

A New Approach of Coding to Improve Speed and Noise Tolerance of On-chip Busses

Sébastien Pillement, Olivier Sentieys
IRISA/INRIA - University of Rennes 1
6, rue de Kerampont
22300 Lannion, France
{pillemen, sentieys}@irisa.fr

Jean-Marc Philippe,
CEA, List,
Boîte Courrier 94,
Gif sur Yvette, F-91191 France;
jean-marc.philippe@cea.fr

Abstract—This paper introduces a new coding scheme that faces simultaneously different issues of interconnection design (power, noise and crosstalk). Based on skewing signals on the link, its implementation is very simple and area-efficient. This scheme permits to double bandwidth and to improve noise tolerance through the use of two error detecting codes. The first one uses temporal redundancy and the second one is a parity-based detecting code. This noise tolerance property enables to decrease the power supply voltage to reduce power consumption.

I. INTRODUCTION

The continuous technology scale-down introduces new sources of errors and constraints on large scale circuits due to the decrease of design parameters. Coming from technology process or from physical effects, these noises will require the design of fault-tolerant chips. This is particularly true at the interconnect level which becomes more and more critical due to growing die size and circuits density in deep submicron technologies. Deviation of standard parameters, such as delay, power consumption or electrical parameters are more and more difficult to both predict and control. This leads to an uncertainty in performance evaluation of interconnects. Noise is one of the main issues in future CMOS technologies [1].

Moreover, computation performance improvement leads to an increasing power consumption due to faster clock rates and a higher number of integrated IP (Intellectual Properties) blocks in modern Systems-on-Chip. Dynamic power consumption reduction can be achieved by managing power supply voltage, capacitance, activity or frequency. The most effective technique is to reduce the power supply voltage at the expense of the noise margin. This leads to issues when designing reliable interconnects since the reduced voltage swing of signals is more sensitive to crosstalk issues. Static power consumption is becoming more and more important because of the voltage swing reduction on busses and the decrease of technology nodes. The used techniques to overcome this issue are physical design methods, such as low-leakage devices which are out of the scope of this work.

Power consumption, crosstalk and noise are the three main limitations that must be taken into account when designing

an interconnect [2]. Some techniques exist to alleviate these different issues, but they address one problem at a time. A great challenge is to combine different techniques in a unified coding scheme to improve the effectiveness of interconnect design.

This paper introduces a simple and very efficient coding scheme addressing the three main issues presented above. The goal is to remove worst-case transitions to face crosstalk impact, as well as to increase noise tolerance of on-chip links. This noise tolerance improvement permits to decrease power supply voltage in order to reduce power consumption. The rest of this paper is organized as follows. Section II introduces the different interconnect issues and presents related works. The proposed unified coding scheme is explained in Section III. Section IV deals with the implementations of the encoder and the decoder. Before concluding we present the experimental results in Section V.

II. BACKGROUND

The growing influence of coupling phenomena (called capacitive and inductive crosstalks) is partly due to the increase of the aspect ratio of wires in deep submicron technologies [3]. Inductive crosstalk is the less important one but its influence is growing due the increase of clock rates. This phenomenon is relatively complex to predict because the influence of a wire on another one can span multiple wires (due to magnetic fields). Capacitive crosstalk is caused by the capacitive coupling between adjacent wires. Besides the increase of the aspect ratio of a wire, the coupling between wires is related to the reduction of wire spacing due to technology evolution. Capacitive and inductive couplings have an opposite influence for the same transition pattern, but the capacitive coupling impact is currently predominant.

Crosstalk increases the propagation delay on busses by introducing a relative delay factor (g), as shown in Table I, where r is the ratio between the coupling capacitance between adjacent wires to the capacitance of a wire to the substrate ($r = \frac{C_c}{C_s}$).

In this table, \uparrow represents a rising transition, \downarrow represents a falling transition and $-$ means that there is no transition. The worst-case delay factor on a line under the crosstalk

coupling is $g = 1 + 4.r$. In the best case, when the three wires are switching in the same direction, the delay on the victim wire is the delay without crosstalk (i.e. $g = 1$). For a plausible situation where $C_c = C_s$, the propagation delay can be multiplied by five [4]. Some studies [5] use a parameter r up to 10.

Noise induced by crosstalk represents a second issue in deep submicron designs. A transition on a wire affects the two adjacent wires by applying to them a voltage peak due to the coupling capacitance [6]. With technology scale-down, crosstalk takes a more and more important part of the general noise level. As a consequence, the voltage peak induced by cross-coupling is more and more important compared to the voltage swing on a bus line.

TABLE I
EFFECTIVE CAPACITANCE (C_{eff}) AND DELAY FACTOR (g) AND THE CORRESPONDING TRANSITION PATTERNS.

C_{eff}	Transition Patterns				g
C_s	($\uparrow, \uparrow, \uparrow$)	($\downarrow, \downarrow, \downarrow$)			1
$C_s + C_c$	($-, \uparrow, \uparrow$)	($-, \downarrow, \downarrow$)	($\uparrow, \uparrow, -$)	($\downarrow, \downarrow, -$)	$1+r$
$C_s + 2.C_c$	($-, \uparrow, -$)	($-, \downarrow, -$)	($\uparrow, \uparrow, \downarrow$)	($\downarrow, \downarrow, \uparrow$)	$1+2.r$
$C_s + 3.C_c$	($-, \uparrow, \downarrow$)	($-, \downarrow, \uparrow$)	($\uparrow, \downarrow, -$)	($\downarrow, \uparrow, -$)	$1+3.r$
$C_s + 4.C_c$	($\uparrow, \downarrow, \uparrow$)	($\downarrow, \uparrow, \downarrow$)			$1+4.r$

The basic method to face this capacitive crosstalk consists in removing worst-case patterns presented in table I. Shielding and duplication are the most famous techniques. Shielding consists in inserting a grounded line between every couple of wires. All the transitions with two adjacent lines switching in opposite directions are removed. An evolution can be found in [7] in which the signals are routing using the pattern $VSGSVSGS\dots$, where V represents a V_{dd} wire, S represents a signal wire and finally G represents a grounded wire. This technique permits to remove some wires. Duplication of each wire is also used to eliminate worst-case transitions. The acceleration of the signals is higher comparing to the shielding technique because all the patterns with two invariant aggressors are removed. But this technique also increases the bus activity. The above techniques lead to additional lines and finally contribute to increase the interconnect area. Moreover the added wires are useless for the transmission.

Coding is another promising approach which basic principle is to transmit codewords instead of the original words. It has been shown in [8] that a 32-bit data word can be transmitted with 53 wires using partial coding. In [5], the authors present a coding scheme which removes all $1+4.r$ and $1+3.r$ patterns. It has 62,5% wire overhead and a delay improvement of 50%. An improvement of this coding scheme removes $1+2.r$ patterns at the cost of 200% area overhead. In [9] six groups of transition patterns with different propagation times and a fast clock are defined. A crosstalk analyzer assigns two consecutive words to one of the six delay groups and adapts the number of transmission cycles needed to send the word. This scheme does not eliminate crosstalk patterns but adapts the length of the transmission to them.

Crosstalk also raises power consumption through the increase of the effective capacitance of the victim wire. This is known as the Miller effect and it is also shown in Table I. The energy consumption can be ten times the one in the best case for short wires due to this phenomenon. This factor decreases with long wires. Since it is generally recognized that interconnects represent up to 50% of the power consumption of a chip [10], optimizing power consumption of interconnects is of high interest.

Low-swing signaling is the most efficient technique to lower power consumption because of its quadratic influence [10]. Decreasing the capacitance that must be loaded by the bus drivers [11] is another technique to control energy consumption. Some studies have been conducted on coding techniques which aim at decreasing the signals activity, such as bus-invert coding [12]. But new investigations have shown that it is not as efficient as it might be due to codecs overhead [13].

Noise influence is becoming a great challenge in deep submicron design because of the power supply voltage reduction and the increase of coupling capacitances. Even if a part of the global noise is reduced thanks to low-swing signaling and crosstalk alleviation methods, external noise caused by soft errors or electromagnetic interferences is increasing compared to the voltage swing. The use of error detecting or correcting schemes will be required, such as parity or Hamming codes [14] [15]. The fact that noise is varying with the environment or with the data patterns that are transmitted on busses makes difficult to design a power efficient robust coding scheme. One solution is to adapt dynamically the coding scheme to the noise level in which the circuit works [16].

Unfortunately these different techniques do not handle the global problem of interconnection network. In [17], the authors present a framework to explore different combinations of the previous techniques to alleviate the different issues. They introduce the concept of on-chip unified coding to produce an area-efficient coding scheme that can decrease the influence of crosstalk, lower power consumption and increase noise tolerance of busses. All the different compared schemes exhibit trade-offs between area overhead, speed-up and energy consumption.

We propose in this paper a new efficient unified coding scheme avoiding worst-case delay patterns and improving noise tolerance. The proposed approach is simple enough to enable area-efficient implementations.

III. DESCRIPTION OF THE PROPOSED METHOD

Crosstalk phenomenon is reduced using signal skewing as it is shown in [18]. This technique consists in shifting adjacent signals in order to avoid critical switching pattern: the remaining patterns in Table I are only ($-, \uparrow, -$) and ($-, \downarrow, -$). The delay factor is greatly reduced compared to the worst-case patterns. In [18], authors shift individual bit of a same bus, we have extended this concept at the bus level. Since classical techniques of crosstalk avoidance increase the bus size with unused wires (shielding, duplication), we propose to

use added signals to transmit a second data-word on the bus. This second data is delayed from the first data words in order to avoid switching of adjacent wires (Fig. 1). The length of the signal skewing between odd and even bus wires is half the transmission clock period. Thus, by setting the transmission clock period to at least twice the propagation time on the bus, the transmitted signal can be changed during the first half of the clock period and will be stable during the rest of the time while adjacent signals can do their own transition. This choice permits a very simple implementation as we will show in the next section. In Figure 1 we show the bus state during a data transmission. Two consecutive words are transmitted: one on even wires of the bus and the other one on odd wires of the bus. In this example, 32-bit data are transmitted.

Two different techniques are used to improve the noise tolerance of the link: a temporal error detecting technique and a spatial one. As the transmission clock period is set at twice the minimum clock period as said above, the total transmission time for a word is equivalent to at least two times the propagation time. We can have two samples for each bit at the decoder side. In fact, transition of the second data word occur between the two samples of the first signal. Due to capacitive coupling this phenomenon can have a great influence on a victim wire: an error on the first sample can be recovered or an error can appear on the second sample (e.g. bit flipping errors). Thus, the temporal error detecting technique consists in comparing the two samples of each bit at the decoder side: if they are not identical, an error is detected and the corresponding word can be retransmitted.

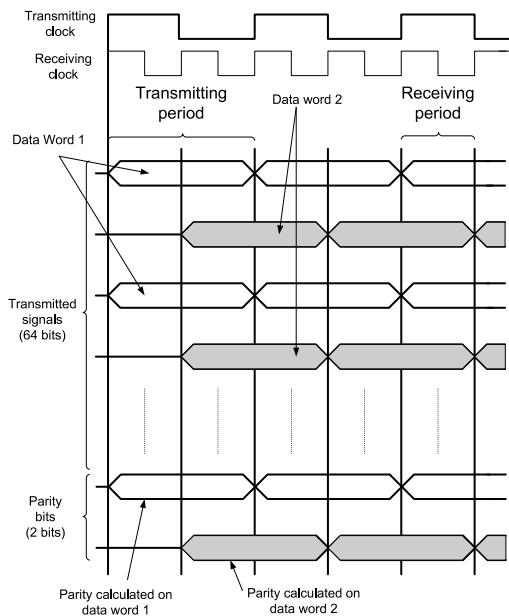


Fig. 1. Illustration of the bus state for some data transmissions.

The second error detecting technique consists in computing a traditional parity bit for each transmitted word as shown in Figure 1. The parity bit corresponding to a data word is transmitted along with the word using the signal skewing technique.

So it does not decrease the speed-up obtained. The parity technique is used because of its area-efficient implementation. But every error detecting or correcting scheme can be used because the crosstalk alleviation method is totally independent from the noise tolerance improvement techniques.

IV. IMPLEMENTATION OF THE PROPOSED TECHNIQUE

Lot of efficient methods proposed in the literature lead to inefficient implementation balancing the gain of the coding scheme. The encoder and the decoder required for our coding scheme are described in this section. The area results are then given for a $0.13\mu\text{m}$ CMOS technology. In the following paragraphs, we present the encoder, the decoder and the hardware implementation results.

The encoder will compute the parity bit of the transmitted word and have to skew the different signals for crosstalk phenomenon alleviation. The architecture of encoder is given in Figure 2.

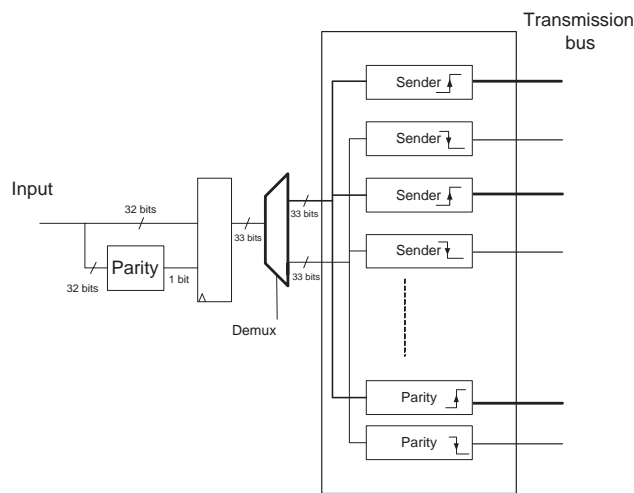


Fig. 2. Encoder scheme.

The parity bit is computed using a simple XOR tree. The implementation of the signal skewing technique is very simple. It consists in using alternatively a positive edge-triggered flip-flop and a negative edge-triggered flip-flop (the sensitivity of the edge-triggered flip-flop is given by the arrows on Figure 2). Since the transmission clock signal has a period that is twice the propagation delay on the considered bus, using rising edge and falling edges of the clock is not too problematic. The computation of the parity bit and the commutation between even and odd wires for the input words need a clock that is two times faster than the transmission one (i.e. running at the nominal frequency). The skewing technique permits to load no more than 33 wires at the fastest clock rate for a 66-bit bus, helping the designer to have quite the same activity at the same clock rate than a standard transmission.

The decoder is a little bit more complex as it must first oversample the input signals to make the acquisition of the redundant information, compare the two samples for each bit,

compute the parity of the transmitted words and compare it with the transmitted one. The decoder architecture scheme is given on Figure 3.

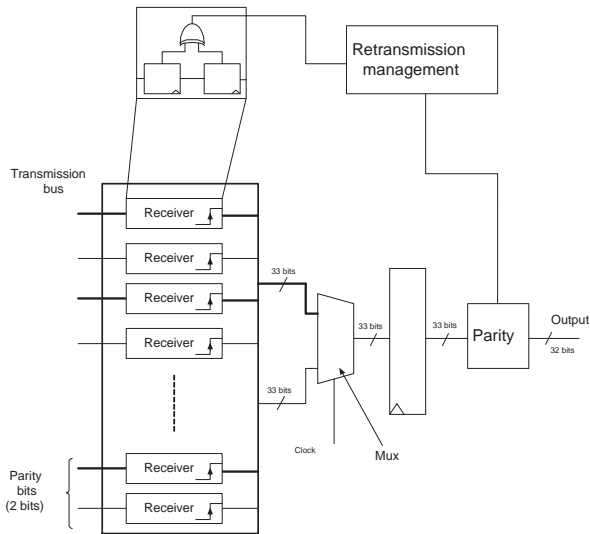


Fig. 3. Decoder scheme.

The decoder oversamples each of the transmitted bits to compare the value of the bit before and after the possible transition on adjacent wires. The receiver is composed of positive edge-triggered flip-flops running at the nominal frequency. If the two samples are identical, the parity is computed by an XOR tree and this result is compared with the transmitted parity bit. If there is a detected error at one of these two steps, then the retransmission process is invoked. This process is specific to the bus or network-on-chip that carries the data. We have implemented a simple OR tree for proof-of-concepts. Using correcting codes instead of detecting ones only will have an impact on the receiver area.

The encoder and the decoder were synthesized using Synopsys Design Compiler and the $0.13\mu\text{m}$ CMOS library from UMC (United Microelectronics Corporation). The critical path is 1.07ns for the encoder and 0.54ns for the decoder. These results are for simple implementation of codecs and can be optimized by pipelining computation.

The area of the codecs are given in Table II. The combinational area of the decoder is larger than the one of the encoder due to the OR tree for the notification of the retransmission process.

Device ($0.13\mu\text{m}$ technology)	Encoder	Decoder
Combinational area (μm^2)	355.9	1762.9
Non combinational area (μm^2)	4944.2	6141.8
Total area (μm^2)	5300.1	7904.7

TABLE II
SYNTHESIS RESULTS FOR THE CODECS.

V. PERFORMANCES

We used a metal-2 bus in the $0.13\mu\text{m}$ technology from UMC. The propagation delays of the signals were obtained using SPICE. The drivers are adapted to the length of the wires giving 1X minimum drive for a 1mm wire and 10x minimum drive for a 10mm wire. The performances of our codec scheme are given regarding the propagation delay, the noise tolerance improvement and the power consumption.

The results in terms of delay are given in Table III. With the signal skewing technique, the only patterns that can appear are either $(-, \uparrow, -)$ or $(-, \downarrow, -)$. Compared to the worst-case delay respectively 1.41ns and 3.29ns (i.e. in un-encoded bus), we obtain a speed-up of 2.35 for a 1mm wire and 2.32 for a 10mm wire. This means that the bandwidth is more than doubled if we double the number of wires on the bus, just by skewing the signals. We can note that speed-up is not dependant of the wire length. An interesting property of this scheme is that we limit the available patterns to those in which the adjacent wires to a victim wire are not doing a transition. This fact permits to control precisely the propagation time on links between routers for example. The well-controlling of electrical parameters is claimed to be very important and an interesting property of NoC [19]. Delay parameters are interesting to control as well.

A1	V	A2	1mm wire [ns]	10mm wire [ns]
\uparrow	\uparrow	\uparrow	0.17	0.32
\uparrow	\uparrow	-	0.24	0.52
\uparrow	\uparrow	\downarrow	0.31	1.36
-	\uparrow	-	0.47	1.37
-	\uparrow	\downarrow	0.80	2.44
\downarrow	\uparrow	\downarrow	1.17	3.21
\downarrow	\downarrow	\downarrow	0.18	0.34
\downarrow	\downarrow	-	0.29	0.55
\downarrow	\downarrow	\uparrow	0.49	1.46
-	\downarrow	-	0.60	1.42
-	\downarrow	\uparrow	1.05	2.51
\uparrow	\downarrow	\uparrow	1.41	3.29

TABLE III
PROPAGATION DELAY ON V (VICTIM) DEPENDING ON TRANSITIONS OF A1 AND A2 (AGGRESSORS).

The residual word error probability can be calculated using the assumptions made in [1] and [14]. In this works

$$\epsilon = Q\left(\frac{V_{dd}}{2\sigma_N}\right) \quad (1)$$

is the probability of having an error at the reception of a symbol, with

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{v^2}{2}} dv \quad (2)$$

V_{dd} is the power supply voltage and σ_N the standard noise deviation.

Assuming that formula, the residual word error probability of an unencoded n-bit word is:

$$P_{Orig} = 1 - (1 - \epsilon)^n \quad (3)$$

For a parity-based error detection scheme, the residual word error probability of an n-bit word (with an additional parity bit) is:

$$P_{Par} = \sum_{i=1}^{n/2} \binom{n+1}{2i} \epsilon^{2i} \cdot (1-\epsilon)^{(n+1-2i)} \quad (4)$$

with

$$\binom{n}{k} = \frac{n!}{(n-k)!k!} \quad (5)$$

Our coding scheme has two error detection parts: the temporal redundancy and the spatial parity. If we consider ϵ as the probability that a sample of a bit can be wrong, the probability of non detection of an error on a bit by the temporal redundancy is ϵ^2 . Then, the parity-based scheme can also detects possible errors. In fact, the total residual word error probability is given by the following equation (for a n-bit word):

$$P_{total} = \sum_{i=1}^{n/2} C_{n+1}^i \epsilon_1^{4i} \cdot (1-\epsilon_1)^{2 \cdot (n+1-2i)} \quad (6)$$

The error probabilities for un-encoded signal, parity based error detection and the proposed coding are plotted on Figure 4. These results are given for an 8-bit word with a standard deviation of 0.2V for the noise signal. As we can see at a power supply of 1.2V, the parity based coding achieves an error probability of 10^{-4} , while our scheme can achieve up to 10^{-10} . Figure 4 shows the noise tolerance improvement provided by the proposed technique in all cases.

The energy consumption due to transitions on the wires is given in table IV. It is noticeable that our scheme eliminates the worst-case high power consuming transition but also the low power consuming transitions. The weighted mean energy consumptions are 143,4fJ for a 1mm wire and 1.83pJ for a 10mm wire. In fact, removing best-case and worst-case patterns has an insignificant impact on energy consumption due to a random signal propagation. The energy consumption per bit of the codecs was computed after synthesis. The energy per transmitted bit for the encoder is about 15fJ/bit and the energy per bit is about 17.5fJ/bit for the decoder.

A1	V	A2	E in fJ (1mm)	E in fJ (10mm)
↑	↑	↑	30	1.14
↑	↑	-	84	1.21
↑	↑	↓	145	1.81
-	↑	-	143	1.77
-	↑	↓	201	2.42
↓	↑	↓	260	2.94

TABLE IV

ENERGY CONSUMPTION OF A VICTIM WIRE DEPENDING ON AGGRESSORS (0.13 μ m PROCESS).

The great improvement in noise tolerance given by our coding scheme enables us to lower power supply voltage in order to lower the power consumption. As it is shown in Figure 4, for the same residual word error probability of 10^{-4}

obtained with a parity scheme with a 1.2V power supply voltage, we can reduce this voltage to less than 0.7V. This will greatly improve the dynamic power consumption of the interconnect.

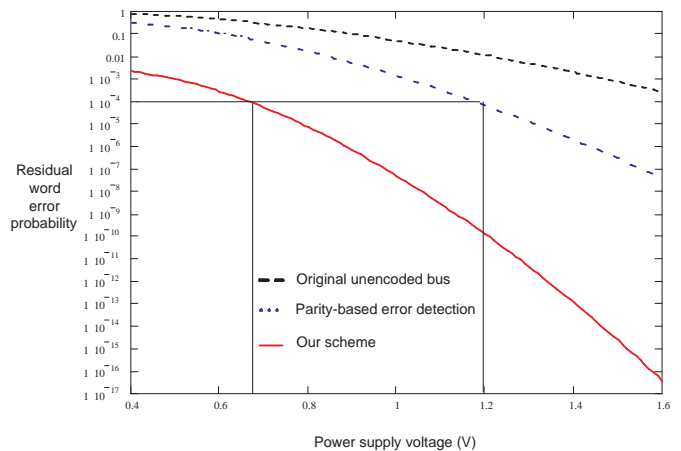


Fig. 4. Residual word error probabilities as a function of the power supply voltage for a standard deviation of 0.2V and an 8-bit word.

VI. CONCLUSION

We have introduced in this paper a new unified coding scheme for bus-based or NoC-based interconnects. Using simple techniques to be area and power-efficient, this scheme permits to improve bandwidth by a factor of more than 2.3 for a doubled number of wires and to have a well-controlled propagation time. Due to its simplicity, this scheme has low area requirements. Its error detection capability enables designers to conceive high-speed and low-power error tolerant links. Additionally, the improvement of noise tolerance enables to dramatically decrease the power supply voltage, and thus to greatly improve dynamic power consumption of interconnects. Future work will consist in adapting this scheme to other error detection codes such as Hamming codes to measure performance improvement.

REFERENCES

- [1] R. Hedge and N. R. Shanbhag, "Toward achieving energy efficiency in presence of deep submicron noise," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 8, no. 4, pp. 379–391, August 2000.
- [2] N. R. Shanbhag, "Reliable and Efficient System-on-Chip Design," *IEEE Computer*, vol. 37, no. 3, pp. 42–50, March 2004.
- [3] ITRS, "http://public.itrs.net/files/2003itrs/home2003.htm," International Technology Roadmap for Semiconductors, Tech. Rep., 2003.
- [4] J. M. Rabaey, A. Chandrakasan, and B. Nikolic, *Digital Integrated Circuits: A Design Perspective*. Prentice Hall, 2002.
- [5] C. Duan and A. Tirumala, "Analysis and Avoidance of Cross-Talk in On-Chip Buses," in *Proceedings of the Symposium on High Performance Interconnects (HOTI '01)*, Stanford, California, USA, 2001.
- [6] A. Devgan, "Efficient Coupled Noise Estimation for On-Chip Interconnects," in *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design (ICCAD'97)*, San Jose, California, USA, November 1997, pp. 147–153.

- [7] S. P. Khatri, R. K. Brayton, and A. L. Sangiovanni-Vincentelli, *Crosstalk Noise Immune VLSI Design using Regular Layout Fabrics*. Kluwer Academic Publishers, 2001.
- [8] B. Victor and K. Keutzer, "Bus Encoding to Prevent Crosstalk Delay," in *Proceedings of the International Conference on Computer-Aided Design (ICCAD '01)*, San Jose, California, USA, 2001, p. 57.
- [9] L. Li, N. Vijaykrishnan, M. Kandemir, and M. J. Irwin, "A Crosstalk Aware Interconnect with Variable Cycle Transmission," in *Proceedings of the IEEE/ACM International Conference on Design, Automation and Test in Europe (DATE'04)*. Paris, France: IEEE Computer Society, 2004, p. 10102.
- [10] H. Zhang, V. George, and J. M. Rabaey, "Low-Swing On-Chip Signaling Techniques: Effectiveness and Robustness," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 8, no. 3, pp. 264–272, June 2000.
- [11] C.-T. Hsieh and M. Pedram, "Architectural power optimisation by bus splitting," *Proceedings of the IEEE/ACM International Conference on Design, Automation and Test in Europe (DATE'00)*, 2000.
- [12] M. R. Stan and W. P. Burleson, "Bus-invert coding for low-power i/o," *IEEE Transactions on VLSI Systems*, vol. 3, no. 1, pp. 49–58, 1995.
- [13] C. Kretschmar, A. K. Nieuwland, and D. Mller, "Why Transition Coding for Power Minimization of on-Chip Buses does not work," in *Proceedings of the IEEE/ACM International Conference on Design, Automation and Test in Europe (DATE'04)*, Paris, France, 2004, pp. 10512–10517.
- [14] D. Bertozzi, L. Benini, and G. DeMicheli, "Low-Power Error-Resilient Encoding for On-Chip Data Buses," in *Proceedings of the IEEE/ACM International Conference on Design, Automation and Test in Europe (DATE'02)*, Paris, France, 2002, pp. 102–109.
- [15] D. Bertozzi, L. Benini, and G. D. Micheli, "Error control schemes for on-chip communication links: the energy-reliability tradeoff," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 24, no. 6, pp. 818–831, June 2005.
- [16] L. Li, N. Vijaykrishnan, M. Kandemir, and M. J. Irwin, "Adaptive Error Protection for Energy Efficiency," in *Proceedings of the International Conference on Computer Aided Design (ICCAD'03)*, San Jose, California, USA, 2003, pp. 2–7.
- [17] S. R. Sridhara and N. R. Shanbhag, "Coding for system-on-chip networks: a unified framework," in *Proceedings of the IEEE/ACM Design Automation Conference (DAC'04)*, San Diego, California, USA, June 2004, pp. 103–106.
- [18] K. Hirose and H. Yasuura, "A Bus Delay Reduction Technique Considering Crosstalk," in *Proceedings of the IEEE/ACM International Conference on Design, Automation and Test in Europe (DATE'00)*. Paris, France: IEEE Computer Society, 2000, p. 441.
- [19] W. J. Dally and B. Towles, "Route packets, not wires: On-chip interconnection networks," in *Proceedings of the IEEE/ACM Design Automation Conference (DAC'01)*, Las Vegas, Nevada, USA, 2001, pp. 684–689.