

# High Throughput and Low Power NoC

Magdy El-Moursy<sup>1</sup>, *Member IEEE* and Mohamed Abdelgany<sup>2</sup>

<sup>1</sup> Mentor Graphics Corporation  
Cairo, Egypt

<sup>2</sup> Electronics Research Institute  
Cairo, Egypt  
German University in Cairo, Cairo, Egypt

## Abstract

The High throughput architecture to achieve high performance Networks-on-Chip (NoC) is proposed. The throughput is increased by more than 38% while preserving the average latency. The area of the network switch is decreased by 18%. The required metal resources for the proposed architecture are increased by less than 10% as compared to the required metal resources for the conventional NoC architecture. Power characteristics of different high throughput NoC architectures are developed. The extra power dissipation of the proposed high throughput NoC is as low as 1% of the total power dissipation. Among different NoC topologies, High Throughput Butter Fat Tree (HTBFT) requires the minimum extra power dissipation and metal resources.

**Keywords:** *Network-on-Chip, Throughput, Power Dissipation, Topology.*

## 1. Introduction

As the number and functionality of intellectual property blocks (IPs) in System on Chips (SoCs) increase, complexity of interconnection architectures of the SoCs have also been increased. Different research articles have been published in high performance SoCs. However, the system scalability and bandwidth are limited. As described in [1]-[5], NoCs are emerging as the best replacement for the existing interconnection architectures. Many NoC topologies have been proposed in the past, e.g., CLICHÉ [1], SPIN [2], Octagon [3] and Butterfly Fat Tree [4]. Different research articles in architectural and conceptual aspects of NoC such as, topology selection, quality of service (QoS) [5], design automation [6], performance evaluation [7], and verification have been reported. NoCs provide different set of constraints in the design paradigm. High throughput and low latency are the desirable characteristics of a multi processing system.

Previous articles have taken a top down approach (a high level analysis of NoC) and they did not touch the issues on a circuit level. However, little research has been reported

on the circuit design issues [8]. Although they were implemented and verified on silicon, they were only focusing on implementing limited set of topologies. In large scale NoCs, power dissipation should be minimized for cost efficient implementation. Many papers have been published in NoCs. They were only focusing on performance and scalability issues rather than power efficiency. Scaling with power reduction is the trend in future technologies. Application specific techniques are required to reduce power dissipation of NoCs.

The main focus of this paper is to present a high throughput interconnect architecture for network on chip. The circuit implementation issues are considered in the proposed architecture. The switch structure along with the interconnect architecture are shown in Figure 1 for 2 IPs and 2 switches. The proposed architecture is applied to different NoCs topologies. Low power switch is also proposed to achieve power-efficient NoC. The efficiency and performance are evaluated.

To the best of our knowledge, this is the first in depth analysis on circuit level to optimize performance of different NoC topologies. The paper is organized as follows: In Section 2, the proposed port architecture is presented. The new High Throughput architecture is described in section 3. In Section 4, closed form expressions for the power dissipation in different high throughput architectures are developed. The performance improvement and circuit overhead of the proposed architecture are provided in Section 5. Finally, conclusions are summarized in section 6.

## 2. Port Architecture

Each port of the switch includes input virtual channels, output virtual channels, header decoder, controller, input arbiter and output arbiter as shown in [4]. The input arbiter consists of a priority matrix and grant circuit. The priority matrix stores the priorities of the requests. A dedicated circuit generates the grant signals to allow only one virtual

channel to access a physical port. The messages are divided into fixed length flow control units (flits). When the granted virtual channel stores one whole flit, it sends a full signal to controller. If it is a header flit, the header decoder determines the destination. The controller checks the status of destination port. If it is available, the path between input and output is established. The flits from more than one input port may simultaneously try to access a particular output port. The output arbiter is used to allow only one input port to access an output port. Virtual channels consist of several buffers controlled by a multiplexer and an arbiter which grants access for only one virtual channel at a time according to the request priority. Once the request succeeds, its priority is set to be the lowest among all other requests.

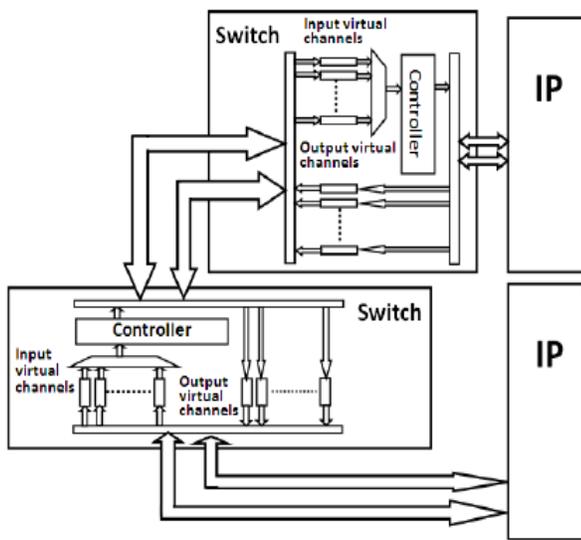


Fig. 1: Proposed high throughput architecture.

In the proposed architecture, rather than using one multiplexer and one arbiter to control the virtual channels, two multiplexers and two arbiters are employed as shown in Figure 2. Using the proposed technique, the virtual channels are divided into two groups; each group is controlled by one multiplexer and one arbiter. Each group of virtual channels is supported by one interconnect bus as described in section 3. However trivial it may look, the proposed port architecture has a great influence on the switch frequency and the throughput of the network. Let us consider an example with 8 virtual channels. In the NoC architecture, 8x8 input arbiter and 8x1 multiplexer are needed to control the input virtual channels as shown in Figure 2 (a). The 8x8 input arbiter consists of 8x8 grant circuit and 8x8 priority matrix. In the proposed architecture, two 4x4 input arbiters, two 4x1 multiplexers, 2x1 multiplexers and 2x2 grant circuit are integrated to

allow only one virtual channel to access a physical port as shown in Figure 2 (b). The 4x4 input arbiter consists of 4x4 grant circuit and 4x4 priority matrix. The values of the grant signals are determined by the priority matrix. The number of grant signals equals the number of requests and the number of selection signals of the multiplexer. The area of two 4x4 input arbiters is smaller than the area of 8x8 input arbiter. Also, the area of two 4x1 multiplexers is smaller than the area of 8x1 multiplexer. Consequently, the required area to implement the proposed switch with the proposed architecture is less than the required area to implement the conventional switch.

In order to divide a 4x1 multiplexer into three 2x1 multiplexers, the 4x4 input arbiter should be divided into three 2x2 input arbiters. The grant signals which are generated by three 2x2 input arbiter (6 signals) are not the same grant signals generated by the 4x4 input arbiter (4 signals). Therefore, the 4x4 input arbiter can not be replaced by three 2x2 input arbiters unless the number of interconnect buses is increased to be equal to the number of virtual channels groups. Therefore, the proposed architecture in Figure 2 (b) is the optimum to allow eight virtual channels in the port. By increasing the number of interconnects, the metal resources and power dissipation are increased as described in Section 6.

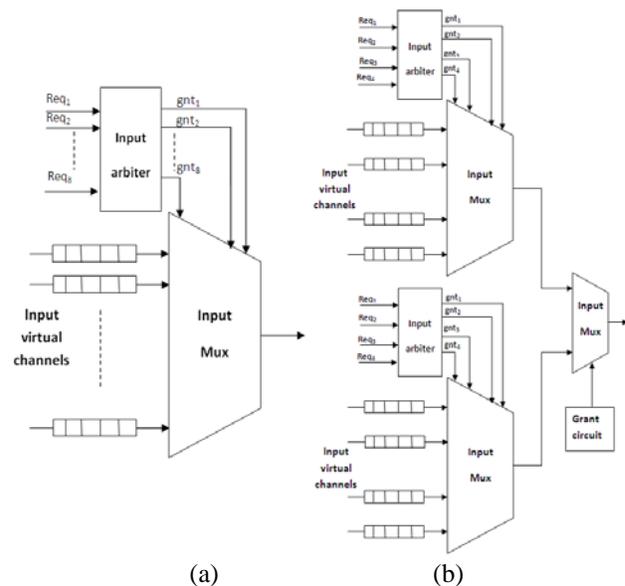


Fig. 2 (a) Circuit diagram of switch port, (b) circuit diagram of High Throughput switch port.

Without circuit optimization, the change in the maximum frequency of the switch with the number of virtual channels in the conventional BFT switch is shown in Figure 3. When the number of virtual channels is increased beyond four, the maximum frequency of the switch is decreased for BFT architecture. Throughput is a parameter

that measures the rate in which message traffic can be sent across a communication network. It is defined by [7]:

$$TP = \frac{(\text{number of messages completed}) * (\text{message length})}{(\text{number of IP blocks}) * (\text{total time})} \quad (1)$$

The throughput is proportional to the number of completed messages. The number of completed messages increases with the number of virtual channels. Total transfer time of messages decreases with the increase in frequency of the switch. Therefore the throughput can be improved by increasing the number of virtual channels or by increasing the operating frequency of the switch. The throughput is saturated when the number of virtual channels is increased beyond four [7]. On the other hand, the average message latency increases with the number of virtual channels. To keep the latency low while preserving the throughput, only four virtual channels are used in [7].

The proposed High Throughput BFT (HTBFT) switch is smaller than the BFT switch. Therefore, the maximum frequency of the switch is improved. The change in the maximum frequency of the proposed switch with the number of virtual channels is shown in Figure 3 for HTBFT architecture. With the proposed switch architecture, the number of virtual channels could be increased up to eight without significant reduction in the operating frequency. The frequency of the network switch is characterized for different network topologies using the proposed architecture as shown in Figure 4. As compared to the conventional architecture, the operating frequency of the proposed architectures is decreased when the number of virtual channels is higher than eight rather than four in the conventional architecture. Doubling the number of virtual channels does not degrade the frequency of the switch (rather than 4 virtual channels, 8 virtual channels could be used in the proposed architecture). However, a severe increase in the number of virtual channels (more than 8) could degrade performance.

Increasing the number of virtual channels would increase the traffic going through the links (interconnects) between the switches, increasing the contention on the bus and increasing the latency that each flit experiences. In order to improve throughput, the links (interconnects) connecting the switches with each other should be increased. Since the number of virtual channels could be doubled (from four in the conventional architecture to eight in the proposed architecture), doubling the number of virtual channels between switches is proposed.

Let us consider an example of BFT architecture. The HTBFT architecture decreases the area of switch by 18%. Consequently, a system with eight virtual channels achieves high throughput, high frequency and low latency while the area of the design is optimized. The architecture of different NoC topologies to achieve high throughput network is described in section 3.

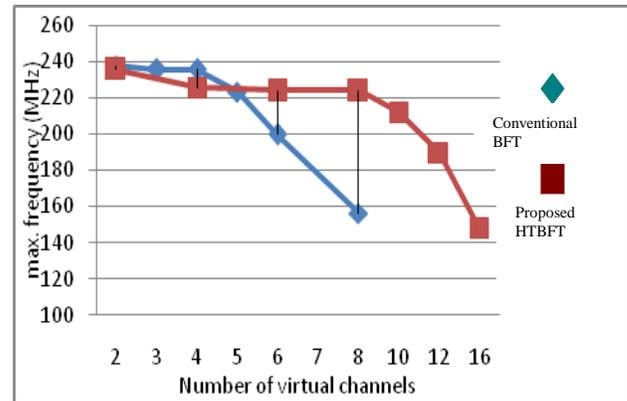


Fig. 3 Maximum frequency of a switch with different number of virtual channels for conventional BFT and proposed HTBFT.

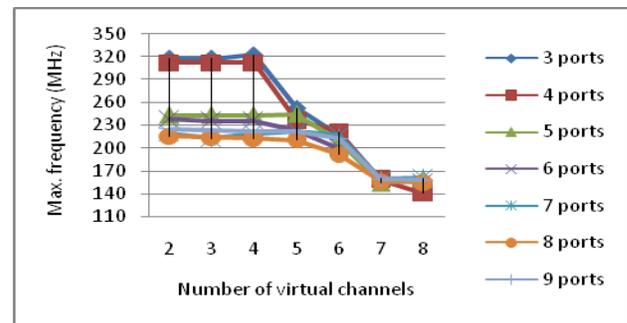


Fig. 4 Maximum frequency of a switch with different number of virtual channels for different NoC topologies of the proposed architecture.

### 3. High Throughput Architecture

To A novel interconnect template to integrate IP blocks in NoC is proposed. In the proposed architecture, rather than using a single interconnect bus between each two elements of NoC (IP block and switch or two switches), two buses are employed. The number of virtual channels can be doubled to get higher throughput. To maintain the average latency, each bus supports half the number of virtual channels. Increasing the number of buses between two switches could improve the throughput by optimizing the design of the switch on the circuit level as shown in Section II. However, using two buses to connect two switches implies using more metal resources and may be silicon area for the repeaters within the long interconnects. The overhead of the proposed architecture is discussed in Section 5.

A novel interconnect template to integrate IP blocks using High Throughput Butter Fly Fat Tree (HTBFT) architecture is proposed. Each group of 4 IPs (no. 0, no. 1, no.2 and no.3) in Figure 5 needs one switch (no.4). Each switch in the first level (no. 4) connects to each switch in the second level (no. 5) by 2 buses. Each bus supports half

the number of virtual channels. Therefore, the throughput can be improved while preserving the average latency.

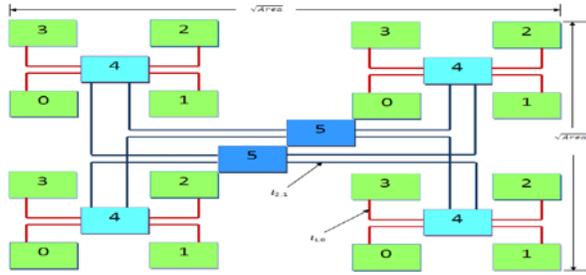


Fig. 5 Interconnect architecture of HTBFT

The interconnect template to integrate IP blocks using High Throughput architecture is implemented to CLICHÉ (to become High Throughput CLICHÉ, HTCLICHÉ), Octagon (to become High Throughput Octagon, HT-Octagon), SPIN (to become High Throughput SPIN, HTSPIN) architectures, in which double the number of interconnects is needed. The throughput improvement is presented in section V for each topology.

Power estimation is very important aspect of NoC design. The average power dissipation of NoC port is obtained. The switch is implemented on the transistor level using ASIC design flow. For different NoC topologies, the average power dissipation of the switch is determined. Closed form expressions are developed for each topology in section 4.

#### 4. Power Characteristics

To Communication network on chip contains three primary components; network switch, interswitch links (interconnects), and repeaters within interswitch links. Including different sources of power dissipation in NoC, the total power dissipation of on chip network is defined as follows:

$$P_{total} = P_{switches} + P_{interconnect} + P_{reps}, \quad (2)$$

$$P_{switches} = P_{switching} + P_{leakage}, \quad (3)$$

where  $P_{total}$  is the total power dissipation of the network.  $P_{switches}$  is the power dissipation in the switches.  $P_{interconnect}$  is the total power dissipation of interswitch links.  $P_{reps}$  is the total power dissipation of the repeaters which are required for long interconnects.  $P_{switching}$  and  $P_{leakage}$  are the switching and leakage power of the switch, respectively. The number of repeaters depends on the length of the interswitch link. According to the topology of NoC interconnects, the interswitch wire lengths, the number of repeaters and the number of switches can be determined a priori.

$$P_{interconnect} = cV_{dd}^2 f, \quad (4)$$

$$P_{reps} = P_{reps-dyn} + P_{reps-SC} + P_{reps-leakage}, \quad (5)$$

$$P_{reps-dyn} = N_{rep} H_{opt} C_0 V_{dd}^2 f, \quad (6)$$

where  $P_{reps-dyn}$  is the total dynamic power dissipation of repeaters,  $N_{rep}$  is the number of repeaters,  $H_{opt}$  is the optimum repeater size,  $C_0$  is the input capacitance of a minimum size repeater,  $V_{dd}$  is the supply voltage and  $f$  is the switching frequency.  $P_{reps-SC}$  is the total short-circuit power of the repeaters.  $P_{reps-leakage}$  is the total leakage power dissipation of the repeaters.  $c$  is the interswitch link capacitance. Closed form expressions for the power dissipation of different high throughput NoC architectures are described in the following subsections.

##### 4.1 High Throughput Butterfly Fat Tree

In the HTBFT, the interconnection is performed on levels of switching. The number of switching levels can be expressed as  $\log_2 N - 3$ , where  $N$  is the number of IP blocks. The total number of switches in the first level is  $N/4$ . At each subsequent level, the number of required switches reduces by a factor of 2 as shown in Figure 5. The interswitch wire length and total number of switches are given by the following expressions:

$$l_{a+1,a} = \frac{\sqrt{Area}}{2^{levels-a}}, \quad (7)$$

$$N_{switches-HTBFT} = \frac{N}{4} \left( \frac{1 - (1/2)^{levels}}{1 - 1/2} \right), \quad (8)$$

where  $l_{a+1,a}$  is the length of the wire spanning the distance between level  $a$  and level  $a + 1$ , where  $a$  can take integer value between 0 and  $(levels - 1)$ . In the HTBFT, the total length of interconnects and the total number of repeaters can be determined from the following equations:

$$l_{tot-HTBFT} = \frac{\sqrt{Area}}{2^{(\log_2 N - 3)}} NX (levels) X 2N_{wires}, \quad (9)$$

$$N_{rep-HTBFT} = 2NN_{wires} \left( \left[ \frac{l_{1,0}}{K_{opt}} \right] + \dots + \frac{1}{2^{N-1}} \left[ \frac{l_{lev,lev-1}}{K_{opt}} \right] \right), \quad (10)$$

where  $K_{opt}$  is the optimum length of the global interconnect [9]. Using the number of switches, the total length of interconnects and the total number of repeaters, the total power dissipation of HTBFT architecture ( $P_{tot-HTBFT}$ ) is determined.

$$P_{tot-HTBFT} = 3 \frac{N}{2} \left( \frac{1 - (1/2)^{levels}}{1 - 1/2} \right) P_{port} + \frac{\sqrt{Area}}{2^{(\log_2 N - 3)}} N X (\log_2 N - 3) X 2N_{wires} cV_{dd}^2 f + 2NN_{wires} \left( \left\lfloor \frac{l_{1,0}}{K_{opt}} \right\rfloor + \dots + \frac{1}{2^{N-1}} \left\lfloor \frac{l_{lev,lev-1}}{K_{opt}} \right\rfloor \right) H_{opt} C_0 V_{dd}^2 f. \quad (11)$$

#### 4.2 High Throughput CLICHÉ

In HTCLICHÉ, the number of switches equals the number of IPs. The interswitch wire lengths can be determined from the following expression:

$$l_{HTCLICHE} = \frac{\sqrt{Area}}{\sqrt{N}}, \quad (12)$$

The number of horizontal interswitch wires between switches equals  $2\sqrt{N}(\sqrt{N} - 1)$ . According to the technology node, the optimum length of global interconnects can be obtained. Therefore, the total length of interconnects and the number of repeaters can be calculated by:

$$l_{tot-HTCLICHE} = 4\sqrt{Area}(\sqrt{N} - 1)N_{wires}, \quad (13)$$

$$N_{rep-TCLICHE} = 4 \left\lfloor \frac{\sqrt{Area}}{\sqrt{N}K_{opt}} \right\rfloor \sqrt{N}(\sqrt{N} - 1)N_{wires}, \quad (14)$$

Using the number of ports, number of switches, total length of interconnects and number of repeaters, the total power dissipation of the HTCLICHÉ architecture can be determined.

$$P_{tot-HTCLICHE} = 5NP_{port} + 4 \left\lfloor \frac{\sqrt{Area}}{\sqrt{N}K_{opt}} \right\rfloor \sqrt{N}(\sqrt{N} - 1)N_{wires} H_{opt} C_0 V_{dd}^2 f + 4\sqrt{Area}(\sqrt{N} - 1)N_{wires} cV_{dd}^2 f. \quad (15)$$

#### 4.3 High Throughput Octagon

For HTOctagon, there are four types of interswitch wire lengths: First [wires which connect nodes (1,5) and (4,8)], second [wires which connect nodes (2,6) and (3,7)], third [wires which connect nodes (1,8) and (4,5)], fourth [wires which connect nodes (1,2), (2,3), (3,4), (5,6), (6,7) and (7,8)]. The interswitch wire lengths can be defined by ( $l_1=3L/4$ ,  $l_2=13w_l N_{wires} + L/4$ ,  $l_3=13L/4$ ,  $l_4=L/4$ ), where  $L$

is the length of four nodes which equals  $\left( 4 * \sqrt{\frac{Area}{N}} \right)$ .  $w_l$  is the summation of the global interconnect width and space. Considering the interswitch wire lengths and the optimum length of global interconnect, the total length of interconnects and number of repeaters can be obtained by:

$$l_{tot-HTOctagon} = (7L + 104w_l N_{wires}) N_{wires} N_{oct-unit}, \quad (16)$$

$$N_{rep-HTOctagon} = \left( 4 \left\lfloor \frac{3L/4}{K_{opt}} \right\rfloor + 4 \left\lfloor \frac{13w_l N_{wires} + L/4}{K_{opt}} \right\rfloor + 4 \left\lfloor \frac{13w_l N_{wires}}{K_{opt}} \right\rfloor + 12 \left\lfloor \frac{L/4}{K_{opt}} \right\rfloor \right) N_{wires} N_{oct-unit}, \quad (17)$$

$N_{oct-unit}$  is the number of basic octagon unit. The total power dissipation of the HTOctagon architecture is obtained by (22).

$$P_{tot-HTOctagon} = 3NP_{port} + (28 \sqrt{\frac{Area}{N}} + 104w_l N_{wires}) N_{wires} N_{oct-unit} c V_{dd}^2 f + \left( 4 \left\lfloor \frac{3L/4}{K_{opt}} \right\rfloor + 4 \left\lfloor \frac{13w_l N_{wires} + L/4}{K_{opt}} \right\rfloor + 4 \left\lfloor \frac{13w_l N_{wires}}{K_{opt}} \right\rfloor + 12 \left\lfloor \frac{L/4}{K_{opt}} \right\rfloor \right) N_{wires} N_{oct-unit} H_{opt} C_0 V_{dd}^2 f. \quad (18)$$

#### 4.4 High Throughput SPIN

An interconnect template to integrate IP blocks using HTSPIN architecture was proposed. In large HTSPIN, the total number of switches is  $3N/4$ . The interswitch wire length can be determined using (7). In HTSPIN, the total length of interconnects and the number of repeaters is defined by:

$$l_{tot-HTSPIN} = 1.75\sqrt{Area}N_{wires}N, \quad (19)$$

$$N_{rep-HTSPIN} = \left( \left\lfloor \frac{\sqrt{Area}}{8K_{opt}} \right\rfloor + \left\lfloor \frac{\sqrt{Area}}{4K_{opt}} \right\rfloor + \left\lfloor \frac{\sqrt{Area}}{2K_{opt}} \right\rfloor \right) 2N_{wires}N. \quad (20)$$

The total power dissipation of the HTSPIN architecture can be determined by

$$\begin{aligned}
 P_{tot-HTSPIN} &= 6NP_{port} + 1.75\sqrt{Area}N_{wires}NcV_{dd}^2f \\
 &+ \left( \frac{\sqrt{Area}}{8K_{opt}} \right) + \left( \frac{\sqrt{Area}}{4K_{opt}} \right) \\
 &+ \left( \frac{\sqrt{Area}}{2K_{opt}} \right) 2N_{wires}NH_{opt}C_0V_{dd}^2f.
 \end{aligned}
 \tag{21}$$

#### 4.5 Power Dissipation for Different High Throughput NoC Architectures

According to (11), (15), (18) and (21), the total power dissipation of the network can be expressed as a function of the number of IP blocks. The change in the power dissipation with the number of IP blocks for different high throughput network architectures is shown in Figure 6. The power dissipation for different NoC topologies increases by different rates as the number of IP blocks increases. The HTSPIN and HTOctagon architectures have much higher rate of power dissipation increase. The HTBFT architecture consumes the minimum power as compared to other NoC topologies making HTBFT more attractive as a power efficient NoC topology.

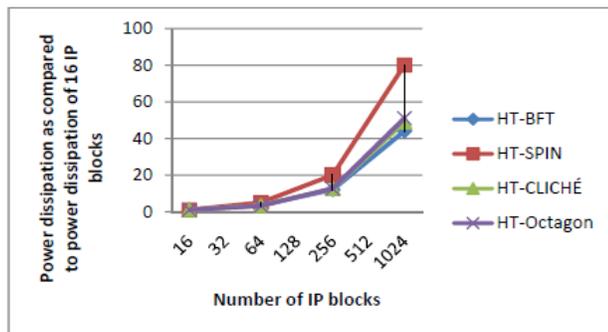


Fig. 6 power dissipation of different NoC topologies

The ratio of the power dissipation in the interconnect links and repeaters as compared to the total power dissipation is shown in Figure 7. For the HTSPIN network, the power dissipation of the interconnect links and repeaters represents 40% of the total power dissipation of the network. For the HTBFT, HTCLICH and HTOctagon, the percent of power dissipation of the interconnect links and repeaters decreases with increasing the number of IP blocks. For future SoC, reducing power dissipation should be focusing on reducing the power of the switches. More detailed results using real example of an SoC are provided in Section 5.

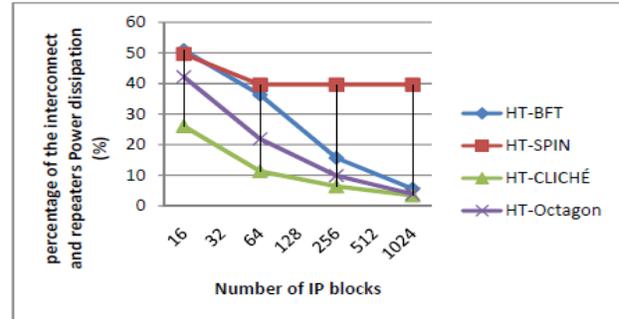


Fig. 7 Power dissipation of interconnect links and repeaters for different NoC architectures.

### 5. Performance and Overhead Analysis

The proposed high throughput architectures are implemented using Application Specific Integrated Circuit (ASIC) design flow (Leonardo Spectrum synthesis tool), with 90nm technology. Under uniform traffic assumption, the throughput for different NoC architectures is calculated. In the following subsections, the throughput and power dissipation are presented.

#### 5.1 Improvement of the Throughput

The proposed high throughput architecture doubles the number of virtual channels to increase the throughput while preserving the average latency. Therefore, the average latency of HTBFT with 8 virtual channels equals the average latency of BFT with 4 virtual channels. Uniform traffic and maximum operating frequency are assumed to determine the throughput of HTBFT. The change in the throughput with the number of virtual channels for HTBFT and BFT is shown in Figure 8. In the proposed architecture, when the number of virtual channels is increased beyond eight, the throughput saturates. The architecture increases the throughput of the network by 38%. The increase in the throughput for different architectures is presented in Table 1. The maximum improvement is achieved in HTCLICHÉ. The increase in the throughput for HTSPIN is the minimum as compared to the other high throughput architectures.

#### 5.2 Overhead of High Throughput Architecture

With the advance in technology, the number of metal layers increases every generation. Considering a chip size of 20 mm x 20 mm, technology node of 90 nm, and a system of 256 IP blocks, the length of interconnect links for different NoC topologies is obtained. Given the optimum global interconnect width  $W_{opt}$  of 935 nm, optimum global interconnect spacing  $S_{opt}$  of 477 nm [9], the global interconnect pitch is 1.412  $\mu\text{m}$  ( $W_{opt} + S_{opt}$ ). Accordingly,

the number of global interconnects  $N_{gi}$  per layer equals

$$\frac{\sqrt{Area}}{W_{opt} + S_{opt}}$$

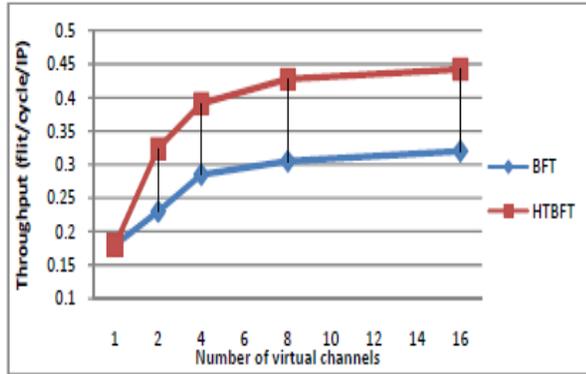


Fig. 8 Throughput for different number of virtual channels

Table 1: The percentage of increase in the throughput for different high throughput architectures

Architecture	Increase in throughput (%)
HT-BFT	38
HT-CLICHÉ	40
HT-Octagon	17
HT-SPIN	12

Using the critical interconnect length of the target technology as 2.54 mm and the optimum repeater size as 174 [9], the number of repeaters is determined. The butterfly fat tree can be laid out in  $O(N)$  active area (IPs and switches) and  $O(\log(N))$  wiring layers [10]. The basic strategy for wiring is to distribute tree layers in pair of wire layers; one for horizontal wiring  $H_{a+1,a}$  and one for vertical wiring  $V_{a+1,a}$ . The length of horizontal part  $H_{a+1,a}$  equals the length of vertical part  $V_{a+1,a}$  given that the chip is squared. More than one tree layer can share the same wiring trace.

High throughput architecture has the same number of switches, but the number of wires and repeaters is doubled. The length of interswitch interconnects depend on the number of levels, which depends on the system size. In the circuit implementation of HTBFT, a bus between each two switches has 12 wires, 8 for data and 4 for control signals. Considering a system of 256 IP blocks, the length of  $H_{a+1,a}$  and  $V_{a+1,a}$  are calculated. The number of wiring levels is seven. The number of repeaters equals 960. The area of the repeaters equals  $20880 \mu\text{m}^2$  (it is double the area of the repeaters in the conventional BFT). The power dissipation is presented in Table 2. The power dissipation is increased by 6%.

Table 2: Power dissipation of repeaters and switches for HT-/BFT

Architecture	Number of repeaters	Power dissipation in interswitch links (%)	Power reduction (%)
BFT	960	8.5	
HT-BFT	1920	16.2	6

The horizontal wiring is distributed in the metal layer no. 11 and the vertical wiring is distributed in the metal layer no. 12. The total length of horizontal wires equals 4800 mm (it is 5% of the total metal resources available in metal 11). Similarly, the total length of vertical wires is 5% of the total metal resources available in metal 12. For the proposed design, double the number of interswitch links is required to achieve the communication between each two switches. Therefore, the total metal resource to implement the proposed architecture is 10%. The extra metal resources to achieve the proposed architecture are negligible as compared to the available metal resources.

Considering the same die size of 20mm x 20mm and the system size of 256 IPs, the power dissipation and the required metal resources of other NoC topologies are shown in Table 3. Since the interswitch links is short enough in CLICHÉ, there is no need for repeaters within the interconnects. By applying the proposed high throughput architecture, the HTBFT topology requires the minimum area and power dissipation as compared to the other NoC topologies.

Table 3: Power dissipation and metal resources for different NoC

Architecture	Number of repeaters	Power dissipation of interswitch links and repeaters (%)	Metal resources (%)
CLICHÉ	0	5.4	7
HT-CLICHÉ	0	10.5	14
Octagon	3810	5.2	8
HT-Octagon	7680	10.2	16
SPIN	12288	24.8	28
HT-SPIN	24576	40.4	56

As feature size decreases, more IPs could be integrated in a single chip. System overhead is determined for the adopted architectures for different technology nodes as shown in Table 4. The extra power dissipation is 1% of the total power dissipation of the BFT architecture for 45 nm.

With the advance in technology, the available metal resources in the same die size increases. The number of switches is also increased. The required metal resources to implement the HTBFT are increased by smaller rate than the rate of increase of the available metal resources with the advance in technology. The extra metal resources and power dissipation to implement the HTBFT decrease. The extra metal resource for HTBFT is 3% of the available

metal resources. The HTBFT is becoming more efficient as technology advances.

Table 4: Power dissipation of interswitch links and repeaters for different technology nodes

Technology node	Number of IPs	Power dissipation of interswitch links and repeaters (%)			
		HT-BFT	HT-CLICHÉ	HT-Octagon	HT-SPIN
130 nm	361	17.1	14.1	14.1	58.1
90 nm	729	8.3	9.1	9.1	50.3
65 nm	1849	4.6	5.2	5.4	49.1
45nm	5625	1.2	2.7	3.1	43.8

For SPIN, the extra power dissipation to achieve the proposed HTSPIN architecture is 22% of the total power dissipation. The extra metal resources are more than 100% of the available metal resources (metal 11 and metal 12). Two more metal layers are needed to layout the proposed architecture. Therefore, the overhead in the HTSPIN is high. Applying the high throughput architecture on the SPIN topology is not recommended.

However the proposed architecture has an overhead in power dissipation and metal resources, the overhead decreases as technology advances. The proposed architecture is efficient in improving the network throughput. In the future technologies, the proposed architecture is becoming more power efficient as well as throughput efficient. In the following section, an efficient power reduction technique is proposed to make the proposed architecture further efficient from the power dissipation point of view.

## 6. Conclusions

In this paper, high throughput NoC architecture is proposed. The proposed architecture is applied to different NoC topologies. The area of the switch is decreased by 18% as compared to the area of conventional NoC switch. The total metal resources to implement the proposed high throughput NoC is increased by less than 10%. It is shown that optimizing the circuit can increase the number of virtual channels without degrading the frequency. The throughput of different NoC topologies is improved with the proposed architecture. Throughput is increased by up to 40%.

The power characteristics of different high throughput NoC topologies are presented. The extra power dissipation to achieve the proposed high throughput architecture is as low as 1% of the total power dissipation of the network. The power dissipation of NoC switches is more than 60% of the total power dissipation of the on chip network. The percent of power dissipation of the interswitch links and repeaters decreases with increasing the number of IP blocks. Reducing power dissipation should be focused on

reducing the power dissipation of the switches. The proposed switch and network architecture are becoming more efficient as technology advances. Power overhead decreases with the future technologies.

## Acknowledgments

The authors would like to thank Prof. Dr. Mohamed Ismail for his advices and directions.

## References

- [1] S. Kumar et al., "A Network on Chip Architecture and Design Methodology," *The Proc. of the IEEE Computer Society Annual Symposium on VLSI*, Apr. 2002, pp. 117-124.
- [2] P. Guerrier and A. Greiner, "A Generic Architecture for On-Chip Packet Switched Interconnections," *The Proc. of Design, Automation and Test in Europe Conference and Exhibition*, Mar. 2000, pp. 250-256.
- [3] F. Karim, A. Nguyen, and Sujit Dey, "An Interconnect Architecture for Networking Systems on Chips," *IEEE Micro*, vol.22, no.5, Sep. 2002, pp. 36-45.
- [4] P.P. Pande, C. Grecu, A. Ivanov, and R. Saleh, "Design of a Switch for Network on Chip Applications," *The Proc. of The 2003 International Symposium on Circuits and Systems*, vol.5, May 2003, pp. 217220.
- [5] E. Bolotin, I. Cidon, R. Ginosar and A. Kolodny, "QNoC: QoS Architecture and Design Process for Network on Chip," *Journal of Systems Architecture*, vol.50, no.23, Feb. 2004, pp. 105-128.
- [6] D. Bertozzi, A. Jalabert and S. Murali et al., "NoC Synthesis Flow for Customized Domain Specific Multiprocessor Systems On Chip," *IEEE Transactions on Parallel and Distributed Systems*, vol. 16, no. 2, Feb. 2005, pp. 113-129.
- [7] P. P. Pande, C. Grecu, M. Jones, A. Lvanov, and R. Saleh, "Performance Evaluation and Design Trade Offs for Network on -Chip Interconnect Architectures," *IEEE Transaction on Computers*, vol. 54, no. 8, Aug. 2005, pp. 1025-1040.
- [8] K. Lee, S.J. Lee, and H.J. Yoo, "Low Power Networks on -Chip for High Performance SoC Design," *IEEE Transactions on Very Large Scale Integration Systems*, vol. 14, no. 2, Feb. 2006, pp.148-160.
- [9] X.C. Li, J.F. Mao, H.F. Huang, and Y. Liu, "Global Interconnect Width and Spacing Optimization for Latency, Bandwidth and Power Dissipation," *IEEE Transactions on Electron Devices*, vol. 52, no. 10, Oct. 2005, pp. 2272-2279.
- [10] A. Dehon, "Compact, Multilayer Layout for Butterfly Fat Tree," *The Proc. of The ACM Symposium on Parallel algorithm Architectures*, Jul. 2000, pp. 206-215.

**Magdy A. El-Moursy** was born in Cairo, Egypt in 1974. He received the B.S. degree in electronics and communications engineering (with honors) and the Master's degree in computer networks from Cairo University, Cairo, Egypt, in 1996 and 2000, respectively, and the Master's and the Ph.D. degrees in electrical engineering in the area of high-performance VLSI/IC design from University of Rochester, Rochester, NY, USA, in 2002 and 2004, respectively. In summer of 2003, he was with STMicroelectronics, Advanced System Technology, San Diego, CA, USA. Between

September 2004 and September 2006 he was a Senior Design Engineer at Portland Technology Development, Intel Corporation, Hillsboro, OR, USA. During September 2006 and February 2008 he was assistant professor in the Information Engineering and Technology Department of the German University in Cairo (GUC), Cairo, Egypt. Dr. El-Moursy is currently a Technical Lead in the Mentor Graphics Corporation, Cairo, Egypt. His research interest is in Networks-on-Chip, interconnect design and related circuit level issues in high performance VLSI circuits, clock distribution network design, and low power design. He is the author of more than 30 papers, four book chapters, and one book in the fields of high speed and low power CMOS design techniques and high speed interconnect.

**Mohamed Abdelgany** was teaching assistant in the Information Engineering and Technology Department of the German University in Cairo (GUC), Cairo, Egypt. He got his Ph.D. in 2009. His research interest is in Networks-on-Chip. He has many papers in the field.