Transition Pattern Coding: An approach to reduce Energy in Interconnect

Paul P. Sotiriadis
Department of EECS
Massachusetts Inst. of Technology
Cambridge, MA 02139
pps@mit.edu

Alice Wang
Department of EECS
Massachusetts Inst. of Technology
Cambridge, MA 02139
aliwang@mtl.mit.edu

Anantha Chandrakasan
Department of EECS
Massachusetts Inst. of Technology
Cambridge, MA 02139
anantha@mtl.mit.edu

ABSTRACT

In this paper, we present a comprehensive energy dissipation model for data busses assuming a complete set of distributed RLC parasitics. Inter-wire parasitic capacitance contributes to significant energy loss during transitions. In this paper, we present a coding strategy (termed Transition Pattern Coding) that modifies the transition profiles to reduce switching energy. Using TPC, certain types of transitions are favored relative to others. We show that in the presence of capacitive coupling between the wires, our approach significantly outperforms previous coding schemes. Using TPC, up to 50% of energy saving is possible with sub-micron technologies.

INTRODUCTION

The energy cost of on-chip communication is significant in deep-submicron technologies. The interconnect energy can be reduced at all levels of system abstraction from advanced materials and processing (e.g., use of air as the dielectric) to circuit design (e.g., low swing transmission) and innovative new architectures (e.g., integrated memory/logic or 3D tiling). Another approach for reducing energy is to explore trade-offs in computation and communication. For example, bus coding techniques add redundancy to the transmitted data with the goal of reducing communication costs. One previously published coding technique to reduce switching energy under the basic lumped bus model, is the bus-invert technique in which the data bus is conditionally inverted to reduce the overall transitions [1]. If more than 50% of the bits change, the entire bus is inverted. Therefore, in addition to the data, an extra bit must be transmitted to indicate if the bus is inverted. This approach reduces transition activity by up to 25% for independent and uniformly distributed inputs.We present a complete model for the bus communication energy that includes the distributed nature of wires. Using this model, we propose a new coding strategy, termed Transition Pattern Coding, that reduces communication costs. In fact, an interesting result of the model is that minimizing the average number of transitions is not necessarily the best approach to reduce energy. Instead, the model suggest that certain types of transitions must be favored for minimal energy. We present the results of a coding scheme that exploits this behavior.

1. A COMPLETE MODEL FOR THE BUS LINES

A general model for busses is presented here. Let n be the number of lines in the bus and let L_p be their physical length. Figure 1 shows an elementary part of the bus of length Δx. For every $i = 1, 2, \ldots, n$, $I_i(x, t)$ is the current running through the *i-th* line at the point $x \in [0, L_p]$ and time $t \geq 0$. Similarly, $V_i(x, t)$ is the potential of that point with respect to the ground.

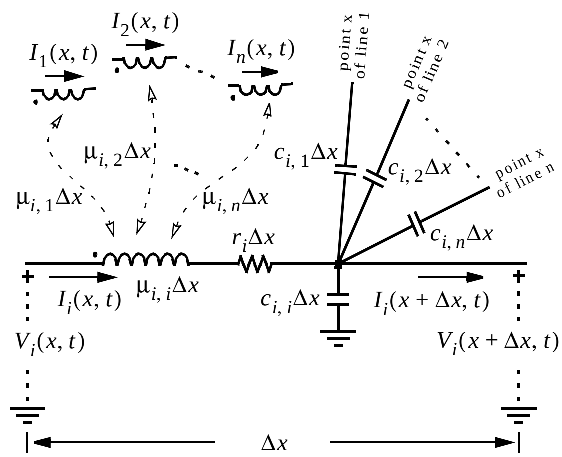

Figure 1: Elementary part of the *i-th* line.

All the important parasitic elements among the lines and between the lines and the ground (shielding or return path) are included in the model. In Figure 1, $r_i(x) \cdot \Delta x$ is the serial resistance of the elementary part, $c_{i,i}(x) \cdot \Delta x$ and $c_{i,j}(x) \cdot \Delta x$ are the capacitances to ground and to the *j-th* line respectively. The self inductance of the elementary part is $\mu_{i,i}(x) \cdot \Delta x$ and its mutual inductance with the *j-th* line is $\mu_{i,j}(x) \cdot \Delta x$. Kirchoff's laws for the elementary part give Equations 1. Dividing them by Δx and taking the limit as $\Delta x \to 0$ we get,

$$\left\{\begin{aligned} I_i(x,t) &= \left[c_{i,i}(x)\cdot\frac{\partial V_i(x,t)}{\partial t} + \sum_{j\neq i} c_{i,j}(x)\cdot\frac{\partial(V_i(x,t)-V_j(x,t))}{\partial t}\right]\cdot\Delta x + I_i(x+\Delta x,t) \\ V_i(x,t) &= \left[\mu_{i,i}(x)\cdot\frac{\partial I_i(x,t)}{\partial t} + \sum_{j\neq i}\mu_{i,j}(x)\cdot\frac{\partial I_j(x,t)}{\partial t} + r_i(x)\cdot I_i(x,t)\right]\cdot\Delta x + V_i(x+\Delta x,t)\end{aligned}\right\} \quad (1)$$

$$\left\{\begin{array}{l} -\dfrac{\partial I_i(x,t)}{\partial x} = \sum_{j=1}^{n} c_{i,j}(x)\cdot\dfrac{\partial V_i(x,t)}{\partial t} - \sum_{j\neq i} c_{i,j}(x)\cdot\dfrac{\partial V_j(x,t)}{\partial t} \\ -\dfrac{\partial V_i(x,t)}{\partial x} = r_i(x)\cdot I_i(x,t) + \sum_{j=1}^{n}\mu_{i,j}(x)\cdot\dfrac{\partial I_j(x,t)}{\partial t} \end{array}\right\} \quad (2)$$

Defining $V = [V_1, V_2, \ldots, V_n]^T$ and $I = [I_1, I_2, \ldots, I_n]^T$, the system of the coupled P.D.Es can be written compactly as

$$-\frac{\partial I}{\partial x}(x,t) = C(x)\cdot\frac{\partial V}{\partial t}(x,t) \quad (3)$$

$$-\frac{\partial V}{\partial x}(x,t) = M(x)\cdot\frac{\partial I}{\partial t}(x,t) + R(x)\cdot I(x,t) \quad (4)$$

for $x \in [0, L_p]$ and $t \geq 0$. The $n \times n$ matrix functions R, M, C are the resistance, inductance and capacitance matrices respectively. They have the following forms, $R(x) = diag[r_1(x), \ldots, r_n(x)]$, $M(x) = [\mu_{i,j}(x)]_{i,j=1,\ldots n}$ and $C(x) = C^L(x) + \sum_{i<j} C^{IL}_{i,j}(x)$. The diagonal matrix $C^L(x) = diag[C_{1,1}(x), \ldots, C_{n,n}(x)]$ corresponds to the parasitic capacitances to ground. The matrices $C^{IL}_{i,j}$, $i<j$ correspond to the $n(n-1)/2$ inter-line capacitance densities. Each matrix has (at most) four nonzero elements (for every *x*) in the entries i,i , i,j , j,i , j,j and is of the form,

$$C^{IL}_{i,j}(x) = c_{i,j}(x)\cdot\begin{bmatrix} 0 & \ldots & 0 & \ldots & 0 & \ldots & 0 \\ \vdots & & \vdots & & \vdots & & \\ 0 & \ldots & 1 & \ldots & -1 & \ldots & 0 \\ \vdots & & \vdots & & \vdots & & \\ 0 & \ldots & -1 & \ldots & 1 & \ldots & 0 \\ \vdots & & \vdots & & \vdots & & \\ 0 & \ldots & 0 & \ldots & 0 & \ldots & 0 \end{bmatrix}$$

It is important to comment that the distributions of the inductances, the capacitances and the serial resistances of the lines are completely **general**. Also, lumped parasitic elements can always be modeled as limiting cases of the distributed ones. Therefore the above model implicitly includes the possible lumped capacitive **loads** at the ends of the lines.

1.1 Drivers and Energy Consumption

To calculate the energy consumption caused by a transition in the data bus, we need a model for the drivers of the lines. We adopt the following one (see also ref. [3] and [4]).

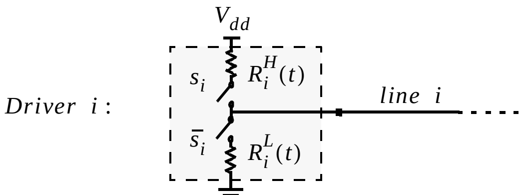

Figure 2: Model for Bus Drivers

The resistors $R_i^H(t)$ and $R_i^L(t)$ are the *ON* resistors of the Mosfets of the drivers and they can be arbitrary function of time. To calculate the energy cost of a transition we need to do three reasonable assumptions that will simplify the analysis. First we assume that *at time t<0 the voltages across the lines are in equilibrium, equipotencial positions*, say

$$V_1(x,0) = V_1^i, \ldots, V_n(x,0) = V_n^i \quad (5)$$

for every $x \in [0, L_p]$ and the voltages $V_1^i, V_2^i, \ldots, V_n^i$ are either 0 or V_{dd} depending on the *last* (*t<0*) *logical* values of the lines (s_i , $i = 1, \ldots, n$ open or closed respectively). Now let $V_1^f, V_2^f, \ldots, V_n^f$ be the voltages corresponding to the new data to be transmitted. V_j^f is either 0 or V_{dd} for $j = 1, \ldots n$ corresponding to the appropriate configuration of the switches s_j , $\bar{s}_j$. At the time limit $t \to \infty$ we have

$$V_1(x,\infty) = V_1^f, \ldots, V_n(x,\infty) = V_n^f \quad (6)$$

for every $x \in [0, L_p]$. We do the assumption that *by the end of the clock period the voltages across the lines have achieved their final equilibrium values given by Equation 6.*

Finally, the one end of each line is connected to its driver, the other end is free (as it was mentioned before the possible capacitive loads of the lines can be integrated into the capacitive distribution C). The free ends imply the boundary conditions,

$$I_1(L_p,t) = 0, \ldots, I_n(L_p,t) = 0, \quad t \geq 0 \quad (7)$$

for the system of Equations 3. The energy consumed by the *i-th* driver during the transition is $E_i = \int_0^\infty V_i^f \cdot I_i(0,t)dt$ and so the total energy during the transition is given by $E = \int_0^\infty (V^f)^T \cdot I(0,t)dt$, where $V^f = [V_1^f, V_2^f, \ldots, V_n^f]^T$. Integrating both sides of Equation 3 over *x* we get,

$$-\int_0^{L_p}\frac{\partial I}{\partial x}(x,t)dx = I(0,t) - I(L_p,t) = \int_0^{L_p} C(x)\cdot\frac{\partial V}{\partial t}(x,t)dx \quad (8)$$

Using the boundary conditions 7 and Equation 8 we take,

$$E = (V^f)^T\int_0^\infty\left(\int_0^{L_p} C(x)\cdot\frac{\partial V}{\partial t}(x,t)dx\right)dt \quad (9)$$

Our third assumption is that *the matrix function* $C(x)\cdot\frac{\partial V}{\partial t}(x,t)$ *is Lebesque integrable within its domain of integration*. This allows the use of Fubini's theorem to interchange the order of integration in Equation 9. Therefore,

$$E = (V^f)^T\int_0^{L_p} C(x)\left(\int_0^\infty\frac{\partial V}{\partial t}(x,t)dt\right)dx$$

$$= (V^f)^T \int_0^{L_p} C(x)(V(x,\infty) - V(x,0))dx$$

$$= (V^f)^T \left(\int_0^{L_p} C(x)dx \right) \cdot (V^f - V^i)$$

Defining the *total capacitance matrix* C_T as $C_T = \int_0^{L_p} C(x)dx$ the energy is given by the following compact formula,

$$E = (V^f)^T \cdot C_T \cdot (V^f - V^i) \qquad (10)$$

Here are some comments on the energy function given by Equation 10. i) The three assumptions are very mild, practically always true, and they appear in similar energy analysis problems in the literature. ii) The energy does not depend on the particular values of the resistances or the inductances. iii) It does not depend on the trajectories of the voltages $V_1, V_2, \ldots, V_n$ (as functions of time) but only on their initial and final conditions. This is like a conservation law. iv) It does not depend on the time variations of the resistors of the drivers, $R_i^H(t), R_i^L(t)$, $i = 1, \ldots, n$ as long as the three assumptions are not violated. v) It depends only on the total capacitance matrix, C_T.

1.2 A Special Case of Data Bus

An common special case of data busses is when the lines are parallel, coplanar and lay between parallel ground planes. Then the capacitive interaction is basically restricted among adjacent lines. Also the distributed capacitances are pretty much uniform along the lines and constant between the lines. Under these conditions the total capacitance matrix C_T becomes,

$$C_T = \begin{bmatrix} 1+\lambda & -\lambda & 0 & \ldots & 0 \\ -\lambda & 1+2\lambda & -\lambda & : & 0 \\ 0 & -\lambda & & : & : \\ : & : & : & 1+2\lambda & -\lambda \\ 0 & 0 & \ldots & -\lambda & 1+\lambda \end{bmatrix} \cdot C_L$$

where $C_L = \int_0^{L_p} c_{i,i}(x)dx$, $C_I = \int_0^{L_p} c_{i,i+1}(x)dx$, $\lambda = C_I / C_L$.

2. TRANSITION PATTERN CODING FOR LOW POWER

From the above analysis it is clear that different transitions in the bus can result in very different amounts of energy dissipated. The following table demonstrates this in the case of a bus with two lines when their voltages change from $[V_1^i, V_2^i]$ to $[V_1^f, V_2^f]$. Then the energy dissipation is $E = E^n \cdot V_{dd}^2 \cdot C_L$. E^n is the normalized energy of the transition (i.e. $V_{dd} = 1$ and $C_L = 1$) and is given in Table 1. Therefore a way to save energy on the data bus is to favor the transitions with *low energy cost*. We call this tailoring of the transition pattern, **Transition Pattern Coding** (**TPC**).

E^n		$[V_1^f, V_2^f]$ 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

Table 1: Normalized energy dissipation

If it is desirable that the data (sequence $\{D(k)\}$) are encoded, transmitted and decoded "instantaneously", in other words the data are available at the output of the receiver the same clock cycle (time k) they are feeded into the encoder, the class of TPC coding schemes is shown in Figure 3.

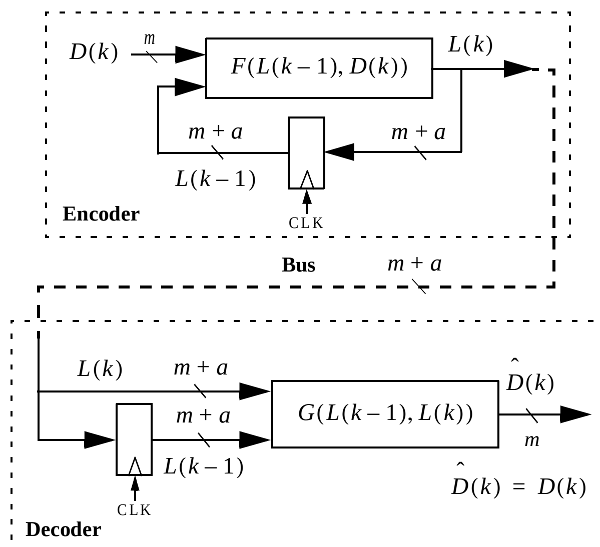

Figure 3: The general class of coding schemes

The "original" data bus of m lines has now been expanded to a bus of m+a lines. Let $L(k)$ be the vector of the logical values of these m+a lines. $L(k)$ is restricted to lay within a subset $W = \{w_1, w_2, \ldots, w_M\}$ of $\{0, 1\}^{m+a}$. We call the M elements of W, the *codewords* of the coding scheme. Then the functions F and G are of the form, $F : W \times \{0,1\}^m \to W$ and $G : W \times W \to \{0,1\}^m$ respectively. The decoder can recover the original data i.e. $\hat{D}(k) = D(k)$ if and only if the following condition holds,

$$G(w, F(w, d)) = d, \forall d \in \{0,1\}^m, \forall w \in W \qquad (11)$$

Relation (11) implies that for every fixed $w \in W$ the mapping, $d \to F(w, d)$ is injective. Therefore the $M \times M$ *transition matrix* $T = [t_{i,j}]_{i,j=1}^M$ associated with function F,

$$t_{i,j} = \begin{cases} 1 & if\ \exists\ \mathrm{d} \in \{0,1\}^m \text{ s.t. } F(w_i, d) = w_j \\ 0 & otherwise \end{cases}$$

has exactly 2^m ones in every one of its rows. The *transition graph* G_T, is defined as

$$G_T = \{(w, F(w, d)) : w \in W,\ d \in \{0, 1\}^m\} \tag{12}$$

2.1 An Illustrative Example

Suppose the "original" bus has 2 lines. Extending it by adding one more line and applying the TPC corresponding to the transition graph shown in Figure 4, we get the energy

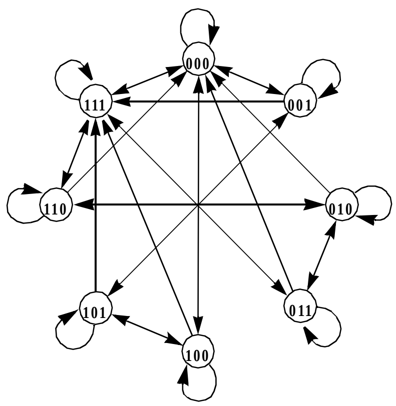

Figure 4: Transition Graph of a TPC scheme with $m = 2$, $a = 1$, $M = 8$. Designed for $\lambda = 2$

savings of Figure 5 (parametrized by λ). As it was men-

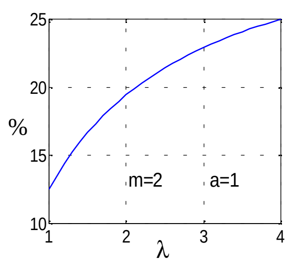

Figure 5: Energy savings of the TPC with Transition Graph shown in Figure 4

tioned before, the energy saving is due to the modification of the transition pattern of the bus. The expected number per cycle of each possible transition of the lines of the "original" bus is shown in Table 2. It is interesting to compare Table 2 with Table 3 which gives the expected number of each possible transition of pairs of adjacent lines for TPC. The expected transition energy per cycle for each case is given by multiplying pointwise the elements of the tables with the elements of Table 1 and summing the products.

$\frac{1}{16}\times$		new values 00	01	10	11
old values	00	**1**	**1**	**1**	**1**
	01	**1**	**1**	**1**	**1**
	10	**1**	**1**	**1**	**1**
	11	**1**	**1**	**1**	**1**

Table 2: Expected transitions without coding

$\frac{1}{24}\times$		new values 00	01	10	11
old values	00	**4**	**1**	**1**	**2**
	01	**1**	**2**	**0**	**1**
	10	**1**	**0**	**2**	**1**
	11	**2**	**1**	**1**	**4**

Table 3: Expected transitions with TPC (between all pairs of adjacent lines)

2.2 Results

For λ varying from 0 to 10, for $m = 2, 3, 4, 5$ and a being $1, 2, 3, 4$ or 5, a TPC coding scheme was designed with these parameters and its energy saving efficiency was calculated (exactly) compared with that of Bus Invert coding. The results are shown in Figure 6.

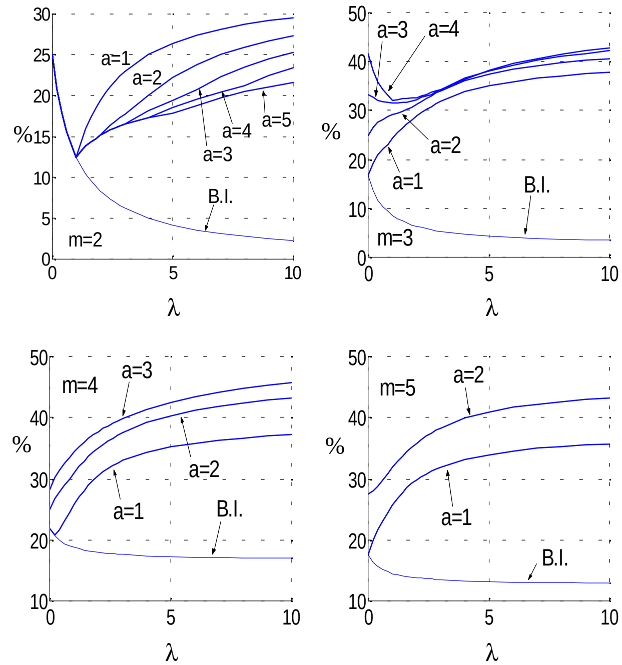

Figure 6: Energy Savings of TPC and Bus Invert

Conclusions

Minimizing transition activity is not necessarily the best approach to reduce energy dissipation when the effect of inter-wire capacitance is significant. An accurate energy model has enabled the development of efficient Transition Pattern Coding strategies using an elaborate distributed model for the wires. The overall energy dissipation can be reduced by a factor of 2.

References

[1] M. Stan, W. Burleson, "Low-power encodings for global communication in cmos VLSI," *IEEE Transactions on VLSI Systems*, pp. 49-58, Vol. 5, No. 4, Dec. 1997.

[2] S. Ramprasad, N. Shanbhag, I. Hajj, "A coding framework for low-power address and data busses," *IEEE Transactions on VLSI Systems*, pp. 212-221, Vol. 7, No. 2, June 1999.

[3] J.M. Rabaey, *Digital Integrated circuits*. Prentice Hall 1996.

[4] H.B. Bakoglou, *Circuits, Interconnections, and Packaging in VLSI*. Addison-Wesley Pub. Co., c1990