Optimal Rate Control in Wireless Network on Chip (*WiNoC*)

Farhaad Raad
Islamic Azad Unversity
Iran
raadfarhaad@gmail.com

ABSTRACT: *With increasing the number of cores in the network on chip infrastructure, high latency and power consumption are emerged due to multi-hop communication between two distant cores. As a result, wireless NoC (WiNoC) architectures, have been recently proposed. In the WiNoC, wireless links have higher bandwidth, less delay and higher flexibility rather than those in wire links in the NoC. In this paper, the rate control problem in the WiNoC has investigated using nonlinear optimization theory and can be written as a utility-based optimization problem. Each core is associated with a utility which defined as a function of its rate. The aim of this study is to maximize the aggregate utility of all cores in the WiNoC, subject to capacity constraints of links. Finally, this primal problem is converted to its dual problem and solved based on the gradient projection method. A synchronous iterative algorithm is proposed to control the rate of cores and can be implemented by a centralized controller with low overhead.*

Keywords: Wireless Network on Chip, Rate Control, Utility Based Optimization, Congestion Control

Received: 20 August 2017, Revised 28 September 2017, Accepted 4 October 2017

1. Introduction

Network-on-chip (*NoC*) is a new communication network with set of tiles (router + IP core) that enables a high degree of integration of multi-cores in systems-on-chip (*SoC*) [1]. Sending a packet between two distant cores in *NoCs* can suffers from high latency and power consumption. So reducing latency and power consumption can be a problem to be solved. This problem can be handled by adding wireless links with high bandwidth and low delay in NoC. This idea provides a new opportunity for the design of wireless NoCs (*WiNoCs*) with on-chip antennas in base routers with suitable area and power consumption [2].

The WiNoC architectures can be classified into four classes depending on their characterized antenna and frequency range operations [13]. In *Ultra Wide Band* (*UWB*) [3], an *UWB* transceiver based on carrier free impulse has been proposed for *WiNoCs*. The antenna used meander type dipole antenna with 1mm data transmission range. 1.16 Gb/s data rate can be reached for a single channel at a central frequency of 3.6 G23Hz. The *mm-wave WiNoCs* is proposed in [2]. Their results have shown that the on-chip antenna for the *mm-wave WiNoCs* had the best power gain with smallest area overhead. Designing miniature antennas and simple transceivers that operate in the *sub-THz* frequency range for wireless communication in network on chip has been demonstrated in [4]. The authors in that paper predicted that it is possible to have 16 channels without overlapping in the frequency range of 100–500 GHz for the *sub-THZ WiNoC*. Channel rate is 10– 20 Gb/s in 32-nm CMOS process. In [5], carbon nanotubes (*CNTs*) in the THz/optical frequency range has been investigated for antenna both theoretically and experimentally. These antennas have a bandwidth of around 500 GHz, whereas antennas operating in the millimeter wave range achieve bandwidths of 10GHz. Hence, *CNT antenna-based wireless* link is proposed for *WiNoC* design. Design of suitable on-chip antennas, transceiver circuits and wireless routers enable efficient wireless links with such a low delay and high bandwidth that can be used optimally to maximize overall network performance without excessive overhead. In [2], the authors classify the *WiNoC* architectures into two categories; mesh-based network topology where the wireless links are in top of a regular mesh and small-word-based network based that is a hybrid topology. Our *WiNoC* architecture that is used in this paper is based on mesh topology.

2. Related Works

Congestion in *NoC* causes throughput to be dropped and packet latency to be increased. Hence, congestion should be controlled to avoid network traffic reaching saturation point. Congestion control in on-chip networks is a novel issue. In [6-8], this problem has been shown. In [6], a prediction-based congestion control strategy for on-chip networks has been proposed where each router bases on buffer occupancies decide detect future congestion problems. In [7], the link utilization has been used as congestion measure and the controller determines the appropriate loads for the Best-Effort sources. DyAD algorithm in [8] uses a hybrid routing algorithm for solving the local congestion problem. In [11], a distributed congestion control mechanism is proposed for the NoC. It uses the number of packets in the port buffer to detect congestion. In [9-10], a global congestion control mechanism models the source rates as the solution to a Delay-sum or rate utility maximization problem, which is solved by a proposed iterative algorithm. Also, stability and throughput-fairness in this algorithm has been addressed. A distributed flow control and buffer management strategy to improve the *WiNoC* endto- end performance is proposed in [12].

In this paper, a fast forwarding by prioritized MAC contention and dynamic virtual output queuing with buffer-sharing is proposed. In the previous studies the optimization based rate control problem in the *WiNoC* has not been investigated completely until now. Therefore, the aim of this study is to propose a solution to this problem.

This paper is organized as follows. In Section III, we present the system model and formulate the underlying optimization problem for congestion control. In Section IV, we solve the optimization problem using an iterative algorithm and propose the solution as a centralized congestion control algorithm to be implemented as a controller. In Section V, we analyze the convergence behavior of the proposed algorithm and prove the underlying theorem of its convergence and present the simulation results. Finally, the section VI concludes the paper.

3. System Model

Our *WiNoC* architecture is constructed by adding antenna in some of routers in mesh topology. These routers have wireless link, in addition to the original wired links. Therefore, wireless routers are capable of transferring packets via both wireless channels and wire signals. Fig.1 shows an example of 5*5 *WiNoC* that three wireless routers are randomly distributed in a mesh topology. Dotted lines and solid lines represent wireless and wire links that transmit packets between routers. Finding the optimal placement of wireless routers and the number of wireless routers in *The WiNoC* topology can be our future studies.

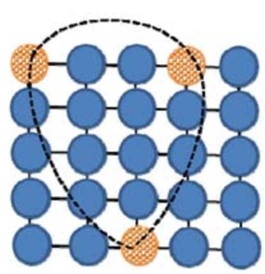

Figure 1. *WiNoC* Topology

The *WiNoC* architecture such as *NoC* is assumed to be lossless, and packets traverse the network on a shortest path using a deadlock free routing algorithm the proposed in next section. Finding the optimal rate allocation in *The WiNoC* can be understood as a utility-based resource allocation problem. Here, each IP core is associated with a utility, which is defined as a function of the IP's rate. The "Aggregated Utility ($\sum U_{IP}(x_{IP})$) represents the sum of all IPs' utilities with regard to a rate allocation.

We model the rate control problem in *The WiNoC* as the solution to a nonlinear optimization problem. A mathematical model just like the ones presented in [14] [9], is presented for *The WiNoC* architecture. Let us suppose that the *WiNoC* consisting of *K* IP's denoted by $K = \{1, 2, ..., K\}$ and *L* wires and wireless network links, denoted by $L = \{1,2,..., L\}$ .The bandwidth capacity of each link $l \in L$ is c_l packets/sec. In our simulation, we assume that the capacities of wireless links and wire links are 2 Gb/s and 1Gb/s respectively. We collect them into a link capacity vector $c = c_l\ , l \in L$ .We denote the set of sources that share link *l* by *K(l)*. Similarly, the set of links that source *k* passes through, is denoted by *L(k)*. Each IP has a rate *x* . We collect them into a rate vector $x = (x_k\ , k \in K)$.

Now, we define an $L*K$ matrix *A*.

$A_{lk} = 1$, if source *k* goes through the link *l*; otherwise $A_{lk} = 0$.

A is the link usage pattern of a network. In our mode, *A* is determined by *routing* in *The WiNoC* network. We know that the sum rate of all IP cores that go through the link *l* should not exceed its capacity C_l. Formally, it is capacity constraint that expressed as follows:

$$A * x \le c \qquad (1)$$

We collect the above notations into Table 1.

Notation	**Definition**
$k \in K = \{1, 2, ..., K\}$	IP's core
$l \in L = \{1, 2, ..., L\}$	Wireless and wire link
$x = (x_k\ , k \in K)$	Rate of $k \in K$
$c = (cl\ , l \in L)$	Link Capacity of $l \in L$
$L(k) \subseteq L$	Set of Link that *k* Goes Through
$K(l) \subseteq K$	Set of Source that Go Through *l*
$A = (A_{lk})_{L*K}$	Link Capacity Constraint Matrix

Table 1. Notation in section III

4. Optimal Rate Allocation

Our objective in *The WiNoC* network is to choose source rates x_k, for each core so that to maximize the sum of utilities of all cores. Hence the maximization problem can be formulated as [14]:

$$\max \sum_{k \in K} U_k(x_k) \qquad (2)$$

subject to :

$$A \cdot x \leq c \tag{3}$$

$$x_k \succ 0 \quad \forall k \in K \tag{4}$$

The assumptions aboutU_k are [14]:

- The utility functions U_k are increasing, strictly concave, and twice continuously differentiable.

- The curvatures of U_k satisfy the following condition:

$$-U_k''(x_k) \geq 1/B_f \succ 0 \quad \forall k \in K \tag{5}$$

U_k is additive, so that the aggregated utility of rate allocation $x = (x_k, k \in K)$ is $\sum_{k \in K} U(x_k)$.

- U_k in the economics literature is referred to as utility function [14-15]; hence Eq. 2 is called a utilitymaximization problem. There are many choices for utility function but one of the popular forms of utility functions is logarithmic, which satisfy *Proportional Fairness* [15].

The remainder of this paper shows that solving this problem and achieves optimality with respect to rate allocations among IP cores. By nonlinear optimization theory, there exists a maximizing value of argument *x* for the above Eq. 2, which can be solved by Lagrangian method [14]. Let us consider the Lagrangian form of this optimization problem:

$$L(x, \rho^{\alpha}) = \sum U_k(x_k) - \rho^{\alpha}(A.x - c) \tag{6}$$

$\rho^{\alpha} = (\rho_l^{\alpha}, l \in L)$ is vectors of Lagrangian multipliers. Eq. 6 can be further derived as follows:

$$L(x, \rho^{\alpha}) = \sum U_k(x_k) - \sum \rho_l^{\alpha} (\sum_{k \in K} A_{lk} x_k - c_l) = \sum U_{\text{k}}(x_k) - \sum x_k \sum \rho_l^{\alpha} A_{lk} + \sum_{l \in L} \rho_l^{\alpha} c_l \tag{7}$$

$$L(x, \rho^{\alpha}) = \sum U_k(x_k) - \sum \rho_l{}^{\alpha} (\sum_{s \in S} A_{lk} x_k) + \sum_{\text{l} \in L} \rho_{\text{l}}^{\alpha} \text{c}_1 = \sum U_k(x_k) - \sum x_k \sum_{l \in L} \rho_l{}^{\alpha} A_{lk} + \sum_{\text{l} \in L} \rho_{\text{l}}^{\alpha} \text{c}_1 \tag{8}$$

We then define new vectors $\lambda^{\alpha} = (\lambda_k^{\alpha}, k \in K)$ as follows:

$$\lambda_l^{\alpha} = \sum_{l \in L} \rho_l^{\alpha} A_{lk} = \sum_{l \in L(k)} \rho_l^{\alpha} \tag{9}$$

For ρ^{α}, ρ_l^{α} can be understood as the *link price* of *l*. Consequently, λ^{α} and λ_l^{α} are the summation of prices of all links that source *k* goes through, or the network price that source *s* has to pay [15]. Eq. 8 becomes

$$L(x, \rho^{\alpha}) = \sum U_k(x_k) - (\lambda^{\alpha} x) + \rho^{\alpha} c \tag{10}$$

According to the Duality Theory [16], each convex optimization problem has a dual whose optimal solution can lead to the optimal solution of the main problem. As the dual problem can be defined in such a way to be unconstrained, solving the dual is much simpler than the primal problem. In next section, we will obtain the dual of Eq.2 and solve it using gradient projection method. In the following, we will propose an iterative algorithm for rate control in IP cores.

4.1 Dual Problem

The dual problem of Eq. 2 is as follows:

$$\begin{aligned} &D : \min D(\rho^{\alpha}) \\ &where \\ &D(\rho^{\alpha}) = \max_{x} L(x, \rho^{\alpha}) \\ &= \max_{x} \sum (U_k(x_k) - \lambda_k^{\alpha}) + \sum_{l \in L} \rho_l^{\alpha} c_l \end{aligned} \tag{11}$$

We define Eq. 12 as below:

$$D(\rho^{\alpha}) = \sum_{k \in K} \max\{U_k(x_k) - \lambda_k^{\alpha})\} + \sum_{l \in L} \rho_l^{\alpha} c_l \tag{12}$$

By assumption that U_k is strictly concave and twice continuously differentiable, we want to maximize:

$$U_k^{'}(x_k) - \lambda_k^{\alpha} = 0 \tag{13}$$

$$x_k(\rho^{\alpha}) = \left[U_k^{'-1}(\lambda_k^{\alpha})\right]_{\min}^{\max} \tag{14}$$

If ρ^{α} is dual optimal, then $x_k(\rho^{\alpha})$ is also primal optimal.

Now, if the optimal price ρ^{α} is available, the optimal rate x_k^* can be achived by solving Eq. 13.

4.2 Rate control Algorithm

The dual problem D is solved by using the gradient projection method. In gradient method, the value of ρ^{α} is adjusted in opposite direction to the gradient $\nabla D(\mu^{\alpha})$ .We have

$$\rho_l^{\alpha}(t+1) = \left[\rho_l^{\alpha}(t) - \psi \frac{\partial D(\rho^{\alpha}(t))}{\partial \rho_l^{\alpha}}\right]^{+} \tag{15}$$

$$\frac{\partial D(\rho^{\alpha}(t))}{\partial \rho_l^{\alpha}} = c_l - \sum_{k \in K(l)} x_k(\rho^{\alpha}) \tag{16}$$

$$\rho_l^{\alpha}(t+1) = \left[\rho_l^{\alpha}(t) + \psi(\sum x_k(\rho^{\alpha}(t) - c_l)\right]^{+} \tag{17}$$

where ψ, is step size. Step size has an important role on the convergence behavior of the Eq. 17. We have used two different scenarios about step-size value. The proposed algorithm can be presented in Table 2.

The proposed algorithm is a synchronous iterative algorithm because the values of rates and link prices should be computed only after all information of the current iteration (time) is signaled through the network. A distributed algorithm that allows both the source rate update and the link price calculation to be used in different time is one of the directions of our future works. Step sizes are chosen as $\psi = 3/(1+t)$ and $\psi = 1/(1+t)$ that be introduce in [17].

4.3 Deadlock Free Routing Algorithm

With increasing bandwidth and decreasing delay in this emerging *WiNoC* technology, a deadlock-free logic based routing scheme is essential for on-chip application [18-19]. In order to apply efficiently routing in *the WiNoC*, we assume that if there exists an XY path to a destination is the shortest one then this route will be selected; otherwise if there are two (or more) wireless routes or wireless addition to wire links are used. An important problem in this routing is cyclic dependency, inducing deadlocks which are not easy to get resolved. To ensure deadlock free routing in *The WiNoC*, we used octagon turn model [18]. The model involves two abstract cycles, a clockwise cycle and a counter clockwise cycle, each formed by eight turns. It prohibits certain turns in

Initialization:

1. *Initialize C_l of all links.*
2. *Set link price vector to zero.*

Main loop

Do Until $\max \left| x_k(t+1) - x_k(t) \right| \prec \theta$

Link Price Update:

a. *Received rates* $x_k(t)$ *from all sources* $k \in K(l)$

b. *Update price by*

$$\rho_l^{\alpha}(t+1) = \left[\rho_l^{\alpha}(t) + \psi(\sum x_k(\rho^{\alpha}(t) - c_l) \right]^{+}$$

c. *Send* $\rho_l^{\alpha}(t+1)$ *to all sources* $k \in K(l)$

Rate Adaption:

a. *Received link prices* $\rho_l^{\alpha}(t)$ *from all links* $l \in L(k)$

b. *Calculate* $\lambda_k^{\alpha}(t) = \sum_{l \in L(k)} A_{lk} \rho_l^{\alpha}$

c. *Adjust rate by* $x_k(t+1) = \left[U_k^{'-1}(\lambda_k^{\alpha}(t)) \right]_{\min}^{\max}$

d. *Send* $x_k(t+1)$ *to all links.*

Table 2. Algorithm

each cycle to break all the cyclic dependencies. The matrix *A*, in section *III* is made by this routing algorithm. Two rules are followed by the octagon turn model [18] is present below:

Rule 1. Any packet is not allowed to make the four turns i.e., *W* → *SE*, *N* →*SW*, *E*→*NW*, and *S* →*NE* at a node as in the clockwise abstract cycle.

Rule 2. Any packet is not allowed to make the four turns i.e., *NE*→ *S*, *NW*→*E*, *SW* → *N*, and *SE*→*W* at a node as in the counter clockwise abstract cycle.

5. Simulation Results

In this section we examine the proposed rate control algorithm, for a typical *The WiNoC* architecture. We have simulated a *WiNoC* with 4×4 Mesh topology which consists of 16router (4 wireless router and 12 wire router) communicating using 30 shared bidirectional links; wireless link has a fixed capacity of 2 Gbps and wire links has 1 Gbps. We assume packets traverse the network on a shortest path using a deadlock free routing that is proposed in section *C*.

In our simulation we have chosen *logarithmic* utility functions because this utility function satisfies fair conditions among cores and is *Weighted Proportionally-Fair* which is an important property in economics [15]. Initial rate all of sources are variable in interval [0-2] and initial link price is zero.

Variation of source rates for some source using two step sizes are shown in Fig. 2. Regarding Figure 2(a), it's apparent that after about 38 iterations, all source rates will be in the vicinity of the steady state point of the algorithm. However, for the second case, Figure 2 (b) reveals that at least 58 iterations needed to have source rates in the vicinity of the optimal point. Comparing Figure 2(a) and 2 (b), we realize that the initial value of the step size directly influences the rate of convergence. The source rates in different times (t = 20, t = 50, t = 150) also shown in Figure 3 for two scenario.

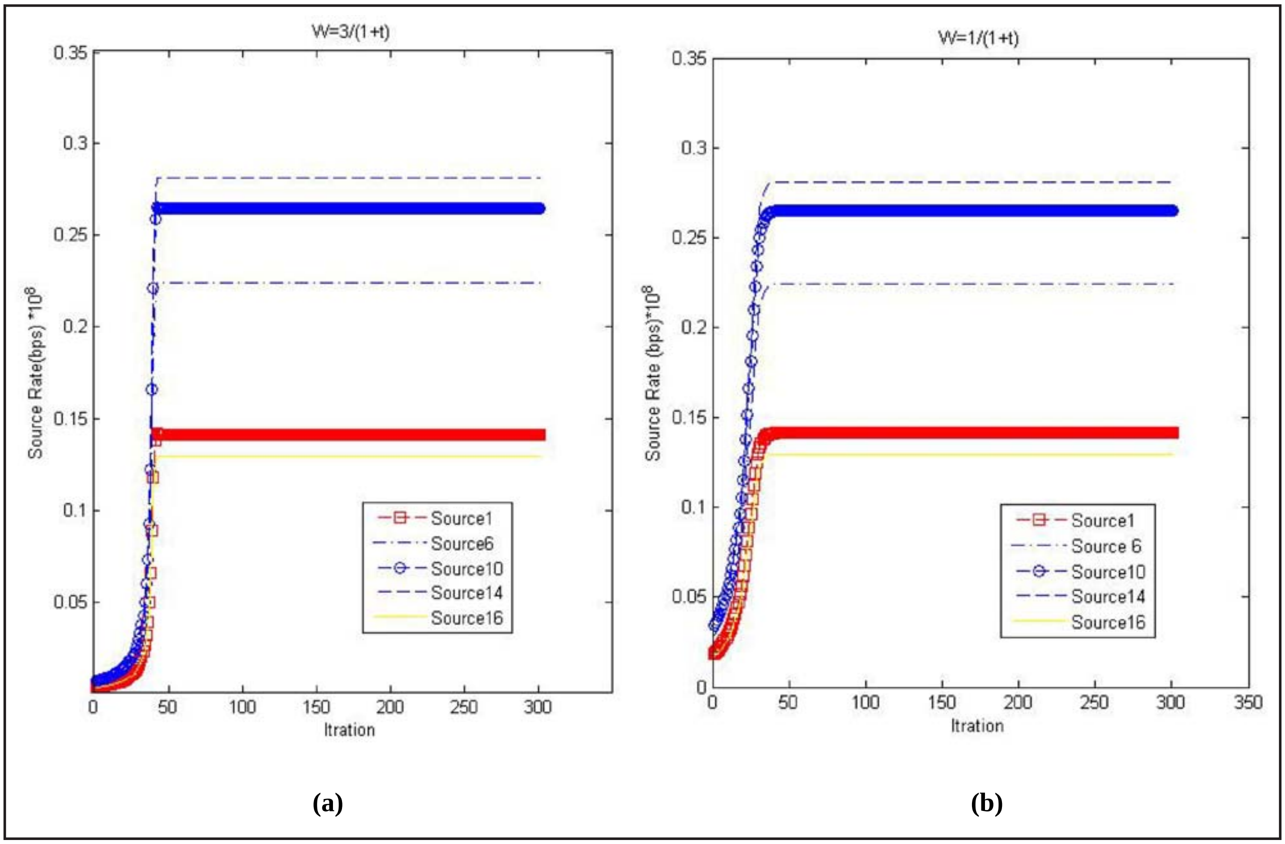

Figure 2. Source rate for (a) $\psi = 3/(1+t)$ and (b) $\psi = 1/(1+t)$

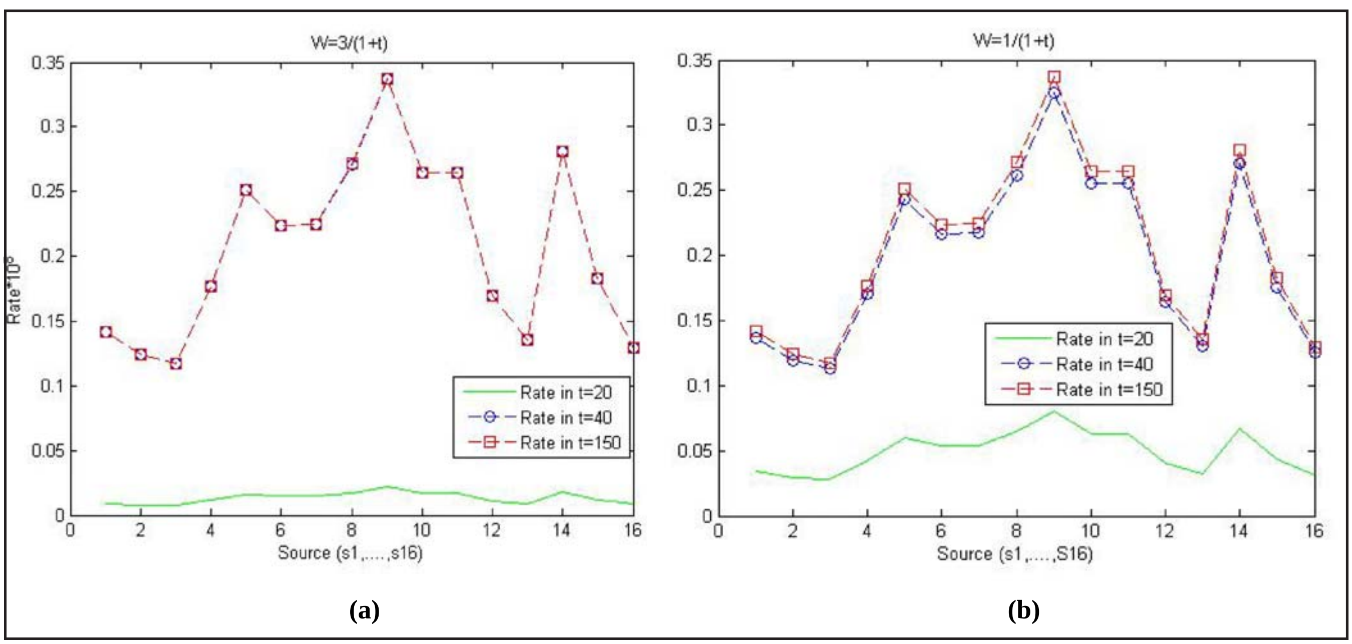

Figure 3. Source rates in different times

For proposed algorithm implementation; we consider a controller in the *WiNoC*. The necessary requirement of such a controller is the ability to accommodate mathematical operations as in Eq.14 and Eq.12 and the allocation of few dedicated links to communicate link price and new rates information to cores with a light load.

5. Conclusion and Future Work

In this paper the problem of rate control in the *WiNoC* has been investigated. The gradient projection method has been applied to solve the rate utility maximization problem. This iterative algorithm, can be used to determine the optimal rates in the *WiNoC* and can be implemented by a controller with low overhead. Further researches in the interest area in this field such as: optimal placement of wireless routers, convergence behavior, dynamic routing, asynchronous algorithm and the effect of different utility functions on the *WiNOC* can be considered as future studies.

References

[1] Benini, L., De Micheli, G. (2002). "Network on Chips: A New SoC Paradigm", *IEEE Computer*, *p* 70-78, January.

[2] Deb, S., Ganguly, A., Chang, K., Pande, P. (2010). "Enhancing performance of network-on-chip architectures with millimeter-wave wireless interconnects". *In*: Proceedings of the IEEE International Conference on ASAP. 73–80.

[3] Zhao, D., Wang, Y. (2008). "SD-MAC: Design and synthesis of a hardwareefficient collision-free QoS-aware MAC protocol for wireless networkon- chip," *IEEE Trans. on Computers Special Sec. on NoC*, 57 (9), p 1230–1245.

[4] Lee, S., Chang, M., Peng, C., Zhang, L. (2009). "A calable micro wireless interconnect structure for CMPs". In *Proceedingsof ACM Annual International Conference on Mobile Computing and Networking (MobiCom '09)*. 20– 25.

[5] Kempa, K., RYBCZYNSKI, J., Huang, Z., Wang, G. (2007)."Carbon nanotubes as optical antennae". *Adv. Mater. 19*, 421–426.2007.

[6] Yorozu, Y., Ogras, M. U. Y., Marculescu, R. (2006). "Prediction-based flow contro for network-on-chip traffic," *Design Automation Conference*, p 839-844.

[7] van den Brand, J. W., Ciordas, C., Goossens, K., & Basten, T. (2007). "Congestion-controlled best-effort communication for networks-onchip". *In*: Proceedings of the design, automation and test in Europe conference and exhibition (DATE) (p 1–6).

[8] Hu, J., Marculescu, R. (2004)." DyAD smart routing for networks on chip". In Proceedings of the 41st design automation conference (p 263–263).

[9] Talebi, M. S., Jafari, F., Khonsari, A. (2007). "A Novel Flow Control Scheme for Best Effort Traffic in NoC Based on Source Rate Utility Maximization," Modeling, Analysis, and Simulation of Computer and Telecommunication Systems, pp. 381-386, 2007.

[10] Talebi, M.S., Jafari, F., Khonsari, A., Yaghmaee, M. H. (2008). "Proportionally-fair best effort flow control in network-on-chip architectures", *International Parallel and Distributed Processing Symposium*, p. 1–8.2008.

[11] Huaxi Gu., Jiang Xu., KunWang. (2010). "A new distributed congestion control mechanism for networks on chip". *Telecommun Syst* 44: 321– 331 . DOI 10.1007/s11235-009-9257-7

[12] Wang, Y., Zhao, D. (2009). "Distributed flow control and buffer managemnent for wireless network on chip". In IEEE ISCAS,.

[13] Deb, V., Ganguly, A., Pande, P. P., Belzer, B., Heo, D. (2012). "Wireless NoC as Interconnection Backbone for Multicore Chips: Promises and Challenges", Ieee Journal on Emerging and Selected Topics in Circuits and Systems, 2 (2), June.

[14] Low, H. (1999). "Optimization Flow Control, I: Basic Algorithm and Convergence", *IEEE/ACM Transactions on Networking*, 7(6): 861-874.

[15] Kelly, F. P., Maulloo, A., Tan, D. (1998). "Rate control for communication networks: Shadow prices, proportional fairness, and stability," *Operational Research Society*, 49 (3), p 237-252.

[16] Boyd, S., Vandenberghe, L. (2004). "Convex Optimization". Cambridge, U.K, Cambridge Univ. Press.

[17] S. Boyd, "Convex Optimization II Lecture Notes", Stanford University, 2006.

[18] Zhao, D., Wu, R. (2012). " Overlaid Mesh Topology Design and Deadlock Free Routing in Wireless Network-on-Chip". Sixth IEEE/ACM International Symposium on Networks-on-Chip.

[19] Wang, C., Hu, W. H., Bagherizadeh, N. (2013)."scalable load balancing congestion-aware network on chip router architecture". *Journal of Computer and System Science* 79. 421-439.