

# Modeling NoC Traffic Locality and Energy Consumption with Rent's Communication Probability Distribution

George B. P. Bezerra  
Dept. of Computer Science  
University of New Mexico  
Albuquerque, NM 87131  
gbezerra@cs.unm.edu

Stephanie Forrest  
Dept. of Computer Science  
University of New Mexico  
Albuquerque, NM 87131  
forrest@cs.unm.edu

Melanie Moses  
Dept. of Computer Science  
University of New Mexico  
Albuquerque, NM 87131  
melaniem@unm.edu

Al Davis  
Dept. of Computer Science  
University of Utah  
Salt Lake City, Utah  
ald@cs.utah.edu

Payman Zarkesh-Ha  
Dept. of Electrical and  
Computer Engineering  
University of New Mexico  
Albuquerque, NM 87131  
payman@ece.unm.edu

## ABSTRACT

In systems on chip, the energy consumed by the Network on Chip (NoC) depends heavily on the network traffic pattern. The higher the communication locality, the lower the energy consumption will be. In this paper, we use the *Communication Probability Distribution* (CPD) to model communication locality and energy consumption in NoC. Firstly, based on recent results showing that communication patterns of many parallel applications follow Rent's rule [6], we propose a Rent's rule synthetic traffic generator. In this method, the probability of communication between cores is derived directly from Rent's rule, which results in CPDs displaying high locality. Next, we provide a model for predicting NoC energy consumption based on the CPD. The model was tested two NoC systems and several workloads, including Rent's rule traffic, and obtained accurate results when compared to simulations. The results also show that Rent's rule traffic has lower energy consumption than commonly used synthetic workloads, due to its higher communication locality. Finally, we exploit the tunability of our traffic generator to study applications with different locality, analyzing the impact of the Rent's exponent on energy consumption.

## Categories and Subject Descriptors

B.7.2 [Integrated Circuits]: Design aids—*simulation*; C.4 [Performance of Systems]: Modeling techniques

## General Terms

Design, Performance, Theory

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SLIP'10, June 13, 2010, Anaheim, California, USA.

Copyright 2010 ACM 978-1-4503-0037-7/10/06 ...\$10.00.

## Keywords

Communication probability distribution, Rent's rule, energy consumption, networks on chip, synthetic traffic generation

## 1. INTRODUCTION

Using derivations based on Rent's rule, the wire length distribution of a VLSI circuit can be estimated from its Rent's exponent and coefficient,  $p$  and  $k$  [4]. This distribution is relevant to VLSI design and implementation because it is related to many properties of the system, such as chip area, signal delay, power consumption, and wire routability [13].

In Systems on Chip (SoC), similar information is provided by the Communication Probability Distribution (CPD) of applications. The CPD describes the probability that packets will travel a certain distance in the Network on Chip (NoC) for a given traffic pattern. This distribution is directly related to the energy consumption of an application, because the larger the distance traveled by packets, the more energy used. Since current NoCs use 30 to 40% of the power budget [14, 7], it is desirable for the distance travelled by packets to be as small as possible in order to minimize this cost.

In this paper, we use the communication probability distribution to study NoC traffic locality and energy consumption. Firstly, motivated by the importance of Rent's rule to VLSI and supported by recent work showing that communication patterns of many parallel applications follow Rent's rule [6], we propose a method for generating Rent's rule traffic patterns. In this method, the probability of communication between processors is derived directly from Rent's rule, leading to CPDs displaying high traffic locality. This method could be used to simulate real traffic as a fast and simple alternative to application-driven workloads.

Based on the CPD, we also propose a model for predicting energy consumption in a network on chip. We tested the model on several synthetic workloads, including Rent's rule traffic, running on two different NoC systems and compared the obtained results with architecture-level simulations. The results show excellent agreement between predicted and experimental values. Our approach does not require simulation and could be used in the primary phases of NoC design, as

well as in the design of energy-efficient applications and better application mapping techniques [8]. Finally, using our traffic generator we also analyze the impact of the Rent's exponent of an application on energy consumption.

This paper is organized as follows. Section 2 presents the methodology used to generate Rent's rule traffic. Section 3 reviews commonly used synthetic traffic patterns and shows their CPD. Section 4 introduces the model for estimating energy consumption. Section 5 presents the experimental results and section 6 concludes the paper with a final discussion.

## 2. RENT'S RULE TRAFFIC PATTERNS

### 2.1 Rent's Rule for Parallel Programs

In VLSI, Rent's rule emerges naturally from circuit placement, in which connections are made as local as possible to minimize wire footprint, power and latency [2]. Similar constraints apply to the communication among processors in multi- and many-core systems. Algorithms used for mapping parallel applications onto cores aim at producing optimized layouts that minimize communication distances.

Greenfield et al. [5] argue that, analogous to circuit placement in VLSI, Rent's rule will naturally arise in multi- and many-core chips from this optimization process. They extended the concept of connection locality in circuits to communication locality among cores, proposing a bandwidth-based version of Rent's rule,

$$B = bN^p \quad (1)$$

where  $B$  is the bandwidth sent or received by a cluster of  $N$  network nodes,  $b$  is the average bandwidth per node, and  $0 \leq p \leq 1$  is the Rent's exponent.

In recent work, Heirman et al. [6] showed that many parallel applications indeed follow Rent's rule. They analyzed 13 popular benchmark applications running on 32 and 64 cores. Using a partitioning algorithm they showed that all of the programs followed Rent's rule with measured values of the Rent's exponent  $p$  ranging from 0.55 to 0.74.

### 2.2 Generating Rent's Rule Traffic Patterns

The discussion above motivates the use of a synthetic generator of traffic that follows Rent's rule. Such a traffic generator could serve as a simple way to evaluate NoCs with workloads that mimic the spatial properties of real traffic. As will be discussed in section 3, many existing synthetic workloads correspond to special case situations used to stress the network and routing algorithm. However, the authors are unaware of work that employs Rent's rule synthetic traffic as a generic model of parallel applications.

In this section we describe a method to generate traffic that follows Rent's rule. In VLSI, the probability of a wire connecting two terminals with Manhattan distance  $d$  apart is given by (adapted from [4]):

$$P(d) = \frac{1}{4d} [(1 + d(d-1))^p - (d(d-1))^p + (d(d+1))^p - (1 + d(d+1))^p], \quad (2)$$

We use the equation above to define the probability of communication between two processors, where  $d$  corresponds to the number of hops in the shortest path between source and destination. Traffic can be generated for each source node by

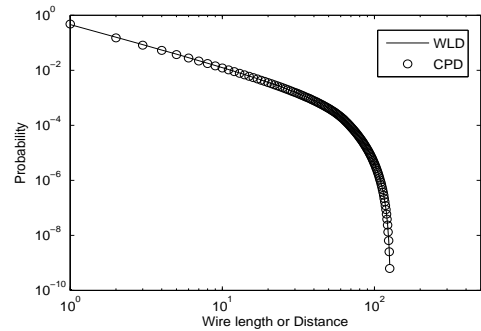


Figure 1: Comparison between the wire length distribution given by [4] and the communication probability distribution produced by the Rent's rule traffic generator.

sampling from the probability in equation 2 for every possible destination node. Repeating this process for all possible source nodes in the network results in traffic that follows Rent's rule.

To validate our method, we generated traffic using equation 2 and measured the resulting CPD. This distribution was then compared to the wire length distribution given by Davis et al. [4], which is derived directly from Rent's rule and is widely used in wire length estimates of real circuits. Figure 1 shows a log-log plot of the comparison between the wire length distribution given by [4] and the CPD produced by our traffic generator. In this figure,  $p = 0.75$ , which is a typical exponent for VLSI architectures, and the network has 1024 nodes. The plot shows a virtually exact match between the two curves.

The formula for the CPD of synthetic Rent's rule traffic can be derived from equation 2 and is given by:

$$CPD(d) = \Gamma P(d) \cdot \sum_{i=1}^{2\sqrt{N}-2} (\sqrt{N} - i) (\sqrt{N} + i - d), \quad (3)$$

$$\text{for } 0 < (\sqrt{N} + i - d) \leq \sqrt{N}.$$

where  $\Gamma$  is the normalization coefficient such that

$$\sum_{d=1}^{2\sqrt{N}-2} CPD(d) = 1.$$

Figure 2(a) shows the CPD produced by our generator on an  $8 \times 8$  mesh network.

An advantage of this method is its ability to generate traffic patterns with arbitrary Rent's exponents. Because the Rent's exponent is related to communication locality and complexity of applications, it is possible to study the NoC under several application scenarios by varying a single parameter in the model.

## 3. SYNTHETIC WORKLOADS

In this section we review some commonly used synthetic traffic patterns and compute their CPD, which is similar to the spatial hop distribution presented in [12]. We compare the obtained distributions with the CPD of Rent's rule traffic.

**Uniform Random Traffic.** In uniform random traffic, each source is equally likely to send packets to each destination. This is the most commonly used traffic pattern for network evaluation because it is straightforward to implement, it makes no assumptions about the application, and it is analytically tractable. Because source nodes do not differentiate between near and distal destination nodes, uniform random traffic does not exploit locality of communication. Figure 2(b) shows the CPD for uniform random traffic on a  $8 \times 8$  mesh network.

**Bit Permutation Traffic.** In permutation traffic, each source  $src$  sends all of its traffic to a single destination,  $des = \pi(src)$ , where  $\pi$  corresponds to a permutation function. Because this type of traffic concentrates load on individual source-destination pairs, they tend to stress the load balance of a topology and routing algorithm. Bit permutations are a subclass of permutations in which the destination address is computed by permuting the bits of the source address. The CPDs of bit transpose, bit complement and bit rotation permutation traffic are shown in figure 2(c), 2(d) and 2(e), respectively. These distributions are considerably different from each other as well as from uniform random traffic. Details on how to generate these traffic patterns are given in [3].

**Nearest Neighbor Traffic.** Nearest neighbor traffic is commonly used to evaluate the impact of communication locality on the performance and power consumption of the network on chip [11]. A fixed percentage of traffic goes to the nearest neighbors with some radius  $r$  and the rest of the traffic is uniform and random. The CPD of nearest neighbor traffic with  $r = 1$  and locality factor of 50% is shown in figure 2(f).

The traffic patterns described above are useful in practice as special cases to analyze the network, but bear little or no resemblance to real traffic. When compared to Rent’s rule traffic (figure 2(a)), most of these workloads display poor communication locality. As will be seen in Section 5, these differences in the CPD have considerable effect on the energy consumption of the NoC.

## 4. MODELING ENERGY CONSUMPTION

The analysis of NoC energy consumption using simulations can be computationally expensive, especially when dealing with application-driven workloads or large system sizes. In this section we provide a simple model for predicting energy consumption based on the CPD, which does not require computer simulations. This model is primarily designed for direct networks in which the length of the wires is the same for every hop, such as mesh and folded torus, but it could be easily extended to other topologies.

The average energy of a flit traversing a path of length  $d$  in the network is given by

$$E_{flit}(d) = d \cdot E_{link} + (d + 1) \cdot E_{router}, \quad (4)$$

where  $E_{link}$  and  $E_{router}$  are the energy consumed by the flit when traversing a link and a router, respectively, and  $d$  is given by the number of hops traversed in the path. The total energy consumed by an application is obtained by first summing  $E_{flit}$  over all communication distances weighted by the probability of a packet traveling that distance. This

value is then multiplied by the number of flits per packet ( $N_{flits}$ ) and the total number of packets ( $N_{packets}$ ):

$$E_{total} = N_{packets} \cdot N_{flits} \cdot \sum_{d=1}^{max} E_{flit}(d) \cdot CPD(d). \quad (5)$$

In equation 5, we assume a constant number of flits per packet. The constants  $E_{link}$  and  $E_{router}$  used in equation 4 can be obtained from architecture-level power models, such as Orion 2 [9].

For traffic that follows Rent’s rule, the model presented above provides a unique advantage over other approaches [8, 10, 11]. Given the Rent’s exponent, the CPD of traffic can be easily obtained from equation 3. As a result, the energy consumption of an application can be predicted directly from the Rent’s exponent. The ability of our model to perform predictions for Rentian traffic based on a single application parameter could significantly simplify and speedup NoC energy analysis.

A potential limitation of this method is that it is based on the assumption that the energy used for communication is proportional to the distance traveled by packets. This is approximately true for most networks on chip and is commonly used in the literature as a simplification step [8, 10, 11]. However, contention in the network may lead to extra dynamic and static energy that are not taken into account in the model and which might introduce errors in the model predictions.

## 5. RESULTS

### 5.1 NoC Energy Consumption

We analyzed the energy consumption of different traffic patterns and tested the predictions of equation 5 on two NoC configurations with different process technologies. The first system is an  $8 \times 8$  mesh network running at 1GHz, on a  $1 \times 1$ cm die, and 65nm technology. Flit size was set to 64 bits and packets have five flits each. The routing algorithm was dimension-order routing with wormhole flow control and 4 virtual channels. Constants for flit energy were obtained using Orion 2 assuming activity factor of 0.5. For each of the traffic patterns, 20,000 packets were injected in the network. The exponents used for Rent’s rule traffic were  $p = 0.55$  and  $p = 0.75$ , corresponding to the two extremes of Rent’s exponents measured in [6].

The energy predictions were compared to computer simulations and the obtained values are shown in figure 3(a). The results show excellent agreement between predicted and experimental energy values, with correlation coefficient of 0.98. Table 1 shows the same results in more detail. The best prediction was obtained for nearest neighbor traffic, with 0.7% error, and the worst for bit transpose, with error of 12.01%. As discussed in section 4, prediction errors can be explained by nonlinear factors in energy consumption and variations in network contention for different traffic patterns.

The second system is a  $10 \times 10$  network, on 45nm process technology and clock frequency of 3GHz. Flits have 32 bits each and the packet size is ten flits. The results are shown in figure 3(b). For this system, there is also a close match between predicted and experimental values, with correlation coefficient of 0.99. The results are shown in detail in table 1. A maximum error of 3.74% was obtained for uniform random traffic and a minimum error of 0.23% for bit complement.

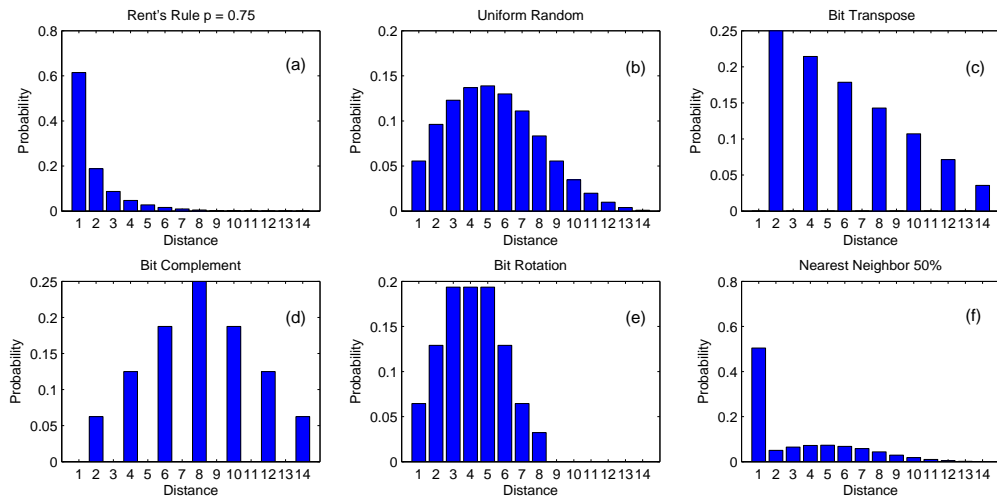


Figure 2: CPDs for a variety of traffic on a  $8 \times 8$  mesh network. (a) Rent's rule with Rent's exponent of 0.75. (b) Uniform random. (c) Bit transpose. (d) Bit complement. (e) Bit rotation. (f) Nearest neighbor with localization factor of 50%.

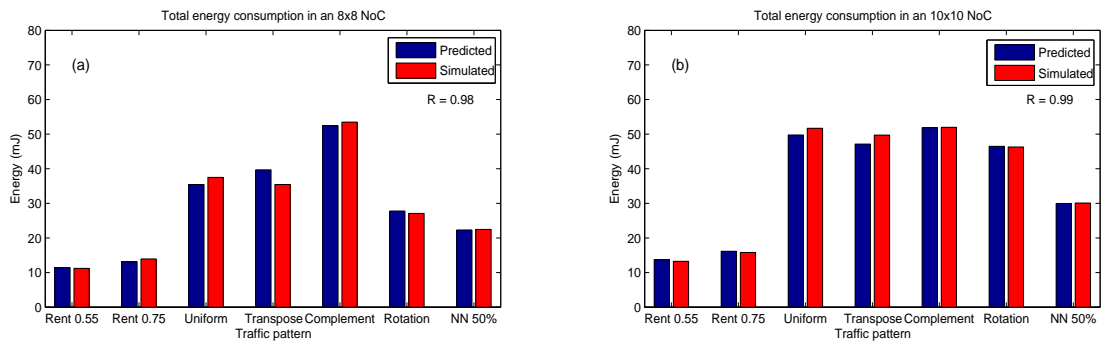


Figure 3: Predicted and simulated energy consumption for (a)  $8 \times 8$  mesh NoC on 65nm and (b)  $10 \times 10$  mesh NoC on 45nm.

The results above show that the proposed model produces accurate results over a wide range of traffic patterns, for different system configurations and also across different technology generations. This methodology could be used as a simple and fast tool for first-order assessment of energy consumption once the communication pattern of an application is known. Figure 3 also shows that Rent’s rule traffic consumes the least energy when compared to the other workloads, especially for the  $10\times 10$  system. This could be predicted from the CPDs in figure 2, since this is the traffic with the most communication locality. It should be expected that Rent’s rule traffic provides a better model of communication locality of real applications than the other synthetic workloads.

## 5.2 Varying the Rent’s Exponent

For VLSI devices, the value of the Rent’s exponent is commonly used as a measure of circuit complexity. Simple, highly regular circuits have small values of the Rent’s exponent, which are associated with high locality of communication. Conversely, the Rent’s exponent is large for more complex circuits in which a significant part of the communication is global. Analogously, in the bandwidth version of Rent’s rule, small values of  $p$  represent simple applications with mostly nearest-neighbor communication, while large values correspond to applications with relatively poor communication locality. In this section we analyze the impact of the Rent’s exponent on the energy used for communication, which could have important implications to application design.

We generated Rent’s rule traffic for a variety of Rent’s exponents and measured the energy consumption for three network sizes:  $6\times 6$ ,  $8\times 8$ , and  $10\times 10$ . The process technology used in the simulations was 45nm for all three systems. The results depicted in figure 4 show a significant increase in the energy consumption as the Rent’s exponent increases in all three networks. The impact of the Rent’s exponent on energy is also stronger for the larger systems. As  $p$  varies from 0.1 to 0.9, there is an increase of 51% in energy for the  $6\times 6$  NoC, 68% for the  $8\times 8$  NoC and 83% for the  $10\times 10$  network.

These results show quantitatively that the price to be paid for communication complexity is high and will tend to increase in the future. As we move towards larger systems with potentially hundreds of cores, the demand for less complex and more energy-efficient applications will increase. Energy-efficient algorithms are an important topic in other fields, such as sensor networks [1], and will likely become a major issue in application design for systems on chip.

These experiments illustrate the flexibility of our synthetic traffic generator and its applicability in the analysis of NoC. By varying the Rent’s exponent, it is possible to generate a continuum of application complexity scenarios, even ones that do not exist yet, and for systems with arbitrary sizes. The analysis presented here would not be possible with conventional execution-driven and trace-driven application workloads, which are limited to existing applications only.

## 6. DISCUSSION

In this paper we used the Communication Probability Distribution to model traffic locality and energy consumption in NoC. We proposed a synthetic traffic generator based on

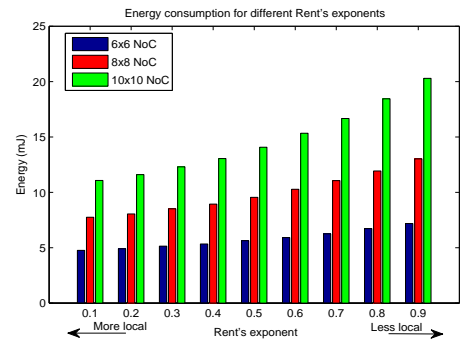


Figure 4: Energy consumption of  $6\times 6$ ,  $8\times 8$ , and  $10\times 10$  NoCs for Rent’s rule traffic as a function of the Rent’s exponent.

Rent’s rule that mimics the CPD of real traffic patterns. This method can be used as simple way to evaluate NoC designs under a variety of application complexity scenarios without having to resort to application-driven workloads.

Although the method is devised to be more realistic than commonly used synthetic traffic patterns, it has some limitations. For example, temporal aspects such as burstiness and variations of the Rent’s exponent over time [6] were not considered. Also, many applications exhibit traffic patterns with a central node, which could be better modeled with a combination of Rentian and hotspot traffic. Extending the model to consider these factors is a good direction for future work.

Based on the CPD, we also proposed a simple model for predicting NoC energy consumption. The model is based on the assumption that energy is proportional to the distance traveled by packets. Our methodology was tested on two system configurations and 6 different traffic patterns and produced accurate results. One advantage of this model is the ability to predict energy directly from the Rent’s exponent for traffic patterns that follow Rent’s rule. The results also showed that the energy consumed by Rent’s rule traffic is less than that of other synthetic workloads, because it has more locality of communication.

Finally, we used Rent’s rule traffic patterns to analyze the impact of the Rent’s exponent on NoC energy consumption. We showed quantitatively that the cost of communication complexity is significant, and will likely become a constraint to the scalability of future networks on chip.

## 7. REFERENCES

- [1] B. Chen, K. Jamieson, H. Balakrishnan, and R. Morris. Span: An energy-efficient coordination algorithm for topology maintenance in ad hoc wireless networks. *Wireless Networks*, 8(5):481–494, 2002.
- [2] P. Christie and D. Stroobandt. The interpretation and application of rent’s rule. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 8(6):639–648, 2000.
- [3] W. J. Dally and B. Towles. *Principles and Practices of Interconnection Networks*. Morgan Kaufman Publishers, San Francisco, 2004.
- [4] J. A. Davis, V. K. De, and J. D. Meindl. A stochastic wire-length distribution for gigascale integration (GSI)

**Table 1: Predicted and simulated energy values for 8×8 and 10×10 NoCs. The uncertainty values arise from the limited number of packets sampled from the CPD.**

Traffic Pattern	8×8 NoC			10×10 NoC		
	Predicted (mJ)	Simulated (mJ)	% Error	Predicted (mJ)	Simulated (mJ)	% Error
Rent ( $p = 0.55$ )	11.43	11.21±0.09	+2.00	13.69	13.25±0.03	+3.32
Rent ( $p = 0.75$ )	13.11	13.92±0.08	-5.78	16.15	15.79±0.03	+2.26
Uniform Random	35.44	37.51±0.15	-5.51	49.76	51.70±0.13	-3.74
Bit Transpose	39.69	35.43±0.11	+12.01	49.18	49.73±0.10	-1.10
Bit Complement	52.43	53.46±0.13	-1.94	51.84	51.97±0.13	-0.23
Bit Rotation	27.77	27.08±0.04	+2.52	47.21	46.29±0.09	+1.97
Nearest Neighbor (50%)	22.30	22.46±0.08	-0.70	29.96	30.09±0.05	-0.44

- Part I: Derivation and validation. *IEEE Transactions on Electron Devices*, VOL 45(3):580–589, 1998.

- [5] D. Greenfield, A. Banerjee, J.-G. Lee, and S. Moore. Implications of Rent’s rule for NoC design and its fault-tolerance. In *Proceedings of the First International Symposium on Networks-on-Chip (NOCS’07)*, 2007.
- [6] W. Heirman, J. Dambre, D. Stroobandt, and J. Campenhout. Rent’s rule and parallel programs: Characterizing network traffic behavior. In *Proceedings of the 2008 International Workshop on System Level Interconnect Prediction, SLIP’08*, 2008.
- [7] Y. Hoskote, S. Vangal, A. Singh, N. Borkar, and S. Borkar. A 5-GHz mesh interconnect for a teraflops processor. *IEEE MICRO*, 27(5):51–61, 2007.
- [8] J. Hu and R. Marculescu. Energy-aware mapping for tile-based NOC architectures under performance constraints. In *Proceedings of ASP-Design Automation Conference*, pages 233–239, 2003.
- [9] A. Kahng, B. Li, L. Peh, and K. Samadi. Orion 2.0: A fast and accurate NOC power and area model for early-stage design space exploration. In *Design, Automation, and Test in Europe*, pages 423–428, 2009.
- [10] J. Palma, C. Marcon, F. Moraes, N. Calazans, R. Reis, and A. Susin. Mapping embedded systems onto NoCs: the traffic effect on dynamic energy estimation. In *Proceedings of the 18th annual symposium on Integrated circuits and system design*, page 201, 2005.
- [11] P. Pande, C. Grecu, M. Jones, A. Ivanov, and R. Saleh. Effect of traffic localization on energy dissipation in NoC-based interconnect. In *ISCA 2005*, pages 1774–1777, 2005.
- [12] V. Soteriou, H. Wang, and L. Peh. A statistical traffic model for on-chip interconnection networks. In *Proceedings of the 14th IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS’06)*, pages 104–116, 2006.
- [13] D. Stroobandt. *A Priority Wire Length Estimates for Digital Design*. Kluwer Academic Publishers, Boston, 2001.
- [14] M. B. Taylor, J. Kim, J. Miller, D. Wentzlaff, F. Ghodrati, B. Greenwald, H. Hoffman, P. Johnson, J.-W. Lee, and W. Lee. The Raw microprocessor: A computational fabric for software circuits and general purpose programs. *IEEE MICRO*, 22(PART 2):25–35, 2002.