Partitioned Bus Coding for Energy Reduction

Lin Xie, Peiliang Qiu

Department of Information Science and
Electronic Engineering
Zhejiang University
Hangzhou, 310027, China
Tel: 86-571-87951820
E-mail: {trthank, qiupl}@zju.edu.cn

Qinru Qiu

Department of Electrical and
Computer Engineering
State University of New York, Binghamton
New York, 13902, USA
Tel: 607-777-4918
E-mail: qqiu@binghamton.edu

Abstract – For VLSI design in deep submicron technology, the bus energy reduction has become more and more important. This paper studies the bus partition scheme for the Transition Pattern Coding (TPC). The genetic algorithm based approach is used. A closed-form expression is derived to calculate the energy dissipation for the partitioned bus with TPC coding. A general bus model with coupling capacitance is considered during the energy estimation and optimization. The resulted partitioned bus coding reduces the encoding and decoding complexity of the original TPC. The experimental results show that the TPC with careful bus partitioned saves up to 16.9% the energy of the TPC with random bus partition.

I. Introduction

As technology scales to deep submicron technology, the bus energy reduction has become more and more important. There are extensive research works on reducing the switching activity on bus. The Gray code [1], T0 code [2] and Beach code [3] are designed for address bus, which usually carries consecutive data. The Bus-Invert code, its variants and the adaptive [4, 5, 6], are designed for data bus, which usually carries random data. When the technology enters deep submicron, the coupling capacitance becomes the dominant factor of on-chip bus capacitance. Recently, minimizing the coupling activity between bit lines has been considered in [7] to reduce the bus energy dissipation. The authors in [8] also bring forward an algorithm to appropriately shuffle the bit lines of the bus to reduce the coupling capacitance.

Transition Pattern Coding (TPC) [9] refers to a set of coding schemes for which the encoders/decoders are time-invariant, finite-state machines (FSM). Ref.[9] proposes an algorithm for TPC that effectively reduces the bus energy based on a general bus model that considers both the coupling capacitance and line-ground capacitance. To reduce the encoder/decoder complexity, the bus can be partitioned into blocks and TPC is applied to each block. This partitioned TPC is here called as PTPC in abbreviation. However, how to partition the bus is not discussed in [9]; only random partition (The lines are sequentially grouped into blocks) is considered. A closed-form expression is mathematically derived to model the bus energy under TPC. The measure is further improved to model the bus energy under PTPC. However, the later is derived based on an implicit assumption that the input distribution of each block is the same. In the worst case, it assumes that each input bit has the same distribution.

In this paper, we study the bus partition schemes for the TPC. We show that by carefully ordering and partitioning the

bus, more energy reduction can be obtained than the random partition. A partition algorithm that is motivated by the relation between the entropy and power in information theory is proposed. We also derive a bus energy model with general input distribution for the PTPC. To distinguish with the random partition, we call such ordered partition for TPC as OPTPC and the random partition for TPC as RPTPC. The experimental results show that the OPTPC saves up to 16.9% energy of RPTPC.

This paper is structured as follows. Section II introduces the TPC and PTPC algorithms. Section III introduces the bus energy model given in [9] and presents our improvements. Section IV presents our bus partition algorithms. Section V gives the experimental results. Finally, Section VI gives our conclusion.

II. Partitioned Transition Pattern Coding Scheme

The basic architecture of the TPC is given in Fig. 1. It contains two FSMs serving as encoder and decoder. Each FSM consists of a set of registers and a combinational logic block (CLB). For a TPC with a bits redundancy, assuming that the original data is an M bit vector, the encoder transforms the M bit original data into an $M+a$ bit vector, which is the actual data that is transmitted on the bus; the decoder receives the $M+a$ bit vector and transforms it back to the M bit original data. The state of the current transmitted data is solely determined by state of the previous transmitted data and value of the original input data. If the original input data is a random variable, the transmitted sequence can be modeled as a Markov process TR. The transition matrix P of the Markov process is determined by the coding scheme and the probability of the input data.

Let $A \bullet B = [a_{i,j} \cdot b_{i,j}]_{i,j}$ be the Hardmard product of two matrices A, B of the same dimensions and let $\underline{1}$ be the (column) vector with all its coordinates equal to one. It can be proved that the time averaged expected bus energy under the TPC could be calculated as the following equation [9].
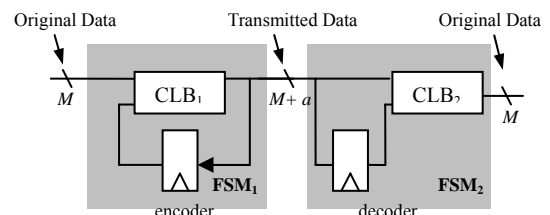

Fig.1 Architecture of the TPC scheme

$$E_{TPC} = b_O^T \cdot (P \bullet E) \cdot \underline{1} \qquad (1)$$

where b_0 is the state probability of the transmitted data, which can be calculated as the left eigenvector of P corresponding to eigenvalue 1; E is the energy cost matrix, its ij-th entry gives the bus energy dissipation when the TR switches from the i-th state to the j-th state. By carefully choosing the coding scheme, we can minimize the E_{TPC} by appropriately setting the values of P and b_0. To reduce the complexity of the encoder and decoder, the M bit input vector is partitioned into n blocks. Fig.2 shows the block diagram of the partitioned TPC. In this figure, m is equal to M/n. For each block, TPC encoding and decoding are applied individually. Later in this paper, we will show how the partition of the input vector affects the final energy dissipation of the PTPC and also give a partition algorithm that is motivated by the relation between the entropy and power in information theory.

III. Bus Energy Dissipation for PTPC scheme Under General Input Distribution

The total time averaged expected bus energy for PTPC (E_{PTPC}) is derived in [9]. Let W be the set of codewords of the transmitted data for each block, $W = \{w_1, w_2, w_3, \ldots, w_L\}$ and $L = 2^{m+a}$. Let W_{0*} and W_{1*} be the subsets of W containing only the codewords whose first bit is 0 and 1, respectively. Similarly, let W_{*0} and W_{*1} be the subsets of W containing only the codewords whose last bit is 0 and 1, respectively. Define h_{*a}^i, and h_{a*}^i as follows

$$h_{*\alpha}^i = \begin{cases} 1, & \text{if } w_i \in W_{*\alpha} \\ 0, & \text{if } w_i \notin W_{*\alpha} \end{cases}$$

$$h_{\alpha*}^i = \begin{cases} 1, & \text{if } w_i \in W_{\alpha*} \\ 0, & \text{if } w_i \notin W_{\alpha*} \end{cases}.$$

Ref. [9] gives the expression of E_{PTPC} as follows.
$$E_T = n b_0^T \cdot (P \bullet E) \cdot \underline{1} + \lambda (n-1) \cdot b_0^T \left(H_{1*} \cdot P \cdot H_{0*} + H_{*1} \cdot P \cdot H_{0*} \right) \cdot \underline{1} \qquad (2)$$

where H_{*a} and H_{a*} are diagonal matrices defined as

$$H_{*\alpha} = diag\left(h_{*\alpha}^1, h_{*\alpha}^2, \ldots, h_{*\alpha}^M \right),$$
$$H_{\alpha*} = diag\left(h_{\alpha*}^1, h_{\alpha*}^2, \ldots, h_{\alpha*}^M \right)$$

λ is called as capacitance factor, the ratio between the coupling capacitance (C_I) and the line-to-ground capacitance (C_L), $\lambda = C_I/C_L$, and P is the transition matrix of all the blocks. Unfortunately, the derivation of E_{PTPC} in (2) is based on an implicit assumption that the input distribution of each block is the same. In the worst case, it assumes that each input bit has the same distribution. Here we improve this derivation to
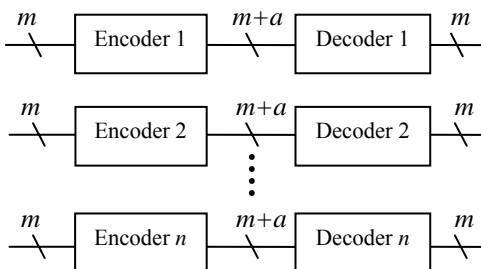

Fig.2 Block diagram of PTPC scheme

consider a general input distribution and give the following theorem.

Theorem: Under general input distribution, the total time average expected energy consumption of the entire coding scheme is given by

$$\begin{aligned} E_T &= \sum_{i=1}^{n} E_i + \sum_{j=1}^{n-1} E_{aj} \\ &= \sum_{i=1}^{n} b_{iO}^T \cdot (P_i \bullet E) \cdot \underline{1} \\ &+ \sum_{j=1}^{n-1} \lambda \cdot \left(b_{jO}^T \cdot H_{*1} \cdot P_j \cdot H_{*0} + b_{(j+1)0}^T \cdot H_{1*} \cdot P_{j+1} \cdot H_{0*} \right) \cdot \underline{1} \end{aligned} \qquad (3)$$

where P_j is the transition matrix of the corresponding Markov process of the encoder for the j-th block, b_{j0} is the corresponding left eigenvector.
The proof is skipped due to the limit of the space.

IV. Our Proposed Algorithm

A. Motivation

From (3), it is clear that the energy consumption of PTPC scheme contains two parts: the total individual energy consumption of all blocks using TPC schemes $E_1 = \sum_{i=1}^{n} E_i$ and the total interaction energy consumption due to the adjacent blocks $E_2 = \sum_{j=1}^{n-1} E_{aj}$. It will be extremely time consuming to find an optimal bus partition that minimizes both E_1 and E_2 at the same time. Here we use the divide-and-conquer algorithm to optimize them alternatively. For simplicity of presentation, the parameters M, a, m and n below are defined as in Section II and III.

From information theory, it is known that the more uniformly the input data is distributed, the greater its entropy and entropy power will be. The lower bound of the energy consumption will increase accordingly.
Definition: Let $p = [p_1, p_2, \ldots, p_K]$ be the input distribution. Non-uniformity of the distribution is measured by the distribution variance D, which is defined as follows:

$$D(p) = \sum_{i=1}^{K} \left(p_i - \frac{1}{K} \right)^2 \qquad (4)$$

Generally, the more uniformly the input data is distributed, the smaller its corresponding distribution variance $D(p)$ will be and the lager the energy will be dissipated.

Experiments have been setup to verify the above property. A large number of 4-bit input sequences with different distributions are generated. Fig. 3 gives the relation between the input distribution variance and the bus energy dissipation after applying TPC schemes. The redundancy (a) of TPC is set to 1 bit and 2 bit for the two figures. It shows the energy consumption after using TPC schemes decreases with the increase of the distribution variance on the whole.
Based on the analysis above, we can reduce E_1 by searching a good partition of bit lines of the bus so that the total distribution variance of all the blocks $D = \sum_{k=1}^{n} D_k$ is as large as possible, where D_k is the distribution variance of the k-th block.
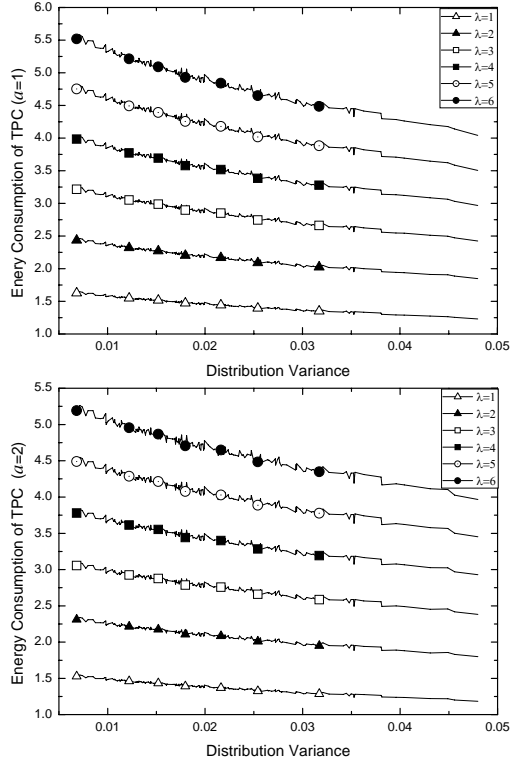
Fig. 3 Energy consumption of TPC schemes corresponding to the distribution variance for K=16, a=1, 2

Besides good partition of bit lines of inside block, the order of the blocks will also affect the interaction energy consumption. The optimal order can be very easily found when the number n is not very large.

Based on the distribution variances, our proposed partitioned bus coding scheme can be defined as follows.

- Partition bit lines of bus into n blocks, with a maximal total distribution variance;
- Apply TPC schemes to each block
- Order the blocks to minimize the total interaction energy consumption;

B. Algorithm

When the bus is more than 16 bit wide, the search space of bus partition becomes extremely huge. Therefore, we use genetic algorithm (GA) to partition the blocks; the optimal order of the blocks can be considered as a minimum weighted path cover (MWPC) problem [10]. We call such partitioned coding scheme using this method as OPTPC.

First, we can consider the bus partition as an ordering problem. After shuffling the lines, the bus is sequentially grouped into several blocks. In the genetic algorithm, six basic components are necessary for the successful implementation.

Chromosome Assume the original bus is M bit wide and the solution is represented by the array G of M elements as
$$G = [l_1, l_2, \ldots, l_M]. \text{ with } l_i \in [1, M], l_i \neq l_j \text{ for all } i \text{ and } j \quad (7)$$

Example: Consider an 8-bit bus $\{l_1, l_2, l_3, l_4, l_5, l_6, l_7, l_8\}$; when a chromosome G is [8, 6, 5, 1, 3, 2, 4, 7]; it means that the final placement order of the bit lines is $\{l_8, l_6, l_5, l_1, l_3, l_2, l_4, l_7\}$.

Fitness function If G indicates the chromosome that maps the final order, the fitness function is defined as the total distribution variance of the bus after ordering.

Selection Criterion We use a proportional criterion, called roulette proportional criterion, guaranteeing that the best individual of the current generation is more possible to pass to the next generation.

Mutation Application of the mutation operator to an order consists of varying the place of a line with a probability equal to the mutation probability $pb1$. The following pseudo code gives the mutation operation.

```
Mutation (G, pb1)
 for i =1: M
    if (Event(pb1)) then
       Generate a random integer j to guarantee that |j-i|<n
       Swap(G(i),G(j));
    endif
 endfor
```

where Event(p) returns true with the probability of p.

Crossover The idea is to choose two elements of the population (G_1, G_2) and to interchange their subclass ($[i, j]$, $i<j$) with a certain probability called crossover probability ($pb2$).

```
Crossover(G1, G2, pb2)
   if (Event(pb2)) then
      Generate two random integers i, j ranging between 1 and M and
      guarantee i > j
      for k =i : j
             Exchange the elements with the index no. from i to j of the
             orders G1, G2
             /Update G1 and G2 to guarantee no lines appear more than
             once/
      endfor
   endif
end
```

Stop Condition We simply design the maximum number of generations of the GA process as the stop condition.

After determining the bit lines of each block and applying TPC schemes for each block, we order the n blocks to further reduce the total interaction energy consumptions. Let $G(V,E)$ be a directed complete graph, where each node in V represents a block and weight $w(u,v)$ on directed edge $(u,v) \in E$ from node u to node v is the interaction energy between blocks u and v. Therefore, there are $n(n-1)/2$ directed edges in G. Since we want to find an optimal order of the blocks so that the sum of interaction energy consumption is minimum, it can be simplified to find an MWPC in $G(V,E)$. The MWPC problem is NP-complete, and we can use a heuristic algorithm, called C-Order [11]. At each step, the edge with the smallest weight is selected, which will not cause a cycle and the degree of a node will be not more than two. The time complexity is bounded by $O(n^2.logn)$.

V. Experiment Results

To evaluate the efficiency of the proposed partitioned bus coding schemes, we did experiments with 60 random

generated data sequences. Half of them are 16 bit and the others are 32 bit. Each sequence has no temporal or spatial correlations. In the genetic algorithm, we consider a population of 20 individuals, a crossover probability of 80% and a mutation probability of 10%. The maximum number of generations is 30.

The experiments were performed to check how much the total energy consumption is reduced using OPTPC and how much the OPTPC optimizes over the RPTPC schemes.

We applied OPTPC schemes to all the 16-bit and 32-bit patterns, and computed the average energy consumption reduction compared to the energy without coding. Fig.4 gives the average reduction for M=16, 32, m=4, a=1, 2, $1 \leq \lambda \leq 6$. It shows that the average power reduction increases with the increase of the capacitance factor and the OPTPC schemes reduce up to 47.2% and 45.8% of the bus energy when M=16 and 32.

We further compared the performances of the RPTPC and OPTPC schemes. Fig.5 gives the results when M=16, m=4, a=2, $1 \leq \lambda \leq 6$. It shows that compared with RPTPC, the OPTPC scheme on average has 1.6% ~ 6.84% less energy dissipation when λ ranges from 1 to 6. For some patterns, the reduction can be much higher. The maximum reduction is 5.32% when λ=1 and 16.9% when λ=6.

VI. Conclusion

In this paper, we addressed the partitioned TPC schemes and introduced an improved bus energy measure. We further propose a bus partition algorithm for the partitioned TPC scheme. The bus lines are shuffled and partitioned in order to minimize the total energy reduction. Experimental results show that the partitioned bus coding schemes using this method can save more bus energy.
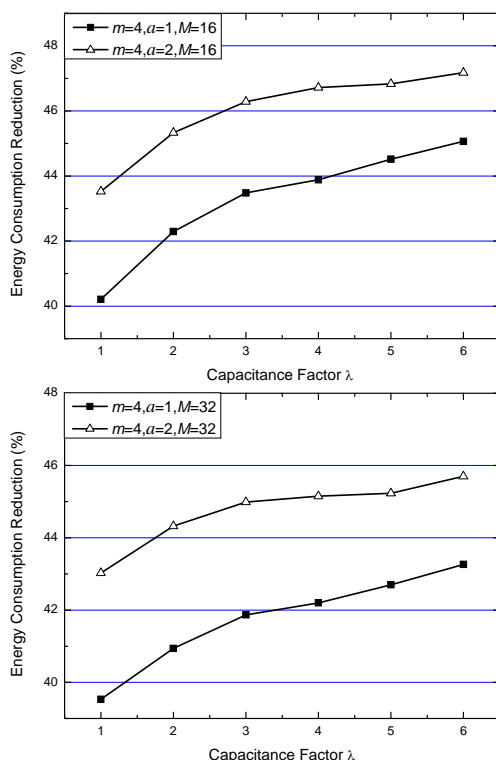

Fig.4 Average energy consumption of OPTPC for $1 \leq \lambda \leq 6$
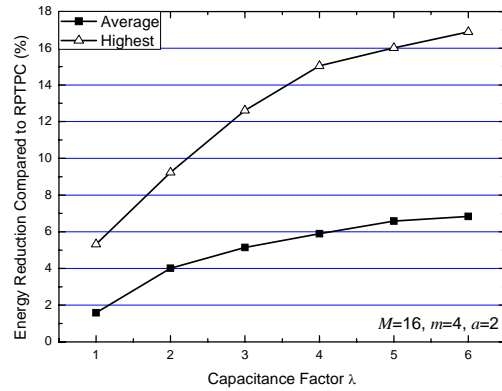

Fig.5 Average and highest energy reduction OPTPC can achieve over RPTPC for M=16, m=4, a=2, $1 \leq \lambda \leq 6$

References

[1] 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*, vol. 11, pp. 24-30, 1994

[2] L. Benni, G. De Micheli, E. Macii, D. Scivto, and C. Silvano, "Asymptotic Zeros-Transition Activity Encoding for Address Busses in Low Power Microprocessor-based Systemes," *Proc. of Great Lakes Symposium on VLSI,* pp. 77-82, 1997

[3] L. Benni, G. De Micheli, E. Macii, and S. Quer, "System-level Power Optimization of Special Purpose Applications, the Beach Solution," *Proc. of ISLPED,* pp. 24-29, 1997

[4] M. Stan and W. Burleson, "Bus-invert coding for low-power I/O," *IEEE Trans. on VLSI Systems*, vol.3, pp.49-58, 1995

[5] Y. Shin, S. Chae and K. Choi, "Reduction of bus-transitions with partial bus-invert coding," *IEE Electronics Letters*, vol.34, pp.642-643, 1998

[6] S. Hong, T. Kim, U. Narayanan and K. S. Chung, "Decomposition of bus-invert coding for low power I/O," *J. Circuits, Syst., Comput.,* vol. 10, pp. 101-111, 2000

[7] K. W. Kim, K. H. Back, N. Shanbhag, C. L. Liu, and S. M. Kang, "Coupling-driven Signal Encoding Scheme for Low-Power Interface Design," *Proc. of ICCAD*, pp. 318-321, 2000

[8] Y. Shin, T. Sakurai, "Coupling-driven bus design for low-power application-specific systems," *Proc .of DAC,* pp. 744-749, 2001

[9] P. Sotiriadis and A. Chandrakasan, "Bus Energy Reduction by Transition Pattern Coding Using a Detailed Deep Submicrometer Bus Model," *IEEE Trans. on Circuits and Systems –I: Fundamental Theory and Application*, vol. 50, pp. 1280-1295, October 2003

[10] C. H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization*, Englewood Cliffs, NJ: Prentice-Hall, 1982

[11] C. Lyuh, T. Kim, and K. Kim, "Coupling-Aware High-Level Interconnect Synthesis," *IEEE Trans. on Computer-Aided Design,* vol. 23, pp. 157-164, January, 2004