A Heterogeneous System-level Power Macro-modeling for Network-on-Chip Router

Ning Wu, Fang Zhou, Fen Ge, and Yifeng Mu

Abstract—A heterogeneous macro-model for power consumption of the Network-on-Chip (NoC) router at system level is proposed, with higher accuracy to overcome the shortcoming of existing architecture-level power simulators, which is aimed to evaluate the network performance rapidly and guide the communication structure design. The power macro-model for each module of the router is extracted in accordance with its own characteristics. Input/output ports, routing calculation unit and crossbar switch are established by multiple linear regression method because of their single data flow state. Arbiter is established based on BP neural network due to its numerous states. Several experiments with different traffic loads and input sequences are carried out to verify the power model. Experimental results show that a significant improvement in evaluation speed is achieved with the proposed power model as compared with the gate-level simulation, while maintaining a minor average error of approximately 5.0%. As an application, we use the proposed model to evaluate the performance of different core mappings for H.264 decoder in system-level low power design.

Index Terms—BP neural network, heterogeneous, multiple linear regression, NoC, power model

I. INTRODUCTION

TO solve the shortcoming of the traditional bus in the area of communication efficiency, interconnect scaling and power consumption, etc., a new on-chip communication structure Network-on-Chip (NoC) has been proposed[1,2]. It aims to provide high performance communication for microprocessors, DSP, memory and other IP cores based on parallel computing and computer network design.

The high-performance communications in NoC is at the cost of an increase in the structure complexity, and an interconnection network dissipates a significant fraction of the total system power budget. Therefore, the communication power consumption has become an important constraint of the NoC design [3]. It is desirable to get detailed trade offs in power and performance early in the design flow, preferably at the system level. Although the commercial EDA tools are accurate in gate level power analysis, but there is limited ability in the system level design to obtain more accurate power consumption data, and the power models are extremely time consuming and intractable as far as power profile for complicated multi-processors are concerned. Therefore, system level power model extraction without going into the hardware details of components is needed. Routers are the most important communication components in NoC, whose power model is the focus of the relevant research.

In this paper, we propose a methodology for power model extraction of the NoC router based on the combination of the linear and nonlinear polynomial representations. The NoC router is decomposed into several components, and a power model is built for each component according to their different characteristics in order to improve the power model's accuracy. Linear regression is used to obtain the power models for the input buffer, the routing calculation unit and the crossbar switch due to their single data flow state. The power model for the arbiter is established based on BP neural network because of its numerous states.

The rest of the paper is organized as follows: Section II surveys the related work and states our contribution. Section III introduces the architecture of NoC router. The NoC power model is described in Section IV and Section V. Section VI presents the experimental results. In Section VII, we use the proposed model to evaluate the performance of different core mappings for H.264 decoder in system-level low power design. Finally, Section VIII concludes the paper.

II. RELATED WORKS

Existing system level power models of the router can be divided into theoretical models and statistical models.

The basic idea of theoretical models is to estimate the effective capacitance of each circuit and the amount of bit-flips in the data stream. In [4], Wang et al. created a network simulator named Orion to estimate dynamic power of router components. This simulator was augmented in [5] to support leakage power. Ye et al. [6] proposed an energy estimation flow to derive bit energy models that are used to evaluate different switch fabrics in network routers. However, the models are tightly coupled with circuit implementations. As such, these models cannot be migrated to different technology libraries without a large amount of re-modeling. Moreover, low level of abstraction (i.e. gate and device level) and the extremely slow simulation make it definitely unsuitable to face with system level SW/HW exploration task.

Ning Wu is with the College of Electrical and Information Engineering, Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing 210016, China (e-mail: wunee@nuaa.edu.cn).

Fang Zhou is with the College of Electrical and Information Engineering, NUAA, Nanjing 210016, China (e-mail: zfnuaa@nuaa.edu.cn).

Fen Ge is with the College of Electrical and Information Engineering, NUAA, Nanjing 210016, China (e-mail: gefen@nuaa.edu.cn).

Yifeng Mu is with the College of Electrical and Information Engineering, NUAA, Nanjing 210016, China (e-mail: zealotmyf@163.com).

Statistical models (i.e. power macro-models) are usually based on multiple linear regression analysis to generate power estimation. Its basic idea is to establish the functional relationship between the statistical properties of the input data and power consumed by routers through the implemented circuit. Power macro-models usually are expressed by the look-up table or empirical polynomial. The authors in [7] automated the extraction of a power model for the STBus, a high performance industrial communication architecture supporting shared buses as well as crossbars, based on regression technique and presented an effective technology to minimize the Design of Experiments (DoE) in [8]. However, a packet switched router contains additional components that are not present in the STBus. Wolkotte et al. [9] presented a cycle accurate power model which is based on the number of flits passing through a router as the unit of abstraction. The work in [10] created power models for routers based on packet switching and circuit switching through calculating the average power consumption per bit traversing on single router by power analysis tool. Penolazzi et al. [11] presented an empirical formulation of the Nostrum NoC, which reflects the impact of different input vectors on the power consumption. Meloni et al. [12] presented power model for xPipes switch [13] with average error of 5%. Chan et al. [14] built a cycle accurate power model by analyzing key signals which influenced power of each functional component. In [15] and [16], there are methods to obtain a good fit between the model parameters and power consumption based on BP neural network.

In this paper, we propose a methodology for system level power model extraction of the NoC router based on heterogeneous macro-models, with higher accuracy to overcome the shortcoming of existing architecture-level power simulators, which is aimed to evaluate the network performance rapidly and guide the communication structure design.

III. ARCHITECTURE OF NoC ROUTER

Fig. 1 shows the block diagram of 2D mesh NoC router.

The NoC router consists of five input/output ports, the routing calculation unit, the arbiter and the crossbar switch. The input ports can be divided into input link controller, the virtual channel buffer and the transmission controller. The input link controller receives the request of communication from the last router, under the control of which the data is written to the virtual channel buffer; The transmission controllers transmits the request to the arbiter and controls flits to be read from the virtual channel buffer. The routing calculation unit judges the packet forwarding direction based on the routing algorithm. X-Y routing algorithm is used in the paper. The arbiter is responsible for steering flits towards appropriate output port. In this paper, polling mechanism is used to process request signals from different input ports.

IV. METHODOLOGY

We create a power model for NoC router based on the theory of macro-model, rather than gate-level circuit. Power
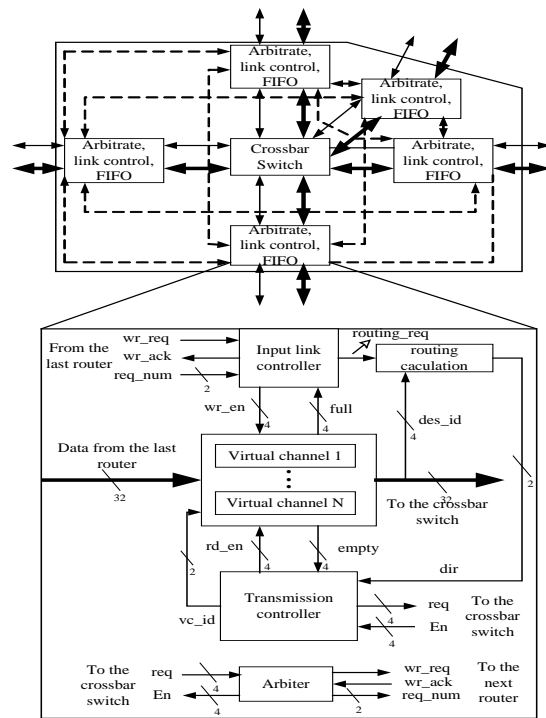

Fig. 1. Block diagram of 2D mesh NoC router

macro model uses flits passing through a router to estimate power consumption, which consists of input/output data statistical property that has a strong correlation to power consumption. Firstly, values of power consumption for different input vectors are get through the gate-level power analysis tool, and then these values are used to obtain a good fit between input data statistical property and average power consumption.

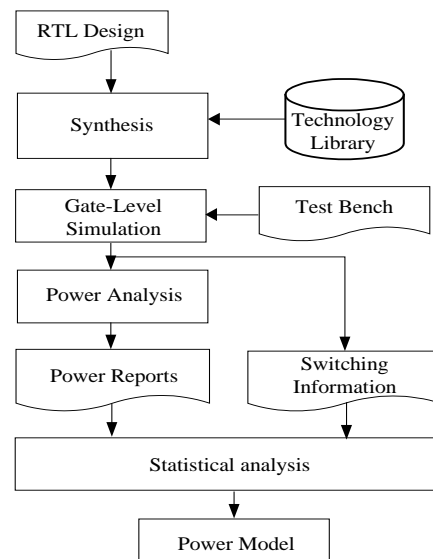The methodology used to create a power macro-model for our work is illustrated in Fig. 2.

Fig. 2. Power model generation methodology

Firstly, we use Synopsys Design Compiler to synthesize the router RTL description to the gate level net-list with technology library. In this step, physical information such as RC parasitic value files (SPEF), standard delay format (SDF)

and design constraints file (SDC), are also generated to be used for gate level power analysis and gate level simulation.

Secondly, traffic patterns are generated to exercise the router under different conditions, which is named test bench. Configuration of different traffic patterns in test bench can improve the accuracy of power model. For the router, we can configure these traffic patterns by varying the traffic load and the amount of bit-flips in the data stream. Adjusting for the traffic load is in order to gain power consumptions under different congestion conditions. Meanwhile, the influence of input data stream on power can be accessed by changing the amount of bit-flips. Test bench and gate level net-list file are used for gate level simulation with Synopsys VCS simulator. During the gate level simulation, a Value Change Dump (VCD) file which has logic switching information for dynamic power measurement is generated.

Thirdly, using the VCD and RC parasitic values, gate-level power analysis is done with the Synopsys PrimeTime PX tool to create cycle accurate power waveform file stored as out file. Finally, the energy consumption and signal values are used as observations in the regression analysis to obtain the power macro model.

V. POWER MODELING OF NoC ROUTER

The cycle energy consumption of a component can be expressed as:

$$P_{total} = \sum_{j=1}^{n} P_{ip,j} + P_{route} + P_{crossbar} + P_{arbiter} \tag{1}$$

where $P_{ip,j}$, P_{route}, $P_{crossbar}$ and $P_{arbiter}$ is the power values derived from the component macro-models for input/output ports, routing calculation unit, crossbar switch and arbiter respectively, and n is the number of input/output ports, which is of value 5 in this paper.

A. Input/Output Ports

Input link controller receives the communication request from the last router and writes data into the virtual channel buffer. Transmission controller sends the request to the arbiter and reads data from the virtual channel buffer. Dynamic power of input/output ports is mainly attributed to the reading and writing of the buffer and clock power. Hence, the power model can be formulated as:

$$P_{ip} = P_{write} + P_{read} + P_{clk} + P_{leak} \tag{2}$$

where P_{write} and P_{read} respectively are the power dissipated reading from and writing to the buffer; P_{clk} is the clock power and P_{leak} is leakage power. During writing to the buffer, the power consumption is in a linear relationship with the frequency of write operation and the Hamming distance of data into the buffer. Hence, the writing power macro-model can be established by multiple linear regression analysis method and it can be further decomposed as:

$$P_{write} = \Psi_{write0} + r_w \Psi_{w\_ctrl} + \alpha_F \Psi_{store} \tag{3}$$

where r_w is the rate of write operation, for example, $r_w = 0.5$ indicates writing to the buffer occurs once every two cycles. α_F is the quotient of hamming distance divided by bit-width. Ψ_{write0} is constant.

Similarly, the reading power macro model can be expressed as follows:

$$P_{read} = \Psi_{read0} + r_r \Psi_{r\_ctrl} + \alpha_F \Psi_{receive} \tag{4}$$

where r_r is the rate of read operation. After the system has been running stably for some time, r_r can be considered to be equivalent to r_w. The following r will represent r_r and r_w. And α_F is the same coefficient in Eqn.(3).

The P_{clk} and P_{leak} are usually constant at a fixed clock frequency, so they can be merged into constant Ψ_{ip0} with Ψ_{write0} and Ψ_{read0}. Then the power of input/output ports can be calculated as:

$$P_{ip} = \Psi_{ip0} + \Psi_{ip1} r + \Psi_{ip2} \alpha_F \tag{5}$$

The variables r and α_F can be obtained by simulation, regression coefficients Ψ_{ip0}, Ψ_{ip1} and Ψ_{ip2} are calculated by multiple linear regression.

B. Routing Calculation Unit

Route calculation takes place in routing calculation unit receives the address information contained in the packet-header flits, which result in the power consumption. The power consumption is in a linear relationship with the number of packet-header flits. Based on these observations, the power macro-model of routing calculation unit can be given as:

$$P_{router} = \Psi_{router0} + n_h \Psi_{compute} \tag{6}$$

where n_h is the number of flit in packet header per cycle. $\Psi_{compute}$ is the regression coefficients of variable n_h, Ψ_{route0} is constant.

C. Crossbar Switch

Router used in the paper is fully connected crossbar network, and crossbar switch power is related to the hamming distance of the input data. Thus, given the bit-flips of input data α_c, the following expression for the power of crossbar switch is resulted in:

$$P_{crossbar} = \Psi_{crossbar0} + \alpha_c \Psi_{dist} \tag{7}$$

D. Arbiter

The arbiter completes the path decision for requests from multiple input ports. Fig. 3 illustrates the state machine of arbiter, which is composed of eight states (s0 - s7). State s0, s2, s4, s6 are wait state, waiting for valid reqn (n=1 - 4) signals from input port. Arbiter uses polling mechanism to process reqn signals from different input ports. State s1, s3, s5, s7 represent the requests have been responded and they are kept until the packet transmission finishes.

The power consumption of arbiter is caused by the change of states, which is related to reqn and the present state. Due to the internal current state being invisible and numerous states, the power modeling for this module is based on BP neural network instead of linear regression.

BP neural network is multilayer feedforward neural network based on error back propagation (BP) algorithm, which consists of input layer, hidden layer and output layers [15]. The structure is shown in Fig. 4.
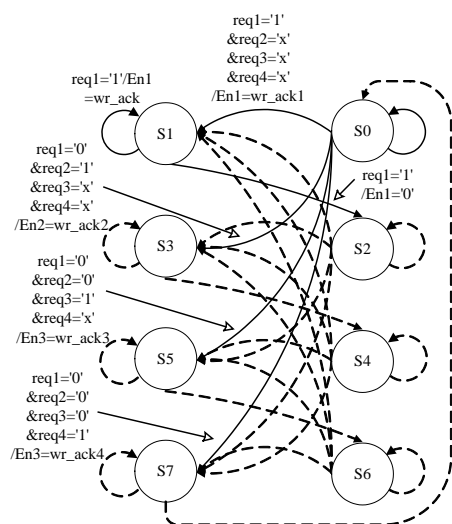
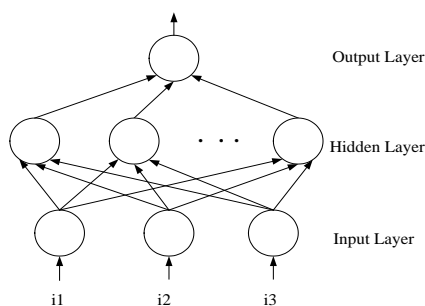Fig. 3. State machine of arbiter

Fig. 4. Structure of BP neural network

The number of neurons nodes in input layer is 3. i1 is the ratio that reqn changes from 0 to 1, and then Enn changes from 1 to 0 in the next cycle. i2 is the ratio that reqn changes from 0 to 1, and then Enn keeps invariable in the next cycle. i3 is the ratio that Enn changes from 1 to 0. Hidden layer has 20 neurons nodes. There is only one neurons node in output layer, which is the average power consumption of arbiter. Power estimation process is regarded as the nonlinear mapping of sequence statistical features onto the circuit power.

The key steps of BP neural network algorithm are as follow:

(1) Generate a certain number of input sequences randomly.

(2) Get the power consumption and output sequences by using gate level simulation.

(3) Calculate the parameters i1, i2, i3 and the average power of the input sequences.

(4) Establish the BP neural network and train it with the data obtained in step (3).

(5) Extract the parameters i1, i2 and i3 of the new input sequences, and estimate the average power consumption by using the established BP neural network in step (4).

VI. EXPERIMENTAL RESULTS

All router designs were synthesized using Synopsys Design Compiler and SMIC 0.18μm standard cell library were used for synthesis. Synopsys PrimeTime PX was used to obtain the power estimates and power waveform at an operating frequency of 200 MHz.

To estimate power more accurately, multiple random input data were imported into router. We obtained the respective average power consumption of components on gate-level power simulations with a separate set of randomly generated traffic patterns (100 traces in total). 50 traces were used to establish the power models through multiple linear regression or BP neural network, and then we verified the accuracy of power models by all of the 100 traces.

Fig. 5 shows the experiment platform. The data source module is connected to router A and the data receive module is connected to router B.
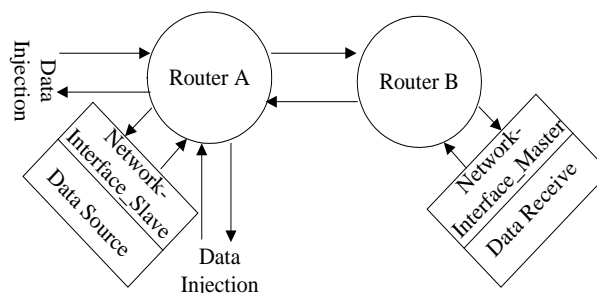

Fig. 5. Experiment platform

Five separate experiments were performed to ensure the validity of our models under different traffic loads, as shown in Tab. 1. We randomly characterized the injection rate and bit-flip of input data to generate 100 sets of traffic patterns. Experiments were performed in Inter Xeon 2.6GHz CPU and Red Hat AS4 operating system, and our power model is at about 600×faster simulation speed over gate-level power analysis by PrimeTime PX.

TABLE I
TRAFFIC LOADS IN FIVE EXPERIMENTS

Experiment No.	Traffic load number	Experiment description
1	1	Data is from the west port of router A to the data receive module.
2	1	Data is from the data source module to the data receive module.
3	2	Data is from the west port of router A to the data receive module using the virtual channels.
4	2	Data is from the data source module, west port and south port of router A to the data receive module.
5	3	Data is from the data source module, west port and south port of router A to the data receive module using the virtual channels.

Fig. 6 shows the actual power consumptions from the gate-level simulation and the expected ones with our system-level power model in five experiments, in which data flip rate is 50% and the flits injection rate is 10%. And Fig. 7 shows the error distribution of power model by statistical data on experiment. As is shown in the figure, the error of the model is below 6% in 70% of the cases. In some rare cases, the error reaches a level above 15%. However, the average error is about 5%.
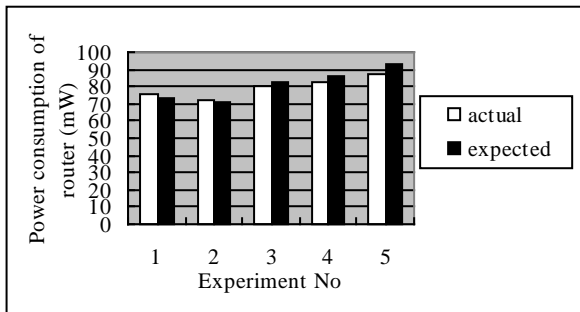

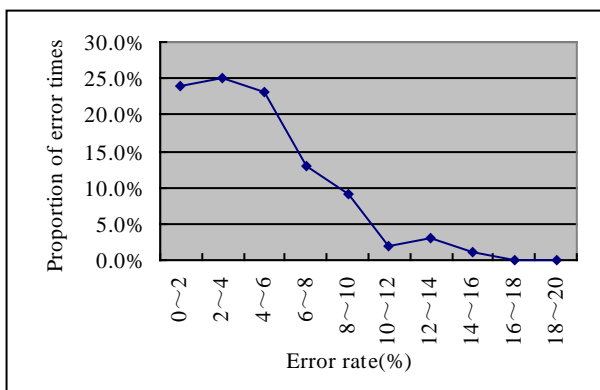Fig. 6. Comparison between the expected power and the actual power with the power model

Fig. 7. Error distribution

VII. SYSTEM LEVEL POWER EXPLORATION

Core mapping is a very important step in NoC design, and the result of mapping has a significant impact on communication power consumption of the system. In this section, system level power estimation is performed using our power model to evaluate the relation between power consumption and different core mappings. H.264 decoder [17] is used to perform core mapping, which is composed of 8 cores suite for 3×3 network. We implemented three different core mapping strategies named linear, random and optimal for 3×3 mesh network. Linear mapping allocates the cores to the same number node. For example, core 0 is mapped to node 0. Optimal mapping allocates the cores using the Ant System Algorithm [18]. Fig. 8 shows the mapping results for H.264 decoder with the three mapping strategies above.

Fig. 8 (c) also shows the traffic loads distribution. The solid arrow represents the loads more than 100Mbs and the hollow arrow represents the ones less than 100Mbs.

We estimate the power consumption of the mapping results using our model, as shown in Fig. 9.
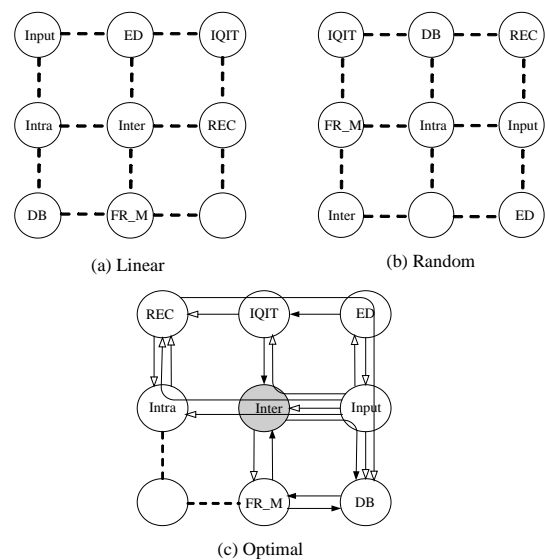

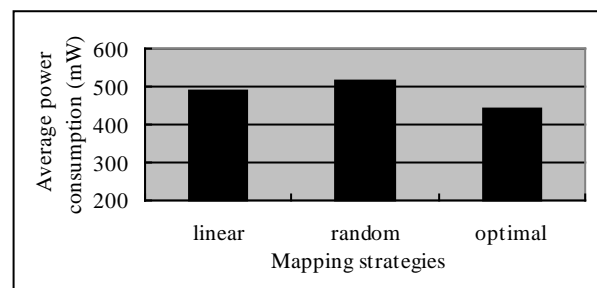Fig. 8. Mapping results with different strategies

Fig. 9. Power consumption of the mapping results

From Fig. 9, it is clear that the power consumptions are different with different mapping strategies. The power of random mapping is the most and the one of optimal is the least.

In this experiment, the estimation time is less than 1s. It proves that our power model can accurately and efficiently evaluate the average power consumption of mapping results with different strategies.

VIII. CONCLUSIONS

In this paper, we propose a methodology for system level power model extraction of the NoC router based on heterogeneous macro-models. The system level power model is built by power macro-models of each component according to their different characteristics: Linear regression is used to obtain the power macro-models for the input buffer, the routing calculation unit and the crossbar switch due to their single data flow state (linear characteristics). The power macro-model for the arbiter is established based on BP neural network because of its multiple states (nonlinear characteristics). Experiment results show that our system level power model demonstrates less than 5.0% average error as compared with the gate level analysis with 600 times speed up. As can be seen, the model can accurately and efficiently evaluate the average power consumption in NoC router, and provide reliable and fast power simulation to designers at system level. In order to demonstrate the utility of our power

model, we applied different core mapping strategies to 3×3 mesh network using H.264 decoder and observed their influence on power consumption. The result of experiments illustrates that the power of random mapping is the most and the one of optimal is the least.

REFERENCES

[1] W. J. Dally and B. Towles, "Router packets, not wires: on-chip interconnection networks," in *Proc. DAC*, Las Vegas, Nevada, 2001, pp 684-689.

[2] S. Kumar, A. Janstech, and J.-P. Soininen, "A network on chip architecture and design methodology," in *Proc. IEEE Computer Society Annu. Symp. on VLSI*, Pittsburgh, 2002, pp 105-112.

[3] K. Lahiri and A. Raghunathan, "Power analysis of system-level on-chip communication architectures," in *Proc. CODES+ISSS*, 2004.

[4] H. Wang, X. Zhu, and L-S. Peh, " Orion: a power-performance simulator for interconnection networks," in *Proc. MICRO*, 2002, pp 294-305.

[5] A. B. Kahng,Bin Li, and Li-Shiuan Peh, "ORION2:A fast and accurate NoC power and area model for early-stage design space exploration," in *Proc. DATE*, 2009, pp 423-428.

[6] T.T Ye, L.Benini, and G.D.Micheli, "Analysis of power consumption on switch fabrics in network routers," in *Proc. DAC*, 2002, pp 524-529.

[7] A. Bona, V. Zaccaria, and R. Zafalon, "System level power modeling and simulation of high-end industrial network-on-chip," in *Proc. DATE*, Washington DC, USA, 2004, pp 303-318.

[8] A. Bona, V. Zaccaria, and R. Zafalon, "Low effort, high accuracy network-on-chip power macro modeling," in *Proc. PATMOS*, 2004, pp 541-542.

[9] Seung Eun Lee, and Nader Bagherzadeh, "A high level power model for Network-on Chip(NoC) router," *Computer and Electrical Engineering*, vol. 35, no.6, pp.837-845, 2009.

[10] P. Wolkottex, G. Smit, and N. Kavaldjiev, "Energy model of networks-on-chip and a bus, system-on-chip," in *Proc. Int. Symp. SoC*, 2005, pp 82–85.

[11] S. Penolazzi and A. Jantsch, "A high level power model for the nostrum noc," in *Proc. DSD*, 2006, pp 673–676.

[12] P. Meloni, I. Loi, and F. Angiolini, " Area and power modeling for networks-on-chip with layout awareness," *VLSI Design*, 2007.

[13] D. Bertozzi and L. Benini, " Xpipes: a network-on-chip architecture for gigascale systems-on-chip," *IEEE Circuits and Systems Magazine*, vol. 4, no.2, pp. 18-31, 2004.

[14] J. Chan and S.Parameswaran and Nocee: energy macro-model extraction methodology for network on chip routers," in *Proc. ICCAD*, 2005, pp 254-259.

[15] W. T. Hsieh, C. C. Shiue, and C. N. Liu, "An efficient power modeling approach of sequential circuits using recurrent neural networks," *IEE Computers & Digital Techniques*, vol. 153, no.2, pp. 78-86, 2006.

[16] QIANG Wei, CAO Yang, and YAN Yuanyuan, "Power Macromodel of CMOS Circuits by BP Neural Network," *Journal of Wuhan University(Natural Science Edition)*, vol. 52, no.3, pp. 353-356, 2006.

[17] Huy-Name Nguyen,Vu-Duc Ngo, and Younghwan Bae, "An QoS aware mapping of cores onto NoC architectures," *Lecture Notes in Computer Science*, pp. 278-288, 2007.

[18] Yang Sheng-guang, Li Li, and Gao Ming-lun, "An Energy- and Delay-aware Mapping Method of NoC," *Acta Electronic Sinica*, vol. 36, no.5, pp. 937-942, 2008.