Bus Energy Reduction by Transition Pattern Coding Using a Detailed Deep Submicrometer Bus Model

Paul P. Sotiriadis and Anantha P. Chandrakasan

***Abstract*—A data-distribution and bus-structure aware methodology for the design of coding schemes for low-power on-chip and interchip communication is presented. A general class of coding schemes for low power, termed transition pattern coding schemes, is introduced. The energy behavior of the schemes is mathematically analyzed in detail. Two algorithms are proposed for deriving such efficient coding schemes, which are optimized for desired bus structures and data distributions. Bus partitioning is proposed and mathematically analyzed as a way to reduce the complexity of the encoder/decoder.**

***Index Terms*—Activity, architecture, buses, coding, deep submicrometer (DSM), digital circuits, energy, estimation, Markov, low power, modeling, power reduction, process, submicron, transition, transition activity, transition activity matrix, very large-scale integration (VLSI).**

I. Introduction

As TECHNOLOGY scales to deep submicrometer (DSM) dimensions, the energy cost of performing computation continues to decrease while the cost of on-chip and interchip communication is not improving. Over the past several years, significant emphasis has been placed on reducing the energy dissipation associated with communication through long on-chip and interchip buses. Numerous schemes have been presented including low-swing signaling, e.g., [1]–[3], charge recycling, e.g., [4]–[7], and data coding, e.g., [8]–[17] including theoretical paper, e.g., [18]. The starting point of these papers is a simple and very popular power estimator, the *transition activity*. The transition activity, T_a, of a circuit node (or line) is a useful power measure when the node (or line) is *decoupled* from any other active node in the circuit. If this is true, then, the transition activity corresponds to a power dissipation given by $P = T_a f C V_{\rm dd}^2/2$ where C is the capacitance between the node (line) and the ground, and f is the operating frequency. This formula is very convenient for estimation of power dissipation. Moreover, such an energy model allows for the design and evaluation of circuits that reduce power dissipation. Unfortunately, as technology scales down, not many lines (not even many nodes) can be regarded as isolated or shielded.

In DSM buses, for example, the coupling between lines is usually much stronger than the coupling between individual lines and ground [19]. This is a result of the high aspect ratio of the wires and the small distance between them. Coupling between nodes (lines) implies that power dissipation depends on the *cross activities* [20] of the nodes (lines) and, therefore, the simple formula $P = T_a f C V_{\rm dd}^2/2$ cannot be used.

Coding techniques for low-power communication DSM buses, comprised of capacitively coupled lines, was introduced in [21]. The energy expression that had to be minimized suggested that minimizing the expected number of transitions is not necessarily the best approach to reduce power. The set of favorable transitions should be selected in a more elaborate way. Several studies followed demonstrating the same principle by presenting schemes specifically designed for DSM data or address buses, e.g., [22]–[27]. An attempt for a generalized approach was introduced in [27].

This paper presents a *method* [transition pattern coding (TPC) schemes and algorithm] for designing coding schemes for low-power communication through buses. The starting point is a general bus model that can accommodate *any* coupling between the lines. A *bus energy model*, which introduces new energy measures in closed form based on the transitions of the entire bus, is derived from this general bus model.

A general class of coding schemes is introduced, termed the TPC schemes. This class contains all schemes for which the encoder and decoder are time-invariant, finite-state machines. The input to the encoder is the new data vector to be transmitted; and, the state of the encoder/decoder is the last vector transmitted through the bus (state of the bus). Many of the schemes presented in the literature are examples of this general class.

The introduced bus-energy model is used to study, in detail, the energy-reduction properties of the class of TPC schemes. Closed-form expressions are derived for the energy reduction using TPC schemes and used in the development of the proposed approximately optimal coding (AOC) and the TPC algorithms. Both algorithms derive efficient coding schemes optimized for desired bus structures and data distributions.

Finally, bus partitioning is mathematically analyzed for the whole class of TPC schemes as a way to reduce the complexity of the encoder/decoder. Examples in the literature (e.g., partial bus invert coding [14]) that use bus partitioning can be analyzed using the presented theory.

The purpose of this paper is not to present a specific coding scheme, but rather to establish a theory of the energy reduction

Manuscript received March 24, 2001; revised September 24, 2001 and May 4, 2002. This work was supported in part by the Interconnect Focus Center Research Program under MARCO and DARPA. The work of P. Sotiriadis was supported in part by the Alexander S. Onassis Public Benefit Foundation, and in part by the Greek Section of Scholarships and Research.This paper was recommended by Associate Editor R. Sridhar.

P. P. Sotiriadis was with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139 USA. He is now with the Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore MD 21218 USA (e-mail: pps@jhu.edu).

A. P. Chandrakasan is with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: anantha@mtl.mit.edu).

Digital Object Identifier 10.1109/TCSI.2003.817765

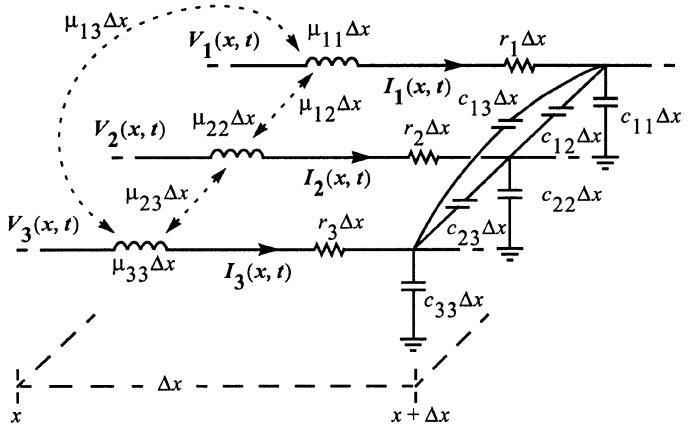

Fig. 1. Elementary segment of DSM bus.

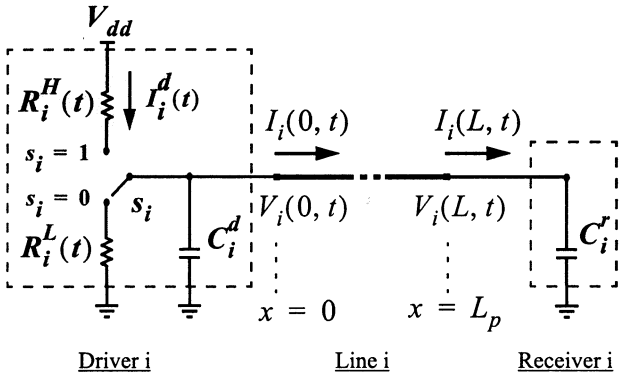

Fig. 2. Driver-line receiver.

behavior of the TPC class as a whole by providing analytical mathematical expressions. Practical issues, for example, power overhead of the encoder/decoder, have to be addressed for specific coding schemes, bus structures and data distributions, and are out of the scope of this paper.

II. General Energy Model for DSM Buses

A bus consists of a segment of parallel wires or a sequence of segments with repeaters and possibly latches between them. The energy required to drive a sequence of segments is approximately the sum of the energy required to drive each segment independently. Therefore, we can consider only the simplest case in our analysis. A general model of the bus segments should include the capacitive and inductive coupling between any two lines. An elementary segment of length is Δx shown in Fig. 1. Details on the model can be found, e.g., [28]–[32].

All the parasitic elements between the lines and ground (shielding or substrate) and between the lines themselves, are included in the model. In the elementary segment above, $r_i(x)\Delta x$ is the serial resistance of the ith line, $c_{i,i}(x)\Delta x$ is the parasitic capacitance between the ith line and ground, $c_{i,j}(x)\Delta x$ is the parasitic capacitance lines i and j, $\mu_{i,i}(x)\Delta x$ is the self inductance of the ith line and $\mu_{i,j}(x)\Delta x$ is the mutual inductance between lines i and j. All parasitic elements are distributed and their densities $r_i(x), c_{i,i}(x), c_{i,j}(x), \mu_{i,i}(x), \mu_{i,j}(x)$ may vary along the length of the wires. Note that lumped parasitics can always be regarded as limiting cases of distributed ones and therefore can be parts of the model.

Let L_p be the total length of the bus segment. Let $I_i(x,t)$ be the current running through the ith line at the point $x, 0 \le x \le L_p$ and time $t \ge 0$. Similarly let $V_i(x,t)$ be the voltage at point x with respect to ground. The bus segment is driven by a driver, which in most cases is a CMOS inverter. Also, the segment has a load that represents a buffer or a latch. A standard simplified model for the driver of the ith line, $i = 1, 2, \ldots, n$ is shown along with the line it drives in Fig. 2, [33].

The capacitor C_i^d models the parasitic capacitance of the driver; and, C_i^r models that of the load. The switch s_i connects the line to $R_i^H(t), (s_i = 1)$ or $R_i^L(t), (s_i = 0)$ depending on whether a one or a zero is transmitted. The resistors $R_i^H(t)$ and $R_i^L(t)$ correspond to the PMOS and NMOS transistors of the inverter.

Now, let T be the period of the clock synchronizing the bus. Suppose that at time $t = 0$, the transmission of the new data values, $V_i^{\text{new}}, i = 1, 2, \ldots, n$ starts (V_i^{new} is 0 or V_{dd}). Since the data are "sampled" at the other end of the bus (segment) at some time $t_s \le T$, it is reasonable to assume that the time period $[0, T]$ is sufficient for the lines to settle to their final values. That is

$$V_i(x, T) = V_i^{\text{new}} \tag{1}$$

for all $x \in [0, L_p]$ and $i = 1, 2, \ldots, n$. Since the operation has period T, (1) implies that

$$V_i(x, 0) = V_i^{\text{old}} \tag{2}$$

for all $x \in [0, L_p]$ and $i = 1, 2, \ldots, n$ (V_i^{old} is 0 or V_{dd}). This is simply due to the fact that previous data must have been established by $t = 0$. This assumption allows us to express the *energy drawn from the power supply* (V_{dd}) by the driver of the ith line during the transition period as

$$E_i = \int_0^T V_{\text{dd}} \cdot I_i^d(t)\, dt. \tag{3}$$

The total energy drawn from the power supply during the transition is of course, $E = \sum_{i=1}^n E_i$. After some algebraic manipulations (details are available in [20]), the transition energy is expressed as

$$E = (\boldsymbol{V}^{\text{new}})^T \cdot \bar{\boldsymbol{C}} \cdot (\boldsymbol{V}^{\text{new}} - \boldsymbol{V}^{\text{old}}) \tag{4}$$

where $\boldsymbol{V}^{\text{new}}, \boldsymbol{V}^{\text{old}}$ are the *old* and the *new* vectors of the voltages of the lines

$$\boldsymbol{V}^{\text{old}} = \begin{bmatrix} V_1^{\text{old}} \\ V_2^{\text{old}} \\ \vdots \\ V_n^{\text{old}} \end{bmatrix} \quad \boldsymbol{V}^{\text{new}} = \begin{bmatrix} V_1^{\text{new}} \\ V_2^{\text{new}} \\ \vdots \\ V_n^{\text{new}} \end{bmatrix}$$

and $\bar{\boldsymbol{C}}$ is the $n \times n$ matrix with elements

$$\bar{\boldsymbol{C}}_{i,j} = \begin{cases} \sum_{k=1}^n \bar{c}_{i,k}, & \text{if } i = j \\ -\bar{c}_{i,j}, & \text{if } i \ne j \end{cases} \tag{5}$$

where $\bar{c}_{i,j}$ is the total capacitance between lines i and j, and $\bar{c}_{i,i}$ is the total capacitance between line i and ground. Note that

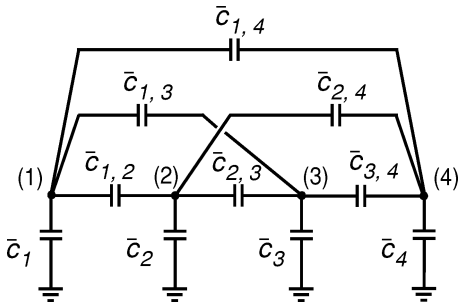

Fig. 3. Energy-equivalent lumped DSM bus model ($n = 4$).

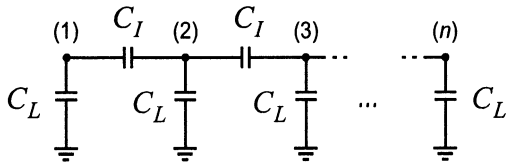

Fig. 4. Simplified energy-wise approximate DSM bus model.

$\bar{c}_{i,i}$ includes the parasitics C_i^d and C_i^r of the driver and load, respectively. Also note that $\bar{\boldsymbol{C}}$ is the conductance matrix of the lumped capacitive network below, where the distributed lines have been replaced by (lumped) nodes interconnected through the total capacitances $\bar{c}_{i,i}$ (for example, $n = 4$).

The network of Fig. 3 can be considered as a limiting case of the original bus (segment) when all capacitances become lumped and all inductances and resistances are eliminated. Therefore, when the above network replaces the bus, driver, and load capacitors, the energy dissipation from $V_{\rm dd}$ remains unchanged for all transitions. This lumped network is, in some sense, *energy-wise* equivalent to the complete distributed bus model of Fig. 1.

Remark: Throughout the paper, we make the convention that *bit value "1"* corresponds to $V_{\rm dd}$-V and *bit value "0"* corresponds to 0 V, i.e., *ground*. Thus, the data vector $\boldsymbol{L} = [l_1, l_2, \ldots, l_n]^T$ corresponds to voltage vector $\boldsymbol{V} = [V_1, V_2, \ldots, V_n]^T$ such that[1] $V_i = l_i \cdot V_{\rm dd}$ for $i = 1, 2, \ldots, n$, or $\boldsymbol{V} = V_{\rm dd} \cdot \boldsymbol{L}$.

During the transition of the bus from, a *logical* state $\boldsymbol{L}(k-1)$ to state $\boldsymbol{L}(k)$, an amount of energy equal to $E(\boldsymbol{L}(k-1), \boldsymbol{L}(k))$ is *drawn form the power supply*.[2] Using (4), the energy is expressed as

$$E(\boldsymbol{L}(k-1), \boldsymbol{L}(k)) = \boldsymbol{L}(k)^T \cdot \bar{\boldsymbol{C}} \cdot [\boldsymbol{L}(k) - \boldsymbol{L}(k-1)]^T \cdot V_{\rm dd}^2. \quad (6)$$

Example: In many structures of DSM buses, the capacitive coupling between nonadjacent lines is very weak compared to that between adjacent lines. Then, the general model simplifies to the circuit shown in Fig. 4, where C_I is the total inter-line capacitance; and, C_L is the total line-to-ground capacitance. For convenience, we define the factor

$$\lambda = \frac{C_I}{C_L}. \quad (7)$$

[1]The bits l_i are considered to be real numbers.

[2]not dissipated.

TABLE I
ENERGY DRAWN FROM $V_{\rm dd}$ BY A TWO-LINE BUS MODELED AS IN FIG. 4

Trans. Energy $C_L \cdot V_{dd}^2$		$[V_1^f, V_2^f]$			
		00	01	10	11
$[V_1^i, V_2^i]$	00	0	$1+\lambda$	$1+\lambda$	2
	01	0	0	$1+2\lambda$	1
	10	0	$1+2\lambda$	0	1
	11	0	λ	λ	0

The matrix $\bar{\boldsymbol{C}}$ (the conductance matrix of the network) becomes

$$\bar{\boldsymbol{C}} = \begin{bmatrix} 1+\lambda & -\lambda & 0 & \cdots & 0 & 0 \\ -\lambda & 1+2\lambda & -\lambda & \cdots & 0 & 0 \\ 0 & -\lambda & 1+2\lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdot & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1+2\lambda & -\lambda \\ 0 & 0 & 0 & \cdots & -\lambda & 1+\lambda \end{bmatrix} \cdot C_L. \quad (8)$$

The energy drawn from the power supply during the transition can be expressed as

$$\begin{aligned} E &= (\boldsymbol{V}^{\rm new})^{\boldsymbol{T}} \cdot \bar{\boldsymbol{C}} \cdot (\boldsymbol{V}^{\rm new} - \boldsymbol{V}^{\rm old}) \\ &= \left\{ \sum_{i=1}^{n} V_i^{\rm new} \left(V_i^{\rm new} - V_i^{\rm old}\right) \right. \\ &\quad + \lambda \sum_{i=1}^{n-1} \left(V_{i+1}^{\rm new} - V_i^{\rm new}\right) \left[\left(V_{i+1}^{\rm new} - V_i^{\rm new}\right) \right. \\ &\quad \left. \left. - \left(V_{i+1}^{\rm old} - V_i^{\rm old}\right) \right] \right\} \cdot C_L. \end{aligned} \quad (9)$$

In standard 0.18-μm technologies, minimum distances between wires, λ can take values as high as six. In 0.13-μm technologies, λ can be greater than eight. In general, technology downscaling results in an increase of the line aspect and thus an increase of the factor λ.

Example: For a bus having two lines, modeled as in Fig. 4, Table I shows the energy drawn from $V_{\rm dd}$ during the transition $[V_1^i, V_2^i] \rightarrow [V_1^f, V_2^f]$.

The energy loss caused by the interaction of the lines through C_I is captured by the "lambdas" in Table I.

Example: In the case of isolated bus lines ($\lambda = 0$ in the above example), the energy formula is reduced to

$$\begin{aligned} E &= (\boldsymbol{V}^{\rm new})^{\boldsymbol{T}} \cdot \bar{\boldsymbol{C}} \cdot (\boldsymbol{V}^{\rm new} - \boldsymbol{V}^{\rm old}) \\ &= \left\{ \sum_{i=1}^{n} V_i^{\rm new} \left(V_i^{\rm new} - V_i^{\rm old}\right) \right\} \cdot C_L \end{aligned} \quad (10)$$

which simply counts the number of lines that transition from ground to $V_{\rm dd}$ (zero to one).

A. Comments on the Energy Model

Independently of the exact bus structure, if equalities (1) and (2) are approximately satisfied (the timing assumption), the energy drawn from $V_{\rm dd}$ during a transition $V^{\rm old} \rightarrow V^{\rm new}$ depends

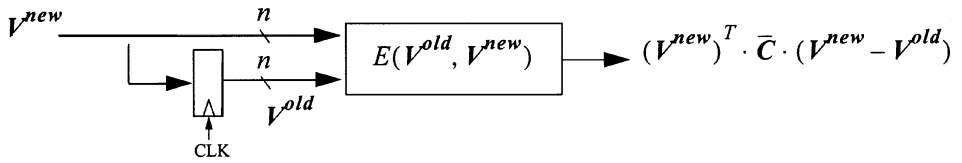

Fig. 5. Energy behavior of a bus.

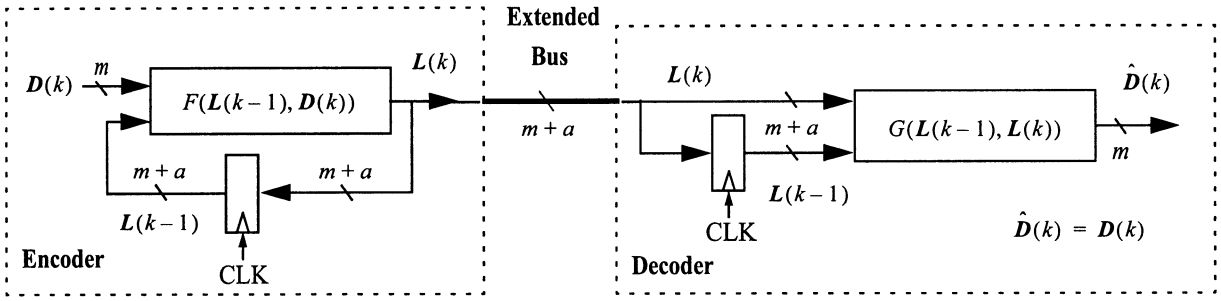

Fig. 6. General class of coding schemes: TPC schemes.

only on the new V^{new} and previous vectors V^{old}. This is expressed by the block diagram shown in Fig. 5 using a memory element (register).

Remark: Although the coding scheme design procedure that is presented in the following sections is applicable to any type of bus, the bus of Fig. 4 is used extensively for the examples.

III. General Class of Coding Schemes

A general class of bus coding schemes is considered in this section. Although many of the techniques proposed in the literature belong to this class, the unifying approach presented here is new and results in exact, closed formulas of power consumption and power reduction.

In Fig. 6, we see the general model of *time-invariant* coding schemes having the following property: the input data vector $\boldsymbol{D}(k) = [d_1(k), d_2(k), \ldots, d_m(k)]^T$ can be encoded, transmitted, and decoded during the same clock cycle k (assuming the bus does not introduce significant delay).

The "original" data bus with m lines has been expanded to a bus with $m + a$ lines in order to accommodate the higher (actual) bit rate due to encoding. The output of the decoder, *bus state* vector $\boldsymbol{L}(k) = [l_1(k), l_2(k), \ldots, l_{m+a}(k)]^T$, contains the *logical values* of these $m+a$ lines. The following relation holds for every time period k:

$$\boldsymbol{L}(k) = F(\boldsymbol{L}(k-1), \boldsymbol{D}(k)). \tag{11}$$

Vector $\boldsymbol{L}(K)$ depends on both the input data vector $D(k)$ as well as the previous state $\boldsymbol{L}(k-1)$ of the bus. The knowledge of the previous state is important because the energy dissipation of the bus is due to the transition $\boldsymbol{L}(k-1) \rightarrow \boldsymbol{L}(k)$. Recursion (11) may force the vector $\boldsymbol{L}(k)$ to take values in a subset $W = \{w_1, w_2, \ldots, w_M\}$ of $\{0,1\}^{m+a}$ (only). The M elements of W are the *codewords* of the coding scheme.

At the other end of the bus, the decoder uses the value of the current and the previous states of the bus to reconstruct the data vector. This information is sufficient and no other earlier states are required. The decoder realizes the relation

$$\hat{\boldsymbol{D}}(k) = G(\boldsymbol{L}(k), \boldsymbol{L}(k-1)). \tag{12}$$

Combining the modeling equations of the encoder and decoder [(11) and (12)] with the desirable property that for every $k, \hat{\boldsymbol{D}}(k) = \boldsymbol{D}(k)$, translates into the requirement that for every $\boldsymbol{w} \in W$ and $\boldsymbol{d} \in \{0,1\}^m$ it is

$$G(\boldsymbol{w}, F(\boldsymbol{w}, \boldsymbol{d})) = \boldsymbol{d}. \tag{13}$$

Therefore, for every fixed $\boldsymbol{w} \in W$, the mapping $\boldsymbol{d} \rightarrow F(\boldsymbol{w}, \boldsymbol{d})$ must be injective. Moreover, for every $\boldsymbol{w}$, the set

$$X_{\boldsymbol{w}} = \{F(\boldsymbol{w}, \boldsymbol{d})\text{: } \mathrm{d} \in \{0,1\}^m\} \tag{14}$$

containing all the possible values of $\boldsymbol{L}(k)$ when the last state was $\boldsymbol{L}(k-1) = \boldsymbol{w}$, must have exactly 2^m elements so that m bits of information can be transmitted each time. The function F gives rise to the *transition matrix* $\boldsymbol{T} = [t_{i,j}]_{i,j=1}^{M}$ with elements

$$t_{i,j} = \left\{ \begin{array}{ll} 1, & \text{if } \boldsymbol{w}_j \in X_{\boldsymbol{w}_i} \\ 0, & \text{otherwise} \end{array} \right\}. \tag{15}$$

The matrix $\boldsymbol{T}$ has exactly 2^m ones in every row and specifies the possible transitions of the bus, i.e., given a state $\boldsymbol{L}(k-1) = w_i$, it gives all the states $\boldsymbol{L}(k) = w_j$ that can follow. The matrix T does *not* provide any information on how *the data values are mapped into the possible transitions*. Finally, we define the *transition graph* G_T, which carries the same information as that of the transition matrix

$$G_T = \{(\boldsymbol{w}, F(\boldsymbol{w}, \boldsymbol{d})) : \mathrm{w} \in W, \quad \mathrm{d} \in \{0,1\}^m\}. \tag{16}$$

Relation (13) defines (the restriction of) the function G on the vertices of the transition graph. The values of G in the set $W \times W - G_T$ are immaterial and can be chosen in any convenient way in order to simplify the hardware implementation of the function. If the transition graph G_T has more than one strongly

connected component [34], then, the coding scheme is degenerate in the sense that some of the codewords are not utilized. To avoid this degeneracy, from now on, we assume that G_T is *strongly connected*. Strong connectivity of G_T is equivalent to the *irreducibility* of the transition matrix T, [34], a property that will be used later.

Remark 1: The sequence of bus states $L(k), k = 1, 2, \ldots$ is defined by (11), the initial state $L(0)$ and the data sequence $D(k), k = 1, 2, \ldots$. If the random vectors $D(1), D(2), D(3), \ldots$ are independent and identically distributed, (i.i.d.) and independent of $L(0)$ as well, then the bus states $L(1), L(2), L(3), \ldots$ form a Markov random sequence. Note that

$$\begin{aligned} &P_r(\boldsymbol{L}(k) \,|\, \boldsymbol{L}(k-1), \boldsymbol{L}(k-2), \ldots, \boldsymbol{L}(1)) \\ &\quad = \sum_{\boldsymbol{D}(k)} P_r(\boldsymbol{L}(k), \boldsymbol{D}(k) \,|\, \boldsymbol{L}(k-1), \boldsymbol{L}(k-2), \ldots, \boldsymbol{L}(1)) \\ &\quad = \sum_{\boldsymbol{D}(k)} P_r(\boldsymbol{L}(k) \,|\, \boldsymbol{D}(k), \boldsymbol{L}(k-1), \boldsymbol{L}(k-2), \ldots, \boldsymbol{L}(1)) \\ &\qquad \cdot P_r(\boldsymbol{D}(k) \,|\, \boldsymbol{L}(k-1), \boldsymbol{L}(k-2), \ldots, \boldsymbol{L}(1)) \\ &\quad = \sum_{\boldsymbol{D}(k)} P_r(\boldsymbol{L}(k) \,|\, \boldsymbol{D}(k), \boldsymbol{L}(k-1)) \cdot P_r(\boldsymbol{D}(k)) \\ &\quad = \sum_{\boldsymbol{D}} \delta_{\boldsymbol{L}(k), F(\boldsymbol{L}(k-1), \boldsymbol{D})} \cdot P_r(\boldsymbol{D}) \end{aligned}$$

where $\delta_{x,y} = 1$ if $x = y$ and $\delta_{x,y} = 0$, otherwise. Therefore, $L(k)$ depends statistically only on $L(k-1)$, which proves the statement. The assumption that the data sequence is i.i.d. is often necessary in order to make the analytical and/or numerical calculations tractable.

Remark 2: Since, in general, the encoder and decoder, given by (11) and (12), are finite-state machines, particular attention must be given to their design so that the possible transmission errors of the bus are not *catastrophic*, i.e., they do not propagate for ever. Alternatively, this can be achieved by using a particular state of the bus, not in W, as a reset signal for the states of both the encoder and decoder.

A. Motivational Example

A simple TPC scheme is presented here that encodes a sequence $\boldsymbol{D}$ of $m = 2$-bit data vectors into a sequence, $\boldsymbol{L}$, of $m + a = 3$ bit vectors. The set of codewords contains the following $M = 6$ elements $W = \{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_6\}$

$\boldsymbol{w}_1 = 000$	$\boldsymbol{w}_2 = 001$	$\boldsymbol{w}_3 = 010$
$\boldsymbol{w}_4 = 101$	$\boldsymbol{w}_5 = 110$	$\boldsymbol{w}_6 = 111$

The coding function F is given as a graph in Fig. 7. For example, if $\boldsymbol{L}(k-1) = 001(\boldsymbol{w}_2)$ and $\boldsymbol{D}(k) = 11$ then $\boldsymbol{L}(k) = 101(\boldsymbol{w}_4)$. The transition matrix resulting from F with columns and rows indexed according to states $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_6$ is

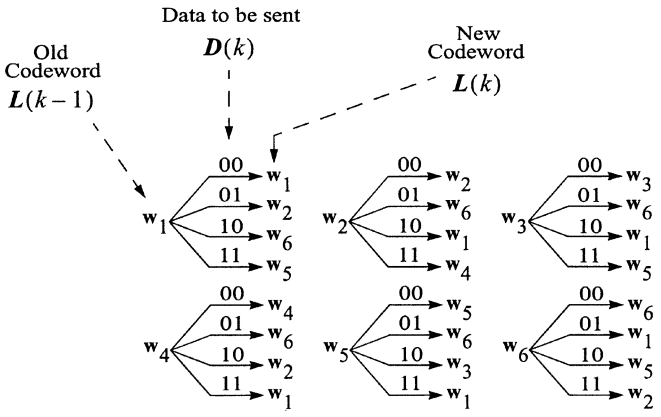

Fig. 7. Graph representation of F.

$$T = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

To indicate the correspondence between the input vectors and the 2^m ones in each row of T, we use the following symbolic matrix that contains all the information of the function F

$$\tilde{T} = \begin{bmatrix} (00) & (01) & 0 & 0 & (11) & (10) \\ (10) & (00) & 0 & (11) & 0 & (01) \\ (10) & 0 & (00) & 0 & (11) & (01) \\ (11) & (10) & 0 & (00) & 0 & (01) \\ (11) & 0 & (10) & 0 & (00) & (01) \\ (01) & (11) & 0 & 0 & (10) & (00) \end{bmatrix}.$$

In this example, *TPC favors transition patterns in which the voltages of neighboring bus lines change values in the same direction*. This reduces the *effective capacitance* between adjacent lines. For a typical 0.18-μm technology, with minimum distance between the wires, the value of λ is about 5. In this case, the coding schemes results in a theoretical energy reduction of about 22%, *even though the number of the bus lines has been increased*. This should not be surprising. In general, the scheme encodes the information $\boldsymbol{D}(k)$ into transitions among codewords that have an average energy cost less than that of the transitions between the original data (only two bits).

In the following sections, we address three issues of TPC: 1) exact calculation of the energy consumption; 2) design of coding schemes; and 3) complexity reduction.

B. Statistical Measures of Energy

Let $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_M$ be the codewords of the coding scheme and let $E(\boldsymbol{w}_i, \boldsymbol{w}_j)$ be the amount of energy *drawn form the power supply*[3] during the transition from codeword $\boldsymbol{w}_i$ to

[3]The energy estimation results that follow remain unchanged if we use the *dissipated energy* instead of the *energy drawn from the power supply*. This is due to the long-term averaging as $N \to \infty$ in definition (18).

codeword $\boldsymbol{w}_j$[4]. We define the *energy cost matrix* of the coding scheme as

$$\boldsymbol{E} = [E(\boldsymbol{w}_i, \boldsymbol{w}_j)]_{i,j=1}^{M}. \tag{17}$$

To quantify the energy behavior of the coding schemes we need to introduce a statistical energy measure that is both physically meaningful and can lead to relatively simple algebraic expressions. It is appropriate to define the *time average expected energy* per transition of the bus E_a as the limit of the average expected energy consumption over N consecutive transitions when N goes to infinity

$$E_a = \lim_{N\to\infty} \frac{1}{N} \cdot \sum_{k=1}^{N} \overline{E(\boldsymbol{L}(k-1), \boldsymbol{L}(k))}. \tag{18}$$

Throughout this paper, the overbar means statistical expectation with respect to all random variables involved in the expression, here, with respect to $\boldsymbol{L}(k-1)$ and $\boldsymbol{L}(k)$. In (18), the quantity

$$\frac{1}{N} \cdot \sum_{k=1}^{N} \overline{E(\boldsymbol{L}(k-1), \boldsymbol{L}(k))}$$

is the average expected energy drawn from V_{dd} during the first N transitions. The limit as $N \to \infty$ provides the time averaging.

C. Energy Consumption Using TPC

To derive the energy properties of TPC schemes, we need to make some assumptions on the statistics of the input data $\{\boldsymbol{D}(k) = (\boldsymbol{d}_1(k), \boldsymbol{d}_2(k), \ldots, \boldsymbol{d}_m(k))^T\}_k$. For presentation purposes, the analysis is done for input data sequences formed out of statistically *independent* and *uniformly distributed* vectors[5] in $\{0,1\}^m$. The result, (33), holds without any change for *independent* data vectors with *any distribution* in $\{0,1\}^m$. It can also be directly extended to the case of Markov data sequences, although this would increase the dimension of the problem significantly.

Suppose for now that $\{\boldsymbol{D}(k)\}_k$ is a sequence of independent and uniformly distributed vectors in $\{0,1\}^m$. Then, $\boldsymbol{L}(k), k = 1, 2, \ldots$ (Fig. 6) is a first-order homogeneous Markov [36] random sequence (see Remark 1 in Section III), and, the conditional probability of the transition $\boldsymbol{w}_i \to \boldsymbol{w}_j$ is given by

$$P_r(\boldsymbol{L}(k) = \boldsymbol{w}_j \,|\, \boldsymbol{L}(k-1) = \boldsymbol{w}_i) = \begin{cases} 2^{\frac{1}{m}} & \text{if, } \boldsymbol{w}_j \in X_{\boldsymbol{w}_i} \\ 0, & \text{otherwise.} \end{cases} \tag{19}$$

The *transition probability matrix* $\boldsymbol{P}$ of the states of the coding scheme is defined as

$$\boldsymbol{P}[P_{i,j}]_{i,j=1}^{M} = P_r(\boldsymbol{L}(k) = \boldsymbol{w}_j \,|\, \boldsymbol{L}(k-1) = \boldsymbol{w}_i)]_{i,j=1}^{M}. \tag{20}$$

From (19), we conclude that

$$\boldsymbol{P} = \boldsymbol{T}/2^m. \tag{21}$$

The probability (row) vector of the state $\boldsymbol{L}(k)$ is

$$\boldsymbol{p}(k) = [P_r(\boldsymbol{L}(k) = \boldsymbol{w}_1), \ldots, P_r(\boldsymbol{L}(k) = \boldsymbol{w}_M)]. \tag{22}$$

[4]Expression (6) can be used if the bus is modeled as in Fig. 1 or Fig. 3

[5]This is a typical case of data streams in telecommunication systems.

If $\boldsymbol{p}(0)$ is the probability distribution of the initial state $\boldsymbol{L}(0)$ of the bus, then, using the Chapman–Kolmogorov formula [36], that is, $\boldsymbol{p}(k) = \boldsymbol{p}(r) \cdot \boldsymbol{P}^{k-r}, k \geq r$, we have

$$\boldsymbol{p}(k) = \boldsymbol{p}(0) \cdot \boldsymbol{P}^k. \tag{23}$$

We use (23) to evaluate the time average expected energy of the general coding scheme of Fig. 6. We start by calculating the expected value of $E(\boldsymbol{L}(k-1), \boldsymbol{L}(k))$. Let $p_i(k)$ be the ith entrance of the probability vector $\boldsymbol{p}(k)$ and $P_{i,j}$ be the (i,j) element of the transition probability matrix $\boldsymbol{P}$. Then

$$\begin{aligned}
&\overline{E(\boldsymbol{L}(k-1), \boldsymbol{L}(k))} \\
&\quad = \sum_{i,j=1}^{M} P_r(\boldsymbol{L}(k) = \boldsymbol{w}_j, \boldsymbol{L}(k-1) = \boldsymbol{w}_i) \cdot E(\boldsymbol{w}_i, \boldsymbol{w}_j) \\
&\quad = \sum_{i,j=1}^{M} P_r(\boldsymbol{L}(k) = \boldsymbol{w}_j \,|\, \boldsymbol{L}(k-1) = \boldsymbol{w}_i) \\
&\qquad \cdot (P_r(\boldsymbol{L}(k-1)) = \boldsymbol{w}_i) \cdot E(\boldsymbol{w}_i, \boldsymbol{w}_j). \\
&\quad = \sum_{i,j=1}^{M} P_{i,j} \cdot p_i(k-1) \cdot E(\boldsymbol{w}_i, \boldsymbol{w}_j).
\end{aligned}$$

Let $\boldsymbol{A} \bullet \boldsymbol{B} = [a_{i,j} \cdot b_{i,j}]_{i,j}$ be the *Hadamard product* [37] of two matrices $\boldsymbol{A}, \boldsymbol{B}$ of the same dimensions and let $\underline{\boldsymbol{1}}$ be the (column) vector with all its coordinates equal to one. Then, we can write

$$\overline{E(\boldsymbol{L}(k-1), \boldsymbol{L}(k))} = p(k-1) \cdot (\boldsymbol{P} \bullet \boldsymbol{E}) \cdot \underline{\boldsymbol{1}} \tag{24}$$

and using (23)

$$\overline{E(\boldsymbol{L}(k-1), \boldsymbol{L}(k))} = p(0) \cdot \boldsymbol{P}^{k-1} \cdot (\boldsymbol{P} \bullet \boldsymbol{E}) \cdot \underline{\boldsymbol{1}}. \tag{25}$$

To continue with the evaluation of the time average expected energy, it is necessary to recall our assumption on the *strong connectivity* of the transition graph G_T. As mentioned before, the strong connectivity of G_T is equivalent to the *irreducibility* of the transition matrix $\boldsymbol{T}$, and the irreducibility of transition probability matrix $\boldsymbol{P}$, due to (21), [37]. Moreover, matrix $\boldsymbol{P}$ is row stochastic by its definition. All the above allow us to use the following modified version of the Perron–Frobenius theorem [37, corollary 8.4.6] .

Theorem 1 (Perron–Frobenius): An irreducible row-stochastic matrix $\boldsymbol{P}$ is similar to a matrix of the block-diagonal form

$$\boldsymbol{\Delta} = \begin{bmatrix} \boldsymbol{\Delta}_1 & 0 \\ 0 & \boldsymbol{Y} \end{bmatrix} \tag{26}$$

where $\boldsymbol{Y}$ is in Jordan form with eigenvalues of modulus less than one and $\boldsymbol{\Delta_1}$ is the diagonal matrix

$$\boldsymbol{\Delta}_1 = \operatorname{diag}\left(1, e^{\frac{2\pi i}{q}}, e^{\frac{4\pi i}{q}}, \ldots, e\frac{2(q-1)\pi i}{q}\right). \tag{27}$$

where q is the number of modulus-one eigenvalues of matrix $\boldsymbol{P}$ and it is always greater than or equal to one. □

Therefore, there exists a nonsingular matrix $\boldsymbol{A}$ and a matrix $\boldsymbol{\Delta}$ as in the theorem such that

$$\boldsymbol{P} = \boldsymbol{A} \cdot \boldsymbol{\Delta} \cdot \boldsymbol{A}^{-1}. \tag{28}$$

Moreover, from (25), it is

$$\overline{E(\boldsymbol{L}(k-1), \boldsymbol{L}(k))} = \boldsymbol{p}(0) \cdot \boldsymbol{A} \cdot \boldsymbol{\Delta}^{k-1} \cdot A^{-1} \cdot (\boldsymbol{P} \bullet \boldsymbol{E}) \cdot \underline{\boldsymbol{1}} \tag{29}$$

and hence

$$\frac{1}{N}\cdot\sum_{k=1}^{N}\overline{E(\boldsymbol{L}(k-1),\boldsymbol{L}(k))}=\boldsymbol{p}(0)\cdot\boldsymbol{A}\cdot\frac{1}{N}\cdot\left(\sum_{k=1}^{N}\begin{bmatrix}\boldsymbol{\Delta_1^{k-1}} & 0\\ 0 & \boldsymbol{Y}^{k-1}\end{bmatrix}\right)\cdot\boldsymbol{A}^{-1}\cdot(\boldsymbol{P}\bullet\boldsymbol{E})\cdot\underline{\mathbf{1}}. \quad (30)$$

Since the spectral radius of $\boldsymbol{Y}$ is less than one, we have $\boldsymbol{Y}^r\to 0$ as $r\to\infty$. Also, for every integer r, it holds that [38]

$$\lim_{N\to\infty}\frac{1}{N}\cdot\sum_{k=1}^{N}e^{\frac{r\pi i}{q}k}=\begin{cases}1, & \text{if r is a mulitiple of } q\\ 0, & \text{otherwise}\end{cases}. \quad (31)$$

From (30), we conclude that

$$\lim_{N\to\infty}\frac{1}{N}\cdot\sum_{k=1}^{N}\overline{E(\boldsymbol{L}(k-1),\boldsymbol{L}(k))}=\boldsymbol{p}(0)\cdot\boldsymbol{A}\cdot\boldsymbol{e}_1^T\cdot\boldsymbol{e}_1\cdot\boldsymbol{A}^{-1}\cdot(\boldsymbol{P}\bullet\boldsymbol{E})\cdot\underline{\mathbf{1}}. \quad (32)$$

Since $\boldsymbol{P}$ is a row-stochastic matrix, its right eigenvector corresponding to the eigenvalue 1 is $\underline{\mathbf{1}}$. Therefore, we can write $(\boldsymbol{A}\cdot\boldsymbol{e_1^T})\cdot(\boldsymbol{e_1}\cdot\boldsymbol{A}^{-1})=\underline{\mathbf{1}}\cdot\boldsymbol{b_0^T}$ where $\boldsymbol{b}_0$ is the left eigenvector of matrix $\boldsymbol{P}$ corresponding to eigenvalue 1, that is $\boldsymbol{b}_0^T\cdot\boldsymbol{P}=\boldsymbol{b}_0^T$, and satisfies $\boldsymbol{b}_0^T\cdot\underline{\mathbf{1}}=1$. Equality (32) can be written as

$$\begin{aligned}\lim_{N\to\infty}\frac{1}{N}\cdot\sum_{k=1}^{N}\overline{E(\boldsymbol{L}(k-1),\boldsymbol{L}(k))}&=\boldsymbol{p}(0)\cdot\boldsymbol{A}\cdot\boldsymbol{e}_1^T\cdot\boldsymbol{e}_1\cdot\boldsymbol{A}^{-1}\cdot(\boldsymbol{P}\bullet\boldsymbol{E})\cdot\underline{\mathbf{1}}\\&=\boldsymbol{p}(0)\cdot\underline{\mathbf{1}}\cdot\boldsymbol{b}_0^T\cdot(\boldsymbol{P}\bullet\boldsymbol{E})\cdot\underline{\mathbf{1}}\\&=\boldsymbol{b}_0^T\cdot(\boldsymbol{P}\bullet\boldsymbol{E})\cdot\underline{\mathbf{1}}\end{aligned} \quad (33)$$

since $\boldsymbol{p}(0)\cdot\underline{\mathbf{1}}=1$. Thus, for the case of i.i.d. and uniform input sequence the time average expected energy is

$$E_a=\boldsymbol{b}_0^T\cdot(\boldsymbol{P}\bullet\boldsymbol{E})\cdot\underline{\mathbf{1}}=\frac{1}{2^m}\cdot\boldsymbol{b}_0^T\cdot(\boldsymbol{T}\bullet\boldsymbol{E})\cdot\underline{\mathbf{1}}. \quad (34)$$

Now, recall that the transition matrix $\boldsymbol{T}$ has a one in the (i,j) position if and only if the transition $\boldsymbol{w}_i\to\boldsymbol{w}_j$ is allowed. There are exactly 2^m ones in every row of the matrix $\boldsymbol{T}$. *In the case of uniformly distributed input data, the time average expected energy is independent of the way that data vectors are mapped into the bus transitions.*

Now consider data sequences of independent but *not uniformly* distributed vectors. Let p_i^d be the probability that[6] $\boldsymbol{D}(k)=i, i=1,2,\ldots,2^m$. Then, every row of the $2^{m+a}\times 2^{m+a}$ transition matrix $\boldsymbol{P}$ is a permutation of $2^{m+a}-2^m$ zeros and the 2^m numbers $p_1^d,p_2^d,\ldots,p_{2^m}^d$.

The analysis presented above is still valid if $\boldsymbol{T}$ is replaced by $2^m\cdot\boldsymbol{P}$. We have

$$E_a=\boldsymbol{b}_0^T\cdot(\boldsymbol{P}\bullet\boldsymbol{E})\cdot\underline{\mathbf{1}}. \quad (35)$$

[6]The obvious identification is used, throughout the paper, between the binary vectors $(0,\ldots,0,0),(0,\ldots,0,1),(0,\ldots,1,0),\ldots,(1,\ldots,1,1)$ and the integers $1,2,3,\ldots,2^m$. This is done in order to use matrix algebra.

TABLE II
TRANSITION ENERGY OF THE 2-LINES EXAMPLES BUS

Old State \ New State	00	01	10	11
00	0	6	6	2
01	0	0	11	1
10	0	11	0	1
11	0	5	5	0

TABLE III
TRANSITION ENERGY OF THE THREE-LINES EXAMPLE BUS

Old State \ New State	000	001	010	011	100	101	110	111
000	0	6	11	7	6	12	7	3
001	0	0	16	6	6	6	12	2
010	0	11	0	1	11	22	1	2
011	0	5	5	0	11	16	6	1
100	0	6	16	12	0	6	6	2
101	0	0	21	11	0	0	11	1
110	0	11	5	6	5	16	0	1
111	0	5	10	5	5	10	5	0

D. Energy-Estimation Examples

The results of the previous section are applied to the example of Section III-1. Consider again a two-line and a three-line buses of structure as in Fig. 4 and $\lambda=5$. For simplicity, we set $C_L=1$ and $V_{\mathrm{dd}}=1$. The transition energies of the 2-line bus are calculated using (4), and shown in Table II.

In the example of Section III-1, two bits of information are transmitted each time. Without using coding and assuming the data is a sequence of independent and uniformly distributed vectors, the expected energy per transmission (or transition) is

$$\begin{aligned}E_a&=\sum_{i,j=1}^{4}P_r(\boldsymbol{L}(k)=\boldsymbol{w}_j,\boldsymbol{L}(k-1)=\boldsymbol{w}_j)\cdot E(\boldsymbol{w}_i,\boldsymbol{w}_j)\\&=\sum_{i,j=1}^{4}P_r(\boldsymbol{L}(k)=\boldsymbol{w}_j)\cdot P_r(\boldsymbol{L}(k-1)=\boldsymbol{w}_i)\cdot E(\boldsymbol{w}_i,\boldsymbol{w}_j)\\&=\left[\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4}\right]\cdot\begin{bmatrix}0&6&6&2\\0&0&11&1\\0&11&0&1\\0&5&5&0\end{bmatrix}\cdot\begin{bmatrix}1/4\\1/4\\1/4\\1/4\end{bmatrix}=3.\end{aligned} \quad (36)$$

Suppose now that we apply the coding scheme presented in Fig. 7. For the three-line bus with $C_L=1, V_{\mathrm{dd}}=1$ and $\lambda=5$, the transition energies are given in Table III.

The codewords used in the example are $\boldsymbol{w}_1=000,\boldsymbol{w}_2=001,\boldsymbol{w}_3=010,\boldsymbol{w}_4=101,\boldsymbol{w}=110$, and $\boldsymbol{w}_6=111$. The *energy-cost matrix* of the scheme, defined by (17), is

$$\boldsymbol{E}=\begin{bmatrix}0&6&11&12&7&3\\0&0&16&6&12&2\\0&11&0&22&1&2\\0&0&21&0&11&1\\0&11&5&16&0&1\\0&5&10&10&5&0\end{bmatrix}. \quad (37)$$

TABLE IV
DISTRIBUTION OF THE INPUT VECTORS

$\boldsymbol{u}$:	00	01	10	11
$P_r(\boldsymbol{D}(k) = \boldsymbol{u})$:	0.1	0.3	0.4	0.2

The transition probability matrix is

$$\boldsymbol{P} = \begin{bmatrix} 1/4 & 1/4 & 0 & 0 & 1/4 & 1/4 \\ 1/4 & 1/4 & 0 & 1/4 & 0 & 1/4 \\ 1/4 & 0 & 1/4 & 0 & 1/4 & 1/4 \\ 1/4 & 1/4 & 0 & 1/4 & 0 & 1/4 \\ 1/4 & 0 & 1/4 & 0 & 1/4 & 1/4 \\ 1/4 & 1/4 & 0 & 0 & 1/4 & 1/4 \end{bmatrix}. \quad (38)$$

The left eigenvector $\boldsymbol{b}_0^T$ of $\boldsymbol{P}$ corresponding to eigenvalue 1 and satisfying the equation $\boldsymbol{b}_0^T \cdot \underline{\mathbf{1}} = 1$ is

$$\boldsymbol{b}_0^T = [0.250, 0.187, 0.062, 0.062, 0.187, 0.250]. \quad (39)$$

Thus, the time average expected energy of the coding scheme, given by (34), is

$$E_a = \boldsymbol{b}_0^T \cdot (\boldsymbol{P} \bullet \boldsymbol{E}) \cdot \underline{\mathbf{1}} = 2.34.$$

Therefore, using this coding scheme, we can achieve energy reductions up to

$$100 \cdot \left(\frac{3 - 2.34}{3} \right) \% = 21.9\%.$$

Now, suppose that the data are stationary but not uniformly distributed. Suppose that for every clock cycle k, the random vector $\boldsymbol{D}(k)$ has the probability distribution shown in Table IV.

If the data is transmitted uncoded through the two-line bus, the time average expected energy is [see (36)]

$$E_a = [0.1, 0.3, 0.4, 0.2] \cdot \begin{bmatrix} 0 & 6 & 6 & 2 \\ 0 & 0 & 11 & 1 \\ 0 & 11 & 0 & 1 \\ 0 & 5 & 5 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0.1 \\ 0.3 \\ 0.4 \\ 0.2 \end{bmatrix} = 3.94. \quad (40)$$

If we use the coding scheme of Fig. 7, the transition probability matrix $\boldsymbol{P}$ is given by

$$\boldsymbol{P} = \begin{bmatrix} 0.1 & 0.3 & 0 & 0 & 0.2 & 0.4 \\ 0.4 & 0.1 & 0 & 0.2 & 0 & 0.3 \\ 0.4 & 0 & 0.1 & 0 & 0.2 & 0.3 \\ 0.2 & 0.4 & 0 & 0.1 & 0 & 0.3 \\ 0.2 & 0 & 0.4 & 0 & 0.1 & 0.3 \\ 0.3 & 0.2 & 0 & 0 & 0.4 & 0.1 \end{bmatrix}. \quad (41)$$

The left eigenvector $\boldsymbol{b}_0^T$ of $\boldsymbol{P}$ corresponding to eigenvalue of one and satisfying the equation $\boldsymbol{b}_0^T \cdot \underline{\mathbf{1}} = 1$ is $\boldsymbol{b}_0^T = [0.251, 0.160, 0.087, 0.036, 0.196, 0.271]$. The time average expected energy of the coding scheme, given by (35), is $E_a = \boldsymbol{b}_0^T \cdot (\boldsymbol{P} \cdot \boldsymbol{E}) \cdot \underline{\mathbf{1}} = 2.74$. Therefore, using the coding scheme, we can achieve energy reductions up to

$$100 \cdot \left(\frac{3.94 - 2.74}{3.94} \right) \% = 30.6\%.$$

IV. TWO ALGORITHMS FOR DERIVING EFFICIENT CODING SCHEMES

Suppose the original bus has m lines and at every clock cycle k the input random vector $\boldsymbol{D}(k)$ takes the value $(0 \ldots 00) \equiv 1, (0 \ldots 01) \equiv 2, (0 \ldots 10) \equiv 3, \ldots, (1 \ldots 11) = 2^m$ with probability $p_1^d, p_2^d, p_3^d, \ldots, p_{2^m}^d$, respectively. We expand the bus by a number a additional lines; so, we have $m+a$ bits to encode the data.

Throughout this section, we assume that all vectors in $\{0,1\}^{m+a}$ are possible states of the expanded bus. In other words, $M = 2^{m+a}$ and $W = \{0,1\}^{m+a}$. Without loss of generality, we can set $\boldsymbol{w}_1 = (0 \ldots 00), \boldsymbol{w}_2 = (0 \ldots 01), \ldots, \boldsymbol{w}_{2^{m+a}} = (1 \ldots 11)$. The transition energy cost matrix $\boldsymbol{E}$ is defined as in (17).

Definition: Given m, a and the distribution $\boldsymbol{p}^d = [p_1^d, p_2^d, \ldots, p_{2^m}^d]$ we define the set $\Pi(m, a, \boldsymbol{p}^d)$ of all $2^{m+a} \times 2^{m+a}$ stochastic matrices $\boldsymbol{P}$ such that every row of $\boldsymbol{P}$ is a permutation of $2^{m+a} - 2^m$ zeros and the 2^m numbers $p_1^d, p_2^d, \ldots, p_{2^m}^d$.

Example: Let $m = 2, a = 1$ and $\boldsymbol{p}^d = [2,1,2,5]/10$. The following three matrices are members of the set $\Pi(m, a, \boldsymbol{p}^d)$:

$$\boldsymbol{P}_1 = \frac{1}{10} \begin{bmatrix} \mathbf{5} & \mathbf{2} & 0 & 0 & \mathbf{1} & 0 & 0 & \mathbf{2} \\ \mathbf{5} & \mathbf{2} & 0 & \mathbf{1} & 0 & 0 & 0 & \mathbf{2} \\ \mathbf{5} & 0 & \mathbf{2} & \mathbf{2} & 0 & 0 & \mathbf{1} & 0 \\ \mathbf{5} & \mathbf{1} & 0 & \mathbf{2} & 0 & 0 & 0 & \mathbf{2} \\ \mathbf{5} & \mathbf{1} & 0 & 0 & \mathbf{2} & 0 & 0 & \mathbf{2} \\ \mathbf{5} & \mathbf{2} & 0 & 0 & \mathbf{2} & \mathbf{1} & 0 & 0 \\ \mathbf{5} & 0 & \mathbf{1} & 0 & 0 & 0 & \mathbf{2} & \mathbf{2} \\ \mathbf{5} & \mathbf{2} & 0 & \mathbf{1} & 0 & 0 & 0 & \mathbf{2} \end{bmatrix}$$

$$\boldsymbol{P}_2 = \frac{1}{10} \begin{bmatrix} \mathbf{5} & \mathbf{2} & 0 & 0 & \mathbf{1} & 0 & 0 & \mathbf{2} \\ \mathbf{2} & \mathbf{5} & 0 & 0 & 0 & \mathbf{1} & 0 & \mathbf{2} \\ \mathbf{1} & 0 & \mathbf{5} & \mathbf{2} & 0 & 0 & \mathbf{2} & 0 \\ \mathbf{2} & 0 & \mathbf{1} & \mathbf{5} & 0 & 0 & 0 & \mathbf{2} \\ \mathbf{2} & 0 & 0 & 0 & \mathbf{5} & \mathbf{1} & 0 & \mathbf{2} \\ \mathbf{1} & \mathbf{2} & 0 & 0 & \mathbf{2} & \mathbf{5} & 0 & 0 \\ \mathbf{2} & 0 & \mathbf{1} & 0 & 0 & 0 & \mathbf{5} & \mathbf{2} \\ \mathbf{2} & 0 & 0 & \mathbf{2} & 0 & 0 & \mathbf{1} & \mathbf{5} \end{bmatrix}$$

$$\boldsymbol{P}_3 = \frac{1}{10} \begin{bmatrix} \mathbf{5} & \mathbf{2} & 0 & 0 & \mathbf{1} & 0 & 0 & \mathbf{2} \\ \mathbf{2} & \mathbf{5} & 0 & 0 & 0 & \mathbf{1} & 0 & \mathbf{2} \\ \mathbf{1} & 0 & \mathbf{5} & \mathbf{2} & 0 & 0 & \mathbf{2} & 0 \\ \mathbf{2} & 0 & \mathbf{1} & \mathbf{5} & 0 & 0 & 0 & \mathbf{2} \\ \mathbf{2} & 0 & 0 & 0 & \mathbf{5} & \mathbf{1} & 0 & \mathbf{2} \\ 0 & \mathbf{2} & 0 & 0 & \mathbf{2} & \mathbf{5} & 0 & \mathbf{1} \\ \mathbf{2} & 0 & \mathbf{1} & 0 & 0 & 0 & \mathbf{5} & \mathbf{2} \\ \mathbf{2} & 0 & 0 & \mathbf{2} & 0 & 0 & \mathbf{1} & \mathbf{5} \end{bmatrix}. \quad (42)$$

Following the discussion in the previous sections, the design of an efficient TPC scheme with the parameters $m, a, \boldsymbol{p}^d$ is equivalent to choosing an appropriate matrix P from the set $\Pi(m, a, \boldsymbol{p}^d)$. Note that $\boldsymbol{P}$ leads directly to the derivation of the time-average expected energy E_a, using (35), that is $E_a = \boldsymbol{b}_0^T \cdot (\boldsymbol{P} \bullet \boldsymbol{E}) \cdot \underline{\mathbf{1}}$ where $\boldsymbol{b}_0$ is the left probability eigenvector of $\boldsymbol{P}$.

A. Approximately Optimal Coding (AOC)

Motivated by (35), where matrix $\boldsymbol{P}$ is involved in the point-wise product with the energy cost matrix $\boldsymbol{E}$, we may attempt achieving a low E_a by minimizing $\underline{\mathbf{1}}^T \cdot (\boldsymbol{P} \bullet \boldsymbol{E}) \cdot \underline{\mathbf{1}}$

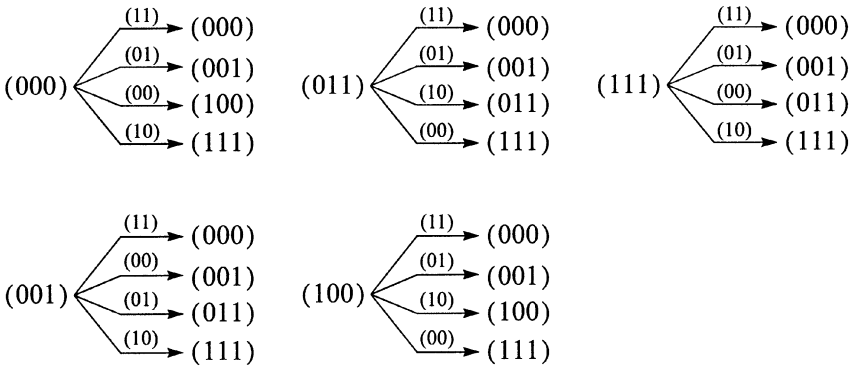

Fig. 8. Example transition diagram.

instead of $\boldsymbol{b}_0^T \cdot (\boldsymbol{P} \bullet \boldsymbol{E}) \cdot \underline{\mathbf{1}}$. Note that this can be done by simply choosing $\boldsymbol{P}$ so that *every rowsum* of $\boldsymbol{P} \bullet \boldsymbol{E}$ is minimal.

Example: Let us use this approach to derive a scheme that encodes a 2-bit sequence with probability distribution $\boldsymbol{p}^d = [2,1,2,5]/10$, into a 3-bit sequence. Let us consider a bus as in Fig. 4 with $C_L = 1, V_{\rm dd} = 1$ and $\lambda = 5$. Using (9) and definition (17), we have

$$\boldsymbol{E} = \begin{bmatrix} 0 & 6 & 11 & 7 & 6 & 12 & 7 & 3 \\ 0 & 0 & 16 & 6 & 6 & 6 & 12 & 2 \\ 0 & 11 & 0 & 1 & 11 & 22 & 1 & 2 \\ 0 & 5 & 5 & 0 & 11 & 16 & 6 & 1 \\ 0 & 6 & 16 & 12 & 0 & 6 & 6 & 2 \\ 0 & 0 & 21 & 11 & 0 & 0 & 11 & 1 \\ 0 & 11 & 5 & 6 & 5 & 16 & 0 & 1 \\ 0 & 5 & 10 & 5 & 5 & 10 & 5 & 0 \end{bmatrix}. \tag{43}$$

We examine the first row, [0, 6, 11, 7, 6, 12, 7, 3] of $\boldsymbol{E}$. An optimal choice for the first row of the stochastic matrix $\boldsymbol{P}$ is $[5,1,0,0,2,0,0,2]/10$. Note that the maximal entry of the probability vector, $\boldsymbol{p}^d = [2,1,2,5]/10$, is matched with the minimal entry of the first row of $\boldsymbol{E}$. The second maximal entry of $\boldsymbol{p}^d$ is matched with the second minimal entry of the first row of $\boldsymbol{E}$, etc. This procedure requires only *sorting* the entries of the rows; and, it guarantees the minimality of every rowsum of $\boldsymbol{P} \bullet \boldsymbol{E}$, and therefore the minimality of $\underline{\mathbf{1}}^T \cdot (\boldsymbol{P} \bullet \boldsymbol{E}) \cdot \underline{\mathbf{1}}$ in $\Pi(m, a, \boldsymbol{p}^d)$. A stochastic matrix achieving the minimum in the example is

$$\boldsymbol{P}^* = \frac{1}{10} \begin{bmatrix} 5 & 1 & 0 & 0 & 2 & 0 & 0 & 2 \\ 5 & 2 & 0 & 1 & 0 & 0 & 0 & 2 \\ 5 & 0 & 2 & 2 & 0 & 0 & 1 & 0 \\ 5 & 1 & 0 & 2 & 0 & 0 & 0 & 2 \\ 5 & 1 & 0 & 0 & 2 & 0 & 0 & 2 \\ 5 & 2 & 0 & 0 & 2 & 1 & 0 & 0 \\ 5 & 0 & 1 & 0 & 0 & 0 & 2 & 2 \\ 5 & 1 & 0 & 2 & 0 & 0 & 0 & 2 \end{bmatrix}. \tag{44}$$

The product $(\boldsymbol{P}^* \bullet \boldsymbol{E}) \cdot \underline{\mathbf{1}}$ is

$$(\boldsymbol{P}^* \bullet \boldsymbol{E}) \cdot \underline{\mathbf{1}} = \frac{1}{10} \begin{bmatrix} 24 \\ 10 \\ 3 \\ 7 \\ 10 \\ 0 \\ 7 \\ 15 \end{bmatrix}. \tag{45}$$

Using matrix $\boldsymbol{P}^*$, we can calculate the time average expected energy of the coding scheme. The left probability eigenvector of $\boldsymbol{P}^*$ is $(\boldsymbol{b}_o^*)^T = [0.500, 0.111, 0, 0.064, 0.125, 0, 0, 0.200]$; and, the time average expected energy is $E_a = \boldsymbol{b}_0^{*T} \cdot (\boldsymbol{P}^* \bullet \boldsymbol{E}) \cdot \underline{\mathbf{1}} = 1.78$. Note that vector $\boldsymbol{b}_o$ has only five nonzero elements. This means that the stochastic matrix $\boldsymbol{P}^*$ corresponds to TPC schemes with only five codewords, $(000) \equiv 1, (001) \equiv 2, (011) \equiv 4, (100) \equiv 5$, and $(111) \equiv 8$. Note also that since the input vectors (00) and (10) both have probability 0.2, the optimal assignment of the inputs to the transitions is not unique. For example, the transitions from state 1 to states 5 and 8 can be assigned to inputs (00) and (10), respectively, or vice versa. In Fig. 8, we see a transition diagram based on $\boldsymbol{P}^*$ with an optimal "input vectors to transitions" assignment.

The expected energy consumption when transmitting the data sequence though a two-line bus uncoded is [see (36)]

$$E_a = [0.2, 0.1, 0.2, 0.5] \cdot \begin{bmatrix} 0 & 6 & 6 & 2 \\ 0 & 0 & 11 & 1 \\ 0 & 11 & 0 & 1 \\ 0 & 5 & 5 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0.2 \\ 0.1 \\ 0.2 \\ 0.5 \end{bmatrix} = 1.90 \tag{46}$$

The energy reduction of the coding scheme achieved using this simple minimization process is

$$100 \cdot \left(\frac{1.90 - 1.78}{1.90} \right) \% = 6.3\%.$$

The energy reduction is small because we ignored the steady-state probability vector $\boldsymbol{b}_0$ during the optimization process. For the same setup, the TPC algorithm presented in the following section generates coding schemes demonstrating possible energy reductions up to 26.5%.

Although AOC may not result in efficient coding schemes for small m, it generates very efficient coding schemes for $m \geq 7$ which is demonstrated in the following section. This efficiency is due to the tendency of the eigenvector $\boldsymbol{b}_0$ to become (more) uniform for large m (or $m + a$). Another desirable property of AOC is its low complexity.

B. Transition Pattern Coding (TPC) Algorithm

The objective of the TPC algorithm is to generate TPC schemes with minimum associated *time average expected*

0: Initialization: $k := 0$, $\xi(0) := \underline{0}_{M\times 1}$, $\boldsymbol{P}^*(0) := [0]_{M\times M}$

1: *Repeat*

2: $k := k + 1$

3: $\boldsymbol{P}^*(k) := \underset{\boldsymbol{P}\in\Pi(m,a,\boldsymbol{p}^d)}{\operatorname{argmin}} [(\boldsymbol{E} + \underline{1}\cdot\xi^T(k-1))\bullet\boldsymbol{P}]\cdot\underline{1}$

4: $\xi(k) := [(\boldsymbol{E} + \underline{1}\cdot\xi^T(k-1))\bullet\boldsymbol{P}^*(k)]\cdot\underline{1}$

5: *Until* $\boldsymbol{P}^*(k) = \boldsymbol{P}^*(k-1)$

Fig. 9. TPC algorithm.

energy, E_a. To do so, the algorithm solves (approximately) the minimization problem

$$\min_{\boldsymbol{P}\in\Pi(m,a,\boldsymbol{p}^d)} \boldsymbol{b}_0^T\cdot(\boldsymbol{P}\bullet\boldsymbol{E})\cdot\underline{1}. \tag{47}$$

By definition, $\boldsymbol{b}_0$ is the left probability eigenvector of $\boldsymbol{P}$. We assume that $\boldsymbol{P}$ is strongly connected; so, all 2^{m+a} codewords are used and $M = 2^{m+a}$. If this is not the case, then, we keep the largest strongly connected component of $\boldsymbol{P}$ and reduce the set of codewords appropriately. Finally, for all $k = 1,2,\ldots,\xi(k) \in [0,\infty)^M$ and $\boldsymbol{P}^*(k) \in \Pi(m,a,\boldsymbol{p}^d)$ we obtain the TPC algorithm shown in Fig. 9.

Note that in the minimization step of the algorithm, each row of $\boldsymbol{P}^*(k)$ is chosen so that the corresponding entry of the vector $[(\boldsymbol{E}+\underline{1}\cdot\xi^T(k-1))\bullet\boldsymbol{P}]\cdot\underline{1}$ is minimal. In this sense, M minimization problems are solved simultaneously and independently.

The algorithm operates "backward" while the variable k counts the backward steps. On step k, the value of $\xi_i(k)$ is the expected energy cost of a (backward) random path of length k, which starts from state i and evolves (backward) with transition probability matrices $\boldsymbol{P}^*(k),\boldsymbol{P}^*(k-1),\ldots,\boldsymbol{P}^*(1)$, as shown in Fig. 10.

For $k = 0, \xi_i(0) = 0$ for every i, since no steps have been encountered. For $k = 1$, the stochastic matrix $\boldsymbol{P}^*(1)$ is chosen within $\Pi(m,a,\boldsymbol{p}^d)$ so that the expected cost of the one step (backward) random path starting from any state $i, i = 1,2,\ldots,M$, is minimal. Since $\xi(0) = 0$, the choice of $\boldsymbol{P}^*(1)$ minimizes every entry of the vector $(\boldsymbol{E}\bullet\boldsymbol{P})\cdot\underline{1}$. On the kth iteration, matrix $\boldsymbol{P}^*(k)$ is chosen to minimize all entries of $[(\boldsymbol{E}+\underline{1}\cdot\xi^T(k-1))\bullet\boldsymbol{P}^*(k)]\cdot\underline{1} = (\boldsymbol{E}\bullet\boldsymbol{P})\cdot\underline{1}+\boldsymbol{P}^*(k)\cdot\xi(k-1)$. We have that

$$\begin{aligned}\boldsymbol{e}_i^T\cdot[(\boldsymbol{E}+\underline{1}\cdot\xi^T)\bullet\boldsymbol{P}^*(k)]\cdot\underline{1}&\\ = \boldsymbol{e}_i^T\cdot(\boldsymbol{E}\bullet\boldsymbol{P}^*(k))\cdot\underline{1} + \boldsymbol{e}_i^T\cdot\boldsymbol{P}^*(k)\cdot\xi(k-1)&\\ = \sum_{j=1}^{M} p_{i,j}^*(k)\cdot\boldsymbol{E}(i,j) + \sum_{j=1}^{M} p_{i,j}^*(k)\cdot\xi_j(k-1).&\end{aligned} \tag{48}$$

The ith entry of $(\boldsymbol{E}\bullet\boldsymbol{P}^*(k))\cdot\underline{1}$ is the expected cost of the backward step from k to $k-1$, starting from state i. The ith entry of $\boldsymbol{P}^*(k)\cdot\xi(k-1)$ is the expected cost of the (backward) random path from $k-1$ to 0. The algorithm stops when the new transition probability matrix equals the previous one. In practice, this condition is reached after a few iterations.

Example: Let $m = 2, a = 1$ and $\boldsymbol{p}^d = [2,1,2,5]/10$, as in the example in the beginning of Section IV. Again, we assume a bus as in Fig. 4 and $\lambda = 5$. The *TPC* algorithm terminates in three steps. For $k = 1,2,3$ it gives the stochastic matrices $\boldsymbol{P}_1^*,\boldsymbol{P}_2^*,\boldsymbol{P}_3^*$ shown in (42). These stochastic matrices result in time average expected energies 1.7854, 1.3981, 1.3962, respectively. For $k = 4, \boldsymbol{P}_4^* = \boldsymbol{P}_3^*$.

Example: Now, let us assume that $m = 8, a = 1$ and a bus, as in Fig. 4 and $\lambda = 5$. The algorithm is run for the following three distributions on the left of Fig. 11, that is *uniform, triangular* and *truncated Gaussian*. The energy savings resulting from using the stochastic matrices $\boldsymbol{P}^*(k), k = 1,2,\ldots$ of the algorithm, are shown on the right. The first step, $k = 1$, corresponds to *AOC*; and, as we see, gives a performance very close to the that of the final stochastic matrix. As it is mentioned before, this is generally the case for $m \geq 7$. For the three input distributions, the algorithm requires 7, 10, and 11 iterations to stop.

C. Results of the TPC Algorithm

The transition pattern coding algorithm was used to derive coding schemes for various combinations of values of the quadruple (m, a, λ, *Input distribution*) assuming the bus model of Fig. 4. In Fig. 12, we see the energy reductions (with respect to the uncoded bus) achievable using the coding schemes derived for $m = 2,4,8, a = 1, 0 \leq \lambda \leq 10$ and with three input distributions *uniform, triangular*, and *truncated Gaussian* of normalized variance: $\sigma/\bar{D} = 0.5$. Each star $(^*)$ in the graphs indicates the energy reduction of a coding scheme for a quadruple (m, a, λ, *Input distribution*). *The energy reduction can be arbitrarily high* and depends strongly on the distribution of the data. For example, the graph in the upper right corner shows reductions up to 80%.

In Fig. 12, we see results attributed to coding schemes with $a = 1$, i.e., with only one additional line. It is true that in most cases larger a (additional lines) results in higher energy reductions and larger additional occupied chip area. In Fig. 13, we see the energy reduction of schemes resulting from the *TPC* algorithm for $m = 4, a = 1,2,3, 0 \leq \lambda \leq 10$ and uniform input distribution. The savings are compared with that of bus invert [8].

D. Reducing Complexity

A way to reduce complexity is to split the input data bits into groups and encode each group independently. The approach is shown in Fig. 14.

The energy dissipation associated with the partitioned coding scheme equals the sum of the energy dissipations of the individual blocks *plus* the energy losses due to interactions at the boundaries of adjacent blocks (assuming a bus structure as that in Fig. 4). These interactions take place between the last line of the first block and the first line of the second block, the last line of the second block and the first line of the third, etc. The calculation of the expected energy dissipation caused by the interactions is presented next.

E. The Interaction Energy

The energy loss caused by the interaction of two consecutive blocks corresponds to the "λ's" of Table I when V_1, V_2 are

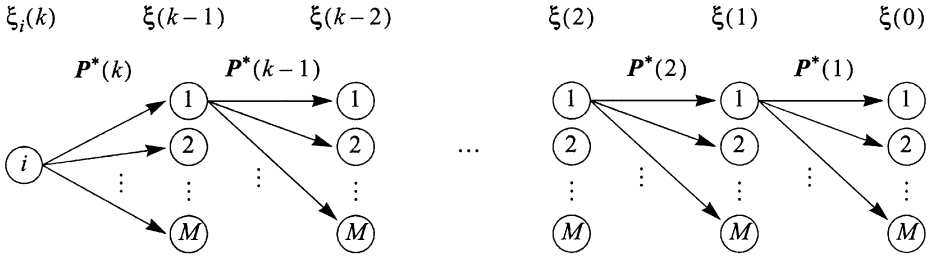

Fig. 10. Expected costs of forward paths using the stochastic matrices derived by the optimization process.

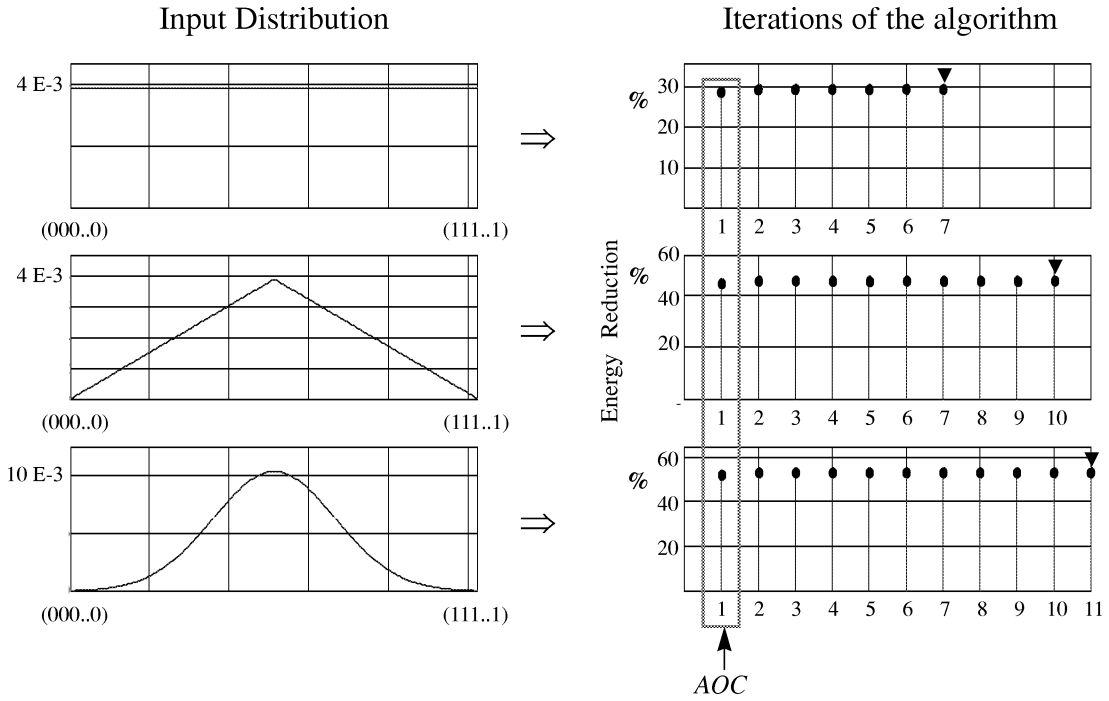

Fig. 11. Energy savings corresponding to the stochastic matrix derived in each iteration of the *TPC* algorithm for different input distributions.

the voltages of the adjacent boundary lines of the two blocks. For simplicity of presentation, all coding blocks are assumed to be identical with parameters m, a, W, M and functions F and G as defined in Section III. Let B and B' be two adjacent coding blocks of the partitioned scheme in Fig. 14. Moreover, suppose that B is "higher" than B', e.g., B and B' are the first and second blocks, respectively, in the figure. Now, let $\boldsymbol{L}(k)$ and $\boldsymbol{L}'(k)$ be the outputs of blocks B and B' at time $k, l(k)$ be the *last* (bottom) bit of $\boldsymbol{L}(k)$, and $l'(k)$ be the *first* (top) bit of $\boldsymbol{L}'(k)$. Thus, $l(k)$ and $l'(k)$ correspond to adjacent lines in the bus. The situation is illustrated in the example of Fig. 15.

From Table I, we extract Table V, which presents the *interaction energy* drawn from the power supply during the kth transition. For convenience, we have set $C_L = 1$ and $V_{\rm dd} = 1$; thus, all quantities are normalized.[7] This amount of energy drawn is due to the inter-line capacitance. Call this (normalized) energy cost $J(k)$.

To proceed in the analysis of the energy behavior of the partitioned coding scheme we have to make some assumptions on the statistics of the input data that allow for a relatively simple presentation of the methodology. We assume that the *partial data vectors* fed into the n encoders, Fig. 14, are statistically independent and identical.[8] The results can be extended to include the case of Markov sources as well.

According to the assumption, the random variables $l(k)$ and $l'(k)$ are independent (because the partial data vectors $\boldsymbol{L}(k)$ and $\boldsymbol{L}'(k)$ are independent) and the expected value of $J(k)$ can be derived using Table V.

$$\begin{aligned}\overline{J(k)} &= \lambda \cdot P_r(l(k)=0, l'(k)=0, l(k+1)=0, l'(k+1)=1)\\ &\quad + \lambda \cdot P_r(l(k)=0, l'(k)=0, l(k+1)=1, l'(k+1)=0)\\ &\quad + 2\lambda \cdot P_r(l(k)=0, l'(k)=1, l(k+1)=1, l'(k+1)=0)\\ &\quad + 2\lambda \cdot P_r(l(k)=1, l'(k)=0, l(k+1)=0,\ l'(k+1)=1)\\ &\quad + \lambda \cdot P_r(l(k)=1, l'(k)=1, l(k+1)=0, l'(k+1)=1)\\ &\quad + \lambda \cdot P_r(l(k)=1, l'(k)=1, l(k+1)=1, l'(k+1)=0).\end{aligned} \tag{49}$$

By grouping one-half of the fourth term with each of the first and the fifth terms, one-half of the third term with each of the second and the sixth terms, and taking into account that the two-bit

[7] To get the real values we have to multiply them by $V_{\rm dd}^2 C_L$.

[8] Spatially and temporally.

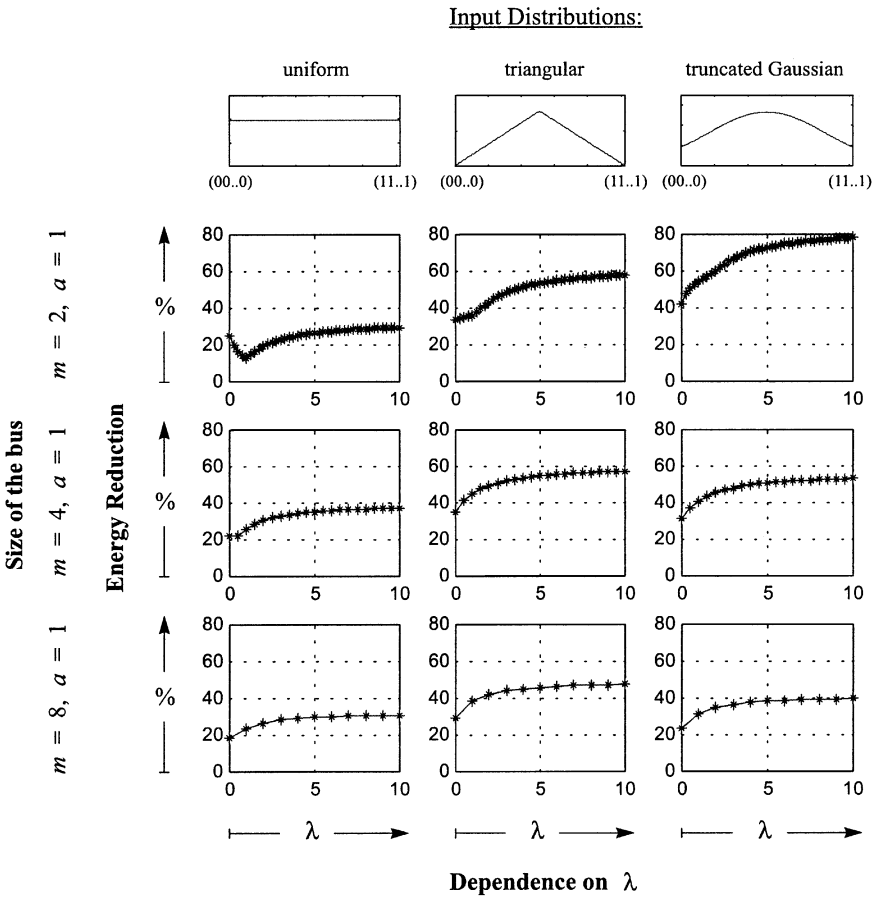

Fig. 12. Energy reduction achievable by coding schemes resulting from the TPC algorithm for $m = 2, 4, 8, a = 1, 0 \leq \lambda \leq 10$ and three input distributions.

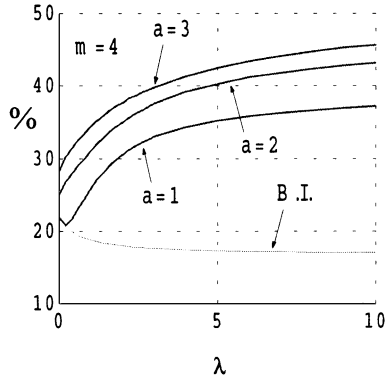

Fig. 13. Energy saving achievable by TPC schemes and the bus invert scheme for uniformly distributed input data, $m = 4, a = 1, 2, 3$ and $0 \leq \lambda \leq 10$.

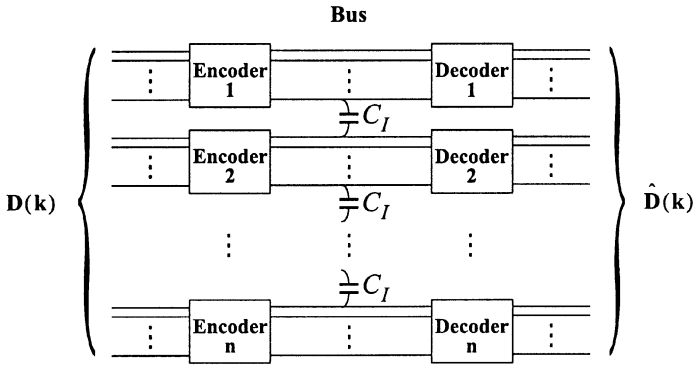

Fig. 14. Partitioned coding scheme: a juxtaposition of simpler blocks.

sequences $\{l(k)\}_k$ and $\{l'(k)\}_k$ are independent (to each other), (49) can be written as

$$\begin{aligned}\overline{J(k)} = {} & \lambda \cdot P_r(l(k+1)=0) \cdot P_r(l'(k)=0, l'(k+1)=1) \\ & + \lambda \cdot P_r(l'(k+1)=1) \cdot P_r(l(k)=1, l(k+1)=0) \\ & + \lambda \cdot P_r(l'(k+1)=0) \cdot P_r(l(k)=0, l(k+1)=1) \\ & + \lambda \cdot P_r(l(k+1)=1) \cdot P_r(l'(k)=1, l'(k+1)=0). \end{aligned} \tag{50}$$

Let W be the set of codewords of every coding block in the bus; thus, $\boldsymbol{L}(k), \boldsymbol{L}'(k) \in W$ for every k. Let W_{0*} and W_{1*} be the

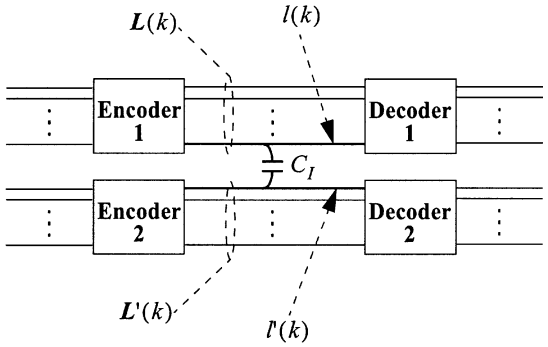

Fig. 15. Example of $B, B', \boldsymbol{L}(k), \boldsymbol{L}'(k), l(k)$ and $l'(k)$.

TABLE V
ENERGY DRAW FROM THE POWER SUPPLY BECAUSE OF THE INTERACTION OF ADJACENT LINES $l(k)$ AND $l'(k)$

$J(k)$		$[l(k+1), l'(k+1)]$			
		00	01	10	11
$[l(k), l'(k)]$	00	0	λ	λ	0
	01	0	0	2λ	0
	10	0	2λ	0	0
	11	0	λ	λ	0

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. See the example that follows.

Since $l(k)$ is the last bit of $\boldsymbol{L}(k)$ and $l'(k)$ is the first bit of $\boldsymbol{L}'(k)$, for α, β in $\{0,1\}$

$$\begin{aligned} &P_r\Big(l(k)=\alpha, l(k+1)=\beta\Big) \\ &= \sum_{\boldsymbol{\nu}\in W_{*\alpha}, \boldsymbol{w}\in W_{*\beta}} P_r\Big(\boldsymbol{L}(k)=\boldsymbol{\nu}, \boldsymbol{L}(k+1)=\boldsymbol{w}\Big) \\ &= \sum_{\boldsymbol{\nu}\in W_{*\alpha}, \boldsymbol{w}\in W_{*\beta}} P_r\Big(\boldsymbol{L}(k+1)=\boldsymbol{w} \,|\, \boldsymbol{L}(k)=\boldsymbol{\nu}\Big)\cdot P_r\Big(\boldsymbol{L}(k)=\boldsymbol{\nu}\Big). \end{aligned} \tag{51}$$

Similarly

$$\begin{aligned} &P_r\Big(l'(k)=\alpha, l'(k+1)=\beta\Big) \\ &= \sum_{\boldsymbol{\nu}\in W_{\alpha*}, \boldsymbol{w}\in W_{\beta*}} P_r\Big(\boldsymbol{L}'(k+1)=\boldsymbol{w} \,|\, \boldsymbol{L}'(k)=\boldsymbol{\nu}\Big)\cdot P_r\Big(\boldsymbol{L}'(k)=\boldsymbol{\nu}\Big). \end{aligned} \tag{52}$$

To get a compact of $\overline{J(k)}$, we need a few more definitions. For $\alpha = 0,1$ and $i = i, \ldots, M$

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

According to the assumptions on the data and Remark 1 of Section III, $\boldsymbol{L}(k), k = 1,2,\ldots$ and $\boldsymbol{L}'(k), k = 1,2,\ldots$ are Markov random sequences with the same statistical properties. The transition probabilities $p_{i,j}$ and the transition probability matrix P are defined as in (20). Then, the joint probabilities (51) and (52) can be written, respectively, as

$$P_r(l(k)=\alpha, l(k+1)=\beta) = \sum_{i,j=1}^{M} P_{i,j}\cdot p_i(k)\cdot h^i_{*\alpha}\cdot h^j_{*\beta} \tag{53}$$

and

$$P_r(l'(k)=\alpha, l'(k+1)=\beta) = \sum_{i,j=1}^{M} P_{i,j}\cdot p_i(k)\cdot h^i_{\alpha*}\cdot h^j_{\beta*}. \tag{54}$$

We also have that

$$P_r(l(k+1)=\alpha) = \sum_{i=1}^{M} p_i(k+1)\cdot h^i_{*\alpha} \tag{55}$$

and

$$P_r(l'(k+1)=\alpha) = \sum_{i=1}^{M} p_i(k+1)\cdot h^i_{\alpha*}. \tag{56}$$

Finally, we define the four diagonal matrices

$$\boldsymbol{H}_{*\alpha} = \operatorname{diag}\left(h^1_{*\alpha}, h^2_{*\alpha}, \ldots, h^M_{*\alpha}\right)$$
$$\boldsymbol{H}_{\alpha*} = \operatorname{diag}\left(h^1_{\alpha*}, h^2_{\alpha*}, \ldots, h^M_{\alpha*}\right).$$

Example: The TPC scheme presented in Fig. 7 has the six codewords $\boldsymbol{w}_1 = 000, \boldsymbol{w}_2 = 001, \boldsymbol{w}_3 = 010, \boldsymbol{w}_4 = 101, \boldsymbol{w}_5 = 110$, and $\boldsymbol{w}_6 = 111$. Therefore

$$W_{*0} = \{\boldsymbol{w}_1, \boldsymbol{w}_3, \boldsymbol{w}_5\} \quad W_{0*} = \{\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3\}$$
$$W_{*1} = \{\boldsymbol{w}_2, \boldsymbol{w}_3, \boldsymbol{w}_6\} \quad W_{1*} = \{\boldsymbol{w}_4, \boldsymbol{w}_5, \boldsymbol{w}_6\}$$

and so

$$\boldsymbol{H}_{*0} = \begin{bmatrix} 1 & & & & & 0 \\ & 0 & & & & \\ & & 1 & & & \\ & & & 0 & & \\ & & & & 1 & \\ 0 & & & & & 0 \end{bmatrix}$$
$$\boldsymbol{H}_{*1} = \begin{bmatrix} 0 & & & & & 0 \\ & 1 & & & & \\ & & 0 & & & \\ & & & 1 & & \\ & & & & 0 & \\ 0 & & & & & 1 \end{bmatrix}$$
$$\boldsymbol{H}_{0*} = \begin{bmatrix} 1 & & & & & 0 \\ & 1 & & & & \\ & & 1 & & & \\ & & & 0 & & \\ & & & & 0 & \\ 0 & & & & & 0 \end{bmatrix}$$
$$\boldsymbol{H}_{1*} = \begin{bmatrix} 0 & & & & & 0 \\ & 0 & & & & \\ & & 0 & & & \\ & & & 1 & & \\ & & & & 1 & \\ 0 & & & & & 1 \end{bmatrix}. \tag{57}$$

From the definitions of $\boldsymbol{H}_{*\alpha}, \boldsymbol{H}_{\alpha*}$ and (53)–(56)

$$P_r(l(k)=\alpha, l(k+1)=\beta) = \boldsymbol{p}(k)\cdot\boldsymbol{H}_{*\alpha}\cdot\boldsymbol{P}\cdot\boldsymbol{H}_{*\beta}\cdot\underline{\mathbf{1}} \tag{58}$$

$$P_r(l'(k)=\alpha, l'(k+1)=\beta) = \boldsymbol{p}(k)\cdot\boldsymbol{H}_{\alpha*}\cdot\boldsymbol{P}\cdot\boldsymbol{H}_{\beta*}\cdot\underline{\mathbf{1}} \tag{59}$$

$$P_r(l(k+1)=\alpha) = \boldsymbol{p}(k+1)\cdot\boldsymbol{H}_{*\alpha}\cdot\underline{\mathbf{1}} \tag{60}$$

$$P_r(l'(k+1)=\alpha) = \boldsymbol{p}(k+1)\cdot\boldsymbol{H}_{\alpha*}\cdot\underline{\mathbf{1}}. \tag{61}$$

Combining (58)–(61) and using the Chapman–Kolmogorov formula [36], $\boldsymbol{p}(k) = \boldsymbol{p}(0) \cdot \boldsymbol{P}^k$, we arrive at (62) of the expected interaction energy

$$\begin{aligned}\overline{J(k)} &= \lambda \cdot [\boldsymbol{p}(0) \cdot \boldsymbol{P}^{k+1} \cdot \boldsymbol{H}_{*0} \cdot \underline{\mathbf{1}}] \cdot [\boldsymbol{p}(0) \cdot \boldsymbol{P}^{k} \cdot \boldsymbol{H}_{0*} \cdot \boldsymbol{P} \cdot \boldsymbol{H}_{1*} \cdot \underline{\mathbf{1}}] \\ &\quad + \lambda \cdot [\boldsymbol{p}(0) \cdot \boldsymbol{P}^{k+1} \cdot \boldsymbol{H}_{1*} \cdot \underline{\mathbf{1}}] \cdot [\boldsymbol{p}(0) \cdot \boldsymbol{P}^{k} \cdot \boldsymbol{H}_{*1} \cdot \boldsymbol{P} \cdot \boldsymbol{H}_{*0} \cdot \underline{\mathbf{1}}] \\ &\quad + \lambda \cdot [\boldsymbol{p}(0) \cdot \boldsymbol{P}^{k+1} \cdot \boldsymbol{H}_{0*} \cdot \underline{\mathbf{1}}] \cdot [\boldsymbol{p}(0) \cdot \boldsymbol{P}^{k} \cdot \boldsymbol{H}_{*0} \cdot \boldsymbol{P} \cdot \boldsymbol{H}_{*1} \cdot \underline{\mathbf{1}}] \\ &\quad + \lambda \cdot [\boldsymbol{p}(0) \cdot \boldsymbol{P}^{k+1} \cdot \boldsymbol{H}_{*1} \cdot \underline{\mathbf{1}}] \cdot [\boldsymbol{p}(0) \cdot \boldsymbol{P}^{k} \cdot \boldsymbol{H}_{1*} \cdot \boldsymbol{P} \cdot \boldsymbol{H}_{0*} \cdot \underline{\mathbf{1}}]. \end{aligned} \tag{62}$$

We define the *time average expected interaction energy* (TAEI) as

$$E_{ai} = \lim_{N\to\infty} \frac{1}{N} \cdot \sum_{k=1}^{N} \overline{J(k)}. \tag{63}$$

We use Theorem 1 and the decomposition $\boldsymbol{P} = \boldsymbol{A} \cdot \Delta \cdot \boldsymbol{A}^{-1}$ of the transition probability matrix in order to calculate the time average expected interaction energy between consecutive blocks of the partitioned coding scheme in Fig. 14.

Without loss of generality, we can assume that $\boldsymbol{A} = [\boldsymbol{a}_0, \boldsymbol{a}_1, \ldots, \boldsymbol{a}_{M-1}]$ with $\boldsymbol{a}_0 = \underline{\mathbf{1}}$ and with the first q eigenvectors corresponding the eigenvalues[9] $1, e^{2\pi i/q}, e^{4\pi i/q}, \ldots, e^{2(q-1)\pi i/q}$ of $\boldsymbol{P}$, see Theorem 1. Also, let $\boldsymbol{A}^{-1} = [\boldsymbol{b}_0^T, \boldsymbol{b}_1^T, \ldots, \boldsymbol{b}_{M-1}^T]^T$ be the inverse of $\boldsymbol{A}$. Then

$$\begin{aligned}\boldsymbol{P}^k &= (\boldsymbol{A} \cdot \Delta \cdot \boldsymbol{A}^{-1})^k \\ &= \boldsymbol{A} \cdot \begin{bmatrix} \boldsymbol{\Delta}_1^k & 0 \\ 0 & \boldsymbol{Y}^k \end{bmatrix} \cdot \boldsymbol{A}^{-1} \\ &= \underline{\mathbf{1}} \cdot \boldsymbol{b}_0^T + \sum_{r=1}^{q-1} e^{\frac{2\pi i}{q} rk} \cdot \boldsymbol{a}_r \cdot \boldsymbol{b}_r^T + \boldsymbol{A} \cdot \begin{bmatrix} 0 & 0 \\ 0 & \boldsymbol{Y}^k \end{bmatrix} \cdot \boldsymbol{B}. \end{aligned} \tag{64}$$

From Theorem 1, we know that $\boldsymbol{Y}^k \to 0$ as $k \to \infty$; and thus, the first term of (62) can be expressed using

$$\begin{aligned}\boldsymbol{p}(0) \cdot \boldsymbol{P}^{k+1} \cdot \boldsymbol{H}_{*0} \cdot \underline{\mathbf{1}} &= \boldsymbol{b}_0^T \cdot \boldsymbol{H}_{*0} \cdot \underline{\mathbf{1}} + \sum_{r=1}^{q-1} e^{\frac{2\pi i}{q} rk} \\ &\quad \cdot \boldsymbol{p}(0) \cdot \boldsymbol{a}_r \cdot \boldsymbol{b}_r^T \cdot \boldsymbol{H}_{*0} \cdot \underline{\mathbf{1}} + \varepsilon_1(k) \\ \boldsymbol{p}(0) \cdot \boldsymbol{P}^{k} \cdot \boldsymbol{H}_{0*} \cdot \boldsymbol{P} \cdot \boldsymbol{H}_{1*} \cdot \underline{\mathbf{1}} &= \boldsymbol{b}_0^T \boldsymbol{H}_{0*} \boldsymbol{P} \boldsymbol{H}_{1*} \underline{\mathbf{1}} + \sum_{r=1}^{q-1} e^{\frac{2\pi i}{q} rk} \\ &\quad \cdot \boldsymbol{p}(0) \boldsymbol{a}_r \cdot \boldsymbol{b}_r^T \boldsymbol{H}_{0*} \boldsymbol{P} \boldsymbol{H}_{1*} \underline{\mathbf{1}} + \varepsilon_2(k) \end{aligned} \tag{65}$$

where $\varepsilon_1, \varepsilon_2 \to 0$ as $k \to \infty$. The rest of the terms in (62) can be expressed similarly. After lengthy algebraic manipulations, we arrive at (66) where $\bar{\boldsymbol{b}}_r$ is the complex conjugate of the eigenvector $\boldsymbol{b}_r$

$$\begin{aligned}E_{ai} = \lambda \Bigg[& \boldsymbol{b}_0^T (\boldsymbol{H}_{1*} \cdot \boldsymbol{P} \cdot \boldsymbol{H}_{0*} + \boldsymbol{H}_{*1} \cdot \boldsymbol{P} \cdot \boldsymbol{H}_{*0}) \underline{\mathbf{1}} \\ & - 4 \sum_{r=1}^{q-1} \sin^2\left(\frac{\pi r}{q}\right) \cdot |\boldsymbol{p}(0) \cdot \boldsymbol{a}_r|^2 \cdot \left(\boldsymbol{b}_r^T \cdot \boldsymbol{H}_{*0} \cdot \underline{\mathbf{1}}\right) \\ & \cdot \left(\bar{\boldsymbol{b}}_r^T \cdot \boldsymbol{H}_{0*} \cdot \underline{\mathbf{1}}\right) \Bigg]. \end{aligned} \tag{66}$$

[9]Theorem 1: q is the number of eigenvalues of modulus one.

When using (66) in numerical calculations, special attention has to be paid to the way matrices $\boldsymbol{A}$ and $\boldsymbol{A}^{-1}$ are formed. The columns (eigenvectors and generalized eigenvectors) of $\boldsymbol{A}$ and its eigenvalues must be arranged according to Theorem 1.

The case of $q = 1$ corresponds to transition pattern coding schemes with connected transition graphs. This is always a desirable property in practice. If $q = 1$, formula (66) simplifies to the more compact

$$E_{ai} = \lambda \boldsymbol{b}_0^T (\boldsymbol{H}_{1*} \cdot \boldsymbol{P} \cdot \boldsymbol{H}_{0*} + \boldsymbol{H}_{*1} \cdot \boldsymbol{P} \cdot \boldsymbol{H}_{*0}) \underline{\mathbf{1}}. \tag{67}$$

Example: Following the example of the TPC scheme in Fig. 7, assuming uniformly distributed data, we have the transition probability matrix $\boldsymbol{P}$ given by (38) whose probability eigenvector $\boldsymbol{b}_0$ is given by (39). Using (57) and (67), with $\lambda = 5$, we arrive at

$$\begin{aligned}E_{ai} &= \lambda \boldsymbol{b}_0^T (\boldsymbol{H}_{1*} \cdot \boldsymbol{P} \cdot \boldsymbol{H}_{0*} + \boldsymbol{H}_{*1} \cdot \boldsymbol{P} \cdot \boldsymbol{H}_{*0}) \underline{\mathbf{1}} \\ &= 5 \cdot \begin{bmatrix} 0.250 \\ 0.187 \\ 0.062 \\ 0.062 \\ 0.187 \\ 0.250 \end{bmatrix}^T \cdot \left\{ \begin{bmatrix} 0 & & & & & 0 \\ & 0 & & & & \\ & & 0 & & & \\ & & & 1 & & \\ & & & & 1 & \\ 0 & & & & & 1 \end{bmatrix} \right. \\ &\quad \cdot \frac{1}{4} \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & & & & & 0 \\ & 1 & & & & \\ & & 1 & & & \\ & & & 0 & & \\ & & & & 0 & \\ 0 & & & & & 0 \end{bmatrix} \\ &\quad + \begin{bmatrix} 0 & & & & & 0 \\ & 1 & & & & \\ & & 0 & & & \\ & & & 1 & & \\ & & & & 0 & \\ 0 & & & & & 1 \end{bmatrix} \cdot \frac{1}{4} \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \end{bmatrix} \\ &\quad \left. \cdot \begin{bmatrix} 1 & & & & & 0 \\ & 0 & & & & \\ & & 1 & & & \\ & & & 0 & & \\ & & & & 1 & \\ 0 & & & & & 0 \end{bmatrix} \right\} \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = 2.19. \end{aligned} \tag{68}$$

F. Total Energy Consumption of the Partitioned Coding Scheme

We have expressions of the time average expected energy, E_a, of the individual coding blocks, as well as expressions of the time average expected interaction energy, E_{ai}, caused by the coupling between adjacent blocks.

The *total time average expected energy consumption* (TTAEE) of the entire coding scheme in Fig. 14 is then given by

$$E_T = n \cdot E_a + (n - 1) \cdot E_{ai} \tag{69}$$

where n is the number of blocks. For the case $q = 1$

$$\begin{aligned}E_T = & n \boldsymbol{b}_0^T \cdot (\boldsymbol{P} \bullet \boldsymbol{E}) \cdot \underline{\mathbf{1}} + \lambda (n-1) \cdot \boldsymbol{b}_0^T (\boldsymbol{H}_{1*} \cdot \boldsymbol{P} \cdot \boldsymbol{H}_{0*} \\ & + \boldsymbol{H}_{*1} \cdot \boldsymbol{P} \cdot \boldsymbol{H}_{*0}) \underline{\mathbf{1}}. \end{aligned} \tag{70}$$

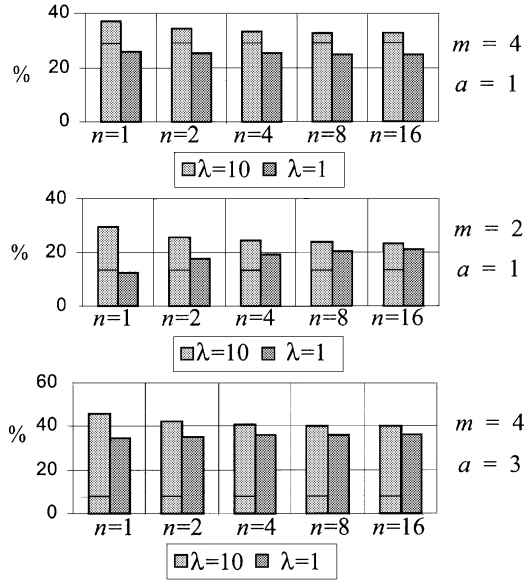

Fig. 16. Percentage of energy savings using partitioned coding schemes n: number of blocks, m: number of bits transmitted per block, and a: additional lines, $\lambda = 1, 10$.

Note that this energy is normalized with respect to $V_{\rm dd}$ and C_L; thus, the actual energy consumed[10] per clock cycle is $V_{\rm dd}^2 \times C_L \times E_T$. Fig. 16 presents the energy savings of partitioned coding schemes in which the coding blocks are TPC schemes designed using the TPC algorithm. The parameters in the graphs are n: number of blocks, m: number of bits transmitted per block, and a: additional lines, λ.

V. Future Directions

This work provides a mathematical framework for the design and evaluation of coding schemes that can be used for power reduction in on-chip or interchip communications through buses. Given a specific bus structure and estimating or measuring the distribution of the data, the two proposed algorithms provide coding schemes that, in theory, can result in significant energy reduction. For practical applications of a specific scheme, or a set of TPC schemes, further work will be required to estimate the complexity and the energy overhead of the encoder and decoder. Both complexity and overhead depend strongly on the particular technology, as well as the specific circuit implementation that will be used. Another issue that has to be addressed is that coding schemes need to be noncatastrophic; or, the encoder/decoder has a periodic resetting mechanism.

VI. Conclusion

A bus-structure and data-distribution-aware methodology for designing and analyzing coding schemes for low-power communication has been presented. A general class of coding schemes for low power, termed TPC schemes, has been introduced and mathematically analyzed in detail. Two algorithms have been proposed for deriving efficient TPC schemes that are optimized for given bus structures and data distributions. Bus partitioning has been mathematically analyzed as a way to reduce the complexity of the encoder/decoder.

[10]The long-term average energy drawn from the power supply equals the long-time average energy consumed, [20].

Acknowledgment

The authors would like to thank M. Khatib for his help to the preparation of the final manuscript.

References

[1] H. Zhang and J. Rabaey, "Low swing interconnect interface circuits," in *Proc. IEEE/ACM Int. Symp. Low Power Electronics and Design, ISPLED'98*, Aug. 1998, pp. 161–166.

[2] Y. Nakagome, K. Itoh, M. Isoda, K. Takeuchi, and M. Aoki, "Sub-1-V swing internal bus architecture for future low-power ULSIs," *IEEE J. Solid-State Circuits*, vol. 28, pp. 414–419, Apr. 1993.

[3] M. Hiraki, H. Kojima, H. Misawa, T. Akazawa, and Y. Hatano, "Data-dependent logic swing internal bus architecture for ultralow-power LSIs," *IEEE J. Solid-State Circuits*, vol. 30, pp. 397–401, Apr. 1995.

[4] E. Kusse and J. Rabaey, "Low-energy embedded FPGA structures," in *Proc. IEEE/ACM Int. Symp. Low Power Electronics Design, ISLPED'98*, Monterey, CA, Aug. 1998, pp. 155–160.

[5] K. Y. Khoo and A. Willson Jr., "Charge recovery on a databus," in *Proc. IEEE/ACM Int. Symp. Low Power Electronics and Design, ISPLED'95*, 1995, pp. 185–189.

[6] B. Bishop and M. J. Irwin, "Databus charge recovery: Practical consideration," in *Proc. IEEE/ACM Int. Symp. Low Power Electronics and Design, ISPLED'99*, Aug. 1999, pp. 85–87.

[7] H. Yamauchi, H. Akamatsu, and T. Fujita, "An asymptotically zero power charge-recycling bus architecture for battery-operated ultrahigh data rate ULSIs," *IEEE J. Solid-State Circuits*, vol. 30, pp. 423–431, Apr. 1995.

[8] M. Stan and W. Burleson, "Low-power encodings for global communication in cmos VLSI," *IEEE Trans. VLSI Syst.*, vol. 5, pp. 49–58, Dec. 1997.

[9] ——, "Two-dimensional codes for low-power," in *Proc. IEEE/ACM Int. Symp. Low Power Electronics Design, ISPLED'96*, Aug. 1996, pp. 335–340.

[10] P. Panda and N. Dutt, "Reducing address bus transitions for low-power memory mapping," in *Proc. Eur. Design and Test Conf.*, Mar. 1996, pp. 63–37.

[11] E. Musoll, T. Lang, and J. Cortadella, "Working-zone encoding for reducing the energy in microprocessor address buses," *IEEE Trans. VLSI Syst.*, vol. 6, pp. 568–572, Dec. 1998.

[12] W. Fornaciari, D. Sciuto, and C. Silvano, "Power estimation for architectural exploration of HW/SW communication on system-level buses," in *Proc. Int. Workshop on Hardware/Software Codesign*, 1999, pp. 152–156.

[13] S. Ramprasad, N. Shanbhag, and I. Hajj, "A coding framework for low-power address and data busses," *IEEE Trans. VLSI Syst.*, vol. 7, pp. 212–221, June 1999.

[14] Y. Shin, S.-I. Chae, and K. Choi, "Reduction of bus transitions with partial bus-invert coding," *Electron. Lett.*, vol. 34, no. 7, 1998.

[15] Y. Zhang, W. Ye, and M. J. Irwin, "An alternative architecture for on-chip global interconnect: Segmented bus power modeling," in *Proc. Asilomar Conf. Signals, Systems and Computers*, 1998, pp. 1062–1065.

[16] 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 *Proc. Great Lakes Symp. VLSI*, 1997, pp. 77–82.

[17] 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.*, 1999, pp. 128–133.

[18] S. Ramprasad, N. R. Shanbhag, and I. N. Hajj, "Information-theoretic bounds on average signal transition activity," *IEEE Trans. VLSI Syst.*, vol. 7, pp. 359–368, Sept. 1999.

[19] T. Sakurai, "Design challenges for 0.1 μm and beyond," in *Proc. Asia and South Pacific Design Automation Conf.*, 2001, pp. 293–296.

[20] P. Sotiriadis and A. Chandrakasan, "A bus energy model for deep submicron technology," *IEEE Trans. VLSI Syst.*, vol. 10, pp. 341–350, June 2002.

[21] ——, "Low power bus coding techniques considering inter-wire capacitances," in *Proc. Custom Integrated Circuits Conf., CICC'00*, May 2000, pp. 507–510.

Anantha P. Chandrakasanreceived the B.S., M.S., and Ph.D. degrees in electrical engineering and computer sciences from the University of California, Berkeley, in 1989, 1990, and 1994 respectively.

Since September 1994, he has been at the Massachusetts Institute of Technology, Cambridge, and is currently a Professor of Electrical Engineering and Computer Science. His research interests include the ultra low-power implementation of custom and programmable digital signal processors, distributed wireless sensors, multimedia devices, emerging technologies, and CAD tools for VLSI. He is a coauthor of the book *Low Power Digital CMOS Design* (Norwell, MA: Kluwer, 1995) and a coeditor of *Low Power CMOS Design* (Piscataway, NJ: IEEE Press, 1998) and *Design of High-Performance Microprocessor Circuits* (Piscataway, NJ: IEEE Press, 2000).

Dr. Chandrakasan received the NSF Career Development Award in 1995, the IBM Faculty Development Award in 1995, and the National Semiconductor Faculty Development Award in 1996 and 1997. He has received several best paper awards including the 1993 IEEE Communications Society's Best Tutorial Paper Award, the IEEE Electron Devices Society's 1997 Paul Rappaport Award for the Best Paper in an EDS publication during 1997 and the 1999 Design Automation Conference Design Contest Award. He held the Analog Devices Career Development Chair from 1994 to 1997 and has served on the Technical Program Committee of various conferences including ISSCC, VLSI Circuits Symposium, DAC, and International Symposium on Low-power Electronics and Design (ISLPED). He has served as a Technical Program Co-Chair for the 1997 ISLPED, 1998 VLSI Design, and the 1998 IEEE Workshop on Signal Processing Systems, and as a General Co-Chair of the 1998 ISLPED. He served as an elected member of the Design and Implementation of Signal Processing Systems (DISPS) Technical Committee of the Signal Processing Society. He was the Signal Processing Subcommittee chair for ISSCC 1999 through 2001. He was the Program Vice-Chair for ISSCC 2002 and the Program Chair for ISSCC 2003. He was an Associate Editor for the IEEE JOURNAL OF SOLID-STATE CIRCUITS from 1998 to 2001.