

Coupling-Driven Signal Encoding Scheme for Low-Power Interface Design*

Ki-Wook Kim, Kwang-Hyun Baek, Naresh Shanbhag, C. L. Liu[†] and Sung-Mo Kang

Coordinated Science Laboratory, Univ. of Illinois at Urbana-Champaign, USA

[†] Dept. of Computer Science, National Tsing Hua University, Taiwan

Abstract

Coupling effects between on-chip interconnects must be addressed in ultra deep submicron VLSI and system-on-a-chip (SoC) designs. A new low-power bus encoding scheme is proposed to minimize coupled switchings which dominate the on-chip bus power consumption. The coupling-driven bus invert method use slim encoder and decoder architecture to minimize the hardware overhead. Experimental results indicate that our encoding methods save effective switchings as much as 30% in an 8-bit bus with one-cycle redundancy.

1 Introduction

Increased coupling effect between interconnects in ultra deep submicron technology not only aggravates the power-delay metrics but also deteriorates the signal integrity due to capacitive and inductive crosstalk noises. Conventional approaches to interconnect synthesis aim at optimal interconnect structures in terms of interconnect topology, wire width and spacing, and buffer location and sizes [3]. In this paper, we study a *signal encoding scheme* to minimize coupling effects between interconnects.

Signal encoding schemes have been proposed to minimize transition activities on buses while ignoring cross-coupled capacitances. When statistical properties are unknown a priori, the bus-invert method [14] and the on-line adaptive scheme [1] can be applied to encode randomly distributed signals. On the other hand, highly correlated access patterns exhibit a spatio-temporal locality which can be exploited for energy reduction [11] in Gray code [9, 16], the T0 method [7], and the working-zone encoding [10]. Lower bounds for minimum achievable transition activity have been derived for noiseless buses in [12] and for noisy buses in [5]. In [17], a segmentation method was introduced to reduce power consumption. Specification transformation approaches were used to reduce the number of memory accesses at the behavioral level [2]. The effectiveness of various encoding schemes was compared at the system level in [4].

Most of the previous bus-encoding schemes were designed to minimize transition activities on each signal line as if each line were isolated from neighboring lines, hence ignoring coupling effects. Such an assumption may be valid for off-chip buses where the impedances of transmission lines are appropriately adjusted. However, this is not the case for long on-chip buses which are particu-

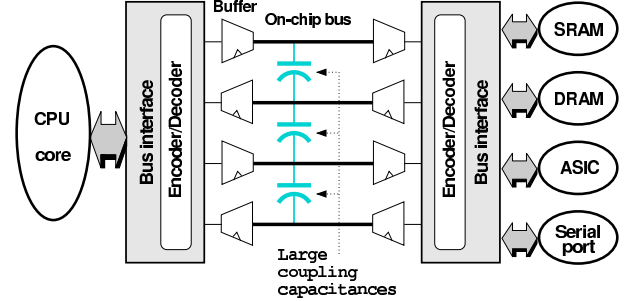


Figure 1: A tightly cross-coupled on-chip buses in a system-level chip design

larly prevalent in a system-on-a-chip. For example, the wire aspect ratio is expected to be over 2.4 for intermediate wiring (namely, third metal layer and fourth metal layer) in 0.18 μm seven-layered metal process [13]. Accordingly, coupling has become an important issue with scaled supply voltage when we consider signal integrity and power dissipated by coupling capacitances, referred to as *coupling power*. Shielding can be a way to avoid crosstalk problem with area overhead.

In this paper, we propose a new encoding scheme for static on-chip bus structure to minimize coupling power. The key idea is that coupling effects could be alleviated by transforming the signal sequences traveling on-chip buses that are closely placed. Small blocks of encoding and decoding logic are employed at the transmitter and receiver of on-chip buses as shown in Figure 1. The encoder and decoder (codec) should have a low-complexity architecture so that the power and delay overhead due to the codec circuitry can be compensated by significant savings in switching activities on tightly coupled buses.

2 Interconnect Power Characteristics

The average energy consumed by a wire with clock frequency $f = \frac{1}{T_{clk}}$ can be [6]

$$E_{av} = \lim_{n \rightarrow \infty} \frac{\int_0^{n \cdot T_{clk}} V_{dd} \cdot I_c(t) dt}{n} = V_{dd} \cdot \Delta Q_{av} \quad (1)$$

where n is the number of clock cycles observed and $I_c(t)$ represents the drawn current due to transitions in a clock period. ΔQ_{av} is the time-averaged charge provided by the power supply to all capacitances of the interconnect and is given by

$$\Delta Q_{av} = p \cdot C_{tot} \cdot \Delta V \triangleq \widehat{C}_{tot} \cdot V_{dd} = (\widehat{C}_s + \widehat{C}_x) \cdot V_{dd} \quad (2)$$

where p denotes the switching probability, and C_{tot} denotes the total lumped capacitance. The effective capacitance \widehat{C}_{tot} is defined by both physical capacitances and switching activities. The effective capacitance accounts for time-averaged charge stored in physical capacitances provided by the power supply.

*Supported in part by National Science Foundation under grant NSF MIP 96 12184, by Intel under grant 5414 and by SRC under 98-HJ-641.

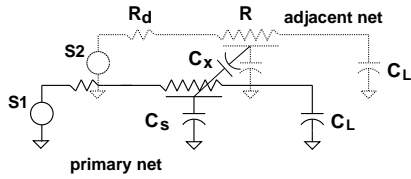


Figure 2: A distributed RC model for the interconnects.

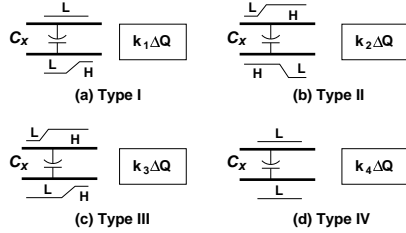


Figure 3: Transition types: (a) Single line switching; (b) both lines switching in opposite direction; (c) both lines switching in the same direction; (d) no switching.

One can model the physical capacitance of a wire as shown in Figure 2. The capacitive components are *self capacitance* C_s and *coupling capacitance* C_x . The self capacitance C_s represents the total lumped capacitance including parallel-plate capacitance and fringe capacitance. For each capacitive component, we define the effective capacitance as

$$\hat{C}_s = Y \cdot C_s \quad (3)$$

$$\hat{C}_x = Z \cdot C_x \quad (4)$$

where Y and Z are the average number of effective transitions per cycle for C_s and C_x , respectively, which are computed as follows.

First, we quantify the *self transition activity* Y for the self capacitance C_s . Let $p_{x,y}^i$ denote the transitional probability that the signal value of i changes from x to y , which can be represented by the signal probability $p(i_{=x}^n)$ and the conditional probability $p(i_{=y}^{n+1} | i_{=x}^n)$ such that $p_{x,y}^i = p(i_{=x}^n) \cdot p(i_{=y}^{n+1} | i_{=x}^n)$, where $x, y \in \{0, 1\}$. If we assume that there is no glitch in the signals, the self transition activity Y is given by

$$Y = p_{0,1}^i \quad (5)$$

since the capacitance C_s will be charged up only when a low-to-high transition takes place.

Next, the *coupled transition activity* Z is computed according to the correlated switchings between physically adjacent interconnects. There are four types of possible transitions when we consider dynamic charge distribution over coupling capacitances as illustrated in Figure 3. We suppose that there are two parallel wires placed with minimum spacing. A type I transition occurs when one of the signals switches while the other stays unchanged such that the coupling capacitance is then charged up to $k_1 C_x V_{dd}$, where the coefficient k_1 is introduced as a reference for other types of transition. In a type II transition, one bus switches from low to high while the other switches from high to low. The effective capacitance will be larger than k_1 by a factor of k_2 the value of which is usually two. In a type III transition, both signals switch simultaneously and C_x will not be charged. However, because of possible misalignment of the two transitions, the amount of power consumption varies according to the dynamic characteristics by a factor of k_3 . In a type IV transition, there is no dynamic charge distribution over coupling capacitance. Thus, we set k_4 to zero.

Let $p_{xy,qr}^{ij}$ denote the joint transitional probability defined by $p_{xy,qr}^{ij} = p(i_{=x}^n \wedge j_{=y}^n) \cdot p(i_{=q}^{n+1} \wedge j_{=r}^{n+1} | i_{=x}^n \wedge j_{=y}^n)$, where $x, y, q, r \in$

$\{0, 1\}$. Each type of transition contributes to the effective coupling capacitance between a wire i and a wire j as follows.

$$Z = k_1(p_{00,01}^{ij} + p_{00,10}^{ij} + p_{11,01}^{ij} + p_{11,10}^{ij}) + k_2(p_{01,10}^{ij} + p_{10,01}^{ij}) + k_3(p_{00,11}^{ij} + p_{11,00}^{ij}) \quad (6)$$

The total lumped capacitance for a bus can be computed according to Equations (5) and (6). Accordingly, the dynamic power consumed by the interconnects and drivers is given by

$$P_{dyn} = (Y(C_s + C_L) + ZC_x) \cdot V_{dd}^2 \cdot f_c \quad (7)$$

We define the *capacitance ratio* $\eta = \frac{C_x}{C_s + C_L}$ for a terminated bus. The capacitance ratio increases as the aspect ratio of the interconnect increases.

3 Low Power Encoding Schemes

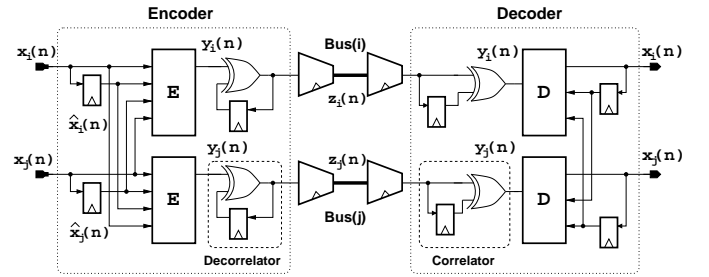


Figure 4: Low-power encoder-decoder (codec) framework

3.1 General Codec Architecture

Figure 4 illustrates a generic codec architecture for two bit signals. The encoder consists of three components: a predictor, an encoding function block E , and a decorrelator. The prediction function $\hat{x}(n)$ is a function of past input values given by $\hat{x}(n) = f(x(n-1), x(n-2), \dots, x(n-K))$. We consider $K = 1$ for low-complexity codec architecture. The combinational function E reduces the average number of self transitions and coupled switchings between $z_i(n)$ and $z_j(n)$. The encoding function E differs from the architectures proposed in [1, 12] in that it takes as input the data on adjacent buses to account for coupling effect. In general, for an input signal $x_i(n)$, the encoding function is given by $y_i(n) = E(x_i(n), \hat{x}_i(n), x_{i-1}(n), \hat{x}_{i-1}(n), x_{i+1}(n), \hat{x}_{i+1}(n))$, where $x_{i-1}(n)$ and $x_{i+1}(n)$ are neighboring signals. The input data $x_i(n)$ and the prediction function $\hat{x}_i(n)$ account for the reduction of self transition activities in $z_i(n)$. Signal integrity and coupling power depend on both the current value of neighboring signals ($x_{i-1}(n)$ and $x_{i+1}(n)$) and the transition histories ($\hat{x}_{i-1}(n)$ and $\hat{x}_{i+1}(n)$). The decorrelator employs a transition encoding scheme, whereby a transition is encoded with the logic value 1 and no transition is encoded with the logic value 0.

As a mirror of the encoder, the decoder consists of three components: a correlator, a decoding function block D , and a register to keep the prediction function $\hat{x}(n)$. The decoding logic function is given by $x_i(n) = D(y_i(n), \hat{x}_i(n), y_{i-1}(n), \hat{x}_{i-1}(n), y_{i+1}(n), \hat{x}_{i+1}(n))$. The decoder D realizes the inverse function of the encoder E .

In choosing a codec scheme, we need to take into account two major issues. The first criterion is a tradeoff between architecture complexity and encoding efficiency. Rent's rule [8] states that there is a simple power-law relationship between the number of I/O terminals for a logic block and the number of gates contained in that

block for a given degree of parallelism. It means that considerations for adjacent input pins are likely to add logic gates to a codec functional block. Increased number of logic gates in turn can induce overhead power consumption and longer propagational delay. Therefore, it should be ensured that benefits from data encoding are large enough to compensate for the overhead of a codec architecture. Secondly, the codec system should guarantee the unique decodability constraints, even in the presence of physical noises. The signal integrity can be ensured by asserting spatial redundancy (extra control lines) or temporal redundancy (extra clock cycles) or by selecting appropriate supply voltage, the size of transmitter and receiver, and the clock frequency.

3.2 The Coupling-Driven Bus-Invert Scheme

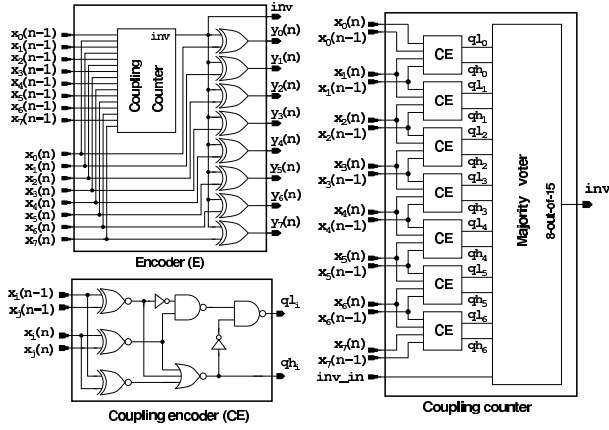


Figure 5: An 8-bit bus encoder for the coupling-driven bus-invert scheme.

The bus-invert method [14] is limited to reduce transition activities while assuming that coupling power contribution can be ignored. However, coupling power becomes a dominant component of dynamic power as wires become thinner and taller. To reflect the technology trend appropriately, we propose a *coupling-driven bus-invert* method to tackle the coupling power reduction problem. The bus-invert method flips the data signal when the number of switching bits is more than half of the number of signal bits. In the same context, we invert the input vector, when the coupling effect of the inverted signals is less than that of the original signals. The problems are then how to accurately account for the coupling effect, and to effectively implement the scheme with low hardware overhead.

Before addressing these issues, we state our assumptions. First, synchronous latches are located at the transmitter side, thus all the transitions shall take place at the same time on the bus. The simultaneous transitions exclude type III transitions by setting $k_3 = 0$. It means that the results we achieve are on the lower end of power saving. Second, statistics on the information source are not given in advance. Hence this scheme is suitable for data bus encoding, where it is difficult to extract accurate probabilistic information off-line.

Enumeration method is employed to represent the coupling effect. If a bus line B_i is located between two other lines, a signal transition on B_i can trigger charge shifts on both coupling capacitances connected to B_{i-1} and B_{i+1} , respectively. In other words, at most two couplings can be initiated by a signal transition. Thus, $2(N-1)$ bits are sufficient to represent the whole set of couplings in an N -bit bus per bus cycle. The encoder architecture is shown in Figure 5.

According to the types of correlated transition between neighboring buses, the coupling encoder generates a codeword as fol-

lows: 00 for a type III or IV transition, 01 for a type I transition, and 11 for a type II transition. The reason that we assign 11 to a type II transition is that switchings in different directions require to change the polarity of the charge stored in the coupling capacitance, hence consuming about twice the amount of charge required for a type I transition. The codeword 11, instead of 10, helps to make a decision on data inversion using a majority voter, because the majority voter outputs high when at least eight input lines are high out of fifteen inputs. The majority voter can be implemented by using either full-adder circuitry or resistors and a voltage comparator [14]. The control signal inv can be transmitted to the receiver using extra bus lines or extra transfer cycles. One problem of additional bus lines for control is the area overhead that may not be allowed due to physical constraints. In some cases, widening the space between signal bus lines can reduce the coupling effects more effectively than introducing extra control lines, because the coupling capacitance is inversely proportional to net space. Temporal redundancy is an alternative using extra clock cycles to transfer control signals. We assume that the input stream is transmitted in burst mode that enables us to accommodate temporal redundancy [15].

The following theorem gives the expected number of couplings per bus cycle with independent data source where each bit switches independently with transitional probability p .

Theorem 3.1 Let $\{B_i\}$ be an N -tuple of 0-1 valued random variables such that $B_i = \langle b_0, b_1, \dots, b_{N-1} \rangle$, in which a bit b_k ($0 \leq k \leq N-1$) switches independently of other bits with transitional probability $p_k = p$ ($0 \leq k \leq N-1$). Suppose $\{B_i\}$ are encoded with the coupling-driven bus-invert algorithm with one-cycle redundancy. Then, the coupled transition activity Z per bus cycle is given by

$$Z = \frac{N-1}{2} - \frac{N \cdot (2N-3)}{2^N \cdot (N-2)} \cdot \left(\frac{N-2}{2} \right) \quad (8)$$

if transitions on $\{B_i\}$ follow the binomial distributions.

Example 3.1: The average number of couplings per bus cycle for an 8-bit bus is evaluated to be 2.484. The average number of couplings per bus cycle for randomly distributed, independent data is 3.5, since $p = 0.5$ and $N = 8$. Hence, the percentage reduction in couplings over raw data transmission is 29% which is calculated as $\frac{3.5-2.484}{3.5} * 100$ when we use the coupling-driven bus-invert method. ■

The coupling-driven bus-invert scheme can employ multiple cycles to notify the receiver whether the transmitted data are inverted or not. Multiple extra cycles for control signals imply that the granularity of unit codeword to be processed becomes finer, hence less couplings are likely to occur on the bus. However, this benefit comes at the cost of extra clock cycles.

4 Experimental Results

The methods proposed were implemented and power consumptions were measured using HSPICE. Applied data streams consist of MPEG video files (V1, V2 and V3 in Table 1), MP3 audio files (A1, A2 and A3) and PDF format files (P1, P2 and P3).

Table 1 presents percentage reduction in transition activities using the bus-invert (BI) method and proposed coupling-driven bus-invert (CBI) method. To compare fairly, we use one redundant cycle to transfer control signals per eight data packets each of which consists of 8 bits. The reason that we compare the CBI scheme with the BI method is that the BI method does not require probabilistic information in advance. The results under the column SP present percentage reduction in self transitions, calculated as:

Table 1: Percentage reduction in transition activity compared to raw data transmission using the bus-invert method (BI) and the coupling-driven bus-invert method (CBI) with one-cycle redundancy for an 8-bit bus.

Data	BI [14]			CBI		
	XP	SP	TP	XP	SP	TP
V1	22.2	25.5	22.7	30.1	16.9	28.2
V2	24.6	26.2	24.8	31.7	18.8	30.0
V3	23.4	25.1	23.6	31.2	17.1	29.4
A1	26.8	25.8	26.7	33.7	19.5	32.0
A2	26.6	26.7	26.6	34.0	20.4	32.3
A3	27.0	25.5	26.8	34.2	19.5	32.4
P1	24.2	23.8	24.2	32.4	16.9	30.5
P2	23.3	21.7	23.1	30.6	16.0	28.8
P3	21.4	19.0	21.1	29.0	13.2	27.0
Avg.	24.3	24.2	24.3	31.8	17.5	30.0

$\frac{SP(X)-SP(X')}{SP(X)} * 100$, where $SP(X)$ denotes the number of self transitions for raw input streams and $SP(X')$ denotes that for encoded data. The results under the columns XP and TP present percentage reduction in coupled switchings and total switchings, respectively. The total switchings account for both self transitions and coupled switchings with a capacitance ratio η . The CBI scheme yields better results compared to the BI method. This is because coupling reductions due to the CBI encoding are larger than that due to the BI encoding. It implies that the CBI scheme saves much more effective switchings by combatting coupled switchings while even sacrificing self transitions in a degree.

The CBI results show 31.8% average coupling reduction which is in accordance with our theoretical bound 29% in Example 3.1. The deviation comes from locality and correlations in input data streams, which are assumed to be none in theoretical analysis.

Table 2 presents the power consumed by both the codec circuitry and interconnects for the CBI scheme. Using $0.18\mu\text{m}$ technology, power is measured for random data streams using HSPICE with 1.6V power supply and the capacitance ratio $\eta = 4$ that is a realistic value from [13]. The results under column $P(I)$ correspond to power in μW consumed by interconnects and the results under column $P(E)$ in μW correspond to the power overhead due to encoder/decoder circuits. Without encoding circuits, all the power consumption contributes to interconnect power $P(I)$. Meanwhile, with suggested codec circuits, total power consumption under column $P(T)$ are the sum of interconnect power $P(I)$ and encoder/decoder power $P(E)$. It should be pointed out that within a realistic range of coupling capacitance, the power overhead $P(E)$ due to the codec are relatively small enough to be compensated by significant reduction in interconnect power $P(I)$. The percentage of power savings are shown in the column “% Red” by using the coupling-driven bus-invert encoding scheme.

5 Conclusions

Tightly coupled on-chip buses in a system-on-a-chip impose new requirements for interconnect power reduction and signal integrity. A novel bus encoding scheme was proposed for reducing power consumed by on-chip buses by decreasing coupled switchings. The coupling-driven bus-invert scheme reduces power consumption about 30% with one-cycle redundancy. Simulation results by using HSPICE indicate that the portion of power consumed by the encoder/decoder logic block for those schemes is fairly small in state-of-the-art technology. Therefore, the overhead due to encoder/decoder circuitry is compensated for by significant interconnect power savings.

Table 2: Power consumed by interconnects $P(I)$ in μW and codec circuitry $P(E)$ in μW under various capacitance values in pF.

C_x	C_s	Raw	CBI			% Red
		$P(I)$	$P(I)$	$P(E)$	$P(T)$	
1.0	0.25	111	66	20	86	22.5
3.0	0.75	319	191	19	210	34.2
5.0	1.25	526	315	21	336	36.1

References

- [1] L. Benini, A. Macii, E. Macii, M. Poncino, and R. Scarsi. Synthesis of low-overhead interfaces for power-efficient communication over wide buses. In *Proc. ACM/IEEE Design Automation Conf.*, pages 128–133, 1999.
- [2] F. Catthoor, F. Franssen, S. Wuytack, L. Nachtergaele, and H. D. Man. Global communication and memory optimizing transformations for low power signal processing systems. In *VLSI Signal Processing VII*, pages 178–187, 1994.
- [3] J. Cong. An Interconnect-centric design flow for nanometer technologies. In *Int. Symp. VLSI Technology, Systems, and Applications*, pages 54–57, June 1999.
- [4] W. Fornaciari, D. Sciuto, and C. Silvano. Power estimation for architectural exploration of HW/SW communication on system-level buses. In *Int. Workshop on Hardware/Software Codesign*, pages 152–156, 1999.
- [5] R. Hegde and N. R. Shanbhag. Energy-efficiency in presence of deep submicron noise. In *Proc. IEEE/ACM Int. Conf. Computer Aided Design*, pages 228–234, 1998.
- [6] S. M. Kang and Y. Leblebici. *CMOS Digital Integrated Circuits: Analysis and Design*. McGraw-Hill, 2nd edition, 1998.
- [7] L. Benini, G. De Micheli, E. Macii, D. Sciuto and C. Silvano. Asymptotic zero-transition activity encoding for address buses in low-power microprocessor-based systems. In *Proc. the Great Lakes Symp. VLSI*, pages 77–82, 1997.
- [8] S. Landman and R. L. Russo. On a pin versus block relationship for partitions of logic paths. *IEEE Trans. Computers*, C-20:1469–1479, 1971.
- [9] H. Mehta, R. M. Owens, and M. J. Irwin. Some issues in Gray code addressing. In *Proc. the Great Lakes Symp. VLSI*, pages 178–180, Mar. 1996.
- [10] E. Musoll, T. Lang, and J. Cortadella. Working-zone encoding for reducing the energy in microprocessor address buses. *IEEE Trans. on VLSI Systems*, 6(4):568–572, Dec. 1998.
- [11] P. R. Panda and N. D. Dutt. Reducing address bus transitions for low power memory mapping. In *Proc. European Design and Test Conf.*, pages 63–37, Mar. 1996.
- [12] S. Ramprasad, N. R. Shanbhag, and I. N. Hajj. Information-theoretic bounds on average signal transition activity. *IEEE Trans. on VLSI*, Sept. 1999.
- [13] Semiconductor Industry Association. International technology roadmap for semiconductors. http://notes.sematech.org/1999-SIA_Roadmap/Home.htm, 1999.
- [14] M. R. Stan and W. P. Bursleson. Bus-invert coding for low-power I/O. *IEEE Trans. on VLSI Systems*, pages 49–58, Mar. 1995.
- [15] M. R. Stan and W. P. Bursleson. Two-dimensional codes for low-power. In *International Symposium on Low-Power Electronics and Design*, pages 335–340, Aug. 1996.
- [16] C. L. Su, C. Y. Tsui, and A. M. Despain. Saving power in the control path of embedded processors. *IEEE Design and Test of Computers*, 11(4):24–30, 1994.
- [17] Y. Zhang, W. Ye, and M. J. Irwin. An alternative architecture for on-chip global interconnect: segmented bus power modeling. In *Asilomar Conf. on Signals, Systems, and Computers*, pages 1062–1065, 1998.