# We are IntechOpen, the world's leading publisher of Open Access books <br> Built by scientists, for scientists 

## 6,900

Open access books available

## 185,000

International authors and editors

Our authors are among the
TOP 1\%
most cited scientists


Downloads


Contributors from top 500 universities

# Interested in publishing with us? Contact book.department@intechopen.com 

Numbers displayed above are based on latest data collected.<br>For more information visit www.intechopen.com



# High Throughput Architecture for High Performance NoC 

Mohamed A. Abd El Ghany,<br>Magdy A. El-Moursy* and Mohammed Ismail**<br>Electronics Engineering Dept., German University in Cairo, Cairo, Egypt Electronics Research Institute, Cairo, Egypt, Mentor Graphics Corporation, Cairo, Egypt*<br>Electrical Engineering Dept., The Ohio State University, Columbus, USA. The RaMSiS Group,<br>KTH, Sweden**

## 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 researches have been published in high performance SoCs; however, the system scalability and bandwidth are limited. Network on Chip (NoC) is emerging as the best replacement for the existing interconnection architectures. NoC is composed of network of interconnects and number of temporary storage elements called switches. The temporary storage element of different NoC architectures has different number of ports. The main component of the port is the virtual channels. The virtual channels consist of several buffers controlled by a multiplexer and an arbiter which grants access for only one buffer at a time according to the request priority. When the number of buffers is increased, the throughput increases. High throughput and low latency are the desirable characteristics of a multi processing system. More research is needed to enhance performance of NoC components (network of interconnects and the storage elements). Many NoC architectures have been proposed in the past, e.g., SPIN (Guerrier \& Greiner, 2000), CLICHÉ (Kumar et al., 2002), Folded Torus (Dally \& Towles, 2001), Octagon (Karim et al., 2002) and Butterfly fat-tree (BFT) (Pande et al., 2003a). Among those, the butterfly fat tree (BFT) has found extensive use in different parallel machines and shown to be hardware efficient (Grecu et al., 2004a). The main advantage of the butterfly fat tree is that the number of storage elements in the network converges to a constant irrespective of the number of levels in the tree network. In the SPIN architecture, redundant paths contained within the fat tree structure are utilized to reduce contention in the network. CLICHÉ (Chip-Level Integration of Communicating Heterogeneous Elements) is simplest from a layout perspective and the local interconnections between resources and storage elements are independent of the size of the network. In the Octagon architecture, the communication between any two nodes takes at most two hops within the basic Octagon unit.
After the NoC design paradigm has been proposed (Dally \& Towles, 2001) ; (Kumar et al., 2002) ; (Guerrier \& Greiner, 2000) ; (Karim et al., 2002) ; (Pande et al., 2003) ; (Benini \&

Micheli, 2002) ; (Grecu et al., 2004a), many researches on architectural and conceptual aspects of NoC have been reported such as topology selection (Murali \& Micheli, 2004), quality of service (QoS) (Bolotin et al., 2004), design automation (Bertozzi et al., 2005) ; (Liang et al., 2004) ; ( Pande et al., 2005a), performance evaluation (Pande et al., 2005b) ; (Salminen et al., 2007) ; (Grecu et al., 2007a) and test and verification (Grecu et al., 2007b) ; (Kim et al., 2004) ; (Murali et al., 2005). These researches have taken a top-down approach (a high level analysis of NoC ) and they didn't touch the issues on a circuit level. However, a little research has reported on design issues in implementation of NoC in the perspective of circuit level (Lee et al., 2003) ; (Lee et al., 2004) ; (Lee \& Kim et al., 2005) ; (Lee et al., 2006) ; (Lee; Lee \& Yoo, 2005). Although, they were implemented and verified on the silicon, they were only focusing on implementation of limited set of architectures.
In large-scale NoC, power consumption should be minimized for cost-efficient implementations. Although different researches 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. Lowering supply voltage is the most effective way to reduce power consumption. With lowering supply voltage, the threshold voltage ( $\mathrm{V}_{\mathrm{TH}}$ ) has to be decreased to achieve high performance requirements. Reducing $\mathrm{V}_{\mathrm{TH}}$ causes significant increase in the leakage component. Different researches have been published in power minimization of high performance CMOS circuits (Khellah \& Elmasry, 1999) ; (Kao \& Chandrakasan, 2000) ; (Kursun \& Friedman, 2004).
In this chapter, different tradeoffs in designing efficient NoC including both elements of the network (interconnects network and storage elements) are described. Building high performance NoC is presented. In addition, a high throughput architecture is proposed. The proposed architecture to achieve high throughput can improve the latency of the network. The circuit implementation issues are considered in the proposed architecture. The switch structure along with the interconnect architecture are shown in Fig. 1 for 2 IPs and 2 switches. The proposed architecture is applied to different NoCs topologies. 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 architectures.
This chapter 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, power characteristics for different high throughput architectures are provided. The performance and overhead analysis of the proposed architecture are provided in Section 5. In Section 6, the proposed design of low power NoC switch is described. Finally, conclusions are provided in Section 7.


Fig. 1. proposed high throughput architecture.

## 2. Port architecture

The switch of different architectures has different number of ports. Each port of the switch includes input virtual channels, output virtual channels, a header decoder, controller, input arbiter and output arbiter as shown in (Pande et al., 2003a). When the number of virtual channel is increased, the throughput increases. The input arbiter is used to allow only one virtual channel to access a physical port. The input arbiter consists of a priority matrix and grant circuits (Pande et al., 2003b).
The priority matrix stores the priorities of the requests. The grant circuits generate the granted 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. All subsequent flits of the corresponding packet are sent from input to output using the established path. 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. In the proposed architecture, rather than using one multiplexer and one arbiter to control the virtual channels, two multiplexer and two arbiters are employed as shown in Fig. 2. The virtual channels are divided into two groups, each group controlled by one multiplexer and one arbiter. Each group of virtual channels is supported by one interconnect bus as described in Section 3. However looks trivial, this port architecture has a great influence on the switch frequency and the throughput of the network.
Let us consider an example with the number of virtual channels of 8 channels. In the NoC architecture, $8 \times 8$ input arbiter and $8 \times 1$ multiplexer are needed to control the input virtual channels as shown in Fig. 2 (a). The $8 x 8$ input arbiter consists of $8 x 8$ grant circuit and $8 \times 8$ priority matrix. In the proposed architecture, two $4 \times 4$ input arbiters, two $4 \times 1$ multiplexers, $2 \times 1$ multiplexers and $2 \times 2$ grant circuit are integrated to allow only one virtual channel to
access a physical port as shown in Fig. 2 (b). The $4 \times 4$ input arbiter consists of $4 \times 4$ grant circuit and $4 \times 4$ priority matrix. The values of the grant signals are determined by the priority matrix. The number of grant signals equals to the number of requests and the number of selection signals of the multiplexer. The area of $8 \times 8$ input arbiter is larger than the area of two $4 \times 4$ input arbiters. Also, the area of $8 \times 1$ multiplexer is larger than the area of two $4 \times 1$ multiplexers. 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 $4 \times 1$ multiplexer into three $2 \times 1$ multiplexers, the $4 \times 4$ input arbiter should be divided into three $2 \times 2$ input arbiters. The grant signals generated by three $2 \times 2$ input arbiter ( 6 signals) aren't the same grant signals generated by the $4 \times 4$ input arbiter ( 4 signals). Therefore, the $4 \times 4$ input arbiter can't be replaced by three $2 \times 2$ input arbiters unless the number of interconnect buses is increased to be equal the number of virtual channels groups. By increasing the number of interconnect buses, the metal resources and power dissipation are increased as described in Section 5.
Without circuit optimization in BFT architecture, the change in the maximum frequency of the switch with the number of virtual channels is shown in Fig. 3. When the number of virtual channels is increased beyond four, the maximum frequency of the switch is decreased. The throughput is saturated when the number of virtual channels is increased beyond four (Pande et al., 2005b) for different number of ports. On the other hand, the average message latency increases with the number of virtual channels. To keep the latency low while preserving the throughput, the number of virtual channels is constrained to four (Pande et al., 2003b),(Pande et al., 2005b). Throughput is a parameter that measures the rate in which message traffic can be sent across a communication network. It is defined by (Pande et al., 2005b):

$$
\begin{equation*}
T P=\frac{(\text { number of messages completed }) *(\text { message length })}{(\text { number of IP blocks }) *(\text { total time })} \tag{1}
\end{equation*}
$$

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 frequency of the switch. Therefore the throughput can be improved by increasing the number of virtual channels or by increasing the frequency of switch (Lee \& Bagherzadeh, 2006). The HT-BFT 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 Fig. 3 for HT-BFT 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 with different number of virtual channels for different network topologies and the proposed architectures as shown in Fig. 4. As compared to the conventional architectures, the operating frequency of the proposed architectures is decreased when the number of virtual channels is higher than eight rather than four. As shown in Fig. 3 and Fig. 4, 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). 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 which each flit will experience. 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 to eight), doubling the number of virtual channels between switches is proposed.


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


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


Fig. 4. Maximum frequency of a switch with different number of virtual channels for different NoC architectures.

Let us consider an example of BFT architecture. The area required to implement the BFT switch and HT-BFT switch is shown with different number of virtual channels in Fig. 5. The HT-BFT 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 design is optimized. The architectures of different NoC topologies to achieve high throughput network is discussed in Section 3.


Fig. 5. Area of a switch for different number of virtual channels.

## 3. High Throughput architecture

A novel interconnect template to integrate IP blocks using NoC architecture is proposed as shown in Fig. 1. 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. Each bus will support half number of virtual channels to maintain the average latency.

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 a consumption of the metal resources and may be silicon area for the repeaters within long interconnect bus. The overhead of the proposed architecture is discussed in Section 5. Applying the proposed high throughput architecture on different NoC topologies is presented in the following subsections.

### 3.1 High Throughput BFT

The interconnect template of butterfly fat-tree topology was proposed in (Pande et al., 2003a). This structure assumes a 4-ary tree with switches connected 4 down links and 2 up links. Each group of 4 leaf nodes needs one switch. At the next level, half as many switches are needed (every 4 switches on the lower level need 2 switches at the next level). This relation continues with each succeeding level.
A novel interconnect template to integrate IP blocks using HT-BFT architecture is proposed as shown in Fig. 6 (a). In the proposed HT-BFT architecture (Abd El Ghany et al., 2009a), rather than using a single interconnect bus between each two switches, two buses are employed. Each group of 4 IPs (no. 0, no. 1, no. 2 and no.3) 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 as shown in Fig. 6 (a). Each bus will support half number of virtual channels. Therefore, the throughput can be improved while preserving the average latency.

### 3.2 High Throughput CLICHÉ

The mesh-interconnect topology called CLICHÉ (Chip-Level Integration of Communicating Heterogeneous Elements) was proposed in (Kumar et al., 2002). The architecture consists of $\mathrm{m} \times \mathrm{n}$ mesh of switches interconnecting the IP blocks. Every switch is connected to four switches and one IP block. At the edges, the switches, except those at the corners, are connected to three switches and one IP block. The number of switches equals to the number of IP blocks. The interconnect template to integrate IP blocks using High Throughput CLICHÉ (HT-CLICHÉ) architecture is shown in Fig. 6 (b) (Abd El Ghany et al., 2009b). The interconnect bus between each two switches consists of two unidirectional links.

### 3.3 High Throughput Octagon

The interconnect template of Octagon topology was proposed in (Karim et al., 2002). The basic unit of Octagon topology consists of eight nodes and 12 bidirectional buses. Each node is associated with an IP block and a switch. Communication between any two nodes takes at most two hops within the basic Octagon unit. The Octagon is extended to multidimensional space for a system of more than eight nodes. The interconnect template to integrate IP blocks using High Throughput Octagon (HT-Octagon) architecture is shown in Fig. 6 (c). For the basic unit of HT-Octagon architecture, number of bidirectional buses equals to 24 rather than 12 bidirectional buses in conventional Octagon architecture.

### 3.4 High Throughput SPIN

The interconnect template called SPIN (Scalable, Programmable, Integrated Network) was proposed in (Guerrier \& Greiner, 2000). This structure assumes a 4 -ary tree with switches
connected 4 down links and 4 up links. Each group of 4 leaf nodes needs one switch. At the next level, the same number of switches are needed (every 4 switches on the lower level need 4 switches at the next level). This relation continues with each succeeding level. The main rationale behind this approach is utilization of the redundant buses by the routers in order to reduce contention in the network. Therefore, SPIN trades area overhead and extra power dissipation for higher throughput. The interconnect template to integrate IP blocks using High Throughput SPIN (HT-SPIN) architecture is shown in Fig. 6 (d). In the proposed HT-SPIN architecture, the double number of buses is needed to connect between each two switches or between an IP block and a switch. Due to the higher usage of on-chip resources by the interswitch links, applying the high throughput architecture on SPIN topology is not efficient for insignificant improvement of throughput as described in Section 5. The power characteristics for different high throughput architectures are provided in Section 4.


Fig. 6. proposed interconnect architectures. (a) HTBFT. (b) HT- CLICHÉ. (c) HT-Octagon. (d) HT-SPIN.

## 4. Power Characteristics

Power dissipation is a primary concern in high speed, high complexity integrated circuits (IC). Power dissipation increases rapidly with the increase in frequency and transistor density in integrated circuits. To achieve power efficient NoC, power dissipation need to be characterized for different topologies. Communication network on chip contains three primary parts; network switch, interswitch links (interconnects), and repeaters within interswitch links as shown in Fig. 7. Including different sources of power consumption in NoC, the total power dissipation of on chip network is defined as follows:

$$
\begin{gather*}
P_{\text {total }}=P_{\text {switches }}+P_{\text {line }}+P_{\text {rep }}  \tag{2}\\
P_{\text {switches }}=P_{\text {switching }}+P_{\text {leakage }} \tag{3}
\end{gather*}
$$



Fig. 7. communication networks on chip
where $P_{\text {switches }}$ is the total power dissipation of these switches forming the network. $P_{\text {switches }}$ is the summation of switching (including dynamic and short circuit) power and leakage power of switches. $P_{\text {line }}$ is the total power dissipation of interswitch links. $P_{\text {rep }}$ is the total power dissipation of the repeaters which are required for long interconnects. The number of repeaters depends on the length of the interswich 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.

The power consumption of interswitch links $P_{\text {line }}$ and the power consumption of repeaters are defined by (El-Moursy \& Friedman, 2004)

$$
\begin{align*}
& P_{\text {line }}=C V_{d d}^{2} f  \tag{4}\\
& P_{\text {rep }}=P_{\text {rep-dyn }}+P_{\text {rep-SC }}+P_{\text {rep-leakage }}  \tag{5}\\
& P_{\text {rep-dyn }}=N_{\text {rep }} H_{\text {opt }} C_{o} V_{d d}^{2} f \tag{6}
\end{align*}
$$

where $P_{\text {rep-dyn }}$ is the total dynamic power dissipation of repeaters, $N_{\text {rep }}$ is the number of repeaters, $H_{\text {opt }}$ is the optimal repeater size and $C_{o}$ is the input capacitance of a minimum size repeater. $P_{\text {rep-SC }}$ is the total short-circuit power of repeaters. $P_{\text {rep-leakage }}$ is the total leakage power dissipation of repeaters. $P_{\text {rep-leakage }}$ and $P_{\text {rep-SC }}$ are negligible as compared to the total dynamic power dissipation of repeaters [32 ]. The closed form equations 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 HT-BFT, the interconnection is performed on levels of switching. The number of switch levels can be expressed as levels $=\log _{2} N-3$, 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 Fig. 6 (a). The interswitch wire length and total number of switches are given by the following expression (Grecu et al., 2004b):

$$
\begin{gather*}
l_{a+1, a}=\frac{\sqrt{\text { Area }}}{2^{\text {levels-a }}}  \tag{7}\\
N_{\text {switches-HTBFT }}=\frac{N}{4}\left(\frac{1-(1 / 2)^{\text {levels }}}{1-1 / 2}\right) \tag{8}
\end{gather*}
$$

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

$$
\begin{gather*}
l_{\text {tot-HTBFT }}=\frac{\sqrt{\text { area }}}{2^{\left(\left(\log _{2} N\right)-3\right)}} N \times(\text { levels }) \times 2 N_{\text {wires }}  \tag{9}\\
N_{\text {repeater-HTBFT }}=N 2 N_{\text {wires }}\left(\left\lfloor\frac{l_{1,0}}{K_{\text {opt }}}\right\rfloor+\frac{1}{2}\left\lfloor\frac{l_{2,1}}{K_{\text {opt }}}\right\rfloor+\frac{1}{4}\left\lfloor\frac{l_{3,2}}{K_{\text {opt }}}\right\rfloor+\ldots \ldots\right. \\
\left.+\frac{1}{2^{N-1}}\left\lfloor\frac{l_{\text {levels,levels }-1}}{K_{\text {opt }}}\right\rfloor\right) \tag{10}
\end{gather*}
$$

Where $K_{o p t}$ is the optimal length of the global interconnect (Li et al., 2005). Using the number of switches, the total length of interconnect and the total number of repeaters, the total power dissipation of HT-BFT architecture ( $P_{\text {tot-HTBFT }}$ ) can be calculated using the following expression:

$$
\begin{align*}
& P_{\text {tot-HTBFT }}=\frac{N}{4}\left(\frac{1-(1 / 2)^{\left(\left(\log _{2} N\right)-3\right)}}{1-1 / 2}\right)\left(6 P_{\text {port }}\right)+\frac{\sqrt{\text { area }}}{2^{\left(\left(\log _{2} N\right)-3\right)}} N \times\left(\left(\log _{2} N\right)-3\right) \times \\
& 2 N_{\text {wires }} c V_{d d}^{2} f+  \tag{11}\\
& \left(\left\lfloor\frac{l_{1,0}}{K_{\text {opt }}}\right\rfloor+\frac{1}{2}\left\lfloor\frac{l_{2,1}}{K_{\text {opt }}}\right\rfloor+\frac{1}{4}\left\lfloor\frac{l_{3,2}}{K_{\text {opt }}}\right\rfloor+\ldots \ldots+\frac{1}{2^{N-1}}\left\lfloor\frac{l_{\left(\left(\log _{2} N\right)-3\right),\left(\left(\log _{2} N\right)-2\right)}}{K_{\text {opt }}}\right\rfloor\right) N 2 N_{\text {wires }} H_{o p t} C_{o} V_{d d}^{2} f
\end{align*}
$$

### 4.2 High Throughput CLICHÉ Architecture

In HT-CLICHE architecture, the number of switches is equal to the number of IPs as shown in Fig. 6 (b). The interswitch wire lengths can be determined from the following expression:

$$
\begin{equation*}
l_{\text {HTCLICHE }}=\frac{\sqrt{\text { Area }}}{\sqrt{N}} \tag{12}
\end{equation*}
$$

The number of horizontal interswitch links between switches equals to $2 \sqrt{N}(\sqrt{N}-1)$, and the number of vertical interswich links between switches equals to $2 \sqrt{N}(\sqrt{N}-1)$. According to the technology node, the optimal length of global interconnect can be obtained (Li et al., 2005). Therefore, the total length of interconnect and the number of repeaters for HT-SPIN can be calculated by:

$$
\begin{gather*}
l_{\text {tot-HTCLICHE }}=4 \sqrt{\operatorname{area}}(\sqrt{N}-1) N_{\text {wires }}  \tag{13}\\
N_{\text {repeaters-HTCLICHE }}=4\left\lfloor\frac{\sqrt{a r e a}}{\sqrt{N} K_{o p t}}\right\rfloor \sqrt{N}(\sqrt{N}-1) N_{\text {wires }} \tag{14}
\end{gather*}
$$

Using the number of ports, number of switches, total length of interconnects and number of repeaters, the total power consumption of the HT-CLICHÉ architecture can be determined by the following expression:

$$
\begin{array}{rl}
P_{\text {tot }- \text { HTCLICHE }}=5 & N P_{\text {port }}+4 \sqrt{\text { area }}(\sqrt{N}-1) \quad N_{\text {wires }} c V_{d d}^{2} f \\
& +4\left\lfloor\frac{\sqrt{\text { area }}}{\sqrt{N} K_{\text {opt }}}\right\rfloor \sqrt{N}(\sqrt{N}-1) N_{\text {wires }} H_{\text {opt }} C_{o} V_{d d}^{2} f \tag{15}
\end{array}
$$

### 4.3 High Throughput Octagon Architecture

For HT-Octagon architecture, there are four types of interswitch wire length as shown in Fig. 6 (c) : First (connecting nodes 1-5 and 4-8), second (connecting nodes 2-6 and 3-7, third
(connecting nodes 1-8 and 4-5), forth (connecting nodes 1-2, 2-3, 3-4, 5-6, 6-7 and 7-8). the interswitch wire lengths can be defined by:

$$
\begin{align*}
l_{1} & =\frac{3 L}{4}  \tag{16}\\
l_{2} & =13 w_{l} N_{\text {wires }}+\frac{L}{4}  \tag{17}\\
l_{3} & =13 w_{l} N_{\text {wires }}  \tag{18}\\
l_{4} & =\frac{L}{4} \tag{19}
\end{align*}
$$

Where $L$ is the length of four nodes; it equals to $\left(4 * \sqrt{\frac{\text { area }}{N}}\right) . w_{l}$ is the summation of the global interconnect width and space. Considering the interswitch wire lengths and the optimal length of global interconnect, the total length of interconnect and number of repeaters can be obtained by:

$$
\begin{equation*}
l_{\text {tol-HTOctagon }}=\left(7 L+104 w_{l} N_{\text {wires }}\right) N_{\text {wires }} N_{\text {oct-units }} \tag{20}
\end{equation*}
$$

$$
\begin{align*}
& N_{\text {repeaters-HTOctagon }}= \\
& \left(4\left\lfloor\frac{\lfloor 3 / 4}{K_{\text {opt }}}\right\rfloor+4\left[\frac{13 w_{l} N_{\text {wires }}+L / 4}{K_{\text {opt }}}\right\rfloor+4\left\lfloor\frac{13 w_{l} N_{\text {wires }}}{K_{\text {opt }}}\right\rfloor+12\left\lfloor\frac{L / 4}{K_{\text {opt }}}\right\rfloor\right) N_{\text {wires }} N_{\text {oct-unit }} \tag{21}
\end{align*}
$$

Where $N_{\text {oct-units }}$ is the number of basic octagon unit. The total power dissipation of the HTOctagon architecture can be determined by the following expression:

$$
\begin{align*}
& P_{\text {tot-HTOctagon }}=3 N P_{\text {port }}+\left(\left(28\left(\sqrt{\frac{\text { area }}{N}}\right)+104 w_{l} N_{\text {wires }}\right) N_{\text {wires }} N_{\text {oct-units }}\right) c V_{d d}^{2} f+ \\
& \left(\left(4\left\lfloor\frac{3\left(\sqrt{\frac{\text { area }}{N}}\right)}{K_{\text {opt }}}\right\rfloor+4\left\lfloor\frac{13 w_{l} N_{\text {wires }}+\left(\sqrt{\frac{\text { area }}{N}}\right)}{K_{\text {opt }}}\right\rfloor+4\left\lfloor\frac{13 w_{l} N_{\text {wires }}}{K_{\text {opt }}}\right\rfloor+\right.\right. \\
& \left.\left.12\left[\frac{\left(\sqrt{\frac{\text { area }}{N}}\right)}{K_{\text {opt }}}\right]\right) N_{\text {wires }} N_{\text {oct-unit }}\right) H_{\text {opt }} C_{o} V_{d d}^{2} f \tag{22}
\end{align*}
$$

### 4.4 High Throughput SPIN Architecture

An interconnect template to integrate IP blocks using SPIN architecture was proposed as shown in Fig. 6 (d). In large SPIN, the total number of switches is $3 N / 4$ (Guerrier \& Greiner, 2000). The interswitch wire length can be determined using eq. (7). In the HT-SPIN, The total length of interconnect and the number of repeaters are defined by:

$$
\begin{gather*}
l_{\text {tot-HTSPIN }}=1.75 \sqrt{\text { area }} N N_{\text {wires }}  \tag{23}\\
N_{\text {repeaters-HTSPIN }}=\left(\left\lfloor\frac{\sqrt{\text { area }}}{8 K_{\text {opt }}}\right\rfloor+\left\lfloor\frac{\sqrt{\text { area }}}{4 K_{\text {opt }}}\right\rfloor+\left\lfloor\frac{\sqrt{\text { area }}}{2 K_{\text {opt }}}\right\rfloor\right) N 2 N_{\text {wires }} \tag{24}
\end{gather*}
$$

The total power dissipation of the network architecture depends on the main three parameters; the number of switches, the total length of interconnect and the number of repeaters. The total power consumption of the HT-SPIN architecture ( $P_{\text {tot-HTSPIN }}$ ) can be determined by:

$$
\begin{gather*}
P_{\text {tot-HTSPIN }}=\frac{3 N}{4}\left(8 P_{\text {port }}\right)+1.75 \sqrt{\text { area }} N N_{\text {wires }} c V_{d d}^{2} f+\left(\left\lfloor\frac{\sqrt{\text { area }}}{8 K_{\text {opt }}}\right\rfloor+\left\lfloor\frac{\sqrt{\text { area }}}{4 K_{o p t}}\right\rfloor+\right. \\
\left.\left\lfloor\frac{\sqrt{\text { area }}}{2 K_{\text {opt }}}\right\rfloor\right) N 2 N_{\text {wires }} H_{o p t} C_{o} V_{d d}^{2} f \tag{25}
\end{gather*}
$$

### 4.5 Power Dissipation for Different NoC Architectures

According to the equations (11), (15), (22) and (25), the total power dissipation of the network can be considered as a function of the number of IP blocks. The change in the power consumption with the number of IP blocks for different network architectures is shown in Fig. 8. The power consumption for different NoC architectures increases by different rates with the number of IP blocks. The SPIN and Octagon architectures have much higher rates of power dissipation. The BFT architecture consumes the minimum power as compared to other NoC architectures.


Fig. 8. power dissipation of different NoC architectures
The percentage of the power dissipation of the interswitch links and repeaters is shown in Fig. 9. For the SPIN and architecture, the power dissipation of the interswitch links and repeaters equals to $25 \%$ of the total power dissipation of the architecture. For the BFT, CLICHÉ and Octagon architectures, the percentage of power dissipation of the interswitch links and repeaters decreases with the number of IP blocks.


Fig. 9. power dissipation of interswitch links and repeaters for different NoC architectures.
The overhead analysis and simulation results are provided in Section 5.

## 5. Performance and Overhead analysis

The proposed high throughput architectures are implemented using the Application Specific Integrated Circuit (ASIC) by Leonardo Spectrum synthesis tool, used for 90nm technology node. Under uniform traffic assumption, the throughput for different NoC architectures is calculated. The comparative analysis focuses on the frequency of the switch, the throughput, the area of the switch and the power consumption is presented in the following subsections.

### 5.1 Improvement of the Throughput

The proposed high throughput architecture trades the double number of virtual channels for higher throughput while preserving the average latency. Therefore, the throughput of using eight virtual channels in the HT-BFT is double the throughput of four virtual channels in BFT. The average latency of HT-BFT with 8 virtual channels equals to the average latency of BFT with 4 virtual channels. Considering the uniform traffic, the Maximum frequency of the switch and the number of completed messages for HT-BFT, the throughput of HT-BFT is determined. The variation of throughput with the number of virtual channels for HT-BFT and BFT is shown in Fig. 10. In our 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 percentage of increasing of throughput for different high throughput architectures is presented in Table 1. The maximum improvement in the throughput is obtained in HT-CLICHÉ and HT-BFT. The increase in the throughput for HTSPIN is the minimum as compared to other high throughput architectures.


Fig. 10. Throughput for different number of virtual channels.

| architecture | The percentage of increase in throughput <br> $(\%)$ |
| ---: | :--- |
| HT-BFT | 38 |
| HT-CLICHÉ | 40 |
| HT-Octagon | 17 |
| HT-SPIN | 12 |

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

### 5.2 Overhead of High Throughput Architectures

With the advance in technology, the number of metal levels increases reaching twelve (ITRS, 2007). Metal resources on chip increase. Considering a chip size of $20 \mathrm{~mm} \times 20 \mathrm{~mm}$ (Area), technology node of 90 nm , and a system of 256 IP blocks, the length of interswitch links for different NoC architectures is obtained. Given the optimal global interconnect width $W_{\text {opt }}$ of 935 nm , optimal global interconnect spacing $S_{o p t}$ of 477 nm (Li et al., 2005), the global interconnect pitch is $W_{\text {opt }}+S_{\text {opt }}$. Assuming all of global interconnects have the same line width and line spacing, then the number of global interconnects $N_{g i}$ per layer equals to $N_{g i}=\frac{\sqrt{\text { Area }}}{W_{\text {opt }}+S_{\text {opt }}}$
According to the NoC architecture, the total length of interswitch links are calculated. Using the critical interconnect length of 2.54 mm , optimal repeater size of 174 (Li et al., 2005), the number of repeaters is determined. The extra area and power required to implement different high throughput NoC architectures are presented in the following subsections.

### 5.2.1 HT-BFT

It is possible to organize the butterfly fat tree so that it can be laid out in $O(N)$ active area(IPs and switches) and $O(\log (N))$ wiring layers (Dehon, 2000). The basic strategy for wiring is to distribute tree layers in pair of wire layers - one for horizontal wiring $\mathrm{H}_{\mathrm{a}+1, \mathrm{a}}$ and one for vertical wiring $\mathrm{V}_{\mathrm{a}+1, \mathrm{a}}$. The length of horizontal part $\mathrm{H}_{\mathrm{a}+1, \mathrm{a}}$ equals to the length of vertical part $\mathrm{V}_{\mathrm{a}+1, \mathrm{a}}$ 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 will be doubled. The length of interswitch wire depends on the number of levels in BFT, which depends on the system size as shown in eq (7).
In the circuit implementation of HT-BFT, 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 $\mathrm{H}_{\mathrm{a}+1, \mathrm{a}}$ and $\mathrm{V}_{\mathrm{a}+1, \mathrm{a}}$ are calculated. The number of BFT levels is seven. Using the critical interconnect length, the number of repeaters equals to 960 repeaters. The area of repeaters required to implement the HT-BFT interswitch links equals to $20880 \mu \mathrm{~m}^{2}$ (it equals to the double area of repeaters required for BFT interswitch links). The power consumption of repeaters and switches required to implement the BFT and HT-BFT is presented in Table 2. The power consumption required to implement HT-BFT is increased by $7 \%$ as compared with the power consumption of BFT.

| Architecture | No. of repeaters | Power dissipation of repeaters and interswitch links (mw) | Power dissipation of switches (mw) | Total power dissipation (mw) | Percentage of power dissipation of repeaters and interswitch links (\%) |
| :---: | :---: | :---: | :---: | :---: | :---: |
| BFT | 960 | 1458.24 | 15663.68 | 17121.92 | 8.5 |
| HT-BFT | 1920 | 2916.48 | 15674.84 | 18591.32 | 15.7 |

Table 2. power consumption of repeaters and switches for BFT and HT-BFT
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 needed equals to 4800 mm (it is $5 \%$ of the total metal resources available in the metal 11). The same for total length of vertical wires, it requires $5 \%$ of the total metal resources available in the metal 12 .

For the proposed design, the double number of interswitch links is required to achieve the communication between each two switches. Therefore, the total metal resources required to implement the proposed architecture will be $10 \%$. The metal resources of HT-BFT architecture equals to the double metal resources of BFT architecture. The extra metal resources required to achieve the proposed architecture is negligible as compared to the metal resources.
The percentage of the metal resources and power consumption of interswitch links and repeaters for different technology node is shown in Table 3. With the advance in technology, the available metal resources in the same die size are increased. Therefore, the number of IPs could be increased. The number of switches is also increased. The required metal resources to implement the BFT and HT-BFT are increased by fewer rates than the rates of increase of the available metal resources with the advance in technology. The extra metal resources and power consumption required to implement the HT-BFT decreases. The extra power consumption required to achieve the proposed architecture is $1 \%$ of the total power consumption of the BFT architecture. Also, the extra metal resources required for HT-BFT is $3 \%$ of the metal resources. The HT-BFT is more efficient with the advance in technology.

| Technology node | No. of IPs | No. of levels | Percentage of power consumption of interswitch links and repeaters for BFT | Percentage of power consumption of interswitch links and repeaters for HT-BFT ( mw ) | Percentage of BFT metal resources | Percentage of HT-BFT metal resources |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 130 nm | 500 | 6 | 10.26\% | 20.5\% | 4.95 \% | 9.89 \% |
| 90 nm | 1000 | 7 | 4.49\% | 8.98\% | 4.02 \% | 8.04 \% |
| 65 nm | 2500 | 9 | 1.32\% | 2.64\% | 2.55 \% | 5.1 \% |
| 45 nm | 7500 | 10 | 0.59\% | 1.19\% | 2.97 \% | 5.94 \% |

Table 3. metal resources and power consumption of interswitch links and repeaters for HTBFT and BFT

### 5.2.2 HT-CLICHÉ

The CLICHÉ architecture with N IP blocks can be laid out in $O(N)$ active area(IPs and switches) and $O(\sqrt{N})$ interswitch links. In the circuit implementation of HT-CLICHÉ, a bus between each two switches has 20 wires, 16 for data and 4 for control signals. Considering a system of 256 IP blocks, the architecture consists of $16 \times 16$ mesh of switches interconnecting the IPs. The length of horizontal links and vertical links equal to 1.25 mm . They are smaller than the critical interconnect length. Therefore, no repeaters are needed within the interswitch links. The power dissipation of the network is presented in Table 4 for CLICHÉ and HT-CLICHÉ. The extra power dissipation required to implement HT-CLICHÉ for 256 IPs equals to $5 \%$.

|  | Power <br> consumption <br> Architecture <br> linkerswitch <br> lepeaters <br> $(\mathrm{mw})$ | Power <br> consumption of <br> switches $(\mathbf{m w})$ | Total power <br> dissipation $(\boldsymbol{m w})$ | Percentage of <br> power dissipation <br> of repeaters and <br> interswitch links <br> $(\%)$ |
| :--- | :--- | :--- | :---: | :---: |
| CLICHÉ | 1398 | 24448 | 25846 | 5.4 |
| HT-CLICHÉ | 2796 | 24471 | 27267 | 10.25 |

Table 4. power consumption for CLICHÉ and HT-CLICHÉ architectures
Using the equation no. 12, the total length of interswitch links is calculated. Distributing the horizontal and vertical interswitch links into metal 11 and metal 12 respectively, the metal resources required to implement the horizontal wires equals to $7 \%$ of the total metal resources available in the metal 11. Also, the metal resources required to the vertical wires equals to $7 \%$ of the total metal resources available in the metal 12 . Therefore, the total metal resources required to implement the HT-CLICHÉ architecture will be $14 \%$. The increasing percentage of the metal resources for HT-CLICHE is negligible as compared to the metal resources.
Since the interswitch links is short enough, there is no need for repeaters within the interconnects, the power and metal resources consumed by CLICHÉ and HT-CLICHÉ are shown in Table 5 for different technology nodes. With the advance in technology, the power dissipation required to implement the HT-CLICHÉ is increased by less than $2 \%$ of the total power consumption of the CLICHÉ architecture. The percentage of metal resources for HTCLICHÉ is increased by $35 \%$ as compared with the metal resources of CLICHÉ. The HTCLICHÉ trades extra metal resources for higher throughput.

| Technology node | No. of IPs | Percentage of power consumption of interswitch links and repeaters for CLICHÉ (\%) | Percentage of power consumption of interswitch links and repeaters for HT-CLICHÉ (\%) | Percentage of CLICHÉ <br> metal resources (\%) | Percentage of HT-CLICHÉ metal resources (\%) |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 130 nm | 361 | 7.6 | 14.1 | 21 | 43 |
| 90 nm | 729 | 4.8 | 9.1 | 22 | 44 |
| 65 nm | 1849 | 2.7 | 5.2 | 28 | 57 |
| 45 nm | 5625 | 1.4 | 2.7 | 36 | 71 |

Table 5. Power consumption of interswitch links and repeaters for HT-CLICHÉ and CLICHÉ

### 5.2.3 HT- Octagon

The HT-Octagon architecture has the same number of switches, but the number of wires and repeaters will be doubled. 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 interswitch links is obtained. According to the critical interconnect length (Li et al., 2005), the number of repeaters equals to 7680 repeaters. The power consumption required to implement the Octagon and HT-Octagon architectures is presented in Table 6. Due to the extra interswitch
links required to implement HT-Octagon architecture, the power consumption is increased by $6 \%$ as compared with the power consumption of Octagon topology.
By distributing the wiring of HT-Octagon architecture into the metal 11, the total length of wires needed equals to 7057.92 mm . The architecture consumes $8 \%$ of the total metal resources available in the metal 11. In the proposed design, the double number of interswitch links is utilized to implement HT-Octagon architecture. Therefore, the total metal resources required to implement the proposed architecture will be $16 \%$.

| Architecture | Power <br> consumption <br> of interswitch <br> links and <br> repeaters <br> $(\mathrm{mw})$ | Power <br> consumption of <br> switches $(\mathrm{mw})$ | Total power <br> dissipation $(\mathrm{mw})$ | Percentage of <br> power dissipation <br> of repeaters and <br> interswitch links <br> $(\%)$ |
| :---: | :--- | :--- | :---: | :---: |
| Octagon | 1094.12 | 19861.04 | 20955 | 5.2 |
| HT-Octagon | 2188.24 | 19844.08 | 22072.3 | 9.9 |

Table 6. power consumption of switches for Octagon and HT-Octagon
The percentage of power consumption and metal resources required to implement the Octagon and HT-Octagon networks in different technologies are shown in Table 7. By increasing the number of IP blocks with the advance in technology, the extra power consumption required to implement the proposed architecture is decreased. The extra power consumption is $2 \%$ of the total power consumption of the Octagon architecture. The percentage of extra metal resources for HT-Octagon is $25 \%$ of the available metal resources.

|  |  | Percentage of <br> power <br> consumption <br> of interswitch <br> Tinks and <br> repeaters for <br> Octagon (\%) | Percentage of <br> power <br> consumption of <br> interswitch <br> links and <br> repeaters for <br> HT-Octagon <br> (\%) | Percentage of <br> Octagon <br> metal <br> resources (\%) | Percentage of <br> HT-Octagon <br> metal <br> resources (\%) |
| :---: | :---: | :--- | :--- | :--- | :--- |
| $\mathbf{1 3 0 ~ \mathbf { ~ m ~ }}$ | 361 | 7.6 | 14.1 | 13 | 26 |
| $\mathbf{9 0 ~ \mathbf { ~ m ~ }}$ | 729 | 4.78 | 9.1 | 13 | 26 |
| $\mathbf{6 5 ~ \mathbf { ~ m ~ }}$ | 1849 | 2.8 | 5.4 | 17 | 35 |
| $\mathbf{4 5 ~ \mathbf { ~ m ~ }}$ | 5625 | 1.6 | 3.1 | 25 | 50 |

Table 7. Power consumption of interswitch links and repeaters for HT-Octagon and Octagon

### 5.2.4 HT-SPIN

By applying the high throughput architecture on SPIN topology, the length of interswitch links and number of repeaters are calculated by eq. (22) and eq. (23) respectively. Considering a system of 256 IP blocks, the number of repeaters equals to 12288 repeaters. The area of repeaters required to implement the HT-SPIN interswitch links equals to 267264 $\mu \mathrm{m}^{2}$ (it equals to the double area of repeaters required for SPIN interswitch links). The horizontal wires and vertical wires are distributed into metal 11 and metal 12 respectively. The length of horizontal wires needed consumes $28 \%$ of the total metal resources available in the metal 11. The vertical wires needed consume $28 \%$ of the total metal resources available in the metal 12. The total metal resources required to implement the proposed HT-

SPIN architecture will be $56 \%$.The power consumption of interswitch links, repeaters and switches required to implement the SPIN and HT-SPIN is presented in Table 8. The extra power dissipation required by the interswitch links and repeaters for HT-SPIN architecture (with 256 IPs) equals to $15 \%$ as compared with the total power dissipation.

| Architecture | No. of <br> repeaters | Power <br> dissipation of <br> repeaters and <br> interswitch <br> links $(\boldsymbol{m w})$ | Power <br> dissipation of <br> switches $(\boldsymbol{m w})$ | Total power <br> dissipation <br> $(\boldsymbol{m w})$ | Percentage of <br> power <br> dissipation of <br> repeaters and <br> interswitch <br> links $(\%)$ |
| :---: | :---: | :--- | :--- | :--- | :--- |
| SPIN | 12288 | 10612.99 | 32263.68 | 42876.67 | 24.75 |
| HT-SPIN | 24576 | 21225.98 | 32280.96 | 53506.94 | 39.67 |

Table 8. power consumption of repeaters and switches for SPIN and HT-SPIN
For different technologies, the power consumption and metal resources required to implement the SPIN and HT-SPIN are shown in Table 9. With the advance in technology, the extra power consumption required to achieve the proposed HT-SPIN architecture is $15 \%$ of the total power consumption of the architecture. The percentage of extra metal resources needed is more than $100 \%$ of the metal resources. Therefore, the overhead in the HT-SPIN is high. Applying the high throughput architecture on the SPIN topology is not recommended.

| Technology node | No. of IPs | Percentage of power consumption of interswitch links and repeaters for SPIN (\%) | Percentage of power consumption of interswitch links and repeaters for HT-SPIN (\%) | Percentage <br> of SPIN <br> metal <br> resources <br> (\%) | Percentage of HT-SPIN metal resources (\%) |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 130 nm | 400 | 41.2 | 58.4 | 21 | 42 |
| 90 nm | 800 | 33.6 | 50.3 | 30 | 59 |
| 65 nm | 2000 | 32.6 | 49.1 | 59 | 118 |
| 45 nm | 6000 | 28.1 | 43.8 | 126 | 253 |

Table 9. Power consumption of interswitch links and repeaters for HT-SPIN and SPIN

Since the proposed architecture increases the power dissipation, a low power NoC switch is proposed in Section 6.

## 6. Low power noc switch design

The switch of BFT has six ports, four children ports and two parent ports. Each port can be used as either input port or output port. If the port considers as input port, the input virtual channels, header decoder and crossbar are active. If the port considers as output port, the output virtual channels are active.
In the proposed design, only one part (input part or output part) is activated as shown in Fig. 11. The stand-by transistors (M1) disconnect the input circuit from the supply voltage
during the output mode. The stand-by transistors (M2) disconnect the output circuit from the supply voltage during the input mode. There is no need for the new control signals to control the stand-by transistors (M1 and M2). The acknowledgment signals (Ack_in and Ack_out) developed by the control unit are used to control the stand-by transistors M1 and M2 respectively. Using the number of virtual channels ( $N_{V C}$ ), The number of stand-by transistors equals to $3+2 N_{V C}$. The number of virtual channels is limited (it is not more 16 virtual channels (Abd El Ghany et al., 2009a)). By comparing the number of stand-by transistors with the total number of transistors required to implement the NoC port (as described is Section 2), the number of stand-by transistors is less than $1 \%$ of the total number of transistors. Therefore, the area overhead in the proposed design is negligible as compared to the area of NoC switch. The total power dissipation can be reduced by using power gating technique.


Fig. 11. proposed design for low power NoC port.
Using the Cadence tools and 90 nm technology node, the proposed low power NoC switch is implemented. The power dissipation of BFT switch is determined. The total power dissipation of the BFT switch equals to 41.29 mW . The total power dissipation of the port during the input mode equals to 6.79 mW . The total power dissipation of the port during the output mode equals to 6.57 mW . In the proposed BFT switch design of one virtual channel, the power dissipation of the main components of the port for the active mode and sleep mode is obtained as shown in Table 10. According to the mode of operation, the activation of the component is determined. In the Input mode, the input FIFO, header decoder and crossbar are activated, while the output FIFO is switched to sleep mode. The power dissipation of the port will be 5.68 mW . In the output mode, the output FIFO is activated while the input FIFO, the header decoder and cross bar are switched to sleep mode. The power dissipation of the port equals to 3.79 mW . Therefore, the average power dissipation of the proposed switch equals to $29,59 \mathrm{~mW}$. The average power dissipation of the proposed BFT switch is decreased by $28.32 \%$ as compared to the average power dissipation of the conventional BFT switch.

| Component | power dissipation in <br> active mode <br> $(\boldsymbol{m W})$ | power dissipation in <br> sleep mode <br> $(\mu \mathrm{W})$ | Percentage of <br> reduction in power <br> dissipation (\%) |
| :---: | :---: | :---: | :---: |
| Input FIFO | 3.618 | 0.1029 | 97.15 |
| Header decoder | 0.955 | 0.2157 | 77.41 |
| Crossbar | 0.473 | 0.1274 | 73.07 |
| Output FIFO | 3.562 | 0.1003 | 97.18 |

Table 10. the power dissipation of the main components of the BFT switch
The power consumption of BFT switch increases with the number of virtual channels as shown in Fig. 12. Applying the leakage power reduction technique on the BFT with different number of virtual channels, the power reduction increases with the number of virtual channels. The percentage of power reduction equals to $28 \%$ when the number of virtual channels equals to one. The percentage of power reduction of BFT switch with 12 virtual channels equals to $45 \%$. Increasing the number of virtual channels can improve the throughput in an on- chip interconnect network. By optimizing the design on the circuit levels, the high throughput can be provided by eight virtual channels (Abd El Ghany et al., 2009a). Using the leakage power reduction technique, the power consumption of BFT switch with 8 virtual channels is reduced by $44 \%$.
With the advance in technology, the number of IPs implemented in the same system size is increased. The effect of power gating technique on the HT-BFT is presented in Fig. 13. The power consumption of HT-BFT architecture using the leakage power reduction technique (HT-BFT-PR) is less than the power consumption of the conventional BFT architecture.


Fig. 12. power dissipation of a switch with different number of virtual channels.


Fig. 13. power dissipation of the HT-BFT using the power reduction technique
The power consumption $P_{\text {switches }}$ of switches for different high throughput architectures is obtained as shown in Table 11. The power consumption of these switches is more than $80 \%$ of the total power consumption of the on chip network. Switching off the power supply is an efficient technique to reduce the total power dissipation of NoC. The minimum power consumption can be obtained by using the BFT architecture as presented in Table 11. Using the leakage power reduction technique, the power consumption for different NoC architectures is determined. The overall power consumption, includes the power consumption of the interswitch links and repeaters, is decreased up to $33 \%$.

| Network <br> architecture | Total <br> power <br> $(\boldsymbol{m W})$ | $\mathbf{P}_{\text {switches }}$ |  | Total power using <br> power reduction <br> technique $(\boldsymbol{m W})$ | Percentage of <br> power <br> reduction |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | $m W$ | $\%$ |  |  |
| HT-BFT | 18591.32 | 15674.84 | 84 | 15104.44 | $19 \%$ |
| HT-SPIN | 53506.94 | 32280.96 | 60 | 46312.7 | $13 \%$ |
| HT-CLICHÉ | 26148.64 | 24471 | 94 | 18608.16 | $29 \%$ |
| HT-Octagon | 22072.32 | 19884.08 | 90 | 16253.04 | $26 \%$ |

Table 11. the total power consumption of different network architectures with 256 IPs

## 7. Conclusions

In this chapter, the high throughput NoC architecture is proposed to increase the throughput of the switch in NoC. The proposed architecture can also improve the latency of the network. The proposed high throughput interconnect architecture is applied on different NoC architectures. The architecture increases the throughput of the network by more than $38 \%$ while preserving the average latency. The area of high throughput NoC switch is decreased by $18 \%$ as compared to the area of BFT switch. The total metal resources required to implement the proposed high throughput NoC is increased by less than $10 \%$ as compared to the metal resources required to implement the conventional NoC design.
The power characterization for different high throughput NoC architectures is developed. The extra power consumption required to achieve the proposed high throughput NoC architecture is less than $15 \%$ of the total power consumption of the NoC architecture. Low power switch design is proposed. The power reduction technique is applied to different high
throughput NoC architectures. The technique reduces the overall power consumption of the network by up to $29 \%$.
The relation between throughput, number of virtual channels and switch frequency is analyzed. The simulation results demonstrate the performance enhancements in terms of throughput, number of virtual channels, switch frequency and power dissipation. It is shown that optimizing the circuit can increase the number of virtual channels without degrading the frequency. The throughput of different NoC architectures is also improved with the proposed architecture. The minimum power consumption and the minimum area can be obtained by using HT-BFT as compared to other high throughput NoC architectures. The extra metal resources required to achieve the proposed HT-BFT is negligible as compared to the metal resources of the network. The extra power consumption required to achieve the proposed HT-BFT is eliminated by using the leakage power reduction technique.

## 8. References

Abd El Ghany, M. A.; El-Moursy, M. \& Ismail, M. (2009a) "High Throughput Architecture for High Performance NoC" Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS), May, 2009 (in publication)
Abd El Ghany, M. A.; El-Moursy, M. \& Ismail, M. (2009b) "High Throughput Architecture for CLICHÉ Network on Chip" Proceedings of the IEEE International SoC Conference, September, 2009
Benini, L. \& Micheli, G. de (2002) "Networks on chips: A new SoC paradigm," IEEE Computer, vol. 35, no. 1, pp. 70-78, Jan. 2002
Bertozzi, D.; Jalabert, A. \& Murali, S. et al., (2005) "NoC synthesis flow for customized domain specific multiprocessor systems-on-chip," IEEE transactions on Parallel and Distributed Systems, vol. 16, no. 2, pp. 113-129, February 2005
Bolotin, E.; Cidon, I.; Ginosar, R. \& Kolodny, A. (2004) "QNoC: QoS architecture and design process for network on chip," Journal of Systems Architecture, vol. 50, no. 2-3, pp. 105-128, February 2004
Dally, W. J. \& Towles, B. (2001) "Route packets, not wires: on-chip interconnection networks", In Proceedings of Design Automation Conference, pp 684-689, June 2001
Dehon, A. (2000) "Compact, Multilayer layout for butterfly fat-tree", In Proceedings of The 12th ACM Symposium on Parallel algorithm Architectures, pp. 206- 215, July 2000
El-Moursy, M. A. \& Friedman, E. G. (2004) "optimum wire sizing of RLC interconnect with repeaters", Integration, the VLSI journal, vol. 38, no. 2, pp. 205-225, 2004
Grecu, C.; Pande, P. P.; Ivanov, A. \& Saleh, R. (2004a) "Structured Interconnect Architecture: A Solution for the Non-Scalability of Bus-Based SoCs," Proceedings of Great Lakes Symposium on VLSI, pp. 192-195, April 2004
Grecu, C.; Pande, P. P.; Ivanov, A. \& Saleh, R. (2004b) "Ascalable Communication-Centric SoC Interconnect Architecture", In Proceedings of IEEE International Symposium On Quality Electronic Design, pp. 22- 24, March, 2004
Grecu, C.; Pande, P.; Ivanov, A.; Marculescu, R.; Salminen, E. \& Jantsch, A. (2007a) "Towards open network-on-chip benchmarks," In Proceedings of the International Symposium on Network on Chip, pp. 205, May 2007

Grecu, C.; Ivanov, A.; Saleh, R. \& Pande, P. (2007b) "Testing network-on-chip communication fabrics," IEEE transactions Computer-Aided Design of Integrated Circuits and Systems, vol. 26, no. 10, pp. 2201-2014, December 2007
Guerrier, P. \& Greiner, A. (2000) "A generic architecture for on-chip packet-switched interconnections", In Proceedings of Design, Automation and Test in Europe Conference and Exhibition, pp. 250-256, March 2000
ITRS 2007 Documents, http://itrs.net/Links/2007ITRS/Home2007.htm
Kao, J. T. \& Chandrakasan, A. P. (2000) "Dual-Threshold Voltage Techniques for Low-Power Digital Circuits ", IEEE Journal of Solid-State Circuits, vol. 35(7), pp. 1009-1018, July 2000
Karim, F.; Nguyen, A. \& Sujit Dey, (2002) "An Interconnect Architecture for Networking Systems on Chips," IEEE Micro, vol. 22, no. 5, pp. 36-45, September 2002
Khellah, M. M. \& Elmasry, M. I. (1999) " Power minimization of high-performance submicron CMOS circuit using a dual-V/sub dd/ dual-V/sub th/ (DVDV) approach" In Proceeding of the International symposium on Low Power Electronics and Design, pp. 106-108, 1999
Kim, J.-S.; Hwang, M.-S \& Roh, S. et al., (2004) "On-chip network based embedded core testing," In Proceedings of the IEEE International SoC Conference, pp. 223-226, September 2004
Kumar, S. et al., (2002) "A Network on Chip Architecture and Design Methodology," In Proceedings of the IEEE Computer Society Annual Symposium on VLSI, pp. 117-124, 2002
Kursun, V. \& Friedman, E. G. (2004) "Sleep switch dual threshold voltage domino logic with reduced standby leakage current," IEEE transactions on VLSI systems, 12(5), pp. 485496, May 2004
Lee, S.-J.; Song, S.-J. \& Lee, K. et al. (2003) "An 800MHz Star-Connected On-Chip Network for Application to Systems on a chip", IEEE Digest of International Solid State Circuits Conference, vol. 1, pp. 468-489, February, 2003
Lee, K.; Lee, S.-J. \& Kim, S.-E. et al. (2004) "A 51mW 1.6GHz On-Chip Network for Low power Heterogeneous SoC Platform", IEEE Digest of International Solid State Circuits Conference, vol. 1, pp.152-518, February, 2004
Lee, S.-J.; Kim, K. \& Kim, H. et al. (2005) "Adaptive Network-on-Chip with Wave-Front Train Serialization Scheme", IEEE Digest of Symposium on VLSI Circuits, pp. 104-107, June, 2005
Lee, S.-J.; Lee, K. \& Yoo, H.-J. (2005)"Analysis and Implementation of Practical CostEffective Network-onChips", IEEE Design $\mathcal{E}$ Test of Computers Magazine (Special Issue for NoC), September 2005
Lee, K.; Lee, S.-J. \& Yoo, H.-J. (2006) "Low-Power Networks-on-Chip for High-Performance SoC Design", IEEE Transactions on Very Large Scale Integration Systems, vol. 14, no.2, pp.148-160, February 2006
Lee, S. \& Bagherzadeh, N. (2006) "Increasing the Throughput of an Adaptive Router in Network-on Chip(NoC)", In Proceedings of $4^{\text {th }}$ International Conference on Hardware/Software Codesign and System Synthesis CODES+ISSS'06, pp. 82-87, Oct. 2006

Liang, J.; Laffely, A.; Srinivasan, S. \& Tessier, R. (2004) "An architecture and compiler for scalable on-chip communication," IEEE transactions on VLSI Systems, vol. 12, no. 7, pp. 711-726, July 2004
Li, X.-C.; Mao, J.-F.; Huang, H.-F. \& Liu, Y. (2005) "Global interconnect width and spacing optimization for latency, bandwidth and power dissipation," IEEE Transactions on Electron Devices, vol. 52, no. 10, pp. 2272-2279, Oct. 2005
Murali, S. \& De Micheli, G. (2004) "SUNMAP: A Tool for Automatic Topology Selection and Generation for NoCs", IEEE Proceedings of Design Automation conference, pp. 914-919, June 2004
Murali, S.; Theocharides, T. \& Vijaykrishnan, N. et al., (2005) "Analysis of error recovery schems for networks on chips," IEEE Design and test, vol. 22, no. 5, pp. 434-442, October 2005
Pande, P. P.; Grecu, C.; Ivanov, A. \& Saleh, R. (2003a) "Design of a Switch for Network on Chip Applications," In Proceedings of The 2003 International Symposium on Circuits and Systems, vol. 5, pp. 217-220, May 2003
Pande, P. P.; Grecu, C.; Ivanov, A. \& Saleh, R. (2003b) "High-Throughput Switch-Based Interconnect for Future SoCs", the $3^{\text {rd }}$ IEEE International workshop on SoC for real-time Applications, pp 304-310, July 2003
Pande, P. P.; Grecu, C.; Ivanov, A. \& Saleh, R. (2005a) "Design, synthesis, and test of networks on chips," IEEE Design and Test of Computer, vol. 22, no. 5, pp. 404-413, Aug. 2005
Pande, P. P.; Grecu, C.; Jones, M.; Ivanov, A. \& Saleh, R. (2005b) "Performance Evaluation and Design Trade-Offs for Network-on-Chip Interconnect Architectures", IEEE Transaction on Computers, vol. 54, no. 8, Aug. 2005
Salminen, E.; Kulmala, A. \& H"am"al"ainen, T. (2007) "On network-on-chip comparison," In Proceedings of the Euromicro conference on Digital System Design Architecture, August 2007, pp. 503-510


## Data Storage

Edited by Florin Balasa

ISBN 978－953－307－063－6
Hard cover， 226 pages
Publisher InTech
Published online 01，April， 2010
Published in print edition April， 2010

The book presents several advances in different research areas related to data storage，from the design of a hierarchical memory subsystem in embedded signal processing systems for data－intensive applications， through data representation in flash memories，data recording and retrieval in conventional optical data storage systems and the more recent holographic systems，to applications in medicine requiring massive image databases．

## How to reference

In order to correctly reference this scholarly work，feel free to copy and paste the following：

Mohamed A．Abd El Ghany，Magdy A．El－Moursy and Mohammed Ismail（2010）．High Throughput Architecture for High Performance NoC，Data Storage，Florin Balasa（Ed．），ISBN：978－953－307－063－6，InTech，Available from：http：／／www．intechopen．com／books／data－storage／high－throughput－architecture－for－high－performance－noc

## INTECH

open science｜open minds

## InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83／A
51000 Rijeka，Croatia
Phone：＋385（51） 770447
Fax：＋385（51） 686166
www．intechopen．com

## InTech China

Unit 405，Office Block，Hotel Equatorial Shanghai
No．65，Yan An Road（West），Shanghai，200040，China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone：＋86－21－62489820
Fax：＋86－21－62489821
© 2010 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License, which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

