Adaptive Low-Power Bus Encoding Based on Weighted Code Mapping

Avnish R. Brahmbhatt, Jingyi Zhang, Qinru Qiu, Qing Wu
Department of Electrical and Computer Engineering
Binghamton University, State University of New York
Binghamton, New York – 13902, U.S.A
{avnish, jzhang5, qqiu, qwu}@binghamton.edu

***Abstract*— In this paper, we propose an adaptive low-power bus encoding algorithm based on weighted code mapping (WCM). The WCM algorithm transforms an original bus data vector to a low-energy code through one-to-one mapping. The code mapping is determined by the data probabilistic distribution in the original sequence. The WCM algorithm considers both the self and coupling capacitance of the bus wires. A window-based adaptive encoding algorithm is proposed to improve the energy saving by adaptively changing the code mapping for different data probabilistic characteristics. Experimental results show that the proposed algorithm outperforms the existing coding algorithms by either significantly lower computation/hardware complexity or higher energy savings.**

I. Introduction

As technology scales down to deep submicron meter, interconnects between on-chip bus lines or off-chip bus lines have increasingly significant impact on the power consumption and performance of circuits. The coupling capacitance in deep submicron technology is comparable to or even exceeds the line-to-ground capacitance of the wire itself, which makes the crosstalk energy the dominant component in bus energy dissipation [1]-[3].

Extensive research works began to focus on the optimization of bus energy dissipated on the coupling capacitances. One popular technique is based on the idea of encoding data before transmission and decoding the code words at the receiver's side. Most of the proposed algorithms need extra bit lines to generate a larger code word space to enable those types of transition with relatively smaller energy dissipation.

The existing bus encoding methods can be divided into two categories: heuristic and probabilistic-based. Among the heuristic algorithms, authors of [2] presented a "forbidden pattern" coding scheme, which reduces crosstalk by eliminating the certain high-energy patterns in the transmitted data. Reference [4] introduces a "forbidden transition" coding scheme for crosstalk reduction, in which no adjacent bits can switch in opposite directions. An even-odd bus inverting algorithm is presented in [3], which divides the data vector into even bits and odd bits. Both two groups of data bits can be inverted independently so as to reduce the crosstalk interactions. The advantage of the heuristic algorithms is that they are simple for implementation, with low hardware and computation overheads. The shortcoming is that they do not consider the probabilistic distribution of the transmitting data, leading to limited energy savings.

In the category of probabilistic-based bus encoding methods, the transition pattern coding scheme (TPC) [10], works effectively on general bus model with both the coupling capacitance and the line-to-ground capacitance. The encoder is essentially an FSM with m-bit input, n-bit output and 2^n states, where m and n are the width of the binary data before and after encoding. Given the input data and the current state of the FSM, the next state, which is also the encoded data to be transmitted, is uniquely defined so as to minimize the average energy dissipation on the bus. The optimal coding scheme is obtained by searching for the optimal mapping from the current state and the input to the next state. The approximately optimal coding (AOC) scheme is introduced as a reduced version of TPC to reduce the runtime of searching. The TPC algorithm provides the optimal bus encoding, given the probabilistic distribution of the data. The shortcomings are that its effectiveness relies on the accurate information about the data sequence, and the hardware and computation overheads are high and increase exponentially with the increase of bus width.

In this paper, we propose a novel low-energy bus coding scheme. The proposed weighted code mapping (WCM) algorithm generates a one-to-one mapping from the original data to the coded data, which is decided by the probabilistic distribution of the original data. We also propose an adaptive algorithm that dynamically adjusts the encoding based on the change of the data probability distribution. The complexity of searching for the code and the hardware/computation overheads are significantly reduced comparing to TPC or AOC. Comparing to the heuristic algorithms, the proposed WCM algorithm provides more energy reduction by both eliminating more crosstalk transitions and considering the stochastic distribution of the input data.

The remaining of the paper is organized as follows. Section II introduces the background on DSM bus model. Our coding algorithm is presented in section III. In section IV experimental results are presented and analyzed in detail. Finally we draw conclusions in section V.

II. BACKGROUND

In general, a bus may consist of one set of parallel lines or more with repeaters between them. The bus lines are usually laid parallel and the capacitive coupling between nonadjacent lines is weak compared to that between adjacent lines. The model of the capacitive parasitic of the on-chip bus lines can be simplified to the circuit shown in Fig.1, where C_i is the total inter-line capacitance, while C_L is the line-to-ground capacitance.

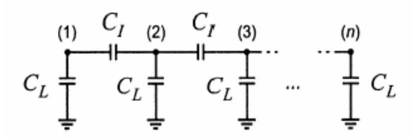

Figure 1 The DSM bus capacitive model.

The conductance matrix of the network can be written as

$$\overline{C} = \begin{bmatrix} 1+\lambda & -\lambda & 0 & \dots & 0 & 0 \\ -\lambda & 1+2\lambda & -\lambda & \cdots & 0 & 0 \\ 0 & -\lambda & 1+2\lambda & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdot & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1+2\lambda & -\lambda \\ 0 & 0 & 0 & \cdots & -\lambda & 1+2\lambda \end{bmatrix} \cdot C_L$$

where λ is the *capacitance factor* which is calculated as $\lambda = C_I / C_L$. Let $E(V^{old}, V^{new})$ be the total energy drawn from the power supply during the transition from data vector V^{old} to V^{new}. $E(V^{old}, V^{new})$ can be calculated by Eqn. (1) [10]

$$E(V^{old}, V^{new}) = (V^{new})^T \overline{C} (V^{new} - V^{old}) \qquad (1)$$

Let $w_1, w_2, \ldots, w_{2^n}$ be the binary vectors transmitted on a *n*-bit bus. An *energy cost matrix* can be defined as

$$\mathbf{E} = \left[E(w_i, w_j)\right]_{i,j=1}^{M}.$$

For example, a 2^3-by-2^3 energy cost matrix with λ=3 can be written as

$$E = \begin{bmatrix} 0 & 4 & 7 & 5 & 4 & 8 & 5 & 3 \\ 0 & 0 & 10 & 4 & 4 & 4 & 8 & 2 \\ 0 & 7 & 0 & 1 & 7 & 14 & 1 & 2 \\ 0 & 3 & 3 & 0 & 7 & 10 & 4 & 1 \\ 0 & 4 & 10 & 8 & 0 & 4 & 4 & 2 \\ 0 & 0 & 13 & 7 & 0 & 0 & 7 & 1 \\ 0 & 7 & 3 & 4 & 3 & 10 & 0 & 1 \\ 0 & 3 & 6 & 3 & 3 & 6 & 3 & 0 \end{bmatrix} \cdot C_L \qquad (2)$$

III. ADAPTIVE ALGORITHM WEIGHTED CODE MAPPING

In this section, we will first introduce the bus encoding algorithm based on weighted code mapping, followed by the window-based adaptive encoding method.

A. Bus encoding using weighted code mapping

By carefully studying the energy cost matrix we found that some column consistently has lower total summation than the others. For example, in the energy cost matrix given by Eqn. (2), the first column or the last column has much lower total summations than the third or the sixth column. It indicates that transmitting certain codeword always costs lower bus energy than transmitting others no matter what is the current vector on the bus. Let *Eng*(*i*) denote the summation of all the entries in the *i*th column of the energy cost matrix plus the summation of all the entries in the *i*th row of the energy cost matrix. The following theorem gives an interesting property of *Eng*(*i*).

Theorem 1: If each of the 2^n binary vector has the same probability to be transmitted in the (*i*-1)th and (*i*+1)th clock cycle, transmitting the vector with the smallest $Eng(i) = \sum_{j=0}^{2^n-1} E(w_j, w_i) + \sum_{j=0}^{2^n-1} E(w_i, w_j)$ in the *i*th clock cycle minimizes the transition energy.

The proof of theorem 1 is straightforward. The average bus energy in the *i*th and (*i*+1)th cycle can be calculated as $E = \sum_{j=0}^{2^n-1} \frac{1}{2^n} E(w_j, w_i) + \sum_{j=0}^{2^n-1} \frac{1}{2^n} E(w_i, w_j) = 2Eng(i)$.
Therefore minimizing *Eng*(*i*) will hence minimize *E*.

Weighted Code Mapping:
1. for i = 1 to 2^n
2. Calculate Eng (i);
3. end
4. Sort the codewords based Eng(i);
5. Select the 2^m codewords with the smallest Eng(i);
6. Sort the input data based on their probability;
7. Map the highest probability data to the code with the smallest Eng(i);

Figure 2 The weighted code mapping Algorithm.

Based on the above observation, we present a heuristic coding algorithm called the Weighted Code Mapping (WCM) algorithm. Let *m* and *n* denote the width of the binary vectors before and after encoding and $n = m + a$ where *a* is the number of redundant lines. Similar as the TPC algorithm, we assume that the distribution of the input data vector is known. How this information can be dynamically obtained will be discussed in the later of this section. The WCM algorithm finds a set of 2^m codewords out from 2^n binary vectors such that the expected steady state bus energy is minimal. At each clock cycle, the encoder maps the input data vector to one of these 2^m codes based on its probability. The pseudo code of the WCM algorithm is given in Figure 2.

Since the coding is a one-to-one mapping procedure, both the encoder and the decoder can be implemented using combinational logic circuit. The hardware complexity of encoder-decoder is much lower than that of TPC, which needs to be implemented using sequential circuit with 2^n states. The complexity of WCM is $O(2^{2n})$ and it is dominated by the complexity of the *for-loop* which calculates the $Eng(i)$ for each column and row of the 2^n-by-2^n matrix. Although this time complexity is significantly lower than that of TPC, as the value of n increases, it still becomes intractable. An improved WCM algorithm is presented next which finds the set of codeword with the minimal $Eng(i)$ based on pattern generation. The runtime complexity is thus reduced to $O(2^m)$.

Let $E_{self}(V^{old}, V^{new})$ and $E_{couple}(V^{old}, V^{new})$ be the bus energy that is due to the line-to-ground capacitance and the coupling capacitance respectively. The total energy drawn from the voltage supply can be calculated as

$$E(V^{old}, V^{new}) = E_{self}(V^{old}, V^{new}) + E_{couple}(V^{old}, V^{new}).$$

The following lemma shows that the $Eng(i)$ is determined only by $E_{cuople}(V^{old}, V^{new})$.

Lemma 1: $\forall w_i, w_j \in \{w_1, w_2, \ldots, w_{2^n}\}$, $Eng(i) > Eng(j)$ iff

$$\sum_{k=0}^{2^n}\left[E_{couple}(w_k, w_i) + E_{couple}(w_i, w_k)\right] >$$

$$\sum_{k=0}^{2^n}\left[E_{couple}(w_k, w_j) + E_{couple}(w_j, w_k)\right]$$

Given an n-bit binary vector $a_0a_1a_2...a_n$, an *intra-vector-switch* (*IVS*) is defined as two adjacent bits a_i and a_{i+1} with $a_i \neq a_{i+1}$. Given a codeword w, its *intra-vector-switch number*, which is denoted as $IVS(w)$, is the number of occurrence of the pattern "01" and "10" in this codeword. For example, $IVS(0000100) = 2$. It can be proved that the $Eng(i)$ is solely determined by the value of $IVS(w_i)$.

Theorem 2: $\forall w_i, w_j \in \{w_1, w_2, \ldots, w_{2^n}\}$, $Eng(i) > Eng(j)$ iff $IVS(w_i) > IVS(w_j)$.

Based on theorem 2, the problem of searching for a set of codewords with the minimal $Eng(i)$ can be transformed into the problem of generating a set of binary vectors with the minimal number of intra-vector-switches. It is easy to see that there are C_{n-1}^{ins} codeword that has *ivs* intra-vector-switches. Hence, the set of codeword with the minimal $Eng(i)$ is a set of binary vectors $A=\{v_1, v_2, \ldots, v_{2^m}\}$ such that $\forall v \in A$, $IVS(v) \leq x$, where $\sum_{i=0}^{x-1} C_{n-1}^{i} \leq 2^m \leq \sum_{i=0}^{x} C_{n-1}^{i}$. Figure 3 gives the pseudo code of the improved WCM algorithm.

It is interesting to note that when the $n \leq 5$ and the input data vectors are uniformly distributed, i.e. $prob(v)=1/2^m$ $\forall v \in \{1, 2, \ldots, 2^m\}$, the WCM algorithm provides the same coding scheme as the "forbidden pattern" algorithm [2]. When the input data is not evenly distributed, the WCM algorithm outperforms the "forbidden pattern" algorithm by assigning the code with lower intra-vector-switches the higher mapping priority. The experimental results show that the WCM algorithm provides about 20% less energy dissipation than the "forbidden pattern" algorithm does when the input follows triangle or normal distribution. Furthermore, the "forbidden pattern" algorithm cannot be used to generate any coding scheme with n>5 and a=1 because there is only $2F_{n+1}$ valid codewords where F_n is the Fibonacci number. However, there is no limitation on the number of codes with the WCM algorithm. Theorem 2 clearly indicates the difference between the WCM algorithm and the "forbidden pattern" algorithm. While the "forbidden pattern" algorithm avoids placing intra-vector-switches adjacent to each other, the WCM algorithm targets at minimizing their occurrence. Since the expected crosstalk energy is proportional to the number of intra-vector-switches, reducing their occurrence directly leads to the lower crosstalk energy.

```
Weighted Code Mapping:-
Set codeword array A[2^m]=0;
ivs=0;
C=0;
id = 0;
While (|A|)<2^m {
   If (C== C^{ivs}_{n-1}) {
      ivs++;
      C=0; }
   Generate the next binary vector w with IVS(w)=ivs;
   A[id++] = w;
}
Sort the input data based on their probability;
Map the highest probability data to the code with the smallest IVS;
```

Figure 3 The WCM algorithm based on pattern generation.

B. *Window-based adaptive bus encoding*

A moving window is used to slide over the input data sequence, and the probability of input data is calculated within each window. The objective of using this moving window is to predict the abrupt changes in input data distributions so that bus encoding process can be applied locally on that particular set of data. In our approach, we compare the probabilistic distributions of two consecutive windows to detect changes in data. If the difference of the distribution exceeds a pre-defined threshold, we will invoke the bus encoding algorithm again based on the new probabilistic distribution.

The pseudo code of the algorithm is given in Figure 4. Different window size will obviously affect the process of detecting data distribution changes, hence affect energy consumption and run time considerably. Experiments show that smaller window size would lead to better coding scheme in terms of energy consumption, while worse in terms of time complexity of the coding method. However, the window size has to be large enough to reflect the probabilistic changes accurately, and to reduce the number of on-the-fly bus encoding.

```
set window size
encode window WinPrev, store coding scheme T
repeat
    move window down to the next window WinCurr
    if ProbDiff > threshold
        encode next window of WinCurr ;
        store new coding scheme T_new, by the probability
                    predicted based on WinCurr;
    else
        encode WinCurr by using T
until data stream terminates
```

Figure 4 Window-based adaptive encoding algorithm.

IV. EXPERIMENTAL RESULTS

In our experimental setups, we first compare the WCM algorithm with other existing heuristic algorithms. Random data sequences of 4, 8 and 16-bit wide are generated following three different types of random distributions: triangular distribution (sequences T-4, T-8 and T-16), uniform distribution (sequences U-4, U-8 and U-16 and normal distribution (sequences N-4, N-8 and N-16). Bus energy comparisons are done for the bus-inverting (BI), the odd-even-inverting (OEI), the forbidden pattern (FP), and the WCM algorithms. The simulation results for a=2, λ=5 are shown in Table 1. In the table, the last column shows the percentage improvement of the WCM algorithm over the best of other algorithms. Experimental results (which are not shown due to the limitation of space) also show that the improvement is even better for high values of λ.

Table 1 Comparisons of bus encoding algorithms.

Sequence	Bus energy per vector				Improvement
	BI	OEI	FP	WCM	(%)
T-4	8.451*C_L	8.451*C_L	7.510*C_L	5.930*C_L	21.0
U-4	8.147*C_L	8.147*C_L	7.295*C_L	5.561*C_L	23.8
N-4	8.580*C_L	8.580*C_L	7.008*C_L	5.643*C_L	19.5
T-8	17.70*C_L	17.70*C_L	17.74*C_L	15.84*C_L	10.5
U-8	17.56*C_L	17.56*C_L	17.60*C_L	13.24*C_L	24.6
N-8	17.98*C_L	17.98*C_L	16.06*C_L	12.18*C_L	24.2
T-16	33.90*C_L	34.06*C_L	33.80*C_L	28.74*C_L	15.0
U-16	33.96*C_L	34.28*C_L	33.85*C_L	30.52*C_L	9.9
N-16	34.43*C_L	34.00*C_L	30.92*C_L	26.40*C_L	14.6

Next, we compare the bus energy consumptions of the "adaptive WCM" algorithm and the "global WCM" algorithm. In this setup, each benchmark sequence contains four subsequences from four different real applications. Note that the "global WCM" algorithm calculates the encoding just once, based on the probabilistic distribution of the whole sequence, while the "adaptive WCM" algorithm may calculate the encoding multiple times, based on the variations of the data probabilistics. Table 2 shows the experimental results for a=1 and λ=5. The results show average saving of about 24% in bus energy for the "adaptive WCM" algorithm.

Table 2 Comparison of adaptive and global algorithms.

Sequence	Energy per vector		# of adaptive recoding	Improvement (%)
	Global_WCM	Adapt_WCM		
Image-4	0.05*C_L	0.02*C_L	6	60.0
Video-4	3.98*C_L	3.57*C_L	8	10.3
Image-8	10.84*C_L	8.05*C_L	16	25.7
Video-8	10.07*C_L	8.66*C_L	16	14.0
Audio-8	2.76*C_L	2.46*C_L	6	10..9

V. CONCLUSIONS

In this paper, we have proposed a low-power bus encoding algorithm that generates one-to-one mapping from the original data to the coded data, based on the probabilistic distribution information. The proposed WCM algorithm has low hardware/computation overheads while achieving better energy savings. A window-based adaptive encoding method is proposed for further energy saving.

REFERENCES

[1] S. R. Sridhara, A. Ahmed, and N. R. Shanbhag, "Area and Energy-Efficient Crosstalk Avoidance Codes for On-Chip Buses,", *Proc. Of IEEE International Conference on Computer Design*, 2004.

[2] C. Duan and A. Tirumala, "Analysis and Avoidance of Cross-talk in On-Chip Buses," *Hot Interconnects 9*, pp. 133-138, Aug. 2001.

[3] N. K. Samala, D. Radhakrishnan and B. Izadi, "A Novel Deep Sub-micron Bus Coding for Low Energy," *Proc. of the International Conference on Embedded Systems and Applications*, June 2004.

[4] B. Victor and K. Keutzer, "Bus Encoding to Prevent Crosstalk Delay," *IEEE/ACM International Conference on Computer Aided Design*, 2001.

[5] S. R. Sridhara, N. R. Shanbhag, "Coding for System-on-Chip Networks, A Unified Framework," *IEEE Transaction on Very Large Scale Integration System*, vol. 13, No. 6, June 2005.

[6] M.R. Stan and W.P. Burleson, "Bus-Invert Coding for Low Power I/O," IEEE Trans. VLSI Syst., pp. 49–58, Mar 1995.

[7] C. Zhang and F. Vahid, " A Power-Configurable Bus for Embedded Systems," IEEE International Symposium on Circuits and Systems, Scottsdale, May 2002, pp.V-809-812.

[8] F. Worm, P. Lenne, P. Thiran and G. Micheli, "An Adaptive Low-Power Transmission Scheme for On-chip Networks," International Symposium on Systems Synthesis, Kyoto, Japan.. pp.92-100,Oct 2002.

[9] M.R. Stan and W.P. Burleson, "Low-power encodings for global communication in CMOS VLSI," IEEE Transactions on VLSI Systems, vol. 5, no.4, pp. 444-455, Dec 1997.

[10] Sotiriadis, P., A. P. Chandrakasan, "Bus Energy Reduction by Transition Pattern Coding Using a Detailed Deep Submicrometer Bus Model," *IEEE Transactions on Circuits and Systems*, pp. 1280-1295, October 2003.

[11] J. Davis and J. Meindl, " Compact distributed RLC interconnect models partII: Coupled line transient expressions and peak crosstalk in multi-level networks,' IEEE Trans. Electron Devices, vol. 47, pp. 2078-2087, Nov. 2000.

[12] Sotiriadis P., A. P. Chandrakasan, " A Bus energy model for deep sub-micron technology," IEEE Trans. VLSI Syst., vol.. 10, pp. 341-350, June 2002.