the related work on previous fault-tolerant routing schemes designed for the faulty links in 3D NoCs. The principle of LOFT is presented in Section 3. Section 4 gives the proof of the livelock-freeness and deadlock-freeness of our scheme. The experimental results are compared and analyzed between LOFT and other relevant schemes in Section 5. Finally, Section 6 gives the conclusion and the future work.

## 2. Related work

The content of related work is divided into two sections. The first section is the introduction of the previous deadlock-free fault-tolerant routing schemes. The second section is the background techniques of our scheme, including the way to implement the routing algorithm and the existing 3D turn models for guiding the routing.

### 2.1. Previous fault-tolerant routing schemes

It is more complicated for the implementation of a deadlock-free fault-tolerant routing scheme in 3D NoCs than in 2D NoCs. The routing rules on the single  $X$ - $Y$  plane need to be extended accordingly to satisfy the 3D scenario. In recent years, several researches started concentrating on the routing schemes for tolerating link faults in 3D NoCs.

In [5], the  $k$ -shortest path algorithm is used to find the replacement paths to improve the fault-tolerance rates of the NoCs in

presence of the invalid vertical links. However, no optimal routing path is chosen in the selectable paths to be applied to route the packets to the destination nodes. For another work, the authors just tried to get a nearest mid-switch with a healthy vertical link at the same horizontal layer to bypass the faulty horizontal and vertical links in [6]. But there are no special constraints on the routing in the horizontal layers, which will probably lead to the congestion and deadlock in 3D NoCs.

In [7], the authors proposed 4NP-First, a fault-tolerant routing scheme for 3D NoCs. This method includes two turn models, 4N-First and 4P-First, which are designed for reverse directions. Only 4N-First will be applied when the fault rate is low. Yet, there is still a requirement of two VCs in each FIFO of routers for the two turn models if the fault rate is high. Another fault-tolerant routing scheme called AFRA was proposed in [8]. AFRA also focused on the faults in the vertical communications. ZXY and XZXY routings, two dimension-order algorithms are used as the routing mechanism. If there is at least one invalid vertical link existing in the straight path from the source node to the destination layer, XZXY will be executed in place of ZXY. However, AFRA needs at least two VCs to ensure the deadlock-freeness when there are more than two faulty vertical links in different directions, or else it only support the single-fault model without VCs in the networks.

Hamiltonian path is introduced in [9] to put forward HamFA. The path is the only way where packets can be sent. The IDs of nodes are sorted in an ascending or descending order in Hamiltonian path. Therefore, HamFA is deadlock-free since no abstract cycle will be formed. However, faults at some special positions are not allowed in HamFA. Once these faults occur, the whole network will not work normally. Associated with a deadlock-recovery technique called random access buffer (RAB), HLAFT was presented in [10]. As a minimal routing, HLAFT is implemented depending on the fault status of the local and neighboring nodes when there is at least one minimal path leading to the destination node. Meanwhile, HLAFT will choose the non-minimal path to continue the fault-tolerant routing if all the selectable output ports are unavailable. In addition, RAB is applied to detect the possible deadlocks of the NoCs and break the dependency of the abstract cycles in the CDGs. In [11], HARS was designed to support both single-fault and multi-fault models with no usage of VCs. A mid-node-searching method is applied by obeying the specific turn model to route the packet to pass around the damaged vertical links from the source layer to the destination layer progressively. However, it is possible that no suitable mid-node can be found in the current layer when the fault rates are extremely high.

From the related work above, we notice that: (1) Lacking detailed operational descriptions, the scheme of neither [5] nor [6] is operable for specific communication tasks of 3D NoCs; (2) Some of the researches above just focus on the faults in the vertical links, like [5,8] and [11], whose fault model is not strong enough; (3) Expensive VCs or similar techniques are needed in the schemes of [6–8] and [10], which is unfavorable for lower systemic overhead; (4) The fairness and traffic flow balance are relative low for the adopted turn constraints in [7–9]. Though the turn model applied in HARS [10] obtained a better effect, it still can be further improved in the turn fairness and traffic flow balance.

Applying no different treatment between the horizontal and vertical faulty links, in this paper, the proposed method LOFT is guided by a new 3D turn model and assisted by an efficient selection strategy to constitute the complete LOFT to get a high-performance, reliable and low-overhead fault-tolerant routing scheme for 3D Mesh NoCs.

Download English Version:

<https://daneshyari.com/en/article/542679>

Download Persian Version:

<https://daneshyari.com/article/542679>

[Daneshyari.com](https://daneshyari.com)