

Energy Reduction through Crosstalk Avoidance Coding in NoC Paradigm

Partha Pratim Pande¹, Haibo Zhu¹, Amlan Ganguly¹, Cristian Grecu²

¹*School of Electrical Engineering & Computer Science
PO BOX 642752
Washington State University
Pullman, WA, USA
(pande, Haibo, ganguly)@eecs.wsu.edu*

²*SoC Research Lab
Department of Electrical and Computer Engineering
University of British Columbia
2332 Main Mall Vancouver, BC, V6T 1Z4 Canada
grecuc@ece.ubc.ca*

Abstract

Commercial designs are currently integrating from 10 to 100 embedded functional and storage blocks in a single SoC and the number is likely to increase significantly in the near future. With this ever increasing degree of integration, design of communication architectures for big SoCs is a challenge. The communication requirements of these large Multi Processor SoCs (MP-SoCs) are conformed by the emerging network-on-a-chip (NoC) paradigm. The basic operations of NoC infrastructures are governed by on-chip packet-switched networks. Crosstalk between adjacent wires is an issue in NoC communication fabrics and it can cause timing violations and extra power dissipation. Crosstalk avoidance codes (CACs) can be used to improve signal integrity and also reduce the coupling capacitance and hence the energy dissipation of a wire segment. By incorporating Crosstalk Avoidance Coding (CAC) in NoC data streams we are able to reduce communication energy, which will help to decrease the energy dissipation as a whole.

Keywords: *networks on chip, crosstalk, crosstalk avoidance codes, interconnect energy, wormhole switching.*

1. Introduction and motivation

The network-on-Chip (NoC) design paradigm is viewed as an enabling solution [1] [2] for the integration of exceedingly high number of computational and storage blocks in a single chip. The common characteristics of NoC interconnect architectures is that the functional blocks communicate with each other with the help of intelligent switches and links. With technology scaling and DSM effects power consumption is the biggest challenge in today's multi-core SoC design. Subsequently in NoC domain significant amount of energy is dissipated by the

communication links. In DSM era the inter-wire spacing decreases rapidly [3], which in turn gives rise to large mutual capacitance. This increases the coupling capacitance between adjacent wires with negative effects on delay, power and signal integrity. Shielding can be used effectively to reduce crosstalk. This involves placing a grounded wire between every pair of signal wires. Although this is effective in preventing crosstalk within a bus, it has the effect of doubling the wire area. In NoC domain the inter-switch link wires have to be routed using higher metal layers, which scale slower than the rest of the layers [4] in order to prevent an unacceptable increase in resistance. Therefore doubling the routing requirements at these levels is difficult to justify. Crosstalk between adjacent wires is an issue in NoC communication fabrics and it can cause timing violations and extra power dissipation. Crosstalk avoidance codes can be used to reduce the coupling capacitance and hence the energy dissipation of a wire segment. By incorporating Crosstalk Avoidance Coding schemes in NoC data streams we expect to enhance the system reliability and at the same time reduce communication energy, which will help to decrease the overall energy dissipation. Crosstalk Avoidance Coding (CAC) reduces the effective coupling capacitance of the wire segments without doubling the number of wires and subsequently reduces the energy dissipation.

CACs reduce the worst-case switching capacitance of a wire by ensuring that a transition from one codeword to another codeword does not cause adjacent wires to switch in opposite directions [5].

It is shown in [6] that in regular NoC architectures the delays in the inter-switch wire segments, assuming worst case coupling capacitance can be fit to be within one clock cycle limit. However from energy dissipation perspective this is not an optimum case since by using different coding techniques total switching capacitance of the wire segments can be reduced. A wide variety of low-power bus encoding schemes are available [7] [8]. According to the analysis in [9] for the specific case of on-chip buses, the bus lines must

be 20 mm or longer in order for these encoding schemes to be energy efficient in practical implementations. In NoC environment the inter-switch wire segments are the longest signal carrying interconnects. Due to the structured nature of the communication fabric of the NoCs these inter-switch wire segments turn out to be significantly shorter than the above mentioned limit [6].

In addition to the reduction of switching capacitances of the inter-switch wire segments due to incorporation of CACs two more energy dissipation sources need to be considered. The first one is the coder-decoder energy and the second one is the extra energy dissipated by the redundant wires introduced by the coding schemes. Our aim in this paper is to study the feasibility of the CACs in the emerging NoC paradigm considering the trade-off between the energy efficiency and silicon area overhead.

2. Related work

In the recent years, there has been an evolving effort in developing on chip networks to integrate increasingly large number of functional cores in a single die [1]. The role of communication infrastructure on energy dissipation is discussed in [10]. Different strategies for power management of NoCs, such as power-aware on-off networks [11], and dynamic voltage scaling [12] have been addressed previously. In addition to these, different low-power encoding techniques have been proposed to reduce power consumption of on-chip buses. According to [9] due to higher power dissipation in the codec blocks these schemes are energy efficient only if the length of the wire segment exceeds a certain limit. Moreover these low power coding techniques reduces only the self-transition in a wire. In DSM processes one of the main sources affecting signal integrity is the crosstalk between adjacent wires. The crosstalk is data dependent and there are different coding schemes that aim to reduce the amount of relative transitions between adjacent wires [5]. This reduces the effective switching capacitance, which in turn speeds up the signal transmission while reducing the energy dissipation. All these coding schemes have been predominantly implemented for bus-based systems. The objective of this paper is to explore the energy savings capability of CACs in NoC domain.

3. Crosstalk avoidance in NoC links

A few NoC interconnect architectures have been proposed by different research groups [10]. The common characteristic of these NoC architectures is that the processor/storage cores communicate with each other through high-performance links and intelligent switches as shown in Fig. 1.

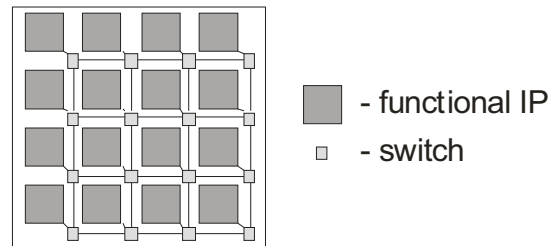


Fig. 1: MESH-based NoC

Data exchange between the functional blocks takes place in the form of packets. The success of the NoC design paradigm relies greatly on the standardization of the interfaces between IP cores and the interconnection fabric. Using a standard interface should not impact the methodologies for IP core development. The *Open Core Protocol* (OCP) [13] is a plug and play interface standard receiving a wide industrial and academic acceptance. Similar to the OCP, the AMBA AXI [14] is another protocol targeted at high performance, platform-based system design. As shown in the Fig. 2 for a core having both master and slave interfaces, the OCP/AXI compliant signals of the functional IP blocks are packetized by a second interface. The network interface has two functions:

- 1: injecting/absorbing the packets leaving/arriving at the functional/storage blocks;
- 2: packetizing/depacketizing the signals coming from/reaching to OCP/AXI compatible cores in form of packets.

In our crosstalk avoidance scheme we propose that the codec be a part of the network interface (NI). Consequently the data is coded at the source NI and is decoded when reaches the destination NI.

The delay of an inter-switch wire in the NoC link depends on the transitions on the wire and wires adjacent to it. The worst case delay of a wire is $(1 + 4\lambda)\tau$, where τ is the delay of a crosstalk-free wire and λ is the ratio of the coupling capacitance to the bulk capacitance [15].

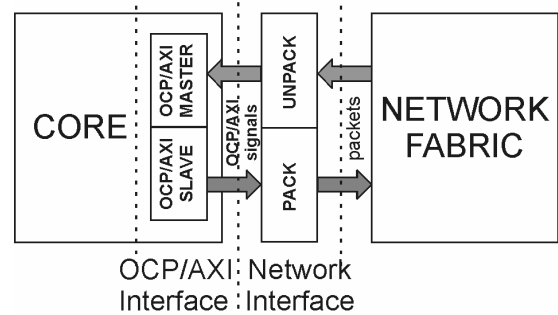


Fig. 2: Interfacing of IP cores with the network fabric

The purpose of the crosstalk avoidance code is to reduce the delay of the line to $(1 + p\lambda)\tau$, where $p=1, 2, \text{ or } 3$ and

it is referred to as the maximum coupling. Consequently the CACs will reduce the energy dissipation per line in a NoC link. However when investigating the overall energy dissipation the effect of redundant wires per link need to be considered as well.

4. Crosstalk avoidance codes

There is a number of crosstalk avoidance codes [16] proposed in literature. Here we consider three representative CACs that achieve different degrees of delay reduction.

Forbidden Overlap Condition (FOC) codes

A wire has the worst-case delay $(1+4\lambda)\tau$ when it executes a rising (falling) transition *and* its neighbors execute falling (rising) transitions. If these worst-case transitions are avoided, the maximum coupling can be reduced to $p=3$. This condition can be satisfied if and only if a codeword having the bit pattern 010 does not make a transition to a codeword having the pattern 101 at the same bit positions. The codes that satisfy the above condition are referred to as Forbidden Overlap Codes (FOC). The simplest method of satisfying the forbidden overlap condition is half-shielding, in which a grounded wire is inserted after every two signal wires. Though simple, this method has the disadvantage of requiring a significant number of extra wires. Another solution is to encode the data links such that the codewords satisfy the FO condition. However, encoding all the bits at once is not feasible for wide links due to prohibitive size and complexity of the codec hardware. In practice, partial coding is adopted, in which the links are divided into sub-channels which are encoded using CACs. The sub-channels are then combined in such a way as to avoid crosstalk occurrence at their boundaries. Considering a 4-bit subchannel the coding scheme is expressed in Table 1.

For coding 32 bits, eight FOC_{4,5} blocks are needed, and as a result of this a 32-bit uncoded link will be converted to a 40-bit coded link. In this case two sub-channels can be placed next to each other without any shielding, as well as not violating the FO condition.

Table 1: FOC_{4,5} coding scheme

Data bits				Code bits				
d3	d2	d1	d0	c4	c3	c2	c1	c0
0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	1	0	0
0	0	1	0	0	0	0	0	1
0	0	1	1	0	0	1	0	1
0	1	0	0	0	0	0	1	1
0	1	0	1	0	0	1	1	1
0	1	1	0	1	0	0	1	1
0	1	1	1	1	0	1	1	1
1	0	0	0	1	0	0	0	0
1	0	0	1	1	0	1	0	0
1	0	1	0	1	0	0	0	1
1	0	1	1	1	0	1	0	1
1	1	0	0	1	1	0	0	0
1	1	0	1	1	1	1	0	0
1	1	1	0	1	1	0	0	1
1	1	1	1	1	1	1	0	1

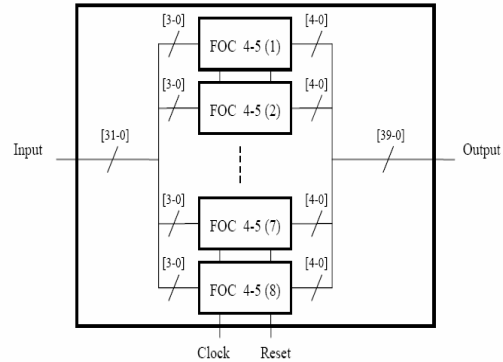


Fig. 3: Bloc diagram of the FOC₃₂₋₄₀ encoder architecture

Forbidden Transition Condition (FTC) codes

The maximum capacitive coupling and, hence, the maximum delay, can be reduced even further by extending the list of non-permissible transitions. By ensuring that the transitions between two successive codes do not cause adjacent wires to switch in opposite directions (i.e., if a codeword has a 01 bit pattern, the subsequent codeword cannot have a 10 pattern at the same bit position, and vice versa), the coupling factor can be reduced to $p=2$. This condition is referred to as Forbidden Transition Condition, and the CACs satisfying it are known as Forbidden

Transition Codes (FTC). Inserting a shielding wire after each signal line can employ the simplest FTC. For wider links, a hierarchical encoding is more suitable, where the inter-switch links are divided into sub-channels that are encoded individually. Considering a 3-bit subchannel the coding scheme is expressed in Table 2.

In this case also we combined the sub channels in such a way that there is no forbidden transition at the boundaries between them. Consequently a 32-bit uncoded link will be converted to 53-bit coded link [5].

Table 2: FTC₃₋₄ Coding scheme

Data bits			Code bits			
d2	d1	d0	c3	c2	c1	c0
0	0	0	0	0	0	0
0	0	1	0	1	0	0
0	1	0	0	0	0	1
0	1	1	0	1	0	1
1	0	0	0	1	1	1
1	0	1	1	1	0	0
1	1	0	1	1	0	1
1	1	1	1	1	1	1

Forbidden Pattern Condition (FPC) codes

The same reduction of the coupling factor as for FTCs (p=2) can be achieved by avoiding 010 and 101 bit patterns for each of the code words. This condition is referred to as Forbidden Pattern Condition, and the corresponding CAC is known as Forbidden Pattern Code (FPC). The simplest FPC code is realized by duplication, where each data bit is transmitted using two adjacent wires. Considering a 4-bit subchannel the coding scheme is expressed in Table 3. While combining the sub channels we made it sure that there is no forbidden pattern at the boundaries. As a result of this similar to the above two CACs FPC also adds redundant bits to the uncoded link and a 32-bit uncoded link is converted to a 52-bit coded link.

In general, when combinational coding/decoding techniques are used to implement CACs, if the uncoded link has n signal lines, and the coded link has k>n wires, the corresponding code is referred to as a (n, k) code. Theoretical limits for the minimum value of k for different crosstalk avoidance techniques were shown in [16].

In this paper we investigate the applicability of these three CAC coding schemes in the NoC domain. Our aim is to study the energy saving characteristics of these schemes at the cost of extra area overhead they introduce.

Table 3: FPC₄₋₅ coding scheme

Data bits				Code bits				
d3	d2	d1	d0	c4	c3	c2	c1	c0
0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	1
0	0	1	0	0	0	1	1	0
0	0	1	1	0	0	0	1	1
0	1	0	0	0	1	1	0	0
0	1	0	1	0	0	1	1	1
0	1	1	0	0	1	1	1	0
0	1	1	1	0	1	1	1	1
1	0	0	0	1	0	0	0	0
1	0	0	1	1	0	0	0	1
1	0	1	0	1	1	0	0	0
1	0	1	1	1	0	0	1	1
1	1	0	0	1	1	1	0	0
1	1	0	1	1	1	0	0	1
1	1	1	0	1	1	1	1	0
1	1	1	1	1	1	1	1	1

5. Energy dissipation in a NoC-based SoC

When flits travel on the interconnection network, both the inter-switch wires and the logic gates in the switches toggle and this will result in energy dissipation. The flits from the source nodes need to traverse multiple hops consisting of switches and wires to reach destinations. Consequently, we determine the energy dissipated in each interconnect and switch hop. The energy per flit per hop is given by

$$E_{hop} = E_{switch} + E_{interconnect} \quad (5.1)$$

The energy dissipated in transporting a flit consisting of n bits through h hops can be calculated as

$$E_{flit} = n \sum_{j=1}^h E_{hop,j} \quad (5.2)$$

Let P be the total number of flits transported, and let $E_{flit,i}$ be the energy dissipated by the i^{th} flit, where i ranges from 1 to P. The average energy per bit $\overline{E_{bit}}$ is then calculated according to the following equation:

$$\bar{E}_{bit} = \frac{\sum_{i=1}^P E_{flit_i}}{P \cdot n} = \frac{\sum_{i=1}^P \left(n_i \sum_{j=1}^{h_i} E_{hop,j} \right)}{P \cdot n} \quad (5.3)$$

In order to quantify the energy dissipation profile for a NoC interconnect architecture, we determine the energy dissipated in each switch, E_{switch} , by running SynopsysTM Prime Power on the gate-level netlist of the switch blocks, including the NIs. To determine interconnect energy, $E_{interconnect}$ the capacitance of each interconnect stage, $C_{interconnect}$ is calculated taking into account the specific layout of each topology and it can be estimated according to the following expression

$$C_{interconnect} = C_{wire} \cdot w_{a+1,a} + n \cdot m \cdot (C_G + C_J)$$

where C_{wire} is the wire capacitance per unit length, and $w_{a+1,a}$ is the wire length between two consecutive switches; C_G and C_J are the gate and junction capacitance of a minimum size inverter, respectively, n denotes the number of inverters (when buffer insertion is needed) in a particular inter-switch wire segment and m is their corresponding size with respect to a minimum size inverter. While calculating C_{wire} without any coding we have considered the worst case switching scenario, where the two adjacent wires switch in the opposite direction of the signal line simultaneously [10]. In the presence of CACs the value of C_{wire} will be reduced according to the coding scheme and this will help in reducing $E_{interconnect}$. On the other hand incorporation of the codec blocks will increase E_{switch} . Our aim is to study the effects of these two together on the overall energy dissipation NoC communication infrastructures.

6. Experimental results and analysis

To evaluate the role of the CAC schemes discussed above on the energy dissipation characteristics of a NoC we consider a system consisting of 64 IP blocks and mapped them onto MESH-based NoC architecture as shown in Fig. 1. Messages were injected with a uniform traffic pattern (in each cycle, all IP cores can generate messages with the same probability). The wormhole routing technique [10], where data packets are divided into fixed length flow control units (flits) is generally adopted. A packet is divided into a header flit containing routing and flow control information, one or more data flits, and a tail flit indicating the end of packet. The routing mechanism used for all simulations was the *e-cube* (dimension order) routing [10].

The energy dissipation of each inter-switch wire segment is a function of λ , the ratio of the coupling capacitance to the bulk capacitance. For a given interconnect geometry, the value of λ depends on the metal coverage in upper and lower metal layers [5]. We vary the value of λ from 1 to 4 [16].

Figs. 4 and 5 show the average bit energy dissipation as a function of the injection load of the NoC under consideration with $\lambda=1$ and $\lambda=4$ at 130 nm technology node respectively. The injection load is expressed as the number of flits injected by each IP per cycle. Each simulation was initially run for 1000 cycles to allow transient effects to stabilize and subsequently it was executed for 20,000 cycles. To calculate average energy, we associate an energy value E_{switch} and $E_{interconnect}$ with each switch and interconnect segment, respectively. The average energy dissipation in transmitting a bit through the NoC is calculated according to equations (5.1) and (5.3).

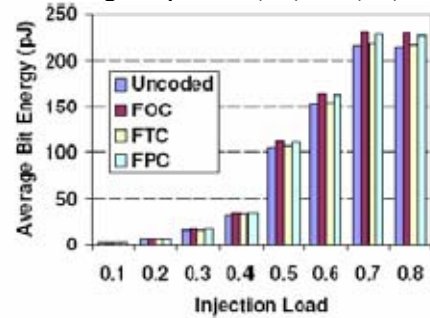


Fig. 4: Bit energy dissipation characteristics for $\lambda=1$

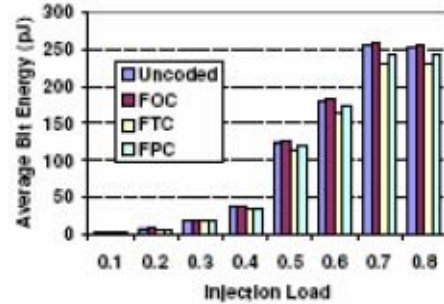


Fig. 5: Bit energy dissipation characteristics for $\lambda=4$

It is evident from Figs. 4 and 5 that when $\lambda=1$, none of the CAC schemes is really energy efficient. For $\lambda=4$, FTC and FPC provide with energy savings, but FOC does not. For $\lambda=1$ the added energy due to the codecs is more than the energy savings arising out of the lowering of the coupling capacitance of the inter-switch wire segments. Table 4 quantifies the energy dissipation characteristics of the CAC schemes for two different values of λ at network saturation.

Table 4: Comparison of Bit Energy Dissipation of CAC schemes

λ	Average Bit Energy Dissipation at Network saturation (pJ)			
	Uncoded	FOC	FTC	FPC
1	215.5	230.4	216.5	228.1
4	253.2	256.6	230.2	241.7

This trend in energy dissipation characteristics for different CAC coding schemes can be explained by

considering the effects of the redundant wires each coding scheme add, on the energy dissipation profile. Though by incorporating CAC the capacitance of each wire is reduced, extra wires are added to the system. We consider an uncoded inter-switch link of 32 bits wide and denote the energy of the whole link by $E_{uncoded}$ and the energy of the coded link including the redundant wires by E_{coded} . The energy dissipated by the codec is denoted by E_{codec} . The energy savings in the interconnect segment arising out of incorporation of coding schemes is denoted by η , where $\eta = (E_{uncoded} - E_{coded})$. We are interested in the situations where $E_{coded} \leq E_{uncoded}$ and η is equal to or greater than E_{codec} as this is the case in which we can achieve net energy savings per link. We plot the ratio

$$\gamma = \frac{\eta}{E_{codec}}$$

as a function of λ for all the three coding schemes under consideration. For those values of λ where $\gamma \geq 1$, we have overall energy savings per link. In these plots we assume that the uncoded and coded link have the same frequency of operation.

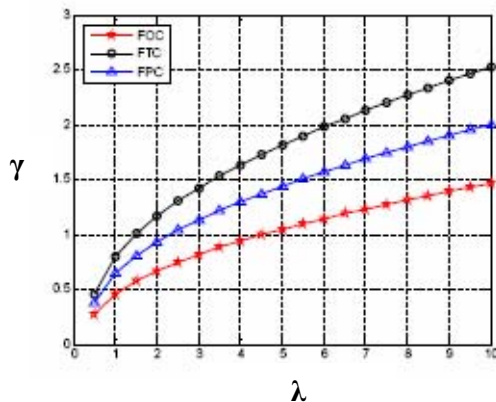


Fig. 6: Energy savings characteristics (CAC schemes are energy efficient when $\gamma \geq 1$)

From Fig. 6 it is clear that for FOC to be energy efficient the value of λ has to be greater than 4. Consequently it is not possible to make the FOC scheme energy efficient even for $\lambda=4$ at 130 nm technology node. On the other hand both FTC and FPC are energy efficient for $\lambda=4$. This explains the characteristics of Figs 4 and 5. It is evident the FTC is the most energy efficient scheme taking all the relevant factors into account.

7. Area overhead

While evaluating the performance of CAC schemes we need to consider the extra silicon area they add to the NoC switch blocks. Through RTL level design and synthesis in 130 nm technology node, we found that the switches, inclusive of the network interface (NI) and without any coding scheme consist of approximately 30K gates. Here, we consider a 2-input minimum-sized NAND structure as a reference gate. In comparison to this the codecs for FOC, FPC and FTC have around 650, 1000 and 770 gates respectively. Consequently the extra area overhead added by the CAC schemes is relatively insignificant.

8. Conclusions and future work

Network on chip is emerging as a revolutionary method to integrate numerous cores in a single SoC. Widespread adoption of NoC paradigm will be possible if it addresses the system level reliability issues in addition to easing the design process. By incorporating Crosstalk Avoidance Codes (CACs) in NoC data stream it is possible to reduce the coupling capacitance of interswitch wire segments and consequently the energy dissipation in communication. Considering a MESH-based NoC We have shown that all the CACs are not energy efficient. Rather the codes for which reduction in interconnect energy, including the redundant wires is more than the additional energy dissipated by the codecs should be used. We aim to make these schemes more energy efficient by modifying the flit structure [10] in such a way that only the header flit needs to be coded/decoded at each intermediate switch between a pair of source and destination IPs. Consequently there would be no codec energy dissipation for the body flits at the intermediate nodes. As a result of this it is expected that there will be more energy savings for the body flits.

In addition to this we are investigating the effect of incorporating Single Error Correction (SEC) codes and CAC together in NoC data stream. CACs help in reducing the coupling capacitance of inter-switch wire segments, but it does not help to protect against any transient malfunction like electromigration, alpha particle hit etc. CACs help to reduce energy in communication and SECs will make the system more robust.

9. References

- [1] L. Benini and G. De Micheli, "Networks on Chips: A New SoC Paradigm," *Computer*, vol. 35, no. 1, Jan. 2002, pp. 70-78
- [2] P. Magarshack and P.G. Paulin, "System-on-Chip beyond the Nanometer Wall," *Proc. 40th Design Automation Conf. (DAC 03)*, ACM Press, 2003, pp. 419-424.

[3] R. Ho, K. W. Mai, M.A. Horowitz, "The Future of Wires", Proceedings of the IEEE, Volume: 89 Issue: 4, April 2001 pp. 490-504

of IEEE International Conference on VLSI Design, 3-7 Jan 2005, pp. 417-422

[4] A. Jantsch, H. Tenhunen (Eds.), *Networks on Chip*, Kluwer Academic Publishers, 2003

[5] S. R. Sridhara, and N. R. Shanbhag, "Coding for System-on-Chip Networks: A Unified Framework", IEEE transactions on very large scale integration (VLSI) systems, vol. 13, no. 6, June 2005, pp. 655-667

[6] C. Grecu, P. P. Pande, A. Ivanov, R. Saleh, "Timing Analysis of Network on Chip Architectures for MP-SoC Platforms", *Microelectronics Journal, Elsevier*, Vol. 36, issue 9, pp. 833-845.

[7] V. Raghunathan, M. B. Srivastava, R. K. Gupta, "A Survey of Techniques for Energy Efficient on-chip Communications", *Proceedings of Design, Automation and Test in Europe (DATE)*, Munich, Germany, June 2003, pp. 900-905

[8] M. R. Stan and W. P. Burlison, "Low-power encodings for global communication in CMOS VLSI", IEEE transactions on very large scale integration (VLSI) systems, Volume 5, Issue 4, Dec. 1997 pp:444 - 455

[9] C. Kretschmar, A. K. Nieuwland, D. Muller, "Why Transition Coding for Power Minimization of on-Chip Buses does not work", Proceedings of the Design, Automation and Test in Europe Conference and Exhibition,(DATE) , 16-20 Feb. 2004 pp. 512-517

[10] P. P. Pande, C. Grecu, M. Jones, A. Ivanov, R. Saleh, "Performance Evaluation and Design Trade-offs for Network on Chip Interconnect Architectures", *IEEE Transactions on Computers*, vol. 54, no. 8, pp. 1025-1040, August 2005.

[11] V. Soteriou and L. S. Peh, "Design-space exploration of power-aware on/off interconnection networks", Proceedings of IEEE International Conference on Computer Design (ICCD), 11-13 Oct. 2004, pp:510 - 517

[12] L. Shang, L. S. Peh and N. K. Jha, "Dynamic voltage scaling with links for power optimization of interconnection networks", Proceedings of the 9th International Symposium on High Performance Computer Architecture (HPCA-9), 8-12 Feb. 2003, pp. 91 - 102

[13] Open Core Protocol, www.ocpip.org.

[14] AMBA Bus specification, <http://www.arm.com>.

[15] P. P. Sotiriadis and A. P. Chandrakasan, "A bus energy model for deep submicron technology", IEEE transactions on very large scale integration (VLSI) systems, Volume 10, Issue 3, June 2002, pp. 341-350

[16] S. R. Sridhara, and N. R. Shanbhag, "Coding for reliable on-chip buses: fundamental limits and practical codes", Proceedings