Spatial-Temporal Enhancement of ACO-Based Selection Schemes for Adaptive Routing in Network-on-Chip Systems

Hsien-Kai Hsin, En-Jui Chang, and An-Yeu (Andy) Wu, *Senior Member, IEEE*

Abstract—Networks-on-Chip (NoC) provides a regular and scalable design architecture for chip multi-processor (CMP) systems. The Ant Colony Optimization (ACO) is a distributed algorithm. Applying ACO to selection models of adaptive routing can improve NoC performance. Currently, ACO-based selection only uses the historical traffic information. While additional temporal and spatial information provides better approximation of network status for global load-balancing. In this paper, we first consider the temporal enhancement of congestion information. We propose the Multi-Pheromone ACO-based (MP-ACO) selection scheme which adopts the concept of *Exponential Moving Average (EMA)* from stock market. We implement a novel ACO system where ants lay two kinds of pheromones with different evaporation rates. The *temporal pheromone variation* can help to capture hidden-state dependencies of upcoming congestion status. Secondly, to acquire the spatial range of congestion information, we propose Regional-Aware ACO-based (RA-ACO) selection to record historical buffer information from routers within two-hop of distances, which helps to extend *spatial pheromone coverage*. Information provided by the proposed two schemes improves the system performance. Simulation results show that MP-ACO and RA-ACO with Odd-Even routing algorithm yields an improvement in saturation throughput over OBL and NoP selection by 14.38 percent and 18.64 percent, respectively. The router architectures for the proposed schemes are also implemented and analyze with small hardware overhead.

Index Terms—Ant Colony Optimization (ACO), Networks-on-Chip (NoC), adaptive routing, selection strategy

1 Introduction

Networks-on-chip (NoC) provides a scalable architecture for designing parallel chip multi-processor (CMP) systems [1], [2], [3]. However, traffic distribution dominates the performance of the NoC. When number of cores scales up, more complex applications cause the on-chip traffic load to become unpredictable and unbalanced. Consequently, congestions in channels increase packet delay and result in serious performance degradation. Therefore, NoC requires an effective adaptive routing algorithm for load-balancing and improving network throughput [4], [5].

To solve this issue, the Ant Colony Optimization (ACO)-based selection function is developed to assist adaptive routing in judging the congestion status of channel with current and historical pheromone information [19], [20], [21], [22], [23]. ACO is a distributed collective-intelligence algorithm, as shown in Fig. 1. In initial state, the ants travel through paths and diffuse pheromone chemical. Through this process, better paths accumulate a higher concentration of pheromone. The selection can reach the optimum without global information. Although ACO can couple with different local selection models and improves the routing efficiency [17], [18], [19], the current ACO-based selection utilizes only historical information in run-time applications and NoC architectures.

In order to provide more precise approximation of the network, we observe pheromone variation and increase the pheromone coverage. We propose two schemes to improve the temporal and spatial awareness of ACO algorithms based on the desired features. For temporal enhancement, based on the design concept in [23], we have rigidly formulated the problem, refined the algorithm, and presented the area/energy analysis in details. Furthermore, we propose the novel spatial enhancement in this work. The proposed temporal and spatial enhancement are as follows:

1. **Temporal enhancement—Multiple-Pheromone ACO-based (MP-ACO) selection**: In Fig. 2a, ant system with single pheromone can determine whether a channel is better than another. However, a traffic burst still reduces the system throughput. If pheromone is designed as in Fig. 2b, multiple pheromones can divide the system into more states and provide extra information to reduce traffic bursts.
   We design an MP-ACO selection scheme adopting the *Exponential Moving Average* (EMA) [33], and implement a system with two pheromones in each channel. The ants lay pheromones with different evaporation rates to capture the buffer state dependency and to be aware of upcoming congestion. With MP-ACO selection, we can better determine the channel congestion states.

• *The authors are with the Graduate Institute of Electronics Engineering, National Taiwan University, Taipei 106, Taiwan. E-mail: {ckcraig, enjui}@access.ee.ntu.edu.tw; andywu@ntu.edu.tw.*

Manuscript received 10 May 2013; revised 14 Nov. 2013; accepted 18 Nov. 2013. Date of publication 4 Dec. 2013; date of current version 16 May 2014. Recommended for acceptance by M.E. Acacio.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TPDS.2013.299

2. **Spatial enhancement—Regional-Aware ACO-based (RA-ACO) selection**: ACO-based selection synthesizes spatial and historical network congestion information. With the accumulation of local model information (i.e., output buffer length (OBL) [24]) in Fig. 3a, the pheromone represents the historical buffer status.
   Many works enhance only on spatial (local model) [11], [12], [13], [14], [15], [16] or temporal awareness (historical pheromone) [21], [22], [23]. Unfortunately, these enhancements may have a marginal effect. To further enhance system performance, we integrate the precise regional-aware congestion model Neighbor-on-Path (NoP) [15] with the ACO historical model, as shown in Fig. 3b. The proposed RA-ACO selection acquires information, including free slots, reservation status, and admissible routes from routers within two-hop of distances. This spatial extension increases the pheromone coverage of the historical congestion status.

The simulation results show that applying MP-ACO/RA-ACO to Odd-Even routing algorithm yields improvements in saturation throughput by 14.38 percent and 18.64 percent, respectively, compared with OBL and NoP selection. We implemented the router architectures of proposed schemes, synthesized them with *TSMC* 90 nm technology. The results showed area efficiency improvements to 1.355 and 1.258, respectively, compared with OBL and NoP selection. Similar results to those presented in this manuscript for *TSMC* 90 nm technology are expected to be obtained with more advanced fabrication technologies.

The rest of this paper is organized as follows: In Section 3, we summarize major contributions regarding the development of routing and ACO-based selection in NoC. In Sections 3 and 4, we introduce the proposed schemes. In Section 5, we present the performance evaluation. In Section 6, we design the router architecture and analyze area overhead, area efficiency, and energy consumption. Finally, Section 7 gives the conclusion of this paper.

2 Review of ACO-Based Selection for Adaptive Routing in NoCs

2.1 Review of Adaptive Routing in NoCs

We summary the key features of adaptive routing in supplemental material I, which is available in the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TPDS.2013.299. As a selection strategy, the OBL selection scheme selects the channel with more buffer occupancy to prevent congestion and reduces the average system latency [24]. The local model was later utilized by Ascia *et al.* to provide NoP selection that offers the buffer free slots and reservations of two-hop-neighbour routers [15]. Performance greatly improved compared to OBL, but the overhead of this method is that it needs extra wiring to transmit the NoP data. In this paper, we use both OBL and NoP for our fundamental network information.

Although many works have explored the routing problem based on different optimization targets and expanding the acquired information to achieve higher approximation of network status, few works discussed how historical information would assist the selection strategy in detecting congestion for an NoC system.

Recently, the propagation-based mechanism has been widely discussed and analyzed. This mechanism acquires network information by propagating through distant hops. However, recent work [16] has shown that the aggregated buffer utilization information exhibits the status of path congestion. By adopting more precise evaluation metrics, [31], [32] further improve the accuracy of this model. GCA [34] proposed a light-weight adaptive routing algorithm by propagating precise congestion information through a separate side-band network. CARS [35] can effectively mitigate the congestion by prioritizing requests based on

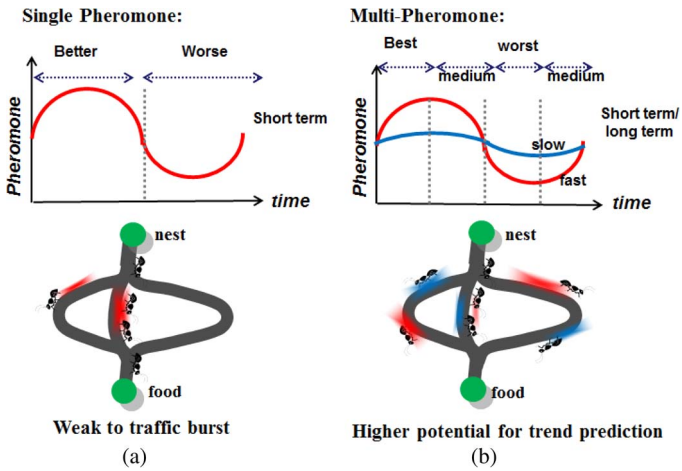

Fig. 2. With multiple pheromone, we can acquire more detailed network information.

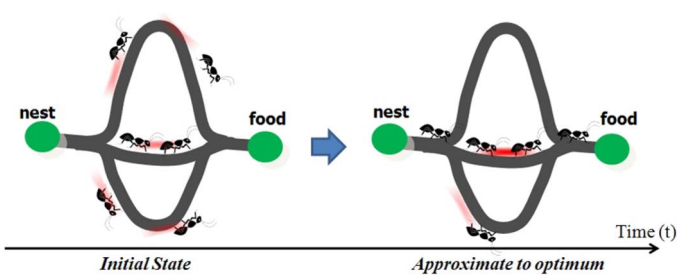

Fig. 1. ACO process helps to explore new paths and find the shortest path with pheromone information.

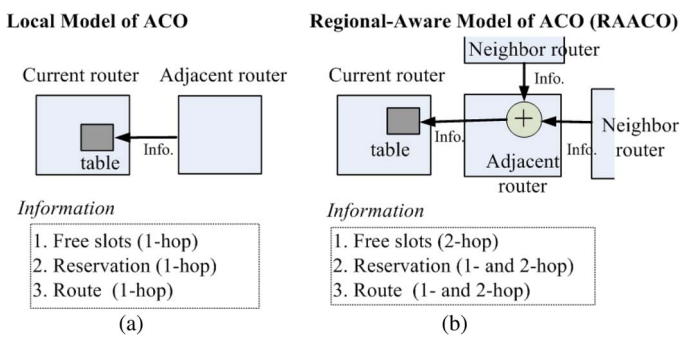

Fig. 3. Information acquiring of (a) ACO with OBL model and (b) RA-ACO with NoP regional-aware model.

TABLE 1
Attributes of Different Routing and Selection Schemes

Techniques	Historical Information	Temporal Variation	Spatial Extension
Buffer Occupancy [15][16][24][31]			√
Hotspot Min. [21]	√		
RACO [22]	√		
Proposed MPACO	√	√	
Proposed RAACO	√		√

* √ denotes the feature that the technique mainly focus on
* Temporal variation denotes the observation of the variation of historical information
* Spatial extension denotes the acquiring and applying process of spatial information

the congestion level information. DeBAR [36] proposed a series of design optimization on minimally buffered NoC routers. Future work can be directed towards the design of adaptive routing on NoC systems where spatial and temporal extension can be applied to the propagation-based acknowledge mechanism to improve performance.

2.2 Concept of Ant Colony Optimization [17]

ACO algorithms can assist the selection strategy on utilizing historical information [17], [18]. The interaction between distributed ants and pheromones are transformed into elements and mechanisms to explore the system environment. Numerous approaches that bring about ant colony characteristics have been developed to support the system design [19], [20]. We shows the basic ACO-based adaptive routing in supplemental material VII available online.

2.3 Regional ACO-Based Selection [22]

In [22], Regional ACO-based (RACO) selection was proposed for reducing ACO table size, making ACO selection feasible for NoC implementation thanks to reduced memory cost, table access time, and power consumption, while maintaining similar performance.

The region setting of RACO-4 is the case of $p = 1$ in 1), with region set $\{NE, NW, SE, SW\}$. N_E is the number of entries with region setting p, which is an integer. The destination nodes are assigned into region sets described by the direction relative to the source node. Since Odd-Even is a minimal routing, we need no more than eight pheromone spaces in the routing table. In 2), the destination index d is replaced by region index R. By taking advantage of the regional characteristic of the NoC, RACO can be feasibly implemented on NoC systems

$$N_E = \begin{cases} 1, & \text{if } p = 0 \\ 4p, & \text{if } p \in N \end{cases} \quad (1)$$

$$P'(j,R) = \frac{P(j,R) + \alpha L_j}{1 + \alpha(N_k - 1)}, \quad R = \{1, 2, \ldots, N_E\}. \quad (2)$$

The routing process of RACO is as follows: First, we set the region relation for the router and table index. At run-time, the Odd-Even routing function provides a list of the admissible channels. The table is updated if the packet is an ant packet. Finally, the output channel is selected based on the pheromone table.

With a smaller table size, the area cost is greatly reduced. Therefore, we can explore the cooperation on spatial and temporal network information in more depth. We describe two methods for exploring network information and achieving load-balancing in the NoC.

2.4 Attributes of Previous Works and This Work

Table 1 shows the attributes of routing and selection schemes. Buffer occupancy schemes mainly focus on spatial extension, in which queue congestion information is acquired, and they may also discuss on how spatial information is applied to problems. However, few prior works have utilized historical information on NoC systems. Works [21], [22] have applied the ACO-based routing on NoC, and the algorithm changes greatly based on the NoC characteristics. This paper continues the NoC-aware ACO research and provides temporal variation and spatial extensions for improving system performance.

The spatial-temporal enhancement attributes in this work are as follows:

- Firstly, we define *temporal information* as the temporal variation of the congestion metrics information. Thus, temporal information requires both the current information and stored historical information. By observing this information, we can predict upcoming congestion.
- Secondly, *spatial information* is defined as the congestion metrics that are acquired from two hops away in router distance. In general, because the coverage of spatial information is higher compared to local information, the spatial information is more approximate to the spatial traffic distribution. Therefore, the selection function can make a better choice by including this spatial information.

3 Temporal Enhancement: Multiple-Pheromone ACO (MP-ACO) for Sensing Pheromone Variation

For a selection scheme, it is crucial to identify the router congestion. According to Ascia *et al.* [15], selecting a congested router can cause the forwarding packet to wait for all the flits of blocking packets to pass and is thus undesirable. This situation also causes long delay.

Unfortunately, the already high degree queuing delay becomes more severe due to the *self-similar* and *bursty* traffic characteristics in the NoC system, which has been discussed in numerous studies [2], [5], [6], [7]. Self-similarity is a statistical phenomenon in which the current traffic is dependent upon previous traffic, while bursty traffic denotes a sudden increase in temporal traffic. If the burst is not detected in the early stages, it starts to form a congestion tree, and blocks subsequent packets. This problem is referred to as *temporal congestion problem*.

We need a dynamic method to observe and resolve this problem. Given two available channels, one is currently congested but will be congestion-free in the near future, while the other is currently non-congested but will become congested. In an indecisive situation, we could choose for

Fig. 4. Stock chart of the 20-day and 50-day moving average (MA) [27].

our immediate benefit, but lose our long-term target at the same time. If we can be aware of the trend of the channels, we can better prevent temporal congestion.

3.1 Inspiration from Stock Market

In the stock market, technical analysis methods are used for forecasting the direction of prices through the study of past market data, primarily price and volume.

The general form of moving average is called the *Simple Moving Average* (SMA). In (3), N is data points and P_t refers to the price on data point t. The period N is selected based on the kind of movement we are concentrating on, such as short, intermediate, or long term. Fig. 4 shows a stock chart with a 20/50-day moving average (denoted as MA20 and MA50, respectively) [27]. System statuses can be represented by the interaction of the moving averages.

$$S_{(t)} = \sum_{k=1}^{N} \frac{P_t}{N}. \tag{3}$$

Although this analysis may deviate from real-world behaviour due to the complexity of financial markets, for on-chip communication, the trend of traffic flow is more predictable due to the regular structure of the NoC.

3.2 Acceleration and Exponential Moving Average (EMA) of Pheromone Information

Since the historical pheromone of ACO performs a weighted average of the historical and the current information, the pheromone can be considered as the *speed* vector. For a conceptual analogy, we can consider the selection schemes with different local models that contain the current status of adjacent routers. The current status corresponds to the *displacement* in physics. The displacement value is an estimation of congestion. Taking the free slots in the buffer as an example, the more free slots a buffer has, the larger its displacement.

For channels with similar pheromone values, we want to determine which channel is prone to congestion. Therefore, we want to know the rate of change of the pheromone, corresponding to the *acceleration*. Assuming this acceleration represents the trend of channels, we give our hypothesis of the system: According to the temporal congestion problem mentioned earlier, self-similar traffic causes the **temporal locality** phenomenon and the induced congestion can result in a decrease of pheromone. With this hypothesis, the trend of pheromones, those are laid and evaporate at different speeds, are observed.

Since the SMA can lag to an undesirable extent and can be disproportionately influenced by old data points, this issue can be solved by giving extra weight to more recent data points. We can achieve this by adopting an iterative method called *Exponential Moving Average* (EMA) in (4). Where α is a weighting between current displacement $D(t)$ and the previous speed $S(\mathrm{t}-1)$. With the EMA, we can estimate the velocity without recording each historical displacement.

$$S_{(t)} = \alpha D_{(t)} + (1-\alpha) S_{(t-1)}. \tag{4}$$

To adopt the concept of EMA, we first rewrite the state transition function (2) in the form of (5). By setting $\hat{\alpha}$ and $\hat{L}_j$ in (6) and substituting them into (5), (5) can be simplified as (7):

$$\begin{aligned} P'(j,R) &= \frac{P(j,R) + \alpha L_j}{1+\alpha(N_k-1)} \\ &= \frac{P(j,R)}{1+\alpha(N_k-1)} + \frac{\alpha(N_k-1)}{1+\alpha(N_k-1)} \frac{L_j}{1+\alpha(N_k-1)} \end{aligned} \tag{5}$$

$$\hat{\alpha} = \frac{\alpha(N_k-1)}{1+\alpha(N_k-1)}, \quad \hat{L}_j = \frac{L_j}{1+\alpha(N_k-1)} \tag{6}$$

$$P'(j,R) = \hat{\alpha}\hat{L}_j + (1-\hat{\alpha}) P(j,R). \tag{7}$$

The pheromone updating mechanism in (7) is identical to the EMA in (4). That is, the ACO-based selection evaluates the historical buffer status of each channel. We can expand (7) to:

$$P_t = \alpha L_t + \alpha(1-\alpha) L_{t-1} + \alpha(1-\alpha)^2 L_{t-2} + \cdots. \tag{8}$$

The window function EMA provides an average of the free buffer slots, and we consider the pheromone as a weighted speed. A higher pheromone value (higher speed) implies that the output channel recently has more free slots.

In our design, we use two pheromones for evaluating the congestion status of each channel. We implement this scheme with two registers of different bit lengths: The fast pheromone (P_{fast}) is designed to have a larger $\hat{\alpha}$ weighing in (7) than the slow pheromone (P_{slow}), thus P_{fast} moves faster than P_{slow}. Therefore, P_{fast} represents the short-term and P_{slow} represents the long-term buffer occupancy [21].

For further evaluating the channels, we define the status of *acceleration*. First, we want to know whether the short-term status is better or worse than the long-term status. Second, we want to know if the short-term status has been increasing or decreasing. In Fig. 5 we show the conceptual diagram of our proposed methodology. The x-axis is the time t, and the y-axis is the pheromone. We define the four network state as *Advance*, *Reaction*, *Decline*, and *Rally* according to the relative position of P_{fast} and P_{slow}, and the slope of P_{fast}. For example, the *Advance* status denotes that P_{fast} is greater than P_{slow} and has a positive slope.

Using a heuristic approach, we prefer a channel that tends to be less congested therefore we assign the selection priority as $Advance > Reaction = Rally > Decline$. Note that, $Reaction$ has a higher pheromone value (implying more vacant buffer space), while having a negative slope (implying increasing traffic). $Rally$ has a lower pheromone value (implying of less vacant buffer space) and a positive slope (implying of decreasing traffic). The slope compensates the effect of the current pheromone status.

In the selection process, since the long-term pheromone P_{slow} represents the long-term buffer occupancy [21]. When the P_{slow} of each channel are with much different values, the P_{slow} shall be taken as the selection index to prevent channel congestion. When the channels are with similar long-term pheromone levels, the *indecision* problem occurs [15]. The proposed acceleration can then be used to help in solving the indecision condition. When more than one channel has the same status of acceleration, we then select the output channel according to the short-term pheromone P_{fast}. Detailed operation and pseudo code of the proposed MP-ACO Algorithm is shown in supplemental material V available online.

4 Spatial Enhancement: Regional-Aware ACO (RA-ACO) for Extending Pheromone Coverage

We have discussed the temporal enhancement. In another aspect, we want to introduce the spatial enhancement. In Fig. 6, the selection function uses network state information to judge the channel congestion status. The network information is acquired with a local model, which is an estimation of the network status. A more appropriate local model can provide more precise knowledge of the network status and enhance the routing efficiency.

For example, when the forwarding packets are blocked, they have to wait for all the flits of blocking packets to pass. Moreover, these packets may continue to block more packets and cause a *congestion tree* condition. If this condition is not solved properly, the latency caused from the congestion tree can be a serious problem. To solve this problem, we have to sense the congestion much earlier in the spatial domain. Therefore, we consider this kind of problem to be a *spatial congestion problem*.

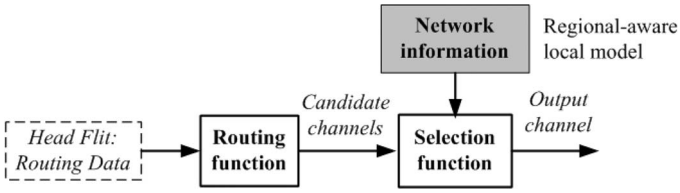

Fig. 6. Routing flow; the network information adopts a regional-aware local model.

In this work, we focus on the *regional-aware local model*, which is a spatial extension of the basic local model by direct wiring, propagation, and some extra computation. In large scale NoCs, appropriate extension of the range of up-to-date information can give the routing process an accurate approximation of the network status. NoP [15] is a spatial extension of the OBL congestion model [24]. NoP collects the queue length and the output channel reservations for routers within two-hop of distances on the path. The result shows that NoP significantly outperforms the original local model. We thus adopt NoP for our regional-aware local model.

4.1 Definition of Information Level (IL)

We define an *Information Level* (*IL*) coordination layer that locates the source of spatial and historical *Router Information* (*RI*), as shown in Fig. 7. The x-axis is the hop count d from the current router, the y-axis is the information retrieved t cycles ago. Combination of grid points can represents the local model. For example, the OBL selection acquires free buffer slots in adjacent routers, and can be represented as below:

$$IL_{OBL} = RI[1, 0]. \tag{9}$$

The above IL of OBL selection is the RI at point [1, 0]. We can index all the available information of the network using IL coordinates on a combination of grid points. We then provide the demand of network information in both historical and spatial domains:

- In the **historical domain**, historical information can help to balance the on-chip traffic load mainly because of the locality of network traffic. Furthermore, the historical local buffer state is correlated with the spatial buffer distribution. This is because the packets staying in local buffer will later be

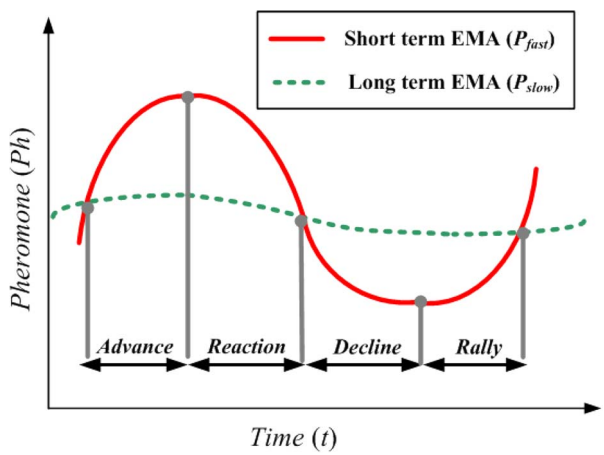

Fig. 5. Definition of the four status of network information. The pheromone is the EMA of free slots, The preferred order of status is: $Advance > Reaction = Rally > Decline$.

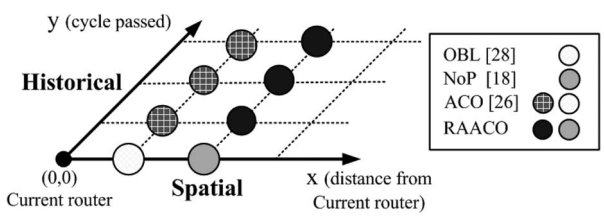

Fig. 7. Illustration of the information level (IL). All the available information of the network can be indexed using the coordinates. Most of the local models can be illustrated on the grid points.

forwarded and added to the spatial congestion. The historical domain IL is shown below:

$$IL_{historical} = \sum_{t=0}^{\infty} W_{\tau}[t] \times RI[d, -t]. \quad (10)$$

The above $IL_{historical}$ keeps a historical weighting window W for the information within past τ cycles. The RI at hop-distance d and t-cycle ago is multiplied by $W_{\tau}[t]$ to accumulate the historical information.

- In the **spatial domain**, we want to be aware of congestion in the paths. By extending the spatial coverage of the IL, we can prevent possible congestion at the distant routers. The implement the spatial domain IL is shown below:

$$IL_{spatial} = \sum_{d=1}^{\infty} W_{\delta}[d] \times RI[d, -t_d]. \quad (11)$$

The above $IL_{spatial}$ keeps a spatial weighting window W for the information within 1-hop to δ-hop. The RI at hop-distance d and t_d-cycle ago is multiplied by $W_{\delta}[d]$ to accumulate the spatial information, where t_d is the information propagation time for router at d.

Note that a larger d increases overhead; besides, since the packet takes much more time to reach a distant router, distant RIs may deviated further from the actual congestion status. To improve the accuracy of (11), we choose t_d carefully to make the congestion value more precise.

4.2 Regional-Aware Model for ACO

As described in [19], ACO has a heuristic correction factor l_n taking into account the state (the length) of the n-th link queue of the current node. l_n represents the instantaneous state of the queue and assumes that the queue serving process is almost stationary or slowly varying. Therefore, l_n also gives a quantitative measurement associated with the queue waiting time. The pheromone routing table value is the outcome of a continual learning process. This table captures both the current and the historical status of the network. With the definitions as (10), the IL of ACO-based selection shown in Fig. 7 can be modelled as

$$IL_{ACO} = \sum_{t=0}^{\infty} W_{\tau}[t] \times RI[1, -t]. \quad (12)$$

After considering previous work on different regional-aware local models and researching the temporal characteristics of ACO, we consider that enhancement to either spatial or temporal domains individually may have only a marginal effect. We aim on exploring the ACO algorithm to jointly compute with a regional-aware local model, and analyze the effect on pheromones.

We observe that the NoP selection scheme reduces the propagation time t_2 to one cycle by direct wiring. With the definitions as (11), The IL of NoP is shown in Fig. 7 and below:

$$IL_{NoP} = RI[2, 0]. \quad (13)$$

Note that the RIs of NoP contain not only the two-hop distant queue state but also the routing information of the next stage router.

We then define the *Regional Aware Model* (RAM) for ACO as: a model that includes information from two or more hops away from the current router: $d \geq 2$. As we extend the coverage of the information, there is a trade-off between the performance and overhead. In this work, we consider a special case and adopt NoP for our regional-aware model. With the definitions as (10) and (11), the IL of Regional-Aware ACO (RA-ACO) of Fig. 7 can be expressed as follows:

$$IL_{RAACO} = \sum_{t=0}^{\infty} W_{\tau}[t] \times RI[2, -t]. \quad (14)$$

The RA-ACO uses the regional-aware model of buffer status and can achieve higher coverage at the information level. The RA-ACO is suitable for spatial long-range patterns.

We now consider the performance of this solution to the network problem. This system offers a set of different local models, and the selection scheme then decides to acquire information based on the regional-aware model, which can be represented at the information level. Wider coverage in the IL provides a better approximation to the network traffic distribution status with a performance saturation point. That is, the highest achievable throughput of a selection scheme with IL_{Φ} is greater than the highest achievable throughput of a selection scheme with IL_{Ω} under saturation, where IL_{Ω} is a subset of IL_{Φ}. For example, the IL_{OBL} is a subset of IL_{ACO} as shown in Fig. 10; therefore, the ACO has the potential to achieve a higher performance than OBL. Detailed operation and pseudo code of the proposed RA-ACO Algorithm in shown in supplemental material VI available online.

5 Experiments and Discussions

5.1 Simulation Setup

We adopted the simulation platform *Noxim* [24], which is an open source SystemC simulator for a mesh-based NoC. This tool provides various traffic scenarios and routing strategies. The tests are conducted under the provided environment settings.

We evaluated the proposed schemes with both synthetic and real traffic scenarios. The environment setting of the simulations is a 16×16 mesh NoC. Traffic sources generate 4-flit packets, with an injection interval following a self-similar distribution. The FIFO buffer size is four flits. Each simulation was initially run for 2,000 cycles to allow network transient effects to stabilize, and then it was executed for 20,000 cycles.

The evaluation index is based on the *packet injection rate* (*pir*), which is the rate that packets are injected into the network. For example, a 0.1 *pir* represents that on average, each node sends a packet every 10 cycles. The instant at which a packet is injected depends on the inter-arrival time distributions. We choose the instant on the basis of a self-similar Pareto distribution.

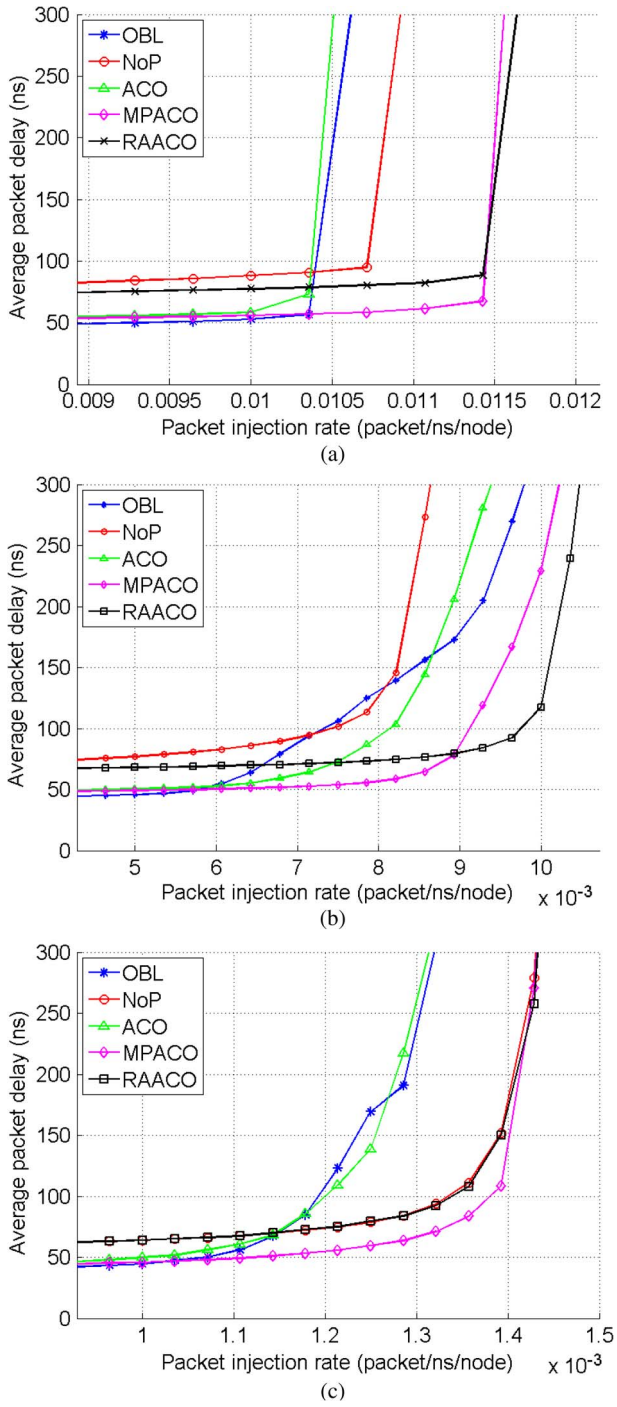

Fig. 8. Average packet delay in a 16 × 16 mesh network with (a) uniform, (b) transpose, and (c) hotspot-centred traffic.

For performance metrics, we evaluate the *delay* of different algorithms. Delay is defined as the time (in cycle periods) that elapses between the event of the head flit injection into the network at the source node and the event when a tail flit is received at the destination node [15]. The average delay, D, is the following performance metric:

$$D = \frac{1}{K}\sum_{i=1}^{K} D_i, \tag{15}$$

where K is the total number of packets reaching their destination nodes, and D_i is the delay of packet i. The judging point is at the *saturation throughput* [11], which is defined as the packet injection rate when the latency is double the zero-load latency.

In our settings, the ratio between ant packet and data packet is 1 : 1. For example, when 2000 packets are injected, 1000 packets are assigned to be ant packets. Both types of packets carry the payload, but only ant packets contain an extra 1-bit identifier in the header flit. Once an ant packet header arrives at a router, the selection process triggers the pheromone table to update itself. On the other hand, data packet selects a path directly according to the pheromone table. When all packets are ant packets, the performance improve marginally due to saturation [19], [20], [21]. The extra ant packets trigger pheromone table update and consume extra energy.

5.2 Experimental Results on Synthetic Traffic

For each traffic scenario and algorithm, we compare the average packet delay with various *pir* values. The selection schemes compared are OBL [24], NoP [15], ACO, MP-ACO, and RA-ACO, all using an Odd-Even routing algorithm [9].

5.2.1 Uniform Traffic

The most commonly used traffic pattern in network evaluation is *uniform* traffic. The results are illustrated in Fig. 8a. As it shows, ACO performs near OBL since historical information is less useful in randomly distributed traffic. NoP acquires better performance than OBL under certain *pir* by employing additional routing functions in order to eliminate the unused network information. The performance of MP-ACO and RA-ACO further improves by considering temporal and spatial domain information.

In uniform traffic, each source is equally likely to send to each destination. Therefore, the dimension-ordered routing algorithm [6] can route the packets in a balance way in the long run. The reason is that the deterministic routing policy (uniform route) is consistent with the long-term traffic distribution (uniform distributed). The performance differences of adaptive routings are caused by transient imbalances. Theoretically, the performance of adaptive routing is bounded by the dimension-order routing in uniform traffic.

5.2.2 Transpose Traffic

Transpose traffic is induced from matrix transpose or corner-turn operations. Because it concentrates the load on individual source-destination pairs, transpose traffic stresses the load balance of a routing algorithm. The results are shown in Fig. 8b. Since the historical data is much more meaningful in this pattern, the ACO has a higher performance. The adaptability of the selection function is also more significant for transpose traffic. MP-ACO and RA-ACO give an improvement of 33.8 percent and 50 percent respectively in saturation throughput compared with OBL and NoP.

5.2.3 Hotspot Centred Traffic

The *hotspot* pattern is a simulation model where the traffic concentrates on several tiles and causes heavy congestion. The hotspot could represent a caching process or frequent

access to certain IPs. In this scenario, four hotspot nodes are located at the centre of the mesh, that is nodes [(7, 7), (7, 8), (8, 7), (8, 8)] with 20 percent hotspot traffic.

The result is shown in Fig. 8c. When traffic flows target a small region of nodes, the blockage situation is severe, thus the working *pir* is almost an order of magnitude lower than uniform and transpose traffic. For algorithms with higher adaptability, the saturation throughputs are similar.

5.3 Experimental Results on Real-Traffic Data

To simulate a realistic traffic load in an NoC system, we perform a complete network analysis, including the Low-Density Parity-Check (LDPC) codes and the MMS [26].

5.3.1 LDPC Traffic

LDPC code is one kind of linear block codes, first proposed in [28]. Due to the excellent error-correcting performance, the LDPC codes have been widely adopted by many advanced wire-line and wireless communication systems. The LDPC decoding algorithm can be represented as a parity-check matrix (H) or a bipartite graph.

To increase the hardware implementation efficiency, a parallel LDPC decoding was proposed on an NoC-based multiprocessor system because of the characteristics of regularity and scalability [29]. Given the LDPC design in [28] and the evaluation method in [30], we evaluate different routing schemes with the data flow of the (1944, 972) LDPC code, which is defined in *IEEE 802.11n*. The mapping of bit node unit (BNU) and check node unit (CNU) onto processing elements can be done by:

$$N_{map} = (BNU/CNUnumber) mod(N_{NoC}), \qquad (16)$$

where N_{map} denotes the mapped node number and N_{NoC} denotes the total number of NoC nodes. We map the (1944, 972) LDPC codes to a 16 × 16 mesh-based NoC system.

Fig. 9 shows the result due to the locality of the LDPC traffic pattern, we observed the MP-ACO outperforms the other selection functions in this pattern. The NoP does not even perform as well as OBL. Because the RA-ACO also adopts the NoP spatial model, the performance is constrained.

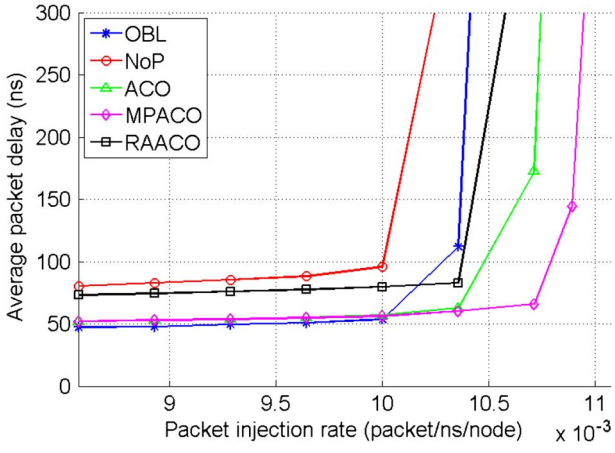

Fig. 9. Average packet delay in low-density parity-check (LDPC) traffic in a 16 × 16 mesh network.

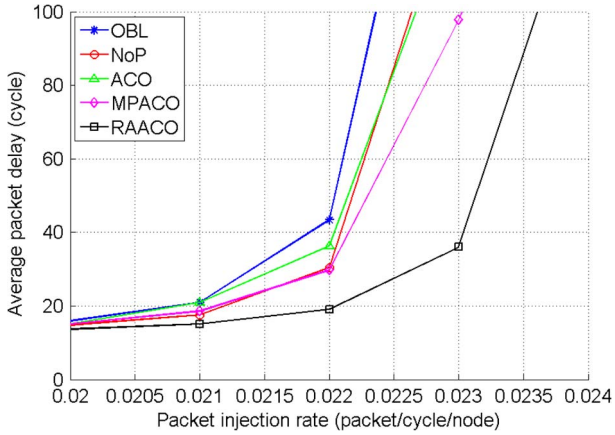

Fig. 10. Average packet delay in multimedia system (MMS) traffic in a 5 × 5 mesh network.

5.3.2 MMS Traffic

Our Multi-Media System (MMS) pattern uses a model from [15] and discussed in [25]. It includes an H. 263 video codec and an MP3 audio codec. The application is partitioned into 40 distinct tasks and assigned on 25 IPs. The mapping of IPs into nodes of a 5 × 5 mesh-based NoC architecture has been obtained by using the method presented in [26]. As we observe from Fig. 10, the result shows that RA-ACO outperforms other algorithms with high information level. Then MP-ACO and NoP have similar performance, and outperform OBL and ACO.

Table 2 presents a summary of the results in terms of saturation throughput, illustrated in Fig. 11. We consider the traffic patterns on 16 × 16 mesh (*Uniform, Transpose1, Hotspot and LDPC*), it shows different levels of improvement of the proposed selection schemes. MP-ACO and RA-ACO have a normalized improvement on the saturation throughput by 14.38 and 18.64 percent.

- **RA-ACO is best-suited for spatial long-range patterns like transpose and MMS**. The reason is that the regional buffer information model of RA-ACO can achieve a higher coverage on the information level.
- **The MP-ACO is suitable for patterns with locality like LDPC**. The reason is that the MP-ACO utilizes the local buffer information model to be aware of the temporal variation.

TABLE 2
Comparison of the Saturation Point of Selection Functions under Different Traffic Scenarios

Enhancement	Temporal			Spatial	
Traffic Scenario	Saturation pir (packets/ns/node)				
	OBL	ACO	MPACO	NoP	RAACO
Uniform	0.0104	0.0104	0.0114	0.0101	0.0114
Transpose	0.0068	0.0078	0.0091	0.0062	0.0093
Hotspot	0.0012	0.0012	0.0014	0.0014	0.0014
LDPC	0.0102	0.0105	0.0108	0.0097	0.0104
MMS	0.0079	0.0079	0.0079	0.0079	0.0081
Average Throughput (10^{-3})	7.15	7.48	8.18	6.84	8.12
Improvement (%)	0.00	4.62	14.38	0.00	18.64

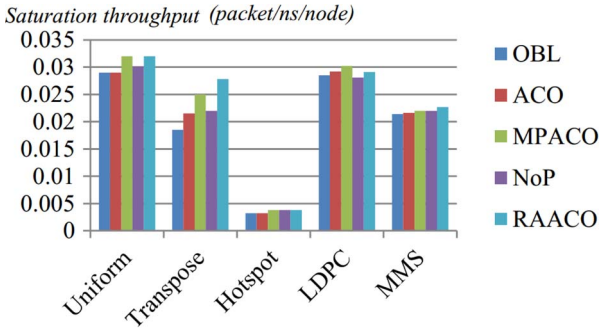

Fig. 11. Comparison of the saturation point of selection functions.

6 Architectural Design of Proposed Schemes

6.1 Enhanced ACO-Based Router Architecture

The fundamental router architecture of this work is as [15]. The routing function is shown in Fig. 12. We presents the overhead and ad-hoc signal analysis in supplemental material III and IV available online.

Fig. 13a shows the framework of the MP-ACO selection function. The ComputeRegion unit uses the destination and current router *id* to identify the region index. The Update unit then uses *free_slots_in* and *table_data_in* to generate updated pheromone information. Next, the ComputeStatus unit uses the status index to map to the current status. The Selection unit then selects a channel based on the status and pheromone value.

Fig. 13b is the framework of the RA-ACO selection function. The difference is that the Update unit uses the aggregated *NoP_data_in* and *table_data_in* to generate updated information. Next, the Selection unit then selects based on the historical regional-aware pheromone value.

The window size of MPACO running averages is achieved by the Update Unit. As we know from [18], [19], [20], the implicit ACO pheromone window size of can be computed with the pheromone bit length and the pheromone weighting. According to [22], [37], the number of pheromone bit length reaches the saturation point at 7 bits. Therefore, we also assign the pheromone to 7 bits to keep the performance in the evaluation process of this work. The weighing are assigned with $\alpha = 1/3$, and $\beta = 1/4$. We compute the window size with (9) and find the window size of short-term and long-term pheromone are 10 and 14 cycles, respectively. We also describe the computation of window size in the supplemental material VIII available online.

6.2 Analysis of Hardware Overhead

Based on the router architecture in Figs. 12 and 13, routers with different selection functions are designed and implemented. Furthermore, each router is synthesized with *TSMC* 90 nm technology. The synthesis results are listed in Table 3.

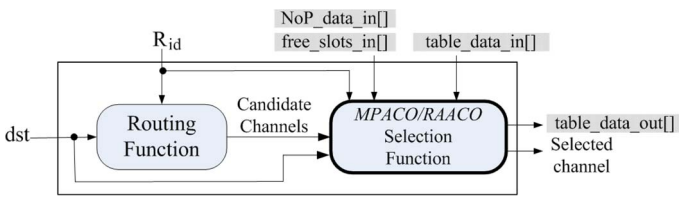

Fig. 12. Implementation of the routing algorithm.

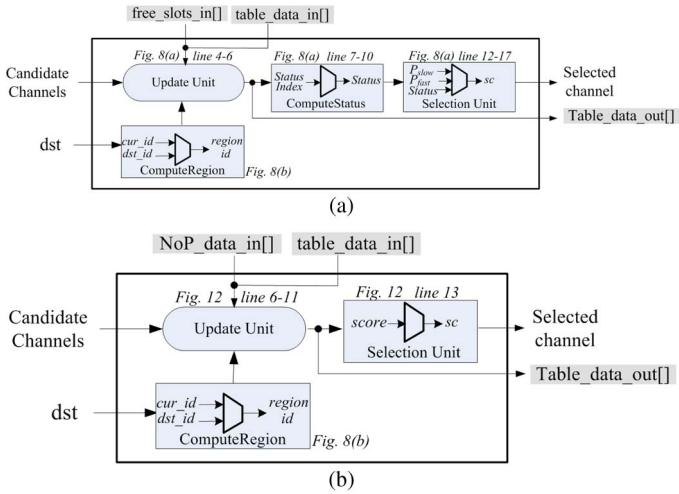

Fig. 13. Implementation of the (a) MP-ACO and (b) RA-ACO selection function.

As we can see, the OBL router consumes the lowest area overhead, since it requires the least information and thus the least computation. As previously described, the routing resource in NoP is larger, and requires extra signal wiring. Therefore, the hardware overhead of NoP and RA-ACO are comparatively high compared to others. ACO requires an ant table to store pheromone information accumulated over time. This is the main hardware overhead, and table cost also increases in MP-ACO. The area cost of MP-ACO and RA-ACO increases by different amounts.

6.3 Discussion on Cycle Time

The influence of cycle time and critical path of ACO is mainly composed of the table lookup, table updating, and the computation of selection.

1. **Table Lookup**: In the original ACO, the table size is $N^2 - 1$, with N denotes the mesh size. In this work, we adopt [22], the table size is reduced to a reasonable 4×2 with the merit of the regular mesh topology and NoC network structure. Therefore, our table does not suffer from the table scaling issue. The smaller table size greatly reduces memory cost, table lookup time, and the corresponding power consumption. The extra computation requirement is the ComputeRegion block as shown in Fig. 8b. The critical path is a comparator and a Mux. The bits of n_c and n_d (current and destination node *id*) increase by when $\log(N)$ system scales to $N \times N$.

TABLE 3
Hardware Synthesis Results

Enhancement	Temporal			Spatial	
Selection Function	OBL	ACO	MPACO	NoP	RAACO
Frequency (MHz)	550	500	500	360	360
Area (μm^2)	58.51	59.60	60.32	62.73	64.52
Area Efficiency (10^{-5})	1.221	1.255	1.355	1.090	1.258

TABLE 4
Total Energy Consumption to Drain 10 M Bytes of Data for Multimedia System and Parallel LDPC System

Real Traffic Data	*pir*	Energy Consumption (mJ)				
		OBL	ACO	MPACO	NoP	RAACO
(1944, 972) LDPC codes	0.027	3.29	3.28	3.27	3.39	3.27
	0.028	3.30	3.29	3.29	3.39	3.28
	0.029	3.37	3.32	3.31	3.67	3.31
	0.030	3.90	3.64	3.35	4.81	3.35
	0.031	4.33	4.07	3.77	4.99	3.58
MMS	0.02	1.47	1.47	1.47	1.47	1.48
	0.021	1.49	1.48	1.47	1.48	1.47
	0.022	1.47	1.47	1.47	1.49	1.47
	0.023	1.48	1.47	1.48	1.47	1.47
	0.024	1.49	1.47	1.48	1.48	1.48

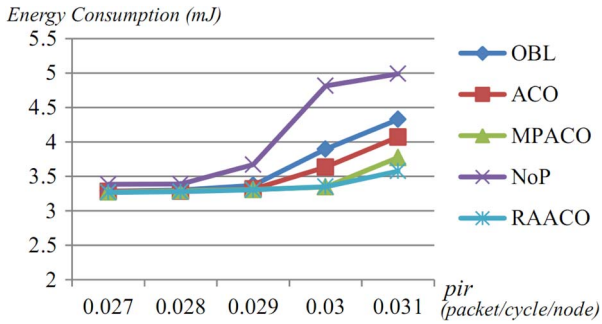

Fig. 14. Total energy consumption to drain 10 M bytes of data for parallel LDPC system.

2. **Table Updating**: For the ACO-based routing algorithm, the routing computation (RC) and switch allocation (SA) are accomplished in first cycle, and table update (TU) and switch traversal (ST) are accomplished in the second cycle. Therefore, the table updating does not influence the critical path of RC. Besides, since our table size does not scale with system, the RC and TU are nearly invariant with system size. That is, the table updating process can be separated from the channel judgment process thus reduce the influence on the computational delay.
3. **Computation of Selection**: For hardware friendly, We implemented the computation of MPACO with shifters and adders. This reduces the hardware resources consumption, delay, and power consumption of the routing process. The status computation has a critical path of two adders, two comparators, and one shifter, which is feasible for the NoC implementation. In the case of RAACO, since the NoP score needed to be computed, thus the computation time is increased.

In Table 3, the OBL router can operate at speeds of 550 MHz because it requires the computation for local buffer occupancy information and Linear Feedback Shift Register (LFSR) selection. Since ACO and MPACO require additional table look up and selection computation for temporal congestion information, it operates at speeds of 500 MHz. Besides, NoP and RAACO require additional computation for look-ahead buffer information; therefore, they operate at speeds of 360 MHz.

6.4 Energy Consumption With Real-traffic Data

To predict the realistic energy consumption of an NoC system, we perform a complete network analysis. We considering congestion effects in draining a fixed workload (10 Mbytes) with two real applications, including the LDPC codes and the MMS [25], as discussed in Section 5.3.

Table 4 shows the total energy consumption of five representative pir under the LDPC codes and MMS traffic. As shown in Fig. 14, at the lower pir of LDPC (i.e., 0.027-0.029), we find that the energy used by NoP, MP-ACO and RA-ACO is comparable to OBL and ACO. Because there is less congestion at these pir values, all of selection schemes can rapidly forward the packets to the destination nodes. Therefore, the amount of energy consumption is directly related to the complexity of routing computation.

On the other hand, at higher pir values of LDPC (i.e., 0.030-0.031), the energy consumption of all selection schemes increases significantly. Due to the high probability of contention under heavy traffic workloads, many packets could be blocked. The routing computation is re-executed which would consume additional energy. Meanwhile, the buffer to store these packets also requires extra energy.

According to [15], ''invoking higher adaptive selection function implies additional energy consumption, this may be balanced by a minor usage of FIFO buffers due to better congestion avoidance.'' Notably, we have also found out that a selection function with better performance can balance its power dissipation overhead, which is consistent with the observation of [15]. When the congestion happens, the extra computation and buffer stall cycles increase the energy consumption and leakage power. The total energy consumption of MP-ACO and RA-ACO are less than the routing process of other selection functions. For example, at the pir value of 0.031, that of LDPC, MP-ACO and RA-ACO schemes have an energy saving of 12.8 percent and 17.4 percent compared to OBL, respectively. Similarly, at the pir value of 0.024 of MMS, MP-ACO and RA-ACO schemes have an energy saving of 0.7 percent compared to OBL.

6.5 Evaluation of Area Efficiency

The network performance and the hardware overhead of each selection function are discussed and evaluated above and shown in Tables 2 and 3, respectively. As we observe, each selection function attains a certain degree of improvement but has the corresponding hardware overhead. In order to evaluate the effectiveness of each selection function, the area efficiency is calculated. Area efficiency is defined as:

$$Area\ Efficiency = \frac{Average\ Throughput}{Area\ Overhead}. \qquad (17)$$

It is calculated by dividing the average network throughput of each selection function by the corresponding hardware overhead. The area efficiency is also tabulated in Table 3. As it shows, MP-ACO and RA-ACO have an

area efficiency equals 1.355 and 1.258, implying that the ratio of improvement gained over overhead consumed is high and the investment is worthwhile compared with the OBL and NoP selection. This shows the acquiring of temporal and spatial network information can improve the system performance.

7 Conclusion

In this work, we propose the Multi-pheromone and Regional-Aware ACO-based selection schemes to explore the temporal and spatial NoC properties. The main purpose is to make use of the temporal pheromone variation and spatial pheromone coverage to sense and predict system congestion. The results show that for a feasible implementation cost, we can substantially improve the network performance of an NoC. We also implemented the proposed ACO router and analyzed the area efficiency and the energy consumption of each schemes.

Acknowledgment

This work was supported by the National Science Council of Taiwan under Grant NSC 100-2221-E-002-001-MY3 and NSC 102-2220-E-002-013.

References

[1] R. Marculescu, U.Y. Ogras, L.-S. Peh, N.E. Jerger, and Y. Hoskote, ''Outstanding Research Problems in NoC Design: System, Microarchitecture,'' *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 28, no. 1, pp. 3-21, Jan. 2009.

[2] U.Y. Ogras, J. Hu, and R. Marculescu, ''Key Research Problems in NoC Design: A Holistic Perspective,'' in *Proc. IEEE/ACM/IFIP Hardw./Softw. Codes. Syst. Synthesis Conf.*, 2005, pp. 69-74.

[3] U.Y. Ogras, P. Bogdan, and R. Marculescu, ''An Analytical Approach for Network-on-Chip Performance Analysis,'' *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 29, no. 12, pp. 2001-2013, Dec. 2010.

[4] W.J. Dally and B. Towles, ''Route Packets, Not Wires: On-Chip Interconnection Networks,'' in *Proc. ACM/IEEE Des. Autom. Conf.*, 2001, pp. 684-689.

[5] L. Benini and G.D. Micheli, ''Network on Chip: A New SoC Paradigm for Systems on Chip Design,'' in *Proc. IEEE Des., Autom. Test Conf.*, 2002, pp. 418-419.

[6] W.J. Dally and B. Towles, *Principles and Practices of Interconnection Networks*. San Mateo, CA, USA: Morgan Kaufmann, 2004, pp. 159-244.

[7] Y.-F. Xu, J.-Y. Zhou, and S.-K. Liu, ''Research and Analysis of Routing Algorithms for NoC,'' in *Proc. IEEE Comput. Res. Dev. Conf.*, 2011, vol. 2, pp. 98-102.

[8] C.J. Glass and L.M. Ni, ''The Turn Model for Adaptive Routing,'' *J. ACM*, vol. 41, no. 5, pp. 874-902, Sept. 1994.

[9] G.-M. Chiu, ''The Odd-Even Turn Model for Adaptive Routing,'' *IEEE Trans. Parallel Distrib. Syst.*, vol. 11, no. 7, pp. 729-738, July 2000.

[10] J.C. Martinez, F. Silla, P. Lopez, and J. Duato, ''On the Influence of the Selection Function on the Performance of Networks of Workstations,'' in *Proc. High Perform. Comput. Conf.*, 2000, pp. 292-299.

[11] L. Shang, L.-S. Peh, and N.K. Jha, ''Powerherd: A Distributed Scheme for Dynamically Satisfying Peak-Power Constraints in Interconnection Networks,'' *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 25, no. 1, pp. 92-110, Jan. 2006.

[12] J. Hu and R. Marculescu, ''DyAD—Smart Routing for Networks-on-Chip,'' in *Proc. ACM/IEEE Des. Autom. Conf.*, June 2004, pp. 260-263.

[13] S. Rodrigo, S. Medardoni, J. Flich, D. Bertozzi, and J. Duato, ''Efficient Implementation of Distributed Routing Algorithms for NoCs,'' *IET Comput. Digit. Techn.*, vol. 3, no. 5, pp. 460-475, Sept. 2009.

[14] T. Mak, P.Y.K. Cheung, K.-P. Lam, and W. Luk, ''Adaptive Routing in Network-on-Chips Using a Dynamic-Programming Network,'' *IEEE Trans. Ind. Electron.*, vol. 58, no. 8, pp. 3701-3716, Aug. 2011.

[15] G. Ascia, V. Catania, M. Palesi, and D. Patti, ''Implementation and Analysis of a New Selection Strategy for Adaptive Routing in Networks-on-Chip,'' *IEEE Trans. Comput.*, vol. 57, no. 6, pp. 809-820, June 2008.

[16] P. Gratz, B. Grot, and S.W. Keckler, ''Regional Congestion Awareness for Load Balance in Networks-on-Chip,'' in *Proc. IEEE Int'l Symp. High Perform. Comput. Architect.*, 2008, pp. 203-214.

[17] M. Dorigo, V. Maniezzo, and A. Colorni, ''The Ant System: Optimization by a Colony of Cooperating Agents,'' *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 26, no. 1, pp. 29-41, Feb. 1996.

[18] M. Dorigo, M. Birattari, and T. Sttzle, ''Ant Colony Optimization,'' *IEEE Comput. Intell. Mag.*, vol. 1, no. 4, pp. 28-39, Nov. 2006.

[19] G. Di Caro and M. Dorigo, ''AntNet: Distributed Stigmergetic Control for Communications Networks,'' *J. Artif. Intell. Res.*, vol. 9, no. 1, pp. 317-365, Aug. 1998.

[20] K.M. Sim and W.H. Sun, ''Ant Colony Optimization for Routing and Load-Balancing: Survey and New Directions,'' *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 33, no. 5, pp. 560-572, Sept. 2003.

[21] M. Daneshtalab and A. Sobhani, ''NoC Hot Spot Minimization Using AntNet Dynamic Routing Algorithm,'' in *Proc. IEEE Appl. Specific Syst., Architect. Processors Conf.*, 2006, pp. 33-38.

[22] H.-K. Hsin, E.-J. Chang, C.-H. Chao, and A.-Y. Wu, ''Regional ACO-Based Routing for Load-Balancing in NoC Systems,'' in *Proc. IEEE Nat. Biolog. Inspired Comput. Conf.*, Dec. 2010, pp. 370-376.

[23] H.-K. Hsin, E.-J. Chang, C.-H. Chao, S.-Y. Lin, and A.-Y. Wu, ''Multi-Pheromone ACO-Based Routing in Network-on-Chip System Inspired by Economic Phenomenon,'' in *Proc. IEEE SOC Conf.*, Nov. 2011, pp. 273-277.

[24] Noxim: Network-on-Chip Simulator. 2008. [Online]. Available: http://sourceforge.net/projects/noxim

[25] J. Hu and R. Marculescu, ''Energy- and Performance-Aware Mapping for Regular NoC Architectures,'' *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 24, no. 4, pp. 551-562, Apr. 2005.

[26] G. Ascia, V. Catania, and M. Palesi, ''Multi-Objective Mapping for Mesh-Based NoC Architectures,'' in *Proc. 2nd IEEE/ACM/IFIP Intl Conf. Hardw./Softw. Codes. Syst. Synthesis*, Sept. 2004, pp. 182-187.

[27] Stock Chart. 2010. [Online]. Available: http://stockcharts.com

[28] R. Gallager, ''Low-Density Parity-Check Codes,'' *IRE Trans. Inf. Theory*, vol. 8, no. 1, pp. 21-28, Jan. 1962.

[29] W.H. Hu, J.H. Bahn, and N. Bagherzadeh, ''Parallel LDPC Decoding on a Network-on-Chip Based Multiprocessor Platform,'' in *Proc. Intl SBAC-PAD*, Oct. 2009, pp. 35-40.

[30] K.-C. Chen, S.-Y. Lin, H.-S. Hung, and A.-Y. Wu, ''Topology-Aware Adaptive Routing for Non-Stationary Irregular Mesh in Throttled 3D NoC Systems,'' *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 10, pp. 2109-2120, Oct. 2012.

[31] S. Ma, N.E. Jerger, and Z.-Y. Wang, ''DBAR: An Efficient Routing Algorithm to Support Multiple Concurrent Applications in Networks-on-Chip,'' in *Proc. Int'l Symp. Comput. Architect.*, 2011, pp. 413-424.

[32] R.S. Ramanujam and B. Lin, ''Destination-Based Adaptive Routing on 2D Mesh Networks,'' in *Proc. ACM/IEEE ISCA*, 2010, pp. 1-12.

[33] C.C. Holt, ''Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages,'' *Int'l J. Forecast.*, vol. 20, no. 1, pp. 5-10, Jan.-Mar. 2004.

[34] M. Ramakrishna, P.V. Gratz, and A. Sprintson, ''GCA: Global Congestion Awareness for Load Balance in Networks-on-Chip,'' in *Proc. Int'l Symp. NoCS*, 2013, pp. 1-8.

[35] M. Daneshtalab, M. Ebrahimi, J. Plosila, and H. Tenhunen, ''CARS: Congestion-Aware Request Scheduler for Network Interfaces in NoC-Based Manycore Systems,'' in *Proc. DATE Conf. Exhib.*, 2013, pp. 1048-1051.

[36] J. Jose, B. Nayak, K. Kumar, and M. Mutyam, ''DeBAR: Deflection Based Adaptive Router With Minimal Buffering,'' in *Proc. DATE Conf. Exhib.*, 2013, pp. 1583-1588.

[37] E.-J. Chang, C.-H. Chao, K.-Y. Jheng, H.-K. Hsin, and A.-Y. Wu, ''ACO-Based Cascaded Adaptive Routing for Traffic Balancing in NoC Systems,'' in *Proc. ICGCS*, 2010, pp. 317-322.

Hsien-Kai Hsin received the BS degree in electronic engineering from the National Taiwan University, Taiwan, in 2009 and is currently pursuing the PhD degree in the Graduate Institution of Electronic Engineering in National Taiwan University, Taipei, Taiwan. His current research interests include Biologically-inspired systems and high-speed interconnects. In particular, his research focuses on swarm intelligence design methodologies for multicore SoCs, with special interest on Network-on-Chip communication architecture.

En-Jui Chang received the BS degree in electrical engineering from National Central University, Zhongli, Taiwan, in 2008 and is currently pursuing the PhD degree in the Graduate Institute of Electronics Engineering, National Taiwan University, Taipei, Taiwan. His research interests include Network-on-Chip (NoC) algorithms/architectures, bio-inspired algorithms/architectures, and fault tolerance algorithms/architectures in VLSI systems.

An-Yeu (Andy) Wu received the BS degree from National Taiwan University in 1987, and the MS and PhD degrees from the University of Maryland, College Park, MD, USA, in 1992 and 1995, respectively, all in electrical engineering. In 2000, he joined the faculty of the Department of Electrical Engineering and the Graduate Institute of Electronics Engineering, National Taiwan University (NTU), where he is currently a Professor. His research interests include low-power/high-performance VLSI architectures for DSP and communication applications, adaptive/multi-rate signal processing, reconfigurable broadband access systems and architectures, and System-on-Chip (SoC)/Network-on-Chip (NoC) platform for software/hardware co-design. He is a Senior Member of the IEEE.