CrossMark

Design and analysis of a mesh-based wireless network-on-chip

**Wen-Hsiang Hu · Chifeng Wang ·
Nader Bagherzadeh**

Abstract Network-on-chip (NoC) architecture is regarded as a solution for future on-chip interconnects. However, the performance advantages of conventional NoC architectures are limited by the long latency and high power consumption due to multi-hop long-distance communication among processing elements. To solve these limitations, we employed on-chip wireless communication as express links for transferring data so that transfer latency can be reduced. A hybrid NoC architecture utilizing both wired and wireless communication approaches is proposed in this paper. We also devised a deadlock-free routing algorithm that is able to make efficient use of the incorporated wireless links. Moreover, simulated annealing optimization techniques were applied to find optimal locations for wireless routers. Cycle-accurate simulation results showed a significant improvement in transfer latency. Area and power consumption analysis demonstrates the feasibility of our proposed NoC architecture.

1 Introduction

During the past decade, single chip integration has witnessed major paradigm shifts to meet challenging design requirements for computation-intensive applications and

W.-H. Hu · C. Wang (✉) · N. Bagherzadeh
Department of Electrical Engineering and Computer Science,
University of California, Irvine, CA 92697, USA
e-mail: chifengw@uci.edu

W.-H. Hu
e-mail: wenhsiah@uci.edu

N. Bagherzadeh
e-mail: nader@uci.edu

highly integrated low-power solutions. However, on-chip interconnects carrying signals from one block to another will be the bottleneck to system performance and reliability. Network-on-chip (NoC) architecture is considered as a promising technology to tackle design challenges faced by the conventional bus architecture by using network-like interconnection among intellectual property (IP) blocks [2]. Nevertheless, NoCs still face limitations such as long transmission latency and high power consumption due to planar multi-hop communications.

Therefore, alternative communication approaches such as optical interconnections [21], RF Interconnect (RF-I) transmission lines [3] and CMOS Ultra-wideband (UWB) wireless interconnect technology [24] have been proposed. The basic idea of these alternatives is to deploy express communication links to reduce transmission latency and power dissipation. However, on-chip optical interconnections have technological challenges such as design of efficient transmitter and receiver components, integration of on-chip photonic components and high manufacturing cost which will prevent its commercial adoption. Although multi-band RF-I can be implemented by silicon-based CMOS technology, it requires additional, physically overlaid transmission lines which serve as wave guides to enable data communication. To achieve high throughput, RF-I based systems must utilize multiple high-frequency oscillators and high-precision filters to validate their feasibility. CMOS UWB wireless links enable on-chip multi-hop communication by the use of embedded wireless channels. It uses existing and well-understood CMOS technology to replace multi-hop wired communication links with single-hop long-range wireless channels so that transmission performance and power consumption problems in conventional, wired NoCs can be addressed simultaneously.

In this work, we proposed a hybrid on-chip communication infrastructure which makes use of on-chip wireless interconnect working with existing wired NoC. We refer to this hybrid architecture as Wireless NoC (WNoC). The backbone interconnect is based on a 2-D mesh NoC architecture divided into rectangular subnets. Wireless links are inserted between subnets to form express communication links by replacing baseline wired routers with routers having wireless communication capabilities. We applied simulated annealing (SA) [15] optimization technique to find optimal locations for wireless routers (WRs) so that average traversal distance is minimized. A virtual channel-based, deadlock-free routing scheme capable of balancing the utilization between wired and wireless networks is also presented. Experimental results from cycle-accurate simulation showed significant improvement in transfer latency. Feasibility analysis of WNoC including area and power consumption analysis is also addressed in this work.

The rest of this paper is organized as follows: Sect. 2 introduces the related work in this paper. Section 3 presents WNoC architecture and the SA-based router placement algorithm. Experiment results and cost analysis are presented in Sect. 4. Finally, brief statements conclude this paper in the last section.

2 Related work

NoC architectures were proposed to replace conventional, bus-based communication schemes and provide a scalable backbone interconnect. 2-D mesh networks are the

most popular topology because of their feasibility resulting from short channel length and low router complexity. However, the two-dimensional floorplan limits the performance improvement expected from NoC. For example, the network diameter of a 2-D mesh NoC grows linearly with the network dimension. For a 2-D mesh NoC with 100 cores, the network diameter is 18, which means there are up to 18 hops between source and destination pairs. This long multi-hop transmission leads to long transfer latency and high power consumption.

Several on-chip interconnect alternatives have been proposed to alleviate these problems. 3-D NoCs [19] make use of the advantages obtained from both 3-D ICs and NoCs to improve latency, throughput and power consumption. Nevertheless, the increased level of power density poses a heat dissipation concern for 3-D NoCs. On-chip optical interconnects [21] are expected to give very high throughput with low latency. The integration of optical wave guides and photonic devices is still a difficult issue for chip fabrication using current technologies. RF interconnects [3] modulate data on a carrier frequency to deliver data over an on-chip RF wave guide instead of sending baseband signals on a parallel bus, but there are several implementation issues with this technology as well [4].

On-chip wireless interconnect was first used to distribute global clock signals in [14]. In [24], a wireless NoC was implemented based on UWB technology, and it utilized wired signals to realize a synchronous and distributed medium access control (MAC) protocol. Their work achieved 1 mm transmission range with antennas of length 2.98 mm. The peak bandwidth on a single channel is 10 Gbps in 0.18 µm technology. It is reported in [17] that the cut-off frequency for NMOS transistors is predicted to be 500 GHz by 2013. At such a high operating frequency, the typical bandwidth efficiency of 1 bps/Hz will lead to a data rate of hundreds of Gbps. From the implementation cost point of view, high operating frequency shrinks the antenna size to (sub)millimeter [6], which results in low-cost communication solutions. [16] presented a scalable wireless interconnect structure utilizing on-chip wireless communication. It used a two-tier wired/wireless architecture to demonstrate the advantages of employing long-range wireless links. Architecture and performance evaluation for various wireless network topology types and sizes are discussed. Networks divided into moderate subnet size and number of subnets can achieve better performance [8,16]. In addition to silicon-based on-chip antenna solutions, the feasibility of applying carbon nanotubes (CNT) to implement on-chip antennas was discussed and analyzed in [7,9,18]. CNT-based antenna is able to provide better emission and absorption characteristics than traditional materials, and this makes it a good candidate for on-chip antenna elements for optical frequencies. Various physical layer designs ranging from UWB, millimeter-wave, sub-terahertz to terahertz are discussed and compared. Reliability issue was addressed and results were compared among wired and wireless platforms in terms of throughput, packet energy dissipation and bit error rate with and without error control coding [7]. With either silicon-based or CNT-based on-chip antennas, we will have key technologies available for wirelessly transmitting data inside a chip at a rate of tens to hundreds of Gbps in the near future.

3 WNoC architecture

3.1 On-chip wireless interconnect

The design of a 324 GHz oscillator using 90 nm CMOS technology [11] and a 410 GHz oscillator using 65 nm CMOS technology [20] has been reported for on-chip wireless communications. Based on these technologies, the level of output power of the on-chip millimeter-wave generator is predicted to be as high as -1.4 dBm in 32 nm CMOS process [16], which enables on-chip short distance communication. With recent advances in CMOS mm-wave circuits, hundreds of GHz of bandwidth will be available in the near future. An on-chip antenna deposited in the polyimide layer to minimize substrate loss was proposed to improve wireless transmission bit-error rate (BER) to a reliable level [16]. At a distance of 10 mm, the BER will be less than 10^{-14} which is sufficient to serve as a reliable transmission medium. Based on the estimated 500 GHz switching rate of a CMOS transistor for the 32 nm CMOS process, we can implement many high-frequency bands for the on-chip wireless network. The maximum available bandwidth is empirically predicted to be 10 % of the carrier frequency. This scenario can accommodate up to a total of 16 available channels for the on-chip wireless network at the range from 100 to 500 GHz, where each channel can transmit at around 20 Gbps. Besides bandwidth capacity, the on-chip wireless network requires a simple wireless transceiver architecture to achieve low power design so as to satisfy the stringent requirements of future chip design. Optical Mach–Zehnder modulation at data rates up to 10 Gbps demonstrated with low RF power consumption of only 5 pJ/bit [10] are currently commercially available. A simple on-off-keying (OOK) system suffices to satisfy these requirements in such high-frequency range.

3.2 Topology

The WNoC architecture is based on a conventional, wired 2-D mesh NoC architecture consisting of 5-port baseline routers (BRs). WNoC is constructed by first dividing the 2-D mesh into rectangular subnets and replacing one of the BRs in each subnet with a WR, which has additional wireless links to WRs at neighboring subnets. Therefore, WRs are capable of transferring packets via both wired and wireless channels. The WR in one subnet is responsible for providing wireless communication for routers in the same subnet. Figure 1 depicts an example of WNoC with 225 routers divided into nine 5×5 subnets. Solid lines and dotted lines in the figure represent wired and wireless channels, respectively, for transmitting packets between routers. The location of WRs is decided by a placement algorithm, which optimizes WR location by taking traffic distribution patterns into account and will be discussed in the next section. The frequency division multiple access (FDMA) technique is adopted for channelization, so each wireless transmitter and receiver pair uses an independent carrier frequency. This allows simultaneous multiple wireless data transfers between WRs.

We use wormhole packet switching for data delivery in WNoC because it has advantages of both low transfer latency and low buffer requirement, and packets are composed of 64-bit flits. The first flit of a packet is the header flit, which carries control
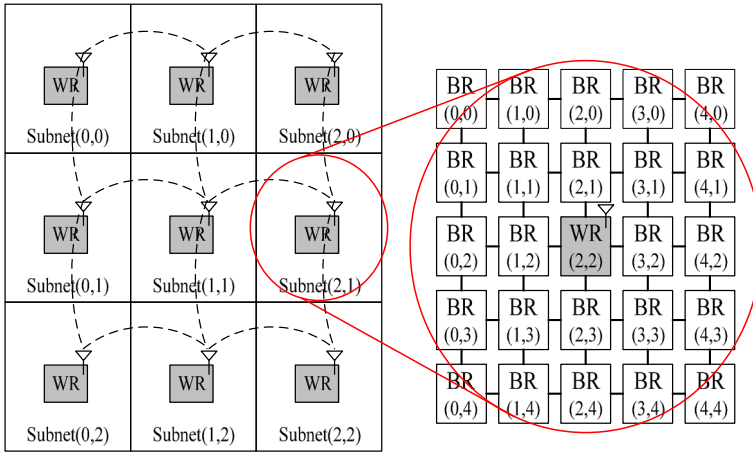
Fig. 1 A 15 × 15 WNoC example

information for packet delivery such as source address, destination address, flit type, payload size, and some control flags. The body flits that follow header flit are the actual payload. Router addresses in WNoC are described by four bit fields: X_{subnet}, Y_{subnet}, X_{local}, and Y_{local}, where the first two fields specify the subnet location and the other two fields identify the router location within a subnet. The separation of subnet and local address fields enables fast routing decision and low hardware complexity at the router.

3.3 WR placement

In WNoC architecture, there is one WR in every subnet which provides express wireless data communication for PEs in the same subnet. It is clear that the location of WRs may affect communication performance. Given a traffic pattern where the amount of data transferred between all source–destination pairs is known beforehand, a WR placement algorithm was devised to minimize communication latency. Because the solution space grows exponentially with network size, we adopted SA metaheuristics [15] to fulfill this purpose. SA is motivated by an analogy to annealing in solids, and the idea of the algorithm is to simulate the cooling of material in a heat bath. SA differs from other hill-climbing optimization algorithms in that, during the optimization process, it allows worse solutions to be chosen some of the times so a local optimum can be avoided. Algorithm 1 shows the WR placement algorithm.

The cost function in the placement algorithm is the average distance between all source and destination pairs, represented in number of hops and calculated by

$$\text{Cost} = \frac{\sum_{s,d \in R} d_{s,d} \times T_{s,d}}{\sum_{s,d \in R} T_{s,d}}, \tag{1}$$

Algorithm 1 *WR placement algorithm*

Input: Network size, subnet size, traffic information
Output: WR locations

f: Cost function
t: Current temperature
α: Temperature reduction function

1: Select an initial solution s_0
2: Set the initial temperature $t_0 > 0$
3: Set the terminating temperature t_f
4: Set the iteration count i_t at each temperature step
5: **while** $t > t_f$ **do**
6:   **for** $i = 1$ **to** i_t **do**
7:     Generate a new solution s
8:     $\delta = f(s) - f(s_0)$
9:     **if** $\delta < 0$ **then**
10:       $s_0 = s$
11:     **else**
12:       Generate a random number x uniformly in the range (0,1)
13:       **if** $x < exp(-\delta/t)$ **then**
14:         $s_0 = s$
15:       **end if**
16:     **end if**
17:   **end for**
18:   $t = \alpha(t)$
19: **end while**
20: Return s_0

where R is the set of routers, $d_{s,d}$ is the distance between s and d (represented in terms of hop count), and $T_{s,d}$ is the total amount of traffic transferred from s to d. At Line 1, the selection of the initial solution is done by choosing the set of center routers of every subnet. The temperature reduction function is calculated by a geometric decrement: $t = t \times 0.95$. The way to generate a new solution at Line 7 is by randomly picking up a WR and exchanging it with one of its neighboring routers. If a new solution results in a lower cost, it is accepted by the algorithm unconditionally. If a new solution increases cost, it is accepted using the probability equation at Line 13. The probability of accepting a worse solution is a function of the current temperature and the change in the cost (δ).

3.4 Routing algorithm

WNoC architecture consists of both wired and wireless networks. When transferring data between PEs in WNoC, a packet may be transmitted by the wired links, the wireless links, or a mix of the two. Thus, we need efficient decision criteria to choose a proper path for transmitting packets. We view WNoC as a hybrid network formed by adding expressways (wireless links) to 2-D mesh NoC, so the path decision problem is whether a packet uses the expressway or not. For packets whose source node and destination node are located in the same subnet, there is no need to use wireless links. For packets having source node and destination node at different subnets, the path

---

Algorithm 2 *Path decision algorithm*

Input: A incoming packet and buffer usage on local WR
Output: Routing path

s: source node
d: destination node
WR_s: source WR
WR_d: destination WR
H_B: traveling distance between s and d without using wireless links
H_W: traveling distance between s and d using wireless links

1: **if** s and d in the same subnet **then**
2:    Route packets from s to d using wired path
3: **else**
4:    **if** $H_W + \Delta < H_B$ **then**
5:       Route packets in three steps:
6:       (a) Route packets from s to WR_s using wired links
7:       (b) Route packets from WR_s to WR_d using wireless links
8:       (c) Route packets from WR_d to d using wired links
9:    **else**
10:       Route packets from s to d using only wired path
11:    **end if**
12: **end if**

---

selection scheme is a function of traveling distance (expressed in terms of hop count) and an adjusting parameter. The path decision algorithm is shown in Algorithm 2.

Path decision algorithm is executed when a packet is injected into the network. Since a WR is shared by all PEs in the same subnet, it is vulnerable to congestion. At Line 4, if we just use ($H_W < H_B$), H_W stands for traveling distance between source and destination using wireless links and H_B stands for traveling distance between source and destination without using wireless links, to make a decision, WRs may be overused and congestion will occur at WRs. Therefore, an adjusting parameter Δ is introduced to balance the utilization of wired and wireless networks. Δ values are dependent on network size and the utilization of wireless links. At each router, there is a table that keeps Δ's corresponding to different network conditions. In general, the larger the network size or the higher the link utilization, the larger the Δ's. This mechanism enables the dynamic adjustment of wireless link usage so that congestion is prevented. In addition, at light traffic load, more long-distance packets can exploit the benefits of wireless express links. Making a decision for the choice of Δ is a heuristic process. Our experimental results demonstrate that larger Δ reduces hot spot possibility, but it also influences utilization of WRs. On the contrary, small Δ easily causes congestion, resulting in transmission performance deterioration. Through exhausting simulations and comparisons, Δ set to six is adopted in 10×10 networks, eight and ten are used in 15×15 and 20×20 networks, respectively, to maintain both the best performance and resource utilization. The Δ is a configurable control factor of WRs and can be fine-tuned to specific traffic patterns and setups in order to get the best performance.

For the routing algorithm, since the wired network of WNoC is a 2-D mesh and the wireless part itself can be viewed as another 2-D mesh network from the subnet view point, XY routing algorithm is used in both networks. XY routing guarantees freedom
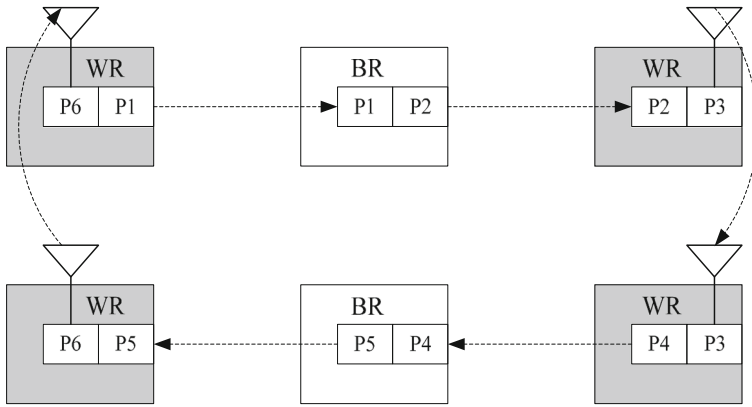
Fig. 2 A deadlock situation in WNoC

of deadlock when packets are traveling within the wired network or the wireless network. However, when packets are allowed to traverse across two networks, there is possibility to cause deadlock. An example of deadlock situation is depicted in Fig. 2. The circular wait condition is established because 90° turns are made from vertical directions to horizontal directions, when packets travel from wireless network to wired network. This situation should not exist in XY routing. To prevent deadlock situations, virtual channels (VCs) are adopted. At each wired input port of both BRs and WRs, two sets of VCs are deployed. One set of VCs is dedicated for traffic moving from source nodes to local WRs (Line 6 of Algorithm 2), called C^{UP}. The other set, referred to as C^{DOWN}, is for traffic traveling from s to d using wired path (Line 2) or traffic traveling from WR_d to d (Line 8). VCs ensure deadlock freedom because they separate the allocation of physical channels from buffers, so physical channels are surrendered to another packet when one packet is blocked. We prove the proposed routing algorithm for WNoC is deadlock-free.

Theorem 1 *The routing algorithm in WNoC is deadlock-free.*

Proof It is shown in [5] that a routing algorithm is deadlock-free if the channels in a direct network can be numbered so that the algorithm always routes the packets along the channels in a strictly decreasing or increasing order. Without loss of generality, we assume WNoC is an $N \times N$ network composed of n $k \times k$ subnets where $N = \sqrt{n} \times k$. Given a router with address $(X_{\text{subnet}}, Y_{\text{subnet}}, X_{\text{local}}, Y_{\text{local}})$, the x and y coordinates are calculated as:

$$x = X_{\text{subnet}} \times k + X_{\text{local}}, \tag{2}$$

$$y = Y_{\text{subnet}} \times k + Y_{\text{local}}. \tag{3}$$

Then we number the output channels of BRs and WRs as depicted in Fig. 3 where $C_{1,i}^{\text{W}}$'s are wireless channels. Note that the routing using only one set of $C_{0,i}^{\text{UP}}$'s, $C_{1,i}^{\text{W}}$'s or $C_{2,i}^{\text{DOWN}}$'s is deadlock-free because XY routing is utilized. For packets which take wired path, they use only $C_{2,i}^{\text{DOWN}}$'s, so there is no deadlock. For packets which take
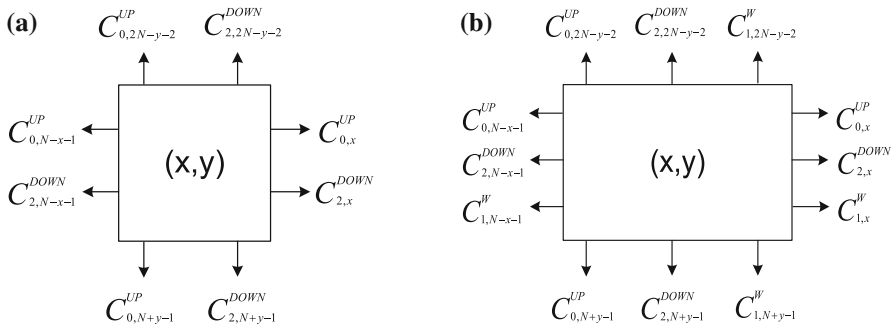
(a)

$C_{0,2N-y-2}^{UP}$ $C_{2,2N-y-2}^{DOWN}$

$C_{0,N-x-1}^{UP}$ ← | (x,y) | → $C_{0,x}^{UP}$

$C_{2,N-x-1}^{DOWN}$ ← | | → $C_{2,x}^{DOWN}$

$C_{0,N+y-1}^{UP}$ $C_{2,N+y-1}^{DOWN}$

(b)

$C_{0,2N-y-2}^{UP}$ $C_{2,2N-y-2}^{DOWN}$ $C_{1,2N-y-2}^{W}$

$C_{0,N-x-1}^{UP}$ ← | (x,y) | → $C_{0,x}^{UP}$

$C_{2,N-x-1}^{DOWN}$ ← | | → $C_{2,x}^{DOWN}$

$C_{1,N-x-1}^{W}$ ← | | → $C_{1,x}^{W}$

$C_{0,N+y-1}^{UP}$ $C_{2,N+y-1}^{DOWN}$ $C_{1,N+y-1}^{W}$

Fig. 3 The numbering of output channels of routers: **a** baseline router, **b** wireless router

wireless path, they traverse through the network only in the strict order of $C_{0,i}^{UP}$'s, $C_{1,i}^{W}$'s and $C_{2,i}^{DOWN}$'s. There is no switching from $C_{2,i}^{DOWN}$'s to $C_{0,i}^{UP}$'s allowed in WNoC. By our channel labeling, $C_{2,i}^{DOWN} > C_{1,i}^{W} > C_{0,i}^{DOWN}$, $\forall\, i$. Therefore, packets are routed along channels in a strictly increasing order, and the routing algorithm is deadlock-free.

□

4 Evaluation

4.1 Experiment setup

A System-C based cycle-accurate simulator was built to analyze the performance of WNoC architecture. The width of wired links is assumed to be 64 bits, which is the size of a flit. Both synthetic traffic and application traffic were applied for performance analysis. For synthetic traffic, the traffic pattern is uniform random, and all packets are 4-flit long. For application traffic, we applied the 3-tuple traffic generation technique proposed in [22]. In [22], it is shown that the spatial and temporal characteristics of the traffic of real-world applications can be captured by three parameters: (1) burstiness, (2) injection distribution and (3) hop distance. Burstiness models how often packet bursts are injected into network and how large these bursts are. This factor reflects the self-similarity of on-chip traffic and is modeled by the Hurst parameter, $0.5 < H \leq 1.0$. H being closer to 1.0 means a higher level of burstiness in traffic. Two levels of burstiness are modeled in simulation: moderate ($H = 0.65$) and high ($H = 0.9$). Injection distribution models how packet injection is distributed among processing nodes. Two modes are considered in this category: evened-out and hot-spot. Evened-out injection means 20 % of the nodes receive 68 % of the total traffic, while hot-spot injection is modeled as 10 % of the nodes receive the same amount of traffic. Hop distance models how far packets travel from source nodes to destination nodes. Therefore, local traffic and long-distance traffic are both considered. In our simulation, local traffic is defined as the traffic where 20 % of the total traffic has a traversal distance greater than four hops. For long-distance traffic, 80 % of the total traffic traverses more than eight hops.

The permutation of three parameters yields eight traffic cases listed in Table 1. Three different network sizes are evaluated: 100 (WNoC_100: 10 × 10), 225 (WNoC_225:

Table 1 3-Tuple traffic categories

	Burstiness	Injection	Hop distance
3tc0	Moderate	Hot-spot	Local
3tc1	High	Hot-spot	Local
3tc2	Moderate	Evened-out	Local
3tc3	High	Evened-out	Local
3tc4	Moderate	Hot-spot	Long-distance
3tc5	High	Hot-spot	Long-distance
3tc6	Moderate	Evened-out	Long-distance
3tc7	High	Evened-out	Long-distance

15 × 15), and 400 (WNoC_400: 20 × 20). The size of subnets is defined as 5 × 5, so there are 4, 9, 16 subnets for each network size evaluated. WNoC networks are compared with their wired counterparts with the same network sizes. We use a standard interconnection network measurement setup where packets are stored in an infinite queue before they are injected. This mechanism isolates the packet generation from the network behavior. Each simulation has a 10,000-cycle warm-up phase and continues for another 100,000 cycles during which performance and power consumption measurements are conducted.

4.2 Performance evaluation

WNoC architecture is compared with its wired counterpart, which is referred to as NoC in our experiment. We use Algorithm 1 to find the locations for WRs.

Figure 4 shows the execution results for 3tc4 traffic in WNoC_400. At the beginning of execution, WRs are set to the centers of every subnet. As can be seen from the figure, average hop gradually decreases with iteration count and finally converges.

The comparison of average inter-node distance for 100, 225 and 400 nodes is shown in Figs. 5, 6 and 7, respectively. We can see that our routing algorithm makes efficient use of wireless links so that the inter-node distance is reduced in all traffic patterns. For random traffic, the reduction in hop count in WNoC_100, WNoC_225 and WNoC_400 is 25, 39 and 47 %, respectively. It is clear that using wireless express links effectively reduces the traveling distance between source and destination. Networks of greater dimension benefit more from wireless links.

For 3-tuple traffic, the improvement in traveling distance is mainly affected by the hop distance traffic parameter. Local traffic (3tc0–3tc3) results in about 8, 16 and 30 % of reduction in hop count for WNoC_100, WNoC_225 and WNoC_400, respectively, whereas, long-distance traffic (3tc4–3tc7) leads to a reduction of about 27, 32 and 48 %, respectively. The results imply that WNoC architecture is more beneficial for long-distance traffic and large networks.

To understand the benefits gained from the proposed placement algorithm, we also compare the average hop count from the configuration where WRs are placed at the center of subnets. It is demonstrated that the placement algorithm could lead to a lower hop count by 2.7–11.3 % for 3-tuple traffic. For uniform random traffic, a slight improvement of about 0.5 % is observed.
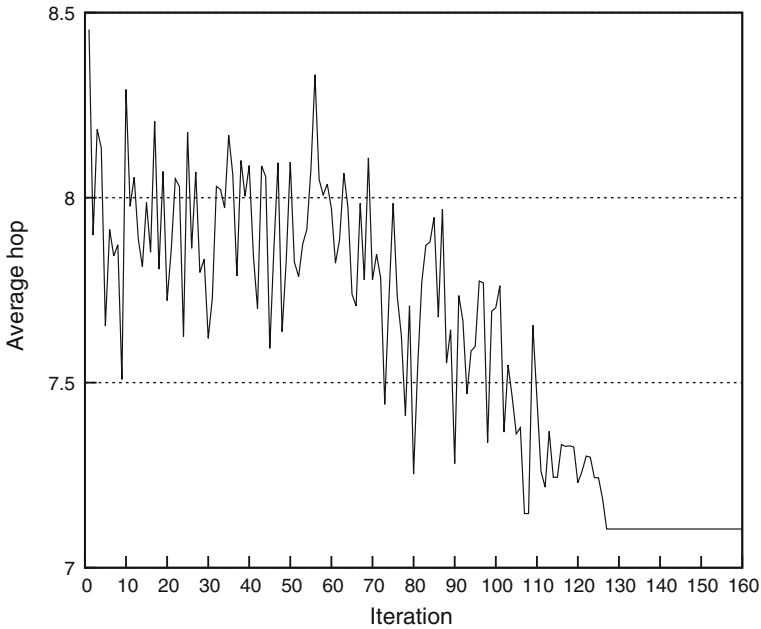
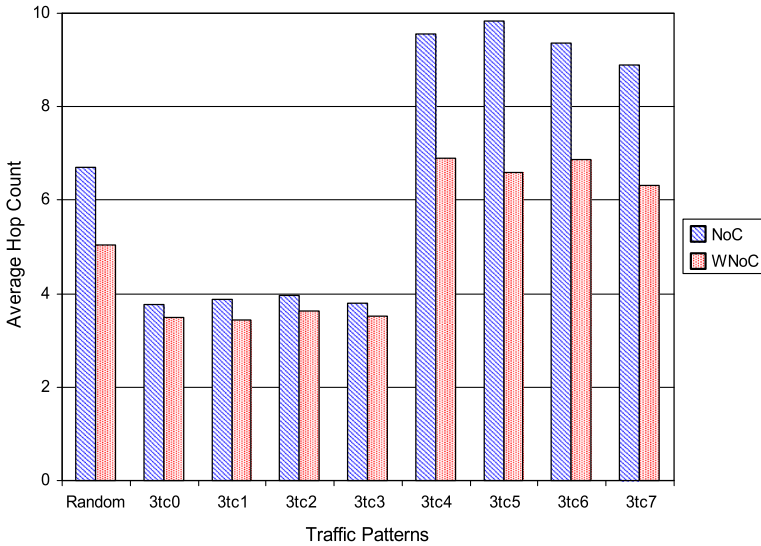**Fig. 4** Execution results of WR placement algorithm

Fig. 5 Comparison of hop count in 10×10 WNoC

Figure 8 shows the comparison of average transfer latency normalized to the latency of NoC architecture. WNoC_100 is compared with NoC_100; WNoC_225 is compared with NoC_225; and WNoC_400 is compared with NoC_400. To simplify the plot, NoC_100, NoC_225 and NoC_400 are normalized as 1 and labeled as NoC in
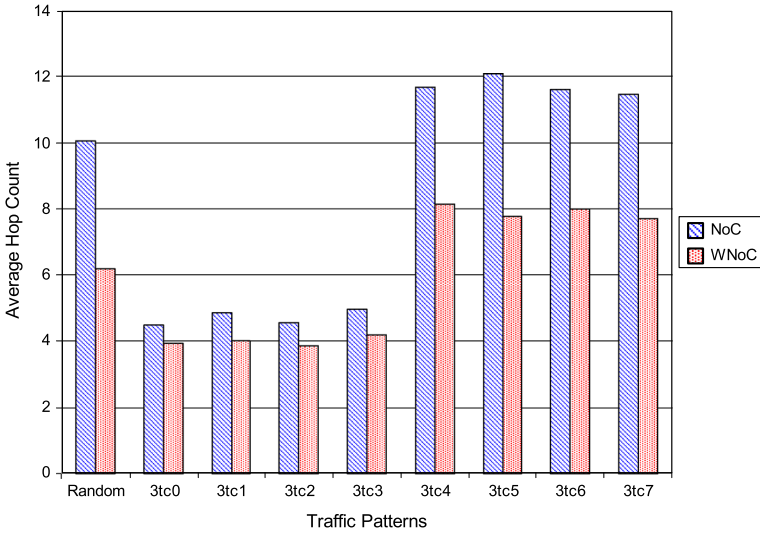
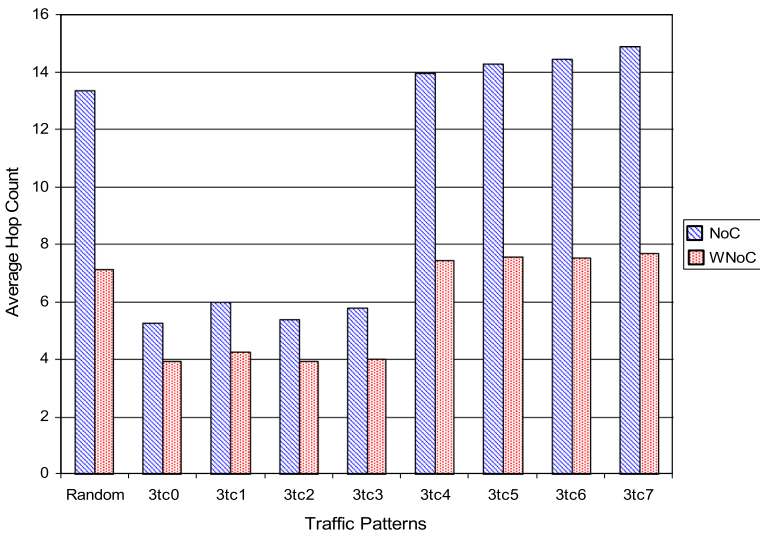**Fig. 6** Comparison of hop count in 15×15 WNoC

Fig. 7 Comparison of hop count in 20×20 WNoC

the plot. Since WNoC_100, WNoC_225 and WNoC_400 have better latency values, we observe latencies that are fractions of the original normalized NoC. For random traffic, the improvement in 100, 225 and 400 nodes is 13, 17 and 18 %, respectively. For 3-tuple traffic, long-distance traffic tends to have more decrease in latency than local traffic, which demonstrates the advantages of employing wireless links. The experimental results also show that WNoC architecture has higher saturation load than NoC. The improvement ranges from 2.1 to 7.6 % in all the test cases, not as much
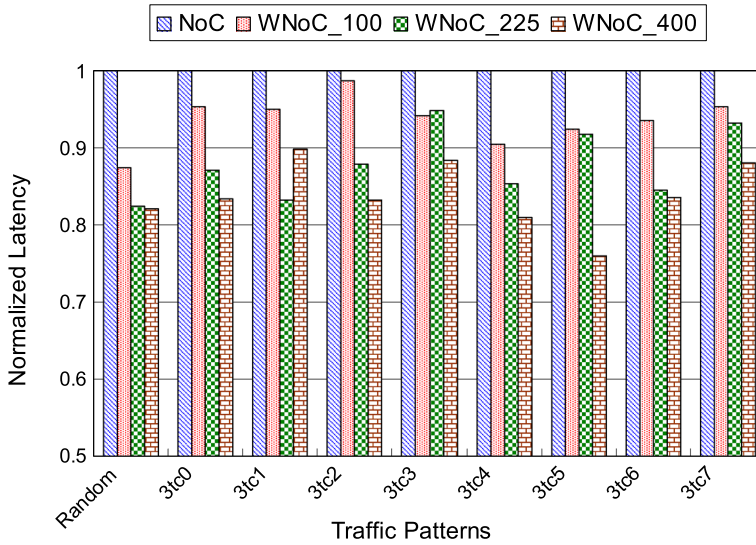
Fig. 8 Comparison of average latency in baseline NoC and WNoC networks

as expected. The major reason is that congestion is observed around WRs because only the utilization of wireless links is considered in routing. Taking the congestion information of wired links into account and devising a congestion-aware adaptive routing algorithm will be our future work.

In Algorithm 2, $H_W + \Delta < H_B$ is used as a criterion to determine whether packets utilize wireless network or not. The less the Δ value is, the more packets are transmitted through wireless routers. Reduced transmission latency and power consumption are expected because long-distance packets utilize WRs shortcuts to improve transmission efficiency. However, hot spots generated by the express links result in poor transmission performance and reduction in tolerated throughput. Performance optimization is conducted by balancing traffic between hybrid networks. Figure 9 shows throughput comparison among different Δ setups under the random traffic profile. WNoC_D0 denotes parameter $\Delta = 0$. Small Δ easily results in congestion situations as traffic load increases. Congestion deteriorates system performance even when congestion management is activated. Large Δ might lower WR utilization so as to sacrifice overall network performance. It is observed that the Δ value ranging between 5 and 6 maintains the best performance while sustaining the most throughput in the WNoC architecture.

4.3 Feasibility evaluation

WNoC feasibility is considered by estimating hardware implementation cost and evaluating system power consumption. The area and power models of routers and links from Orion 2.0 [13] are used to calculate WNoC router area, router and link power consumption. The parameters used for analysis are listed in Table 2.
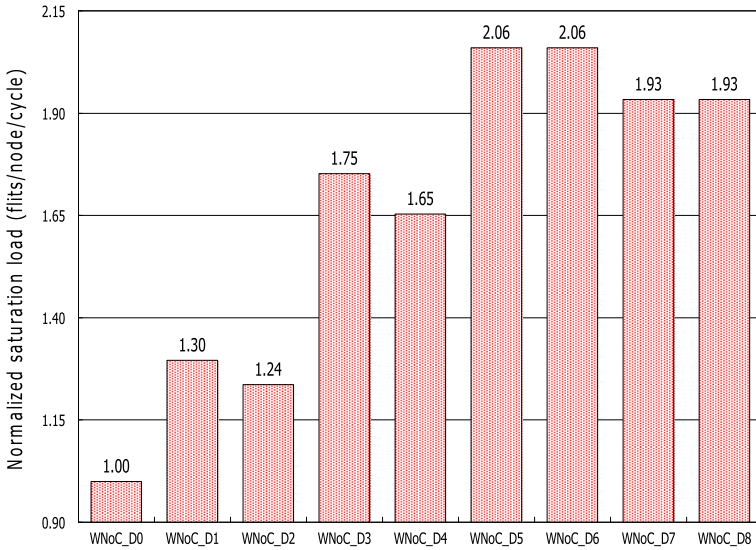
Fig. 9 Comparison of normalized saturation load for various delta settings in 10×10 WNoC

Table 2 Simulation parameters

Parameter	Setting
Technology	65 nm
V_{dd}	1.0 V
$V_{threshold}$	0.7 V
Frequency	1 GHz
Transistor type	NVT
Number of cores on chip	100, 225, 400 cores
Switching technique	Wormhole
Baseline mesh routing	XY routing
Number of virtual channels	2 VCs
Number of wireless routers	4, 9, 16 WRs
Flit width	64 bits

4.3.1 Implementation cost evaluation

The area of WNoC routers is collected from different router configurations. The BRs in WNoC is a five-port router with two virtual channels at each input port and are estimated to occupy an area of 0.108 mm^2. The area of WRs is estimated to be 0.175 mm^2 which consists of a hybrid six-port router and a wireless base station, which is estimated to be 18,332 μm^2 in 65 nm technology, obtained by scaling a 130 nm design from [24]. Note that only one WR is deployed in each subnet although it occupies more area than BRs. When comparing the area of the proposed WRs and widely used processing elements such as ARM11MPCore and PowerPC 405 processor, which occupy 0.938

and 1.4 mm^2 in 65 nm technology, respectively [1, 12], the area overhead imposed by the WNoC routers would be within the margin of design, demonstrating the area cost feasibility.

4.3.2 Power consumption evaluation

Power consumption of WNoC is mainly attributed to router power, wired links power and wireless links power [23], listed as:

$$P = P_{\text{ROUTER}} + P_{\text{BR}_{\text{LINK}}} + P_{\text{WR}_{\text{LINK}}} \qquad (4)$$

where P_{ROUTER} is associated with the number of router ports and traffic workload. $P_{\text{BR}_{\text{LINK}}}$ is associated with port numbers of routers, link length and transmission probability. $P_{\text{WR}_{\text{LINK}}}$ is associated with transceiver power and transmission distance.

P_{ROUTER} is calculated by different configurations for BRs and WRs and traffic workload among them. By adding wireless routing resources, transmission load between wired networks is reduced to accelerate data transfer and eliminate congestion situations. $P_{\text{BR}_{\text{LINK}}}$ link model is listed as:

$$P_{\text{BR}_{\text{LINK}}} = \alpha \cdot C_1 \cdot V_{\text{dd}}^2 \cdot f_{\text{clk}} \qquad (5)$$

where α, C_1, V_{dd} and f_{clk} denote the activity factor, load capacitance, supply voltage, and frequency, respectively. Using 65 nm ARM11MPCore which is 0.938 mm^2 for an embedded design of WNoC platform, the link between processors is assumed to be 1,050 µm. Because of transfer collaboration of hybrid network, we observed that WNoC consumes less link power because wireless links lower traveling hop counts so as to decrease wired link transition during the communication tasks. For wireless links power consumption, it is evaluated in an analytical way. In the scenario for different WNoC platforms, we assume that each subnetwork embraces 24 BRs with one WR, and each core occupies an area of 1 mm^2, which includes a processor and a router. The largest wireless transmission distance is estimated to vary from 7.1 to 14.1 mm diagonally. Therefore, an on-chip wireless antenna with transmitter power −10 dBM can provide reliable transmission performance and energy consumption is estimated to be 4.5 pJ/bit [16].

System power consumption is collected based on introduced power model and performed under various network scenarios. Figure 10 depicts power consumption comparison between baseline NoC and WNoC networks. Three different network size scenarios, including 100, 225, 400 cores (WNoC_100, WNoC_225, WNoC_400) and nine traffic patterns including random, 3tc0 to 3tc7 are compared. The results show WNoC has better power efficiency than its counterpart. WNoC architecture is superior because it utilizes long-distance wireless links to reduce latency. Packets go through less hops and access fewer FIFOs in WNoCs so that the overall power consumption decreases accordingly. Different network sizes result in the same conclusion where the improvement is up to 15 %. The saving is significant especially in traffic patterns with more long-distance transmission tasks such as 3tc4–3tc7 and in larger network sizes which manifest WNoC's advantages. The cases of 3tc0–3tc3 demonstrate moderate
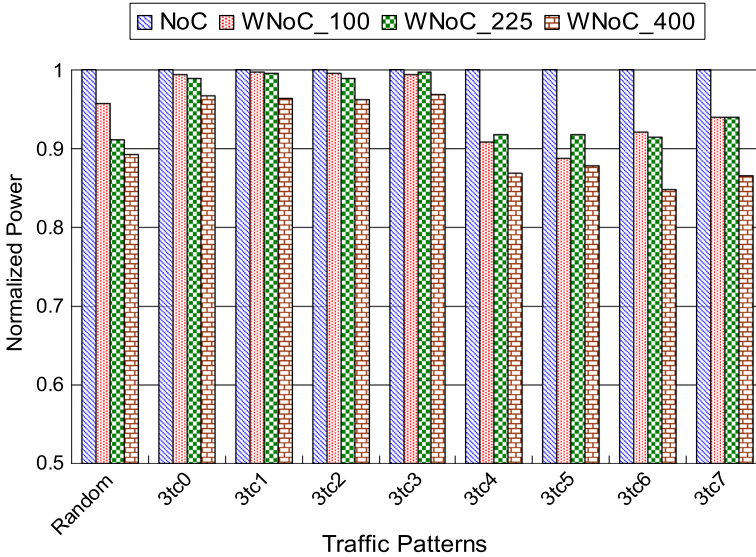
Fig. 10 Comparison of average power consumption in baseline NoC and WNoC networks

improvement which is attributed to local traffic behaviors. By providing better power efficiency and enhanced transmission latency, adopting wireless links improves performance and energy consumption of tasks themselves. On the other hand, a shorter active time allows the system to enter the power-saving mode earlier to save energy for the rest of data transmission time.

5 Conclusion

We developed a novel WNoC architecture which is composed of a hybrid wired and wireless network. An intelligent routing algorithm was devised to balance the utilization of two networks so that network congestion is avoided. We employed virtual channels in WNoC architecture to ensure deadlock-free routing. It is also shown that applying SA optimization techniques can effectively find optimal locations for wireless routers so that inter-node distance is minimized. By applying long-distance wireless links to bypass wired multi-hop communications, simulation results demonstrated that communication latency and overall network throughput were improved. Evaluation of hardware cost and power consumption also shows the feasibility of WNoC architecture.

References

1. ARM. ARM11 MPCore. http://www.arm.com. Accessed Sept 2007
2. Benini L, De Micheli G (2002) Networks on chip: a new paradigm for systems on chip design. In: Proceedings of design, automation and test in Europe conference and exhibition. pp 418–419

3. Chang M, Cong J, Kaplan A, Naik M, Reinman G, Socher E, Tam S-W (2008) CMP network-on-chip overlaid with multi-band RF-interconnect. In: High performance computer architecture, 2008. IEEE 14th International Symposium on HPCA 2008. pp 191–202

4. Chang MCF, Socher E, Tam S-W, Cong J, Reinman G (2008) RF interconnects for communications on-chip. In: Proceedings of the 2008 international symposium on physical design, ISPD '08. ACM, New York, pp 78–83

5. Dally W, Seitz C (1987) Deadlock-free message routing in multiprocessor interconnection networks. IEEE Trans Comput C-36(5):547–553

6. Deb S, Chang K, Yu X, Sah SP, Cosic M, Ganguly A, Pande PP, Belzer B, Heo DH (2013) Design of an energy-efficient CMOS-compatible NoC architecture with millimeter-wave wireless interconnects. IEEE Trans Comput 62(12):2382–2396

7. Deb S, Ganguly A, Pande PP, Belzer B, Heo DH (2012) Wireless NoC as interconnection backbone for multicore chips: promises and challenges. IEEE J Emerg Sel Topics Circuits Syst 2(2):228–239

8. Ganguly A, Chang K, Deb S, Pande P, Belzer B, Teuscher C (2010) Scalable hybrid wireless network-on-chip architectures for multi-core systems. IEEE Trans Comput PP(99):1

9. Ganguly A, Chang K, Pande P, Belzer B, Nojeh A (2009) Performance evaluation of wireless networks on chip architectures. In: Quality of electronic design, 2009. ISQED 2009. pp 350–355

10. Green W, Rooks M, Sekaric L, Vlasov Y (2007) Ultra-compact, low RF power, 10 Gb/s silicon Mach–Zehnder modulator. Optics Express 15(25):17106–17113

11. Huang D, LaRocca T, Chang M-C, Samoska L, Fung A, Campbell R, Andrews M (2008) Terahertz CMOS frequency generator using linear superposition technique. IEEE J Solid-State Circuits 43(12):2730–2738

12. IBM. IBM PowerPC405 Embedded Core. http://www.ibm.com. Accessed Sept 2007

13. Kahng A, Li B, Peh L-S, Samadi K (2009) ORION 2.0: a fast and accurate NoC power and area model for early-stage design space exploration. In: Design, automation test in Europe conferenceexhibition, 2009. DATE '09. pp 423–428

14. Kim K, Yoon H, Kenneth KO (2000) On-chip wireless interconnection with integrated antennas. In: International electron devices meeting, 2000. IEDM Technical Digest. pp 485–488

15. Kirkpatrick S, Gelatt C Jr, Vecchi M, McCoy A (1983) Optimization by simulated anneling. Science 220(4598):671–679

16. Lee S, Tam S, Pefkianakis I, Lu S, Chang M, Guo C, Reinman G, Peng C, Naik M, Zhang L, et al (2009) A scalable micro wireless interconnect structure for CMPs. In: Proceedings of the 15th annualinternational conference on Mobile computing and networking. pp 217–228. ACM, New york

17. Kenneth KO, Kim K, Floyd B, Mehta J, Yoon H, Hung C-M, Bravo D, Dickson T, Guo X, Li R, Trichy N, Caserta J, Bomstad WR I, Branch J, Yang D-J, Bohorquez J, Seok E, Gao L, Sugavanam A, Lin J-J, Chen J, Brewer J (2005) On-chip antennas in silicon ICs and their application. IEEE Trans Electron Devices 52(7):1312–1323

18. Pande P, Ganguly A, Chang K, Teuscher C (2009) Hybrid wireless network on chip: a new paradigm in multi-core design. In: 2nd international workshop on network on chip architectures, 2009. NoCArc 2009. pp 71–76

19. Pavlidis V, Friedman E (2006) 3-D topologies for networks-on-chip. In: 2006 IEEE international SOC conference. pp 285–288

20. Seok E, Cao C, Shim D, Arenas D, Tanner D, Hung C-M, Kenneth KO (2008) A 410 GHz CMOS push-push oscillator with an on-chip patch antenna. In: IEEE international solid-state circuitsconference, 2008. ISSCC 2008. Digest of Technical Papers. pp 472–629

21. Shacham A, Bergman K, Carloni L (2008) Photonic networks-on-chip for future generations of chip multiprocessors. IEEE Trans Comput 57(9):1246–1260

22. Soteriou V, Eisley N, Wang H, Li B, Peh L-S (2006) Polaris: a system-level roadmap for on-chip interconnection networks. In: International conference on computer design, 2006. ICCD 2006. pp 134–141

23. Wang C, Hu W-H, Bagherzadeh N (2011) A wireless network-on-chip design for multicore platforms. In: Proceedings of 19th Euromicro international parallel, distributed and network-basedprocessing (PDP) conference. pp 409–416

24. Zhao D, Wang Y (2008) SD-MAC: design and synthesis of a hardware-efficient collision-free QoS-aware MAC protocol for wireless network-on-chip. IEEE Trans Comput 57(9):1230–1245