

High-Level Modeling Approach for Analyzing the Effects of Traffic Models on Power and Throughput in Mesh-based NoCs

S. Koohi⁺, M. Mirza-Aghatabar⁺, S. Hessabi^{*}, M. Pedram[†]

⁺^{*}Sharif University of Technology, Tehran, Iran, [†]University of Southern California, CA, USA

⁺{Koohi,Aghatabar }@ce.sharif.edu, ^{*}Hessabi@sharif.edu, [†]Pedram@usc.edu

Abstract

Traffic models exert different message flows in a network and have a considerable effect on power consumption through different applications. So a good power analysis should consider traffic models. In this paper we present power and throughput models in terms of traffic rate parameters for the most popular traffic models, i.e. Uniform, Local, HotSpot and First Matrix Transpose (FMT) as a permutational traffic model. We also select Mesh topology as the most prominent NoC topology and validate the presented models by comparing our results against simulation results from Synopsys Power Compiler and Modelsim. From the comparison, we show that our modeling approach leads to average error of 2% for power and 2.8% for throughput modeling.

1. Introduction

NoC is an efficient on-chip communication architecture for SoC architectures that is structured, reusable, scalable, and has high performance [1][2]. Nowadays, power issues are becoming of primary design constraints for even very high-end microprocessors. Different applications impose different traffic patterns into the network, which have a direct effect on behavior of power consumption. So, modeling this behavior in terms of traffic patterns leads to power consumption changes under traffic variation.

While network performance analysis due to different message flow (which stems from different traffic pattern) based on queuing and markov models, has been studied rigorously in the past [3][4][5], but network power analysis has not been explored.

N.Eisley proposed a framework in [6] that takes as input, message flows, and derives a power profile of the network fabric. Although this framework provides accurate power estimation, it is not always applicable. For instance when the generated traffic by application is not specified in terms of message flows, this method is not applicable. After partitioning and mapping stages, which lead to process allocation, we can specify nearly accurate traffic load between different nodes which can be summarized in terms of some popular traffic models. For example, if we have a memory node which acts as a global memory for all other nodes, we can predict

that the traffic behavior is a certain percentage of the hotspot traffic model. In all, specifying traffic patterns in terms of some predefined traffic models is easier than determining message flow, which enforces us to implement, or at least simulate the whole application.

None of these works has modeled power consumption and throughput in *high level of abstraction*, and therefore, enforces designers to provide a detailed knowledge on generated traffic. In this paper, we present high level power and throughput models in terms of traffic parameters for the most popular traffic models, i.e. Uniform, Local, HotSpot and FMT, in Mesh topology as the most prominent NoC topology. In Section 2, fundamental concepts such as traffic models and evaluation metrics are presented. Section 3 presents experimental results, and finally in Section 4, we conclude our work and give the summary.

2. Fundamental Concepts

In this section, we briefly explain power and throughput as our modeled performance criteria, and also present selected traffic models.

2.1. Power and Throughput

We have modeled power and throughput in the mesh architectures under different traffic models with XY routing algorithm.

Definition 2.1: In message passing systems, throughput (TP) may be defined as average number of flits arrived at their destination node for each node in each cycle [7].

Power consumption in NoCs consists of two components, the power consumed in routers, and the one associated with links. We calculated router's power consumption using Synopsys Power Compiler. Link's power consumption is determined as follows:

$$P_{link} = \alpha C_{wire} V_{DD}^2 f \quad (2.1)$$

We have calculated the power dissipated in links of each router, and used UMC18 [8]. The switching activity of each link is extracted from simulation. The length of metal wires is estimated as 2mm for the mesh topology. Based on the above data and technology's parameters [9], we have $C_{wire} = 0.7$ pf.

2.2. Traffic Models

Based on address distribution method [10], we define our traffic models as follows:

Uniform: each node sends its messages to any other node with equal probability.

HotSpot: each node sends specific portion of its generated messages to the hot node.

Local: each node sends specific portion of its generated messages to its neighbors within a predefined distance, called neighborhood radius.

FMT (First Matrix Transpose): In this traffic model the destination address for each node is fixed [11]. Each node with address of (i,j) in an $M \times N$ topology sends its message to a node with address of $(M-1-j, N-1-i)$. This traffic is generated by different applications such as Fast Fourier Transform (FFT).

Usually, generated traffic is composed of multiple address distribution methods, and therefore it is necessary to consider their composite effect while modeling power and throughput of the NoC.

3. Power and Throughput Analysis

The goal of power and throughput modeling under different traffic models is to predict power and throughput at high level of abstraction and their behavior with traffic pattern changes. Our presented power and throughput models are based on described traffic models in Section 2. Therefore, generated traffic patterns in our mesh-based NoC consist of these traffic models with l (*local*), h (*hotspt*), f (*FMT*), and u (*uniform*) percentages which vary in the range of $(0,1)$ and satisfy the relation $l+h+f+u = 1$.

Definition 3.1: *Pure* traffics are defined as traffics with one nonzero parameter among f , l , and h . For example, in a pure local traffic with $l=0.8$, in each node, 80% of packets are sent locally and 20% of them are sent uniformly.

Definition 3.2: Power and throughput models presented in this paper are defined as $Power = P(f, l, h)$, $Throughput = T(f, h, l)$, where $f, h, l \in (0,1)$.

Definition 3.3: *Mixed* traffic is defined as a traffic pattern in which more than one parameter among f , l , and h is nonzero.

So, we present our models in a 3D space with l, f and h axes. The key steps of our modeling strategy are as follows:

1. First, power and throughput on three l, f and h axes (pure traffics) are extracted through simulation.
2. Mutual effects between different traffic models are analyzed.
3. Finally, power and throughput models are obtained using extracted results and analysed effects presented in the previous steps, respectively.

Now, we go through the detailed description of each step of our modeling strategy.

3.1. Power and throughput for pure traffics

Figure 1 shows throughput and power consumption of a 6×6 mesh topology under pure traffic models with XY routing algorithm. These values are extracted by simulation at the brink of saturation, because these points show maximum throughput that an NoC can deliver and maximum power that should tolerate.

Figure 1(a) shows power consumption under pure traffics versus traffic model percentage in the range of $(0,1)$. Usually, with percentage increment, blocking probability in the network increases, which leads to reduction of switching activity, and so the power consumption drops as shown in Figure 1(a). But under different traffic models, this reduction occurs with a different behavior. We base our power model on these behaviors which are described below.

Since our topology is a 2D Mesh, when traffic percentage increases under local traffic models, the number of messages sent to neighbors of each node will increase and so the average distance between source and destination will decrease, which leads to reduction of switching activity in both X and Y dimensions. This reduction in two dimensions results in quadratic power reduction in terms of traffic percentage. Under FMT traffic model, the destination address for each node is fixed, so the XY routing algorithm as a deterministic routing always selects a specific path between each two nodes. It means that when FMT percentage increases in the network (which leads to uniform traffic reduction), the traffic load between each two nodes increases in only one path (as opposed to two dimensions which was the case for local traffic model). Therefore, we can conclude that traffic load increases linearly in terms of traffic percentage, which leads to linear increase in blocking time and decrease in power consumption. When percentage increases under hotspot traffic model, traffic load around hot node increases in both X and Y dimensions. On the other hand, when a flit is blocked in a buffer (e.g. the buffer size is equal to one flit), the blocking easily propagates in the network due to the wormhole switching nature and results in a hot area in the network. Therefore, when the probability of blocking occurrence in this traffic model increases with hot percentage increment, hot area's diameter grows linearly. Quadratic increase of traffic load due to traffic load increment in two dimensions around hot node, together with linear increase of hot area, lead to cubic increase in blocking time and so cubic decrease in switching activity, which leads to cubic power reduction in terms of traffic percentage. Based on these

behaviors, following power models are proposed for pure traffics:

$$P_h = \alpha_1 h^3 + \alpha_2 h^2 + \alpha_3 h + \alpha_4 \quad \alpha_1 < 0 \quad (3.1)$$

$$P_r = \beta_1 l^2 + \beta_2 l + \beta_3 \quad \beta_1 < 0 \quad (3.2)$$

$$P_f = \gamma_1 f + \gamma_2 \quad \gamma_1 < 0 \quad (3.3)$$

where $f, h, l \in (0,1)$ are traffic percentages and P_h, P_l and P_f are power associated with pure Hotspot, Local and FMT traffic models, coefficients α_4, β_3 and γ_2 represent power consumed when $h=l=f=0$. Assuming power consumption under uniform traffic as P_u :

$$\alpha_4 = \beta_3 = \gamma_2 = P_u \quad (3.4)$$

Coefficients of (3.1), (3.2) and (3.3) are calculated using MATLAB's regression functions, and are based on real values extracted for power consumption under pure traffic models using Synopsys Power Compiler.

Figure 1(b) shows throughput of NoC under pure traffics versus traffic model percentage in the range of $(0,1)$. To model throughput versus traffic percentage, first we should analyse its behavior under different traffic models in terms of traffic percentage.

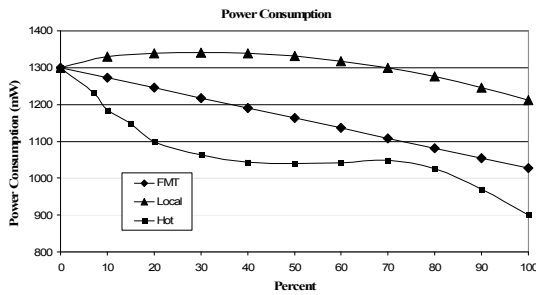
Under local traffic model, when locality increases, the number of messages sent to neighbors of each node will increase and so the distance between source and destination will decrease which leads to smaller delay and greater throughput. Average distance between source and destination nodes decreases linearly with traffic percentage increment which leads to linearly increasing throughput. Let us consider d as an average distance for uniform messages and 1 (one) for local traffic. Then, the average distance in an $N \times N$ network ($AvgD$) is:

$$AvgD = l \times 1 + (1-l) \times d = (1-d) \times l + d \quad (3.5)$$

and d is extracted as following:

$$d = \frac{\sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \left(\frac{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (|x-i| + |y-j|)}{N^2 - 1} \right)}{N^2} \quad (3.6)$$

In a 6×6 mesh topology, d is 4. It is clear from (3.5) that



(a) Power Consumption

average distance increases linearly with traffic percentage increment. On the other hand, latency has a direct relation with average distance, while throughput from (2.1) has a reverse relation with latency:

$$Throughput \propto \frac{1}{AvgD} = \frac{1}{(1-d) \times l + d} \quad (3.7)$$

We can write (3.7) as follows:

$$Throughput \propto \frac{1}{1-k} ; k = \frac{(d-1)}{d} \times l \quad (3.8)$$

In (3.8), we know that $l < 1$ and $(d-1)/d$ is lower than one, therefore, $k < 1$ and using Taylor series from (3.9), we can extend (3.10) from (3.7):

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + x^4 + \dots = \sum_{n=0}^{\infty} x^n : |x| < 1 \quad (3.9)$$

$$Throughput \propto 1 + k + k^2 + k^3 + k^4 + \dots \quad (3.10)$$

We showed that $k < 1$, so we can ignore the terms with high orders of k in (3.10) and rewrite this equation as follows:

$$Throughput \propto (1+k = 1 + \frac{(d-1) \times l}{d}) \propto 1 + 0.75l \quad (3.11)$$

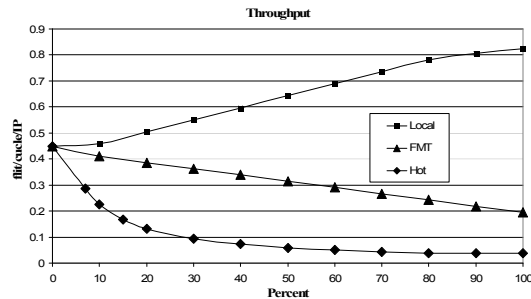
Relation (3.11) shows that throughput has a linear relation with traffic percentage which is clear from Figure 1(a), too.

On the other hand, in FMT and Hotspot traffic models, traffic percentage increment leads to more blocking time and so switching activity reduction, which results in more delay and smaller throughput. As mentioned before, under FMT traffic model, blocking time increases linearly in terms of traffic percentage which results in linear decrease in throughput. Similarly, under FMT traffic model, we can show that throughput has a linear relation with traffic percentage as follows. Let us consider d and r as average distances for uniform and FMT traffic models, respectively. Then we have:

$$AvgD = f \times r + (1-f) \times d = (r-d) \times l + d \quad (3.12)$$

The value of d is 4 as calculated in (3.6). For calculating value of r , we can extract (3.13) from the definition of FMT traffic model as follows:

$$r = \frac{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (2|N-i-j+1|)}{N^2} = 4.22 \quad (3.13)$$



(b) Throughput

Figure 1: Power consumption and Throughput for mesh topology under pure traffic models versus traffic percentages

From (3.8) we can conclude that throughput is related to r and d as follows:

$$\text{Throughput} \propto \frac{1}{1-k} ; k = \frac{(d-r)}{d} \times f \quad (3.14)$$

The value of $|k|$ is lower than one too, so we can use (3.9) again to write throughput as (3.10) and, due to small value of k we can ignore k^2 , k^3 , etc and therefore (3.15) is derived as follows:

$$\text{Throughput} \propto (1+k = 1 + \frac{(d-r) \times f}{d}) \propto 1 - 0.05 f \quad (3.15)$$

Under hotspot traffic model, percentage increment leads to more blocking time, which results in throughput reduction like FMT. But unlike FMT, in hotspot traffic model, relative change of throughput versus percentage is not constant; i.e., with percentage increment, slope of throughput versus traffic percentage decreases. Because when percentage traffic increases from low to high, number of blocking increases noticeably. In traffic percentage closer to 100%, due to high number of blocking around hot node, blocking get closers to its maximum value and therefore slope of throughput reduces to its minimum value. This leads to slight throughput variation. So, we infer that throughput has an inverse relation with traffic percentage.

Based on these behaviors, following throughput models are proposed for pure traffics:

$$T_h = \frac{1}{\alpha_5 h + \alpha_6} \quad \alpha_5 > 0 \quad (3.16)$$

$$T_l = \beta_4 l + \beta_5 \quad \beta_4 > 0 \quad (3.17)$$

$$T_f = \gamma_3 f + \gamma_4 \quad \gamma_3 < 0 \quad (3.18)$$

where $f, h, l \in (0,1)$ are traffic percentages and T_h, T_l and T_f are throughputs associated with pure Hotspot, Local, and FMT traffic models, respectively. When $h=l=f=0$, the values of $1/\alpha_6, \beta_5$ and γ_4 are equal to throughput of the network under uniform traffic (T_u):

$$\frac{1}{\alpha_6} = \beta_5 = \gamma_4 = T_u \quad (3.19)$$

Values of all coefficients presented in (3.16), (3.17) and (3.18) are calculated using MATLAB's regression functions, based on real values extracted for throughput under pure traffic models using Modelsim.

3.2. Analyzing mutual effects between different traffic models on power and throughput

As mentioned before, generated traffic is usually a mixture of multiple address distribution methods (mixed traffic). Therefore, before power and throughput modeling, we should investigate **composite effects** of various traffic models in a mixed traffic. As we know:

“Superposition principle is not applicable to power and throughput of a traffic pattern; i.e., power

consumption (throughput) of mixed traffics is not the sum of power consumed (throughput) of each traffic model.”

In step 1, we have extracted power and throughput formulas for pure traffic models, but as stated, we cannot simply sum up these formulas to extract the final power and throughput models for the mixed traffics. Hence, we should analyze the mutual effects of traffic models on each other.

We know that switching activity, which determines the value of power consumption, depends on the average number of blocking occurred in the network. In presence of all the mentioned traffic models; i.e., Local, FMT, Hotspot and Uniform models, number of blocking depends on traffic load generated by all these traffics. As mentioned before, number of blocking will significantly increases with hotspot percentage increment (step1), while this effect does not appear under local traffic, and is weak under FMT traffic (compared to hotspot traffic model). Therefore, presence of hotspot traffic model compared to any other traffic in a mixed traffic will increases the number of blocking and so decreases switching activity, which leads to power reduction. We explain this effect with an example.

Example1: Assume that we have two different traffic models T1 and T2. T1 is a pure local traffic model with $l=0.2$ ($u=0.8$) and T2 is a mixture of local and hot traffic models with $l=0.2$ and $h=0.4$ ($u=0.4$). Power consumed by 20% messages sent locally in both traffic patterns, strongly depends on the blocking situations and traffic load formed by other messages in the network. Since the hotspot percentage in T2 is 0.4, we have increased the number of blocking, which leads to message blocking even for locally sent messages. So, the power consumed by packets sent locally decreases (while local percentage is fixed). The same statement holds for simultaneous presence of FMT and hotspot traffic models. On the other hand, the messages sent locally do not increase the number of blocking in the network, and so, do not increase blocking probability for messages sent under FMT traffic mode, which leads to no power change for FMT messages, and vice versa.

From the above explanation, it is obvious that there is a mutual effect between hotspot and other traffic models. Therefore, in presence of hotspot traffic model, we cannot apply superposition principle to power, and similarly throughput values. We will present our final power and throughput formulas base on this mutual effect.

3.3. Final power and throughput models

In previous steps, we have described basic power and throughput models for pure traffics and their mutual effects in a mixed traffic. In this section, we will present our final power and throughput models using these parts.

As mentioned in step2, presence of hotspot traffic model reduces the slope of power (throughput) variation versus traffic percentage in Local and FMT traffic models. With hot percentage close to 1, this effect is stronger and power (throughput) of messages sent under other traffic models slightly changes with percentage increment. This effect stems from high blocking occurrence for high hot percentage. In the other words, as the simplest form, we can model the effect of hot percentage increment on the slope change of power, and similarly throughput values, versus traffic percentage under FMT or local traffic models, as follows:

$$\frac{dP}{dt} = k_1(1 - \sqrt{h}) ; \quad h, t \in (0,1) \quad (3.20)$$

$$\frac{dT}{dt} = k_2(1 - \sqrt{h}) ; \quad h, t \in (0,1) \quad (3.21)$$

where P and T are power and throughput values, respectively, for local or FMT traffic models. h is the hot percentage and t is FMT or Local traffic percentage in the mixed traffic model. k_1 and k_2 are constants in term of h . From (3.1)-(3.4) and (3.20), we conclude that the formula for power consumption under mixed traffic model is:

$$\begin{aligned} P_{total} &= P'_h + P'_l + P'_f \\ &= \alpha_1 h^3 + \alpha_2 h^2 + \alpha_3 h + \overbrace{(\beta_1 l^2 + \beta_2 l)}^{k_1} (1 - \sqrt{h}) \\ &\quad + \underbrace{\gamma_1 f}_{k_1} (1 - \sqrt{h}) + P_u \end{aligned} \quad (3.22)$$

where P_{total} is the total power consumed under mixed traffic, P'_h , P'_l and P'_f are power consumed by messages sent under hotspot, local and FMT traffic models considering mutual effects between different traffics. Coefficients α_i $i=1..3$, β_j $j=1,2$ and γ_1 are defined in (3.1)-(3.3), h , l and f are hotspot, local and FMT traffic percentages, respectively. P_u is the power consumption under uniform traffic model (when $f=l=h=0$). Similarly, from (3.16)-(3.19) and (3.21), we conclude the following formula for throughput under mixed traffic model:

$$\begin{aligned} T_{total} &= T'_h + T'_l + T'_f \\ &= \frac{1}{\alpha_5 h + \alpha_6} - \frac{1}{\alpha_6} + \overbrace{\beta_4 l}^{k_2} (1 - \sqrt{h}) \\ &\quad + \underbrace{\gamma_3 f}_{k_2} (1 - \sqrt{h}) + T_u \end{aligned} \quad (3.23)$$

where T_{total} is the total throughput under mixed traffic, T'_h , T'_l and T'_f are throughput values for messages sent under Hotspot, Local and FMT traffic models considering mutual effect between different traffics. Coefficients α_i $i=5,6$, β_4 and γ_3 are defined in (3.16), (3.17) and (3.18), h , l and f are Hotspot, Local and FMT traffic percentages, respectively. T_u is the throughput value under uniform traffic model (when $f=l=h=0$). From (3.19) we know that $1/\alpha_6 = T_u$, therefore we can derive (3.24) from (3.23) as follows:

$$T_{total} = \frac{1}{\alpha_5 h + \alpha_6} + \beta_4 l (1 - \sqrt{h}) + \gamma_3 f (1 - \sqrt{h}) \quad (3.24)$$

(3.22) and (3.24) present our final power and throughput models. The main important point in these models is their simplicity, which leads to rapid power and throughput calculation under mixed traffic, while calculating these values using simulation needs considerable time.

Beside power and throughput as two performance criteria for evaluating NoC architecture, energy consumption as the product of these factors plays an important role for NoC evaluation. Therefore, as a direct conclusion from power and throughput modeling, we can obtain the model for power per throughput ratio as follows:

$$\text{Energy} \propto \frac{P}{T} = \frac{\alpha_1 h^3 + \alpha_2 h^2 + \alpha_3 h + (\beta_1 l^2 + \beta_2 l)(1 - \sqrt{h}) + \gamma_1 f (1 - \sqrt{h}) + P_u}{\frac{1}{\alpha_5 h + \alpha_6} + \beta_4 l (1 - \sqrt{h}) + \gamma_3 f (1 - \sqrt{h})} \quad (3.25)$$

3.4. Validation and Results

In this part, we validate our power and throughput models presented in (3.22) and (3.24) by comparing against simulation results from Power Compiler and Modelsim.

As definition 3.2 shows, our power and throughput models are 4D functions whose domains are 3D space (f, l, h) . According to the range of traffic parameters, this 3D space is a unit cube. Validating (3.22) and (3.24) by computing all points in this volume and then comparing them with true values obtained from simulation, is impractical. Therefore, we selected a subset from this volume and validated the model upon it. In the $1 \times 1 \times 1$ 3D space, power consumption and throughput values associated to 12 lines and a 3D cube (consisting of about 110 points with 0.1 steps along each axis) are calculated by proposed models and compared against true values from simulation. Selected cube is associated to 3D subset $f \in [0, 0.3]$, $l \in [0, 0.3]$ and $h \in [0, 0.3]$.

Simulation values for 12 lines (on three planes, including Hot_FMT, FMT_Local and Local_Hot) are

Table 1- Comparing modeling values against real values in 3D cube

Traffic Percentage			Power (mW)			Throughput		
<i>h</i>	<i>l</i>	<i>f</i>	Sim.	Mod.	Err	Sim.	Mod.	Err
0.1	0.1	0.3	1157	1175	1.5	0.252	0.249	1.1
0.1	0.2	0.2	1140	1177	3.2	0.240	0.245	2.2
0.1	0.3	0.2	1126	1107	1.7	0.239	0.231	3.2
0.2	0.1	0.2	1118	1149	2.8	0.162	0.165	1.8
0.2	0.2	0.1	1112	1105	0.6	0.157	0.161	2.6
0.2	0.3	0.2	1108	1110	0.2	0.157	0.152	3.5
0.3	0.1	0.1	1092	1118	2.4	0.118	0.111	6.0
0.3	0.1	0.2	1090	1110	1.8	0.122	0.117	4.0

compared against their power and throughput values extracted from proposed models. But because of their huge volume, only resulted average error for each line is shown in Figure 2. As seen from the figure, maximum average error obtained is about 3.04% for throughput and 3.74% for power. Also, the maximum value for error's standard deviation is equal to 1.98. Computing average error for all 12 selected lines leads to total average value of 2% for power and 2.8%. Since 4D functions cannot be drawn, we were enforced to present the extracted results for the simulated cube in Table 1 (for brevity results are shown for some selected points). In this table, columns named as *Mod* and *Sim* refer to data obtained from modeling and simulation and column named *Err* refers to modeling error, for each triple (l, f, h) . As seen from this table, modeling in 3D cube results in accurate values with average error of 1.8% for power and 2.8% for throughput. Another interesting conclusion can be drawn by comparing modeling errors from Figure 1 for 2D planes and those from Table 1 for 3D cube which show 2% against 1.8% error for power, and 2.8% error in both cases for throughput. From this comparison, we conclude that our modeling approach remains accurate by dimension increment from 2D to 3D parameter space.

4. Conclusions

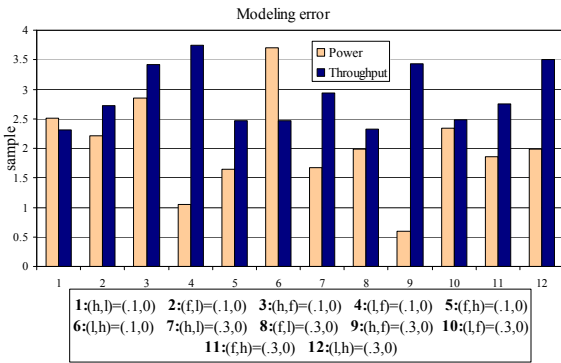


Figure 2: Modeling error, $(h, f) = (0.1, 0)$ means percent of Hotspot and FMT is 10%, and 0% respectively, and percent of local is varying from 0% to 90%.

In this paper, we proposed a power and throughput modeling approach that takes as input the traffic parameters of the application, and outputs the values of power and throughput of the network. Although our proposed models are based on mesh topology and some popular traffic models, this modeling approach can be generalized to other topologies with variety of traffic models.

We stated that different traffic models have mutual effects on each other. This fact prevents us from applying superposition principle to power (throughput) of pure traffics for calculating mixed traffic's power (throughput). Based on this observation, and extracted power and throughput values for pure traffics, we presented our power and throughput models for mixed traffics. In the proposed models, mutual effects between different traffics are taken into account.

We validated the proposed models by comparing our results against simulation results from Synopsys Power Compiler and Modelsim. From the comparison, we showed that our modeling approach leads to average errors of 2% for power and 2.8% for throughput values. Low error rates verify our modeling accuracy. Using this modeling approach, we can estimate NoC's behavior under different traffic patterns.

5. References

- [1] L. Benini and G. de Micheli, "Networks-on-Chip: A new Paradigm for System on Chip Design," *Design Automation and Test in Europe*, IEEE computer, vol. 35, no. 1, pp. 70-78, January 2002.
- [2] V. Tiwari et al. Reduction power in high-performance micro processors. In *35th Design Automation Conference*, 1998.
- [3] V. S. Adve, M. K. Vernon, "Performance Analysis of Mesh Interconnection Networks with Deterministic Routing," *IEEE Trans. on Par. and Dist. Syst.*, vol. 5, no. 3, pp. 225-246, March 1994.
- [4] A. Agarwal, "Limits on Interconnection Network Performance," *IEEE Trans. on Par. and Dist. Syst.*, vol. 2, no. 4, pp 398-412, October, 1991.
- [5] W. J. Dally, "Performance Analysis of *k*-ary *n*-cube Interconnection Networks," *IEEE Trans. on Computers*, vol. 39, no. 6, pp. 775-785, June 1990.
- [6] N. Easley and L. Peh. High-level power analysis for on-chip networks. In *CASES*, Sept. 2004.
- [7] Partha Pratim Pande, Cristian Grecu, Michael Jones, André Ivanov, Resve A. Saleh, "Performance Evaluation and Design Trade-Offs for Network-on-Chip Interconnect Architectures," *IEEE Trans. Computers*, vol. 54, no. 8, pp 1025-1040, 2005.
- [8] United Microelectronics Corporation (UMC), <http://www.umc.com/>, Dec, 2006.
- [9] International Technology Roadmap for Semiconductor (ITRS), <http://www.itrs.net/>, Dec, 2006.
- [10] W. Hsh, "Performance issues in wire-limited hierarchical networks," PhD Thesis University of Illinois-Urbana Champaign, 1992.
- [11] K. Hwang, *Advanced computer architecture: parallelism, scalability and programmability*, McGraw-Hill (Ed.), 1993.