

Bus Encoding for Low-Power High-Performance Memory Systems

Naehyuck Chang^{*}
School of CSE, Seoul National
University, Korea
naehyuck@snu.ac.kr

Kwanho Kim
School of CSE, Seoul National
University, Korea
khkim@cslab.snu.ac.kr

Jinsung Cho[†]
Dept. of Comp. Engr., Seoul
National University, Korea
cjs@cslab.snu.ac.kr

ABSTRACT

High-performance memory buses consume large energy as they include termination networks, BiCMOS and/or open-drain output. This paper introduces power reduction techniques for memory systems deliberating on burst-mode transfers over the high-speed bus specifications such as Low Voltage BiCMOS (LVT), Gunning Transfer Logic (GTL+) and Stub Series Termination Logic (SSTL₂) which are widely used. The reduction techniques take both the static and the dynamic power consumption into account because most high-performance bus drivers and end-termination networks dissipate significant static power as well. Extensive performance analysis is conducted through mathematical analysis and trace data-driven simulations. We had reduction of 14% with random data and up to 67.5% with trace data.

1. INTRODUCTION

Processors, memory modules, and I/O controllers still require dense BL (board-level) or SL (system-level) buses in spite of the system-on-a-chip trend because of performance, scalability, resource optimization, time-to-market, marketing and development strategies of the semiconductor vendors, *etc.* It is well known that BL/SL bus drivers consume much more power by orders of magnitude than on-chip gates. This idea leads to various bus encoding techniques: reduction of the bus activity [1, 2, 3] based on the pure CMOS logic and reduction of low state [4] based on end-termination network driven by the pure CMOS logic. The BL/SL buses, unfortunately, introduce signal reflection, ground bounce, cross-talk, and heavy output load as the operating frequency often goes over 100MHz. Practical high-bandwidth BL/SL buses include the BiCMOS technology, termination networks, and/or open-drain/collector output drivers in common which involve power consumption by the bus activity and the duration of low state at the same time.

^{*}Corresponding author.

[†]Currently with Telecommunication R&D Center, Information & Communication Business, Samsung Electronics Co., Ltd., Korea, E-mail: chojs@telecom.samsung.co.kr.

Memory system bus is one of the most dominant BL/SL buses. It often requires high-bandwidth at the cutting edge thus consuming significant power. Microprocessors access DRAM arrays in burst-mode due to embedded cache memories. Memory access patterns thus are far different from the program data flow. Pull-up states are often inserted between the consecutive burst-mode transfers because the access time of the first burst data is about a few times larger than the successive ones. The address line remains in pull-up state during the rest of the clock cycles because only the start address of a burst sequence is driven. The assumption that address the sequence in a program flow is preserved in main memory access patterns [5, 1] is not valid in general for high-performance memory systems. Unless power reduction techniques are based on the accurate model, they can achieve much smaller amount of power reduction than they expect. Sometimes, power consumption may even increase. Nevertheless, existing bus encoding techniques are based on too simplified power consumption models and rule out actual access protocols between DRAMs and a microprocessor.

This paper introduces a power-cost function and reduction techniques that well reflect bus characteristics between a microprocessor and memory devices in view of access protocols as well as high-performance bus specifications. This paper excludes from discussion the bus encoding schemes that require significant additional cost such as limited-weight codes [2]. Address sequences as well as data sequences are regarded as random because of the cache memories and the passive pull-up. So, this paper does not consider correlation-based bus encoding schemes [6, 5, 1], either.

In this paper, empirical characterization achieved power consumption models of the high-performance bus specifications. We evaluated the performance of the existing and the new bus-encoding schemes in three different ways; mathematical analysis derived the expected values, simulation studies based on uniformly distributed random data verified the expected values, and the trace-data-driven simulation demonstrated actual power reduction. We built a prototype with Xilinx 4000XL series LCA to justify the feasibility.

2. PROBLEM STATEMENT

2.1 SDRAM burst-mode transfer

Figure 1 illustrates a burst-mode write transfer transaction. Address signals are driven by two steps in DRAM: asserting row address and its strobe, and then column address and its strobe. SDRAM operates in a similar way using command lines, but the operation is synchronized by the clock rather than the row and the column address strobes themselves. As shown in Figure 1, there has to be a

Permission to make digital/hardcopy 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, the copyright notice, the title of the publication and its date appear, and notice is given that copying is by permission of ACM, Inc. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
DAC 2000, Los Angeles, California
(c) 2000 ACM 1-58113-188-7/00/0006..\$5.00

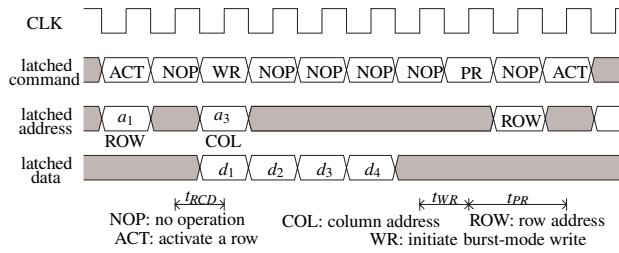


Figure 1: SDRAM burst-mode write transfer transaction (burst length = 4).

few idle clocks between row-active command and column-active command. In addition, there has to be an idle time more than $t_{RCD} + t_{WR} + t_{PR}$ between each burst-mode write transfer where t_{RCD} , t_{WR} , and t_{PR} are RAS to CAS delay, write recovery time, and precharge command period, respectively. There are thus shaded areas in the data and the address streams that represent undefined or high-impedance state. Termination resistors maintain passive pull-up state, or *bus hold* does the previous state, during the shaded period. There is no shaded period in general between the consecutive burst-mode data (b_i and b_{i+1}) due to *latched protocol of GTL/GTL+* [7] and *registered SDRAM* modules¹. Recent high-performance microprocessors, including the Pentium processor, manage cache memories by themselves, and data sequences are not altered by the SDRAM controller. However, additional delay is introduced by page and/or row hit/miss [8]. Address sequence, on the other hand, is altered because row and column addresses are multiplexed.

2.2 Bus-invert coding

The bus-invert coding has been introduced to reduce the bus activity: Hamming distance between the consecutive binary numbers. If the Hamming distance of the two consecutive binary numbers is more than half of the word length, the latter binary number is sent in inverted polarity [3] by asserting an additional signal line that indicates bus inversion. It can be used to reduce the weight (the number of ones or zeros) of the binary numbers if the bus-inversion decision is made when the weight is more than half of the bus width [4]. This paper formally describes the bus-invert coding schemes and the power-cost functions to make a better bus-inversion decision.

Definition 1. Bus-invert coding of \mathbf{B} is defined by $\Phi = (\mathbf{B}, \Xi, \mathbf{E})$ where

- $\mathbf{B} = (b_0, \dots, b_{l+1})$ is an input binary sequence where l is the burst length.
- $\Xi = (\xi_0, \dots, \xi_{l+1})$ is a set of Boolean variables representing bus inversion where

$$\xi_i = \begin{cases} 1, & \text{inversion of } b_i \text{ is disabled} \\ 0, & \text{inversion of } b_i \text{ is enabled} \end{cases}$$
- $\mathbf{E} = (e_0, \dots, e_{l+1})$ is a sequence of encoded binary numbers.

The inversion of a bit is denoted by $\bar{\xi}_i = |\xi_i - 1|$. XOR is denoted by $\xi_i \oplus \xi_j = |\xi_i - \xi_j|$.

¹KMM377S2858AT2 by Samsung Electronics.

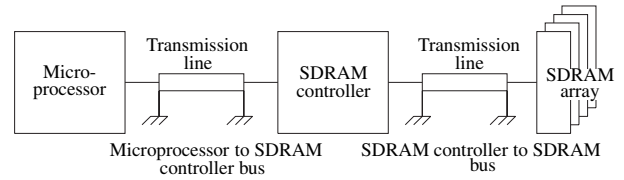


Figure 2: Bus model of a high-performance memory system.

Definition 2. The Hamming-distance and the weight vectors of \mathbf{B} , \mathbf{H} and \mathbf{W} , are defined by

- $\mathbf{H} = (h_1, \dots, h_{l+1})$ is a set of integers representing Hamming distance between b_{i-1} and b_i ($1 \leq i \leq l+1$)
- $\mathbf{W} = (w_0, \dots, w_{l+1})$ is a set of integers representing the number of zeros of b_i ($0 \leq i \leq l+1$).

Burst-mode data sequence is denoted by $\mathbf{D} = (1, d_1, \dots, d_4, 1)^2$ as shown in Figure 1. Bus-invert coding of \mathbf{D} is defined by letting $\mathbf{B} = \mathbf{D}$. The first and the last 1s are the default bus state (usually pull-up). Thus $w_0 = 0$ and $w_5 = 0$. Burst-mode address sequence between a microprocessor and an SDRAM controller is denoted by $\mathbf{C} = (1, a_1, 1, 1, 1, 1)$ because the microprocessor asserts the start address of the burst sequence. Its bus-invert coding is defined by letting $\mathbf{B} = \mathbf{C}$. Thus $h_1 = h_2 = w_1$, and $h_3 = h_4 = h_5 = w_0 = w_2 = w_3 = w_4 = w_5 = 0$. Burst-mode address sequence between the microprocessor and the SDRAM controller is denoted by $\mathbf{S} = (1, a_1, 1, a_3, 1, 1)$ as shown in Figure 1. Its bus-invert coding is defined by letting $\mathbf{B} = \mathbf{S}$. So $h_1 = h_2 = w_1$, $h_3 = h_4 = w_3$, and $h_5 = w_0 = w_3 = w_4 = w_5 = 0$.

2.3 Power-cost function

The power cost function of the original burst-mode sequence is

$$f_c(\mathbf{B}) = c_w \sum_{i=0}^5 w_i + c_h \sum_{i=1}^5 h_i + 9c_s. \quad (1)$$

We assume that $t_{RCD} + t_{WR} + t_{PR} = 5$, without loss of generality, throughout the paper. The power cost function of the encoded burst-mode sequence is given by

$$\begin{aligned} f_c(\mathbf{E}) &= c_w \sum_{i=0}^5 (\xi_i w_i + \bar{\xi}_i (n+1 - w_i)) \\ &+ c_h \sum_{i=1}^5 ((\xi_i \oplus \xi_{i-1}) h_i + (\xi_i \oplus \xi_{i-1}) (n+1 - h_i)) + 9c_s, \end{aligned} \quad (2)$$

where c_w , c_h and c_s are power-cost coefficients associated with the weight cost, the Hamming-distance cost and the common-mode static cost, respectively. The coefficients c_w , c_h and c_s are bus specification-dependent. Section 3 gives a full detail of them. An objective of this paper is to minimize the cost function.

3. POWER CONSUMPTION MODEL

Memory subsystems of common microprocessor-based systems can be illustrated as Figure 2. GTL/GTL+ is designed for high-speed,

²Burst length can be determined to 4, 8, or 16 according to the cache line size.

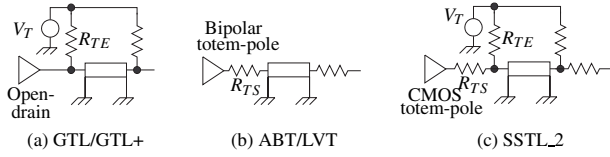


Figure 3: Output and termination structure of high-performance bus specifications.

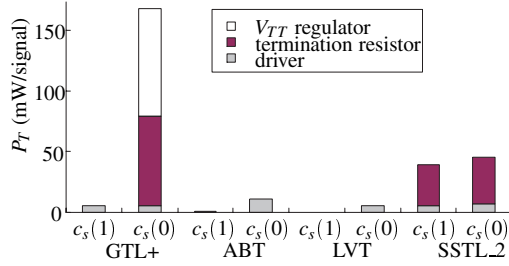


Figure 4: Static power consumption of the GTLP16616, the 74ABT16245, the 74LVT16245 and the SSTL16837 ($c_s(1)$: output is high, $c_s(0)$: output is low. $V_{TTI} = V_{CC}$ for GTL+, $V_{TTI} = V_{TT}$ for SSTL_2).

low power signal transmission by reducing voltage swing [7]. It has been used in Intel's Pentium processors over years and is famous for the counter-connection between a microprocessor and an SDRAM controller. The GTL/GTL+ output structure is open-drain, and each end is terminated to V_{TT} through the termination resistor, $50\Omega R_{TE}$. The R_{TE} draws static current when the logic state is low. A V_{TT} voltage regulator supplies the load current. The regulators are mostly *linear regulators*³. So we have to count in the power consumption in the regulator. The transmission line does not change the power consumption because GTL/GTL+ performs incident wave switching and all the stored charges in the transmission line are transferred to the other.

There is some degree of freedom in choosing buses between *an SDRAM controller and SDRAMs* (Figure 2) according to the bus frequency and the memory capacity. LVCMOS without termination is justified only with a small number of DRAMs, short line length, and slow slew rate. In contrast, as the number of DRAMs increases, the line length becomes inevitably long and the bus drivers should be powerful. Higher bandwidth demands sharper slew rate and thus the ABT (Advanced BiCMOS Technology) or the LVT drivers and series termination resistors (Figure 3 (b)). The BiCMOS drivers draw unsymmetrical static current, and source termination resistors increase dynamic power consumption proportional to the round-trip delay of the transmission line because it allows only reflected wave switching. SSTL_2 is intended to improve operations where buses must be isolated from relatively large stubs [9]. Its termination structure has both a source termination resistor, R_{TS} (25Ω), and end termination resistors, R_{TE} (50Ω at both ends)⁴. High-performance systems (266MHz equivalent bus clock) adopt DDR (Double Data Rate) SDRAMs⁵ to enhance the transfer rate; SSTL_2 is commonly used for DDR SDRAM interface. Unlike

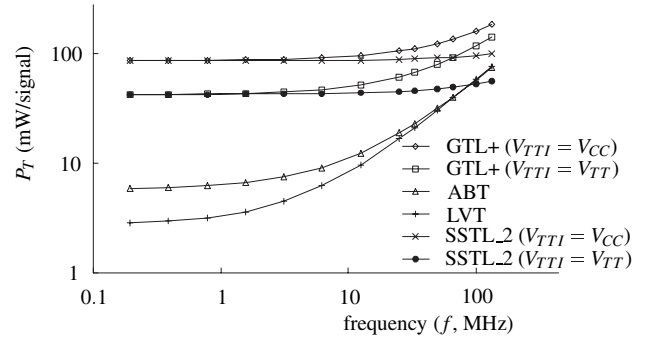


Figure 5: Dynamic power consumption of the GTLP16616, the 74ABT16245, the 74LVT16245 and the SSTL16837 over frequency ($d = 0.5$).

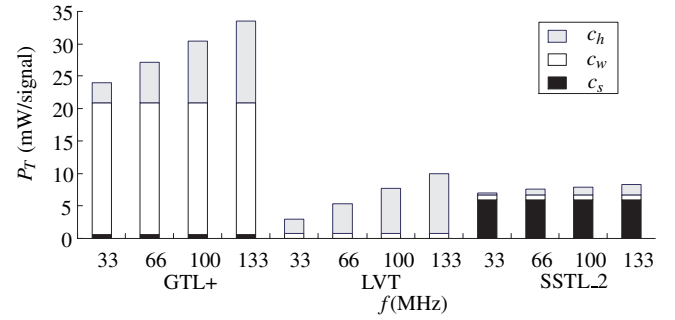


Figure 6: Power consumption coefficients over frequency ($V_{TTI} = V_{TT}$).

the open-drain output, its totem-pole output allows the termination resistors ($R_{TE} + R_{TS}$) to draw current from the output to V_{TT} regulator even when the output is high. The V_{TT} regulator for SSTL_2, therefore, must be able to source as well as sink current⁶ while unidirectional linear regulators are widely used as the V_{TT} regulators for GTL/GTL+; there is significant static current flow when the output is high.

Figure 4 shows the static power consumption by the bus specifications. The bus drivers mostly draw the maximum static power supply current when the output is low. We denote the maximum and the minimum static power as $c_s(0)$ and $c_s(1)$, respectively. The static power consumption is a linear function as given by

$$\frac{c_s(1) + (1-d)(c_s(0) - c_s(1))}{l + p} \quad (3)$$

where p is $t_{RCD} + t_{WR} + t_{PR} = 9$, and d is the duty ratio of the signal. We can denote the power-cost coefficients as follows:

$$c_s = \frac{c_s(1)}{9}, \quad c_w = \frac{c_s(0) - c_s(1)}{9}, \quad c_h = \frac{C_{PD}f}{9} \quad (4)$$

where C_{PD} is power dissipation capacitance, the slope of the dynamic power consumption, and f is the bus clock frequency. Figure 5 illustrates that the dynamic power consumption is proportional to the switching frequency.

GTL+ consumes high static power; it has very large c_w while c_s is very small. It also has comparable c_h as the frequency increases.

⁶ML6553 by Micro Linear.

³LM3460-1.2, -1.5 by National Semiconductor.

⁴Class II

⁵KMM368L3313C by Samsung Electronics.

LVT and ABT have virtually zero c_s and very small c_w while they have significant c_h . SSTL2 draws significant static current as does in GTL/GTL+. However, SSTL2 has large c_s instead of c_w . In addition, it has relatively small c_h .

These imply that it would be important to reduce dynamic power in LVT and ABT while static power in GTL+. However, the bus-invert coding would not be effective in SSTL2 because c_s is large but c_h and c_w are small. More detailed and systematic analysis is described in the following sections.

4. BUS-INVERSION DECISION SCHEMES

4.1 Data bus encoding

4.1.1 Exiting bus-invert coding schemes

When no bus-invert coding is applied to \mathbf{D} , $E[w_i] = E[h_i] = \frac{n}{2}$ because w_i and h_i have independent identical probability distributions for random binary sequences. Thus the expected value of the power cost function is given by

$$E[f_c(\mathbf{E}|\mathbf{B} = \mathbf{D})] = 2nc_w + \frac{5}{2}nc_h + 9c_s. \quad (5)$$

where n is the bus width and is usually an even number⁷. When the bus-inversion decision is made by the Hamming distance, the expected value of the Hamming distance, $E[h_i]$, is reduced to $E[(h_i, n+1-h_i)]$ while the expected value of the weight, $E[w_i]$ is not altered. In other words, $E[w_i]$ is independent on the bus-invert coding that reduces the expected value of the Hamming distance, $E[h_i]$. We derived the expected value taking the inversion bit into account:

$$E[\min(h_i, n+1-h_i)] = \frac{n+1}{2} \left(1 - P\left[X = \frac{n}{2}\right]\right) \quad (6)$$

where X is the binomial random variable, $B\left(n, \frac{1}{2}\right)$, and $P[X = r] = nC_r \left(\frac{1}{2}\right)^n$. Traditional bus-inversion decision reduces the Hamming distance between the consecutive binary numbers [3]. This paper names it *Hamming-distance-based decision*. It has been introduced assuming that $c_w \approx 0$. It is thus only applicable to high-speed CMOS with bus hold function which is suitable to low performance memory systems. When we apply it to the GTL/GTL+, the ABT or the LVT drivers, it ignores the effect of the passive pull-up states and the static power consumption. Thus the expected power consumption is given by

$$E[f_c(\mathbf{E}|\mathbf{B} = \mathbf{D})] = 2c_w(n+1) + \left(\frac{5}{2} - \frac{3}{2}P\left[X = \frac{n}{2}\right]\right)c_h(n+1) + 9c_s. \quad (7)$$

The bus-invert coding can go by the weight of the binary numbers [4]. We call it in this paper *weight-based decision*. It reduces the expected value of the weights as follows:

$$E[\min(w_i, n+1-w_i)] = \frac{n+1}{2} \left(1 - P\left[X = \frac{n}{2}\right]\right) \quad (8)$$

Unlike the Hamming-distance-based decision, it also reduces the Hamming distance at the same time. First, it reduces h_1 and h_5 because $w_0 = w_5 = 0$ and thus $w_1 = h_1$ and $w_4 = h_5$.

⁷We assume throughout the paper that n is an even number.

Lemma 1. *The expected value of the Hamming distance of $(n+1)$ -bit, $\frac{n}{2}$ -limited weight code is given by*

$$E\left[h_i | w \leq \frac{n}{2}\right] = \sum_{j=0}^{\frac{n}{2}} \left(B_j \left(\sum_{k=0}^{\frac{n}{2}} B_k E_{jk} \right) \right)$$

where $B_r = P[X = r] + P[X = n-r]$ and

$$E_{jk} = \frac{1}{nC_k} \sum_{l=0}^{\min(j,k)} jC_l (n-j)C_{k-l} (j+k-2l).$$

PROOF. We left the proof out due to the limited space. \square

In addition, the $E[h_i]$ s for $i = 2, 3, 4$ are reduced as well because the binary numbers of the entire sequence become $(n+1)$ -bit, $\frac{n}{2}$ -limited weight codes after the bus-inversion according to Lemma 1. The expected power consumption is given by

$$E[f_c(\mathbf{E}|\mathbf{B} = \mathbf{D})] = 2c_w(n+1) \left(1 - P\left[X = \frac{n}{2}\right]\right) + 3c_h E\left[h_i | w \leq \frac{n}{2}\right] + c_h(n+1) \left(1 - P\left[X = \frac{n}{2}\right]\right) + 9c_s. \quad (9)$$

$E[h_i | w \leq \frac{n}{2}] = 4.164$ when $n = 9$ while $E[h_i] = 4.5$.

4.1.2 New bus-inversion decision strategy

This paper first takes the passive pull-up state into consideration. If the Hamming-distance-based decision considers the passive pull-up state, the expected power consumption is enhanced such that

$$E[f_c(\mathbf{E}|\mathbf{B} = \mathbf{D})] = \left(2 - \frac{1}{2}P\left[X = \frac{n}{2}\right]\right)c_w(n+1) + \left(\frac{5}{2} - 2P\left[X = \frac{n}{2}\right]\right)c_h(n+1) + 9c_s. \quad (10)$$

The Hamming-distance-based and the weight-based decision take into consideration only the current and the previous values, so they may mislead the bus coding strategy to the local minimum because the sequence always ends with the pull-up state. This paper introduces a *look-ahead* scheme that exploits all the burst-mode data. The controller makes the decision when it fills up the cache line (read) so that there is no performance penalty. The look-ahead bus-inversion decision first determines ξ_1 and ξ_4 by the weight. Secondly, it determines ξ_2 and ξ_3 by Table 1. The first step thus minimizes both the weight and the Hamming distance. This modifies h_2 and h_4 to h'_2 and h'_4 , respectively:

$$h'_2 = \xi_1 h_2 + \overline{\xi_1}(n+1-h_2) \quad (11)$$

$$h'_4 = \xi_4 h_4 + \overline{\xi_4}(n+1-h_4). \quad (12)$$

The symbols, \circ and \times in Table 1, represent the status of the condition flags such that

$$w_i, h_i^c, h_i'^c = \begin{cases} \circ, & w_i, h_i, h_i' < \frac{n+1}{2} \\ \times, & \text{otherwise} \end{cases} \quad (13)$$

Table 1 does not guarantee to find the optimal Ξ . Rather, it reduces the number of b_i and h_i marked with \times . The approximated expected value of the power cost function is given by

$$E[f_c(\mathbf{E}|\mathbf{B} = \mathbf{D})] \approx \frac{n+1}{32} \left(77c_w + 70c_h - (51c_w + 61c_h)P\left[X = \frac{n}{2}\right] \right) + 9c_s. \quad (14)$$

Table 1: ξ_2 and ξ_3 table.

before					Ξ		after				
w_2^c	w_3^c	h_2^c	h_3^c	h_4^c	ξ_2	ξ_3	w_2^c	w_3^c	h_2^c	h_3^c	h_4^c
0	0	0	0	0	1	1	0	0	0	0	0
0	0	0	0	x	1	1	0	0	0	0	x
0	0	0	x	0	1	1	0	0	0	x	0
0	0	0	x	x	1	0	0	x	0	0	0
0	0	x	0	0	1	1	0	0	0	x	0
0	0	x	0	x	1	1	0	0	x	0	x
0	0	x	x	0	0	1	x	0	0	0	0
0	0	x	x	x	1	0	0	x	x	0	0
0	x	0	0	0	1	1	0	x	0	0	0
0	x	0	0	x	1	0	0	0	0	x	0
0	x	0	x	0	1	0	0	0	0	0	x
0	x	0	x	x	1	0	0	0	0	0	0
0	x	x	0	0	1	1	0	x	0	0	0
0	x	x	x	0	1	0	x	0	0	0	x
0	x	x	x	x	1	0	x	0	0	0	0
x	0	0	0	0	1	1	x	0	0	0	0
x	0	0	0	x	1	1	x	0	0	0	x
x	0	0	x	0	0	1	0	0	0	x	0
x	0	0	x	x	0	1	0	0	0	0	0
x	0	x	x	0	0	1	0	0	0	0	x
x	0	x	x	x	0	0	0	0	x	0	x
x	x	0	0	0	0	0	0	0	x	0	0
x	x	0	0	x	0	0	0	0	x	0	0
x	x	0	x	0	0	1	0	0	x	0	0
x	x	x	0	0	0	0	0	0	0	0	x
x	x	x	0	x	0	0	0	0	0	0	0
x	x	x	x	0	0	0	0	x	0	0	0
x	x	x	x	x	0	0	0	0	0	x	0

4.2 Address bus coding

When no bus-invert coding is applied, the expected values of the power cost functions are as follows:

$$E[f_c(\mathbf{C})] = \frac{1}{2}c_w(n+1) + c_h(n+1) + 9c_s. \quad (15)$$

$$E[f_c(\mathbf{S})] = c_w(n+1) + 2c_h(n+1) + 9c_s. \quad (16)$$

Simply the weight-based decision introduces the optimal bus encoding in the burst-mode address sequence:

$$E[f_c(\mathbf{E}|\mathbf{B}=\mathbf{C})] = \frac{n+1}{2}(c_w + 2c_h) \left(1 - P\left[X = \frac{n}{2}\right]\right) + 9c_s \quad (17)$$

$$E[f_c(\mathbf{E}|\mathbf{B}=\mathbf{S})] = (n+1)(c_w + 2c_h) \left(1 - P\left[X = \frac{n}{2}\right]\right) + 9c_s \quad (18)$$

5. IMPLEMENTATION

Figure 7 shows the lookahead data bus encoder for the burst-mode data sequences. We use two *threshold elements*; ($\bar{c} = 0$ if $w_i + \xi_i > \frac{n+1}{2}$). We implemented the encoder using Xilinx FPGA XC4010XL. The pipelined architecture is suitable to flip-flop abundant architecture of the Xilinx XC4000 series FPGA [10]. Data path libraries are composed by LogiBLOX. The lookahead encoder uses 34 CLBs and 39 flip flops; its cost is trivial in view of the complexity of modern digital systems.

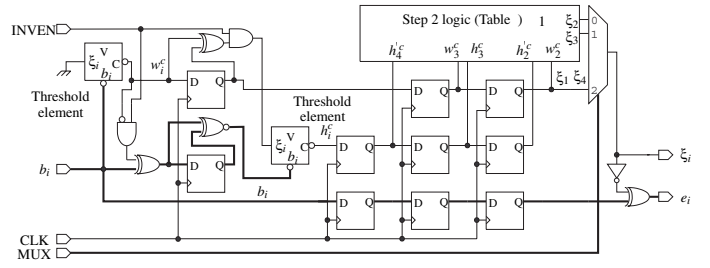


Figure 7: Lookahead bus encoder for burst-mode data sequences.

Table 2: Analytic expected values of power consumption (mW/signal) (no: no encoding, Hm: Hamming-distance-based, W: weight-based, and LA: lookahead).

bus	f (MHz)	no	Hm	W	LA
GTL+	33	49.0	53.6	41.2	43.2
	66	56.8	60.9	48.6	50.3
	100	64.8	68.4	56.2	57.6
	133	72.5	75.7	63.6	64.6
LVT	33	7.2	7.0	6.6	6.5
	66	12.9	12.4	12.1	11.7
	100	18.9	18.0	17.8	17.2
	133	24.7	23.4	23.3	22.4
SSTL_2	33	12.4	12.5	12.1	12.2
	66	13.7	13.7	13.3	13.3
	100	14.9	14.9	14.5	14.4
	133	15.7	15.6	15.2	15.1

6. PERFORMANCE EVALUATION

Performance analysis has been conducted assuming that the bus encoding schemes are applied byte-wise: four encoders for a 32bit bus. Because the simple weight-based decision scheme leads to the optimal bus-invert coding for the burst-mode address sequence, the performance evaluation is performed only to the data bus encoders. Tables 2 and 3 show the analytic expected values and the simulation result with uniformly distributed random data, which are quite close with each other. They show that the Hamming-distance-based decision offers very small amount of power reduction even when $c_w \ll c_h$ because it does not count on the pull-up state between the consecutive burst-mode transfers, and thus misleads the inversion decision of the first data in the sequence. This is more dominant in GTL+ of which power consumption even increases due to the cost of the inversion bit. It justifies that the bus encoding schemes should pay attention to the actual bus-transaction protocols rather than the program flow.

Weight-based decision shows quite a good performance because it also reduces the Hamming distance. Its performance is comparable to others, which are more complex than the weight-based decision, especially for GTL+ in which $c_w > c_h$. In contrast, the performance degradation becomes distinct as the operating frequency increases for LVT or ABT in which $c_w \ll c_h$. The weight-based decision, however, is still a good cost-effective method even when there is only small portion of static power consumption in the case of the SDRAM burst-mode transfer.

The lookahead decision scheme suggested in this paper also shows excellent performance. The lookahead decision scheme does not require tuning by the operating frequency and the types of buses

Table 3: Expected power consumption by random data simulation (mW/signal) (no: no encoding, Hm: Hamming-distance-based, W: weight-based, LA: lookahead, and Opt: optimal).

program	bus	f(MHz)	no	Hm	W	LA	Opt
random	GTL+	33	49.2	53.6	41.1	41.1	41.1
		66	57.0	60.9	48.5	48.3	48.1
		100	65.0	68.4	56.1	55.6	55.3
		133	72.7	75.7	63.5	63.2	62.0
	LVT	33	7.2	7.0	6.6	6.4	6.1
		66	12.9	12.4	12.1	11.5	10.9
		100	18.9	18.0	17.8	16.8	16.0
		133	24.7	23.4	23.2	21.8	20.8
	SSTL ₂	33	12.4	12.5	12.1	12.1	12.1
		66	13.7	13.7	13.3	13.2	13.2
		100	14.9	14.9	14.5	14.3	14.2
		133	15.7	15.6	15.2	15.0	14.9
CRC	GTL+	33	71.1	52.6	23.6	23.6	23.6
		66	78.1	57.9	27.1	27.1	27.0
		100	85.3	63.3	30.8	30.6	30.5
		133	92.3	68.6	34.3	34.2	33.9
	LVT	33	7.4	5.5	3.3	3.2	3.2
		66	12.6	9.4	5.9	6.0	5.7
		100	18.0	13.5	8.7	8.7	8.3
		133	23.2	17.4	11.3	11.2	10.7
	SSTL ₂	33	13.1	12.3	11.1	11.1	11.1
		66	14.2	13.2	11.7	11.7	11.7
		100	15.4	14.0	12.3	12.3	12.3
		133	16.1	14.5	12.6	12.6	12.5
DCT	GTL+	33	66.4	50.5	26.1	26.1	26.1
		66	73.1	55.8	30.0	30.0	29.9
		100	80.1	61.2	34.2	34.0	33.9
		133	86.7	66.4	38.1	38.1	37.6
	LVT	33	7.0	5.4	3.7	3.6	3.5
		66	12.0	9.3	6.6	6.6	6.3
		100	17.2	13.4	9.7	9.6	9.2
		133	22.1	17.3	12.7	12.5	11.9
	SSTL ₂	33	12.9	12.2	11.3	11.3	11.3
		66	14.0	13.1	11.9	11.9	11.9
		100	15.1	13.9	12.6	12.5	12.5
		133	15.7	14.4	12.9	12.9	12.8

while it reduces the weight and the Hamming-distance cost at the same time; it is robust to the ratio of c_w and c_h . It makes possible to compose a unified bus encoder/decoder when different types of buses form a hierarchy (e. g., GTL+ and SSTL₂), which helps to reduce the cost.

Table 3 illustrates the actual performance of each bus-inversion decision scheme. They show similar ranking of performance enhancement among each other while the quantity of the power reduction is much more optimistic (up to 67.5% reduction) than the analytic and the random-data simulation results because the trace data is correlated with each other and there are more zeros than ones on the bus before bus encoding. The power reduction ratio in SSTL₂ is much smaller than that of others because c_s , the common-mode static power, is much larger than c_w and c_h .

7. CONCLUSIONS

This paper exploits bus encoding techniques for low power, high-performance memory systems. We discovered that the static power is significant in high-performance buses. We also demonstrated that the data sequence can be altered by the bus specification.

We introduced new bus-inversion decision schemes based on accurate power consumption models and access protocols which are directly applicable to high performance memory systems such as the Pentium processor-based portable computers. Performance analy-

sis including mathematical analysis and simulation studies showed that the lookahead encoder makes virtually optimal bus-invert decision in burst-mode transfer regardless of the environmental variation (clock frequency, driver type, and termination structure, i. e., the ratio of static and dynamic power consumption).

8. ADDITIONAL AUTHORS

Heonshik Shin (School of CSE, Seoul National University, Korea. E-mail: shinhs@comp.snu.ac.kr).

9. REFERENCES

- [1] Luca Benini, Giovanni De Micheli, Enrico Macii, Massimo Poncino, and Stefano Quer, "System-level power optimization of special purpose applications: The beach solution," in *Low power electronics and design*, 1997, pp. 24 – 29.
- [2] Mircea R. Stan and Wayne P. Burleson, "Limited-weight codes for low-power I/O," in *Proceedings of Int. Workshop on Low Power Design*, Napa CA, USA, Apr. 1994, pp. 209–214.
- [3] Mircea R. Stan and Wayne P. Burleson, "Bus-invert coding for low power I/O," *IEEE Transactions on VLSI*, pp. 49–58, Mar. 1995.
- [4] Mircea R. Stan and Wayne P. Burleson, "Coding a terminated bus for low power," in *Proceedings of Great Lakes Symposium on VLSI*, Buffalo, NY, USA, Mar. 1995, pp. 70–73.
- [5] Huzefa Mehta, Robert Michael Owens, and Mary Jane Irwin, "Some issues in gray code addressing," in *Proceedings of Great Lakes Symposium on VLSI*, Ames, IA, USA, Mar. 1996, pp. 178–181.
- [6] L. Benini, G. De Micheli, E. Macii, D. Sciuto, and C. Silvano, "Asymptotic zero-transition activity encoding for address busses in low-power microprocessor-based systems," in *GLS-VLSI-97: IEEE 7th Great Lakes Symposium on VLSI*, Mar. 1997, pp. 77–82.
- [7] "Gunning transceiver logic (GTL) low-level, high-speed interface standard for digital integrated circuits," *JEDEC standard*, <http://www.jedec.org/download/freestd/jesd8-xx/JESD8-3.PDF>, 1993.
- [8] "Intel 440LX AGPset design guide," <http://www.intel.com/design/chipsets/designex/297651.htm>, Intel, 1998.
- [9] "Stub series terminated logic for 2.5 volts (SSTL₂)," *EIA/JEDEC standard*, <http://www.jedec.org/download/freestd/jesd8-xx/JESD8-9.PDF>, 1998.
- [10] "XC4000E and XC4000X series field programmable gate arrays," *Xilinx data book*, <http://www.xilinx.com/partinfo/4000.pdf>, Xilinx, 1999.