Load-balanced routing algorithm for Network-on-Chip employing fair arbiter

Lu Liu1, Yanfei Yang1, Qianqian Lei1, Huhu Wang1 and Song Lixun1*

1 Department of Microelectronics Science and Engineering, Xi’an Polytechnic University, Xi’an, Shaanxi Province, 710048, China

*Corresponding author’s e-mail: songlixun@xpu.edu.cn

Abstract. Balanced load is one of the most important performance metrics for on-chip communication. To design flexible, adaptive load-balanced routing algorithms, an arbiter that alternated the priority order fairly is employed. With fair arbiter to decide the routing path for network-on-chip, balanced load can result in performance improvement in speed and capacity for both congestion-oblivious and congestion-aware routing schemes.

1. Introduction
Flexible routing algorithms are demanded to improve Network-on-Chip (NoC) performance, high reliable NoC requires balanced network load[1]. Routing algorithms that can adjust communication path according to network load are employed, to obtain balanced load, fair arbiter is applied in stages of both congestion-oblivious and congestion-aware routing algorithms[2]. Congestion-oblivious (CO) routing algorithms decide transmission path according to network information ignoring the load on each path, while congestion-aware (CA) routing algorithms select proper path based on different load conditions[3]. To guarantee balanced load distribution, routing algorithms employ fair arbitration to select transmission path[4].

In this paper, a fair arbiter that alternates priority is introduced and applied in both congestion-oblivious and congestion-aware routing to improve the adaptiveness and performance of NoC routing algorithms.

2. Fair arbitration
Based on the priority order of on-chip arbiter, an arbiter with alternate priority is proved to achieve fairness, and can be applied in adaptive routing algorithm with balanced load.

2.1. On-chip arbiter
Conventional on-chip arbiter is applied in the field of System-on-Chip (SoC), as a solution to conflicts among on-chip processors. As shown in Figure 1, SoC arbiters are applied to multiple processors named “master” applying the same element “slave”, the granted “master” achieve the opportunity to use the contended “slave” to complete SoC communication. The quality of SoC arbiter can decide the performance of SoC communication, fairness is one of the performance metrics to describe the quality of arbiter[5].
As a solution to larger scale integrated circuits, the parallel Network-on-Chip (NoC) with the thought of computer network is applied[6]. The router architecture is shown in Figure 2, with data in forms of flits, on-chip communication can be complete on the network though the crossbar switch inside the router, transmission direction is determined by the switch allocation (SA)[7]. The NoC arbiter is applied to select a proper routing path for delivered flits.

Routing algorithm is the scheme to decide the delivered path on the network, as depicted in Figure 3. for routing scheme with certain flexibility, such as odd-even (OE) turn model, NoC arbiter plays the role of path selection. With NoC arbiter applied, OE routing obtain more candidate paths than fixed XY routing. To balance network load, the fairness of NoC arbiter is the most important performance of on-chip arbiter design[8].

### 2.2. Fair alternate priority arbiter (APA)

The granting mechanism of alterable priority arbiter (APA) is designed according to Table 1. The priority order of APA is alternated according to the increased pointer value, arbiter granting among 4
requests is decided under evenly varied priority order, the alternated method can guarantee the fairness of granting. APA is proposed as a fair arbiter for NoC path selection.

| Status Pointer | Priority order | Granting mechanism |
|---------------|----------------|--------------------|
| 3             | 4n             | g₁ = r₃; g₂ = r₂; g₃ = r₁; g₀ = r₀ |
| 2             | 4n+1           | g₀ = r₁; g₁ = r₀; g₂ = r₃ |
| 1             | 4n+2           | g₀ = r₀; g₁ = r₁; g₂ = r₃; g₃ = r₂ |
| 0             | 4n+3           | g₀ = r₀; g₁ = r₁; g₂ = r₂; g₃ = r₃ |

The fairness of APA and round-robin(RR) arbiter can be compared in Figure 4. The fairness of each arbiter is measured by the standard deviation among granted requests. As shown in Figure 4, the APA obtain lower standard deviation among all granted requests, which indicating a higher fairness than RR arbiter of the equally granted APA. Thus, APA can be a solution to routing algorithm decision for load balanced communication.

3. Load balanced routing algorithm

The fairness of APA can be applied as routing decision at the stages of both Congestion-oblivious(CO) and congestion-aware(CA) routing algorithms.

3.1. APA applied in CO routing

The flowchart of APA applied in CO routing algorithm(OE-FA) is depicted in Figure 5. APA grants the requests standing for directions d₁, d₂ permitted by OE turn model simultaneous, the direction granted by APA is applied as the output for flit. According to the fairness of APA, OE-FA can select proper routing path to improved the adaptiveness of OE model, balanced load can be achieved with higher routing adaptiveness.
3.2. APA applied in CA routing

Different from CO routing ignoring network load status, the flowchart of APA applied in CA routing algorithm (CAOE-FA) is depicted in Figure 6. For the directions $d_1$, $d_2$ permitted by OE turn model, CAOE-FA compare their load and select the light loaded one as output. Nevertheless, when directions $d_1$, $d_2$ share the same congestion index (CI) level, APA is applied to grant the output fairly and produce a load balance CAOE-FA routing algorithm, instead of random selection.

**Figure 5. OE-FA flowchart**

**Figure 6. CAOE-FA flow graph**
4. Result analysis

4.1. Load balanced CO routing

The performance matrix of average packet latency indicating speed and saturation throughput indicating capacity of OE-FA are shown in Figure 7 and Figure 8. Under ordinary traffic patterns, with the introducing of fair APA, OE-FA can improve the routing performance with reduced latency and improved throughput compared to OE.

Figure 7(a). Random latency
Figure 7(b). Transpose1 latency
Figure 7(c). Transpose2 latency
Figure 7(d). Bit reversal latency

Figure 8. Throughput
4.2. Load balanced CA routing

The performance matrix of average packet latency indicating speed and saturation throughput indicating capacity of CAOE-FA are shown in Figure 9 and Figure 10. Under hotspot traffic patterns, with the introducing of fair APA, CAOE-FA can improve the routing performance with reduced latency and improved throughput compared to OE and randomly selected OE-random.

5. Conclusions
Routing algorithm is the technique that determines the communication path of on-chip flits. Since load balancing on routing path is a essential performance of NoC routing, the fairness of path selection plays an important role. In this paper, a fair arbiter that alternates priority orders evenly is proved to grant all directions equally, which can be applied at the stages of congestion-oblivious and congestion-
aware routing algorithms. Results show that the fair APA as path selection can improve the performance metrics in both speed and capacity, since fair arbiter can improve the routing property of load balance.

Acknowledgments
Financial support from: PhD’s research start-up fund BS201937.

References
[1] John N, James H, and Ehud F, et al. Digital atomic scale fabrication an inverse Moore’s Law - A path to atomically precise manufacturing[J]. Micro and Nano Engineering, 2018(1): 1-14.
[2] Kliem Daniel, Voigt Sven-Ole. Scalability evaluation of an FPGA-based multi-core architecture with hardware-enforced domain partitioning[J]. Microprocessors and Microsystems, 2014, 38(8): 845-859.
[3] V Gourisetty, H Mahmoodi, and V Melikyan, et al. Low power design flow based on Unified Power Format and Synopsys tool chain[C]. 3rd interdisciplinary Engineering Design Education Conference, 2013:28-31.
[4] Ezz-Eldin R, El-Moursy MA, and Hamed H F A. Process Variation Delay and Congestion Aware Routing Algorithm for Asynchronous NoC Design[J]. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 2016, 24(3):909-919.
[5] Bahrebar P, Stroobandt D. The Hamiltonian-based odd–even turn model for maximally adaptive routing in 2D mesh networks-on-chip[J]. Computers & Electrical Engineering, 2015, 45(7): 386-401.
[6] Lu Liu, Yadong Sun, Zhangming Zhu, Yintang Yang. A Congestion-aware OE Router Employing Fair Arbitration for Network-on-Chip. Journal of Semiconductor, 2019, 39(12): 1-7.
[7] Bachanna Prashant, Jalad Vivek, and Shetka Sharanbasappa. Design and analysis of high speed low power reusable on chip bus based on wishbone[C]. 5th International Conference on Signal and Image Processing (ICSIP). Bangalore:IEEE, 2014:197-200.
[8] Lu Liu, Zhangming Zhu, and Duan Zhou, et al. A fair arbitration for Network-on-Chip routing with odd-even turn model[J]. Microelectronics Journal, 64(2017):1-8.