Journal of Circuits, Systems, and Computers
© World Scientific Publishing Company

POWER ESTIMATION AND POWER OPTIMAL COMMUNICATION IN DEEP SUB-MICRON BUSSES: ANALYTICAL MODELS AND STATISTICAL MEASURES

PAUL P. SOTIRIADIS*
Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge MA 02139 USA

ANANTHA CHANDRAKASAN
Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge MA 02139 USA

Received June 16th 2002
Accepted Aug. 13th 2002

Reduction of power dissipation in digital circuits is a subject of research in industry and academia. A major component of power dissipation in modern microprocessors is due to their large interconnect networks which are responsible for the distribution of power and clocks as well as for the intra-chip communication. Communication is realized by data and address buses. In this paper we (i) discuss an analytical model for energy estimation in deep sub-micron buses, (ii) present statistical energy measures based on the analytical model, (iii) derive the energy limits of communication through buses, (iv) and introduce energy efficiency measures of communication.

1. Introduction

Power consumption is a major concern in the design of high performance and portable circuit systems. During the past decade, a lot of effort has been devoted in developing low power design methodologies [1,2] at all design levels as well as mathematical models and CAD tools for estimating power consumption [3,4,5].

In CMOS digital circuits, dynamic power dissipation due to capacitive coupling is still the major power component. The transition activity T_a of a circuit node (1/2 the probability of changing value) is a simple statistical measure widely employed in power estimation. Transition activity is translated into power dissipation using $P = T_a f C_L V$, where C_L is the capacitance between the node and ground, V is the voltage swing of the node and f is the frequency of operation. The transition activity alone cannot be used when the node is coupled to other active nodes.

*He is currently with the Department of Electrical and Computer Engineering, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218 USA (e-mail: pps@jhu.edu)

Coupling between nodes implies that power depends also on their cross-activities and therefore the simple power formula above is not valid.

In today's high performance microprocessors with huge interconnect networks, intra chip communication has been associated with significant power dissipation [6,7]. This is a result of technology scaling into deep sub-micron dimensions that resulted in the emerging of new parasitic elements between nodes. In interconnect networks and especially in long buses, where lines are laid out in parallel, and in many cases with minimum distance to each other, inter-line coupling cannot be ignored [7]. Work has been done in developing compact circuit models for interconnect networks and estimating the sizes of their elements [8,9,10,11,12].

The first part of the paper, section 2, is devoted in establishing an analytical energy model for deep sub-micron buses based on the available interconnect circuit models. Statistical measures are developed in Section 3 to estimate energy dissipation as a function of statistical properties of data or address sequences. The transition activity matrix is discussed as an extension of transition activity. As an application of the definitions and the derived expressions, at the end of Section 3 we formulate mathematically a heuristic technique for power reduction presented in [13].

Buses are very basic communication channels that dissipate energy according to the sequence of bit vectors they transmit. In Section 4 we study the relation between information rate and power dissipation in deep sub-micron buses. We derive analytic expressions of the minimum energy that is required to transmit a bit of information through the bus. The results are very general and apply to buses (or other devices) with any energy cost function. Previous work on this problem has discussed the case of one-line buses (which is the same with buses having many decoupled lines) [14]. Finally, the energy formulas of Section 2 can be used for practical cases.

In Section 5 we introduce the energy communication efficiency factor, a measure of the energy redundancy in communication through buses. The factor equals the ratio of the amount of power dissipated when given data or address sequences are transmitted over the minimum amount of power necessary to transmit information through the bus at the same rate. The efficiency factor can be used in characterizing the efficiency of power reduction techniques based on coding (like for example those at [15,16,17,18,19]).

2. An Analytical Bus Energy Model

A bus consists of a set of parallel lines as shown in Figure 1. The lines are connected to the drivers (left) and the receivers (right). In most cases the drivers and the receivers are (chains of) CMOS inverters. Repeaters may exist at certain places along the bus dividing it into segments. In that case we can examine each segment

individually and sum their corresponding energy consumptions.

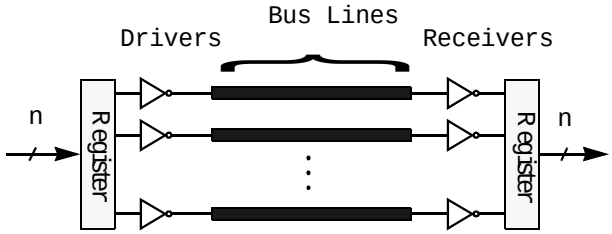

Fig. 1. Bus

For the bus lines we use the model of Figure 2 that has been used extensively for delay estimation as well as signal integrity evaluation [20,21,22,23]. The lines are distributed, laid in parallel along the x axis and have physical length L. They have serial resistance $r_i(x)$, $i = 1, 2, \ldots, n$. The capacitance density between the ith line and ground is $c_{i,i}(x)$ and that between lines i and j is $c_{i,j}(x)$. Moreover, $\mu_{i,i}(x)$ is the density of the self inductance of the ith line and $\mu_{i,j}(x)$ is the density of the mutual inductance between lines i and j. The densities may depend on x. Finally, possible lumped parasitics can be included as limiting cases of distributed ones.

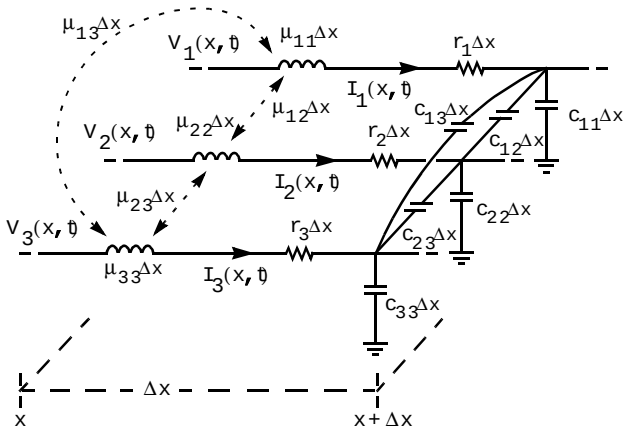

Fig. 2. Distributed model of the bus lines

Let $I_i(x,t)$ be the current running through the ith line at the point $0 < x < L$ and time $t \geq 0$ and let $V_i(x,t)$ be the voltage at that point with respect to ground. If we set $I = [I_1, I_2, \cdots, I_n]^T$ and $V = [V_1, V_2, \cdots, V_n]^T$ then the line model satisfies

the system of partial differential equations:

$$-\frac{\partial I}{\partial x}(x,t) = A(x)\frac{\partial V}{\partial t}(x,t) \tag{1}$$

where $A(x) = [a_{i,j}(x)]$ is the admitance matrix of the distributed capacitance of the network at x, i.e.

$$a_{i,j}(x) = \begin{cases} \sum_{k=1}^{n} c_{i,k}(x) & \text{if} \quad i = j \\ -c_{i,j}(x) & \text{if} \quad i \neq j \end{cases} \tag{2}$$

More details of the electrical characterization of the bus lines and their modeling can be found in [24,25,26].

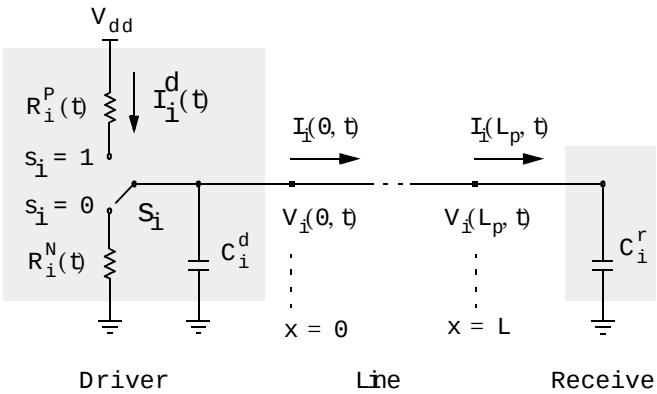

Fig. 3. Driver, line, receiver

The drivers and the receivers of the bus are modeled as in Figure 3. The driver (CMOS inverter) is modelled as a switch connecting the line, either to power supply, or ground, through the PMOS or NMOS transistors respectively. [27]† The parasitic capacitance at the output of the ith driver is C_i^d and the parasitic capacitance at the input of the ith receiver is C_i^r.

The current $I_i^d(t)$ drawn from V_{dd} by the ith driver is 0 if the binary value transmitted is 0 and $I_i(0,t) + C_i^d\, \partial V_i(0,t)/\partial t$ if the binary value is 1. The binary values correspond to the final voltages $V_i^f = 0$ and $V_i^f = V_{dd}$ respectively (superscript f stands for final). We may write:

$$I_i^d(t) = \frac{V_i^f}{V_{dd}} \left\{ I_i(0,t) + C_i^d \, \frac{\partial V_i(0,t)}{\partial t} \right\} \tag{3}$$

At the (right) end of the line, $x = L$, we have the boundary condition:

$$I_i(L,t) = \frac{\partial V_i(L,t)}{\partial t} \tag{4}$$

†This model does not account for the short-circuit currents of the drivers. In most buses short-circuit currents result to negligible energy consumption relatively to that resulting by capacitive parasitics.

Let T be the clock period of the bus. At $t = 0$ the drivers connect the lines to V_{dd} or ground and the transmission of the data vector starts. The data is sampled at the output of the receivers at $t = T$. In most applications it is reasonable to assume that time period T is sufficient for the voltages along the lines to settle to their final values, i.e.:

$$V_i(x,T) = V_i^f \ , \ \ 0 \leq x \leq L \ , \ \ i = 1, 2, \ldots, n \tag{5}$$

The assumption also implies that at $t = 0$, the voltages along the lines correspond to their previous binary values (superscript i stands for initial), i.e.:

$$V_i(x,0) = V_i^i \ , \ \ 0 \leq x \leq L \ , \ \ i = 1, 2, \ldots, n \tag{6}$$

We define the vectors of initial and final voltages as $V^i = (V_1^i, V_2^i, \cdots, V_n^i)^T$ and $V^f = (V_1^f, V_2^f, \cdots, V_n^f)^T$ respectively.

The energy drawn from the power supply, by the ith driver, during the transition period $0 \leq t \leq T$ is given by

$$E_i^{Vdd} = \int_0^T V_{dd}\, I_i^d(t)\, dt \tag{7}$$

The total energy drawn from V_{dd} is $E^{Vdd} = \sum_{i=1}^n E_i^{Vdd}$. An expression for E^{Vdd} is given by the following proposition.

Proposition 1. *With the assumptions stated above, the energy drawn from V_{dd} during the transition of the bus from an initial state $V^i = (V_1^i, V_2^i, \cdots, V_n^i)^T$ to a final state $V^f = (V_1^f, V_2^f, \cdots, V_n^f)^T$ is:*

$$E^{Vdd} = (V^f)^T\, \mathcal{A}\, (V^f - V^i) \tag{8}$$

where $\mathcal{A}$ is the admitance matrix of the lumped capacitive network shown in Figure 4 (where for simplicity $n = 4$):

$$[\mathcal{A}]_{i,j} = \begin{cases} \sum_{k=1}^n C_{i,k} & \text{if} \quad i = j \\ -C_{i,j} & \text{if} \quad i \neq j \end{cases} \tag{9}$$

The capacitance $C_{i,j}$ is the total capacitance between the bus lines i and j if $i \neq j$ or the total capacitance between the ith bus line and ground, including those of the driver and receiver, if $i = j$. Therefore it is:

$$C_{i,j} = \begin{cases} \int_0^L c_{i,i}(x)\, dx + C_i^d + C_i^r & \text{if} \quad i = j \\ \int_0^L c_{i,j}(x)\, dx & \text{if} \quad i \neq j \end{cases} \tag{10}$$

Proof: The calculation of the energy E, as a function of the initial and final voltages, is given in Appendix A.

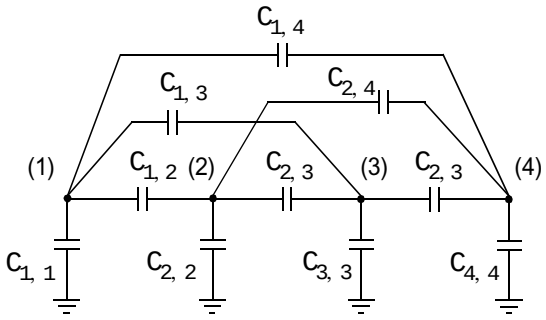

Fig. 4. Equivalent capacitive network ($n = 4$)

Notice that the energy drawn from V_{dd} is different from the energy lost during the transition from V^i to V^f. We have the following proposition.

Proposition 2. *With the setup of Proposition 1, the energy loss during the transition of the bus from an initial state* $V^i = (V_1^i, V_2^i, \cdots, V_n^i)^T$ *to a final state* $V^f = (V_1^f, V_2^f, \cdots, V_n^f)^T$*, is:*

$$E(V^i, V^f) = \frac{1}{2}\,(V^f - V^i)^T\,\mathcal{A}\,(V^f - V^i) \tag{11}$$

Proof: In the beginning of the transition, at $t = 0$, the voltages along the lines are: $V^i = (V_1^i, V_2^i, \cdots, V_n^i)^T$ and so the energy stored in the capacitances of the drivers, lines and receivers is:

$$E^i = \frac{1}{2}\sum_{k=1}^{n} C_{k,k}(V_k^i)^2 + \frac{1}{2}\sum_{k<r} C_{k,r}(V_k^i - V_r^i)^2 = \frac{1}{2}\,(V^i)^T\,\mathcal{A}\,V^i. \tag{12}$$

Similarly, at $t = T$, the voltages along the lines are: $V^f = (V_1^f, V_2^f, \cdots, V_n^f)^T$ and so the energy stored in the capacitances is:

$$E^f = \frac{1}{2}\,(V^f)^T\,\mathcal{A}\,V^f. \tag{13}$$

Energy conservation implies that the energy drawn from the power supply must equal the energy loss, E, plus the difference in the energy stored in the capacitors. Therefore we have:

$$\begin{aligned} E &= E^{Vdd} - (E^f - E^i) \\ &= (V^f)^T\,\mathcal{A}\,(V^f - V^i) - \frac{1}{2}(V^f)^T\,\mathcal{A}\,V^f + \frac{1}{2}(V^i)^T\,\mathcal{A}\,V^i \\ &= \frac{1}{2}\,(V^f - V^i)^T\,\mathcal{A}\,(V^f - V^i) \end{aligned} \tag{14}$$

Example 1. Suppose we want to calculate the energy loss on a four line bus, during the transition from the state $V^i = V_{dd} \cdot (0,1,0,1)$ to the new state $V^f = V_{dd} \cdot (1,0,0,0)$. Suppose the bus has total parasitic capacitances like those of the network in Figure 4. We have:

$$E = \tfrac{1}{2}\,(V^f - V^i)^T\,\mathcal{A}\,(V^f - V^i)$$

$$= \tfrac{V_{dd}}{2}\,(1,-1,0,-1)\cdot \begin{bmatrix} \sum_{k=1}^{4} C_{1,k} & -C_{1,2} & -C_{1,3} & -C_{1,4} \\ -C_{2,1} & \sum_{k=1}^{4} C_{2,k} & -C_{2,3} & -C_{2,4} \\ -C_{3,1} & -C_{3,2} & \sum_{k=1}^{4} C_{3,k} & -C_{3,4} \\ -C_{4,1} & -C_{4,2} & -C_{4,3} & \sum_{k=1}^{4} C_{4,k} \end{bmatrix} \begin{pmatrix} 1 \\ -1 \\ 0 \\ -1 \end{pmatrix}$$

Comments: Expression (11) of the energy dissipation is correct even when some of the grounds in Figure 2 are replaced by V_{dd}. In modern bus design fabrics like that of Figure 5 with groups of three or more signal lines are used to minimize the inductive effects. In this case we can apply formula (11) directly, regarding V_{dd} and ground as identical.

Since the wires in consecutive layers are laid out vertically, the coupling between bus lines and individual wires above and below the bus is relatively very weak. There are many such couplings and so averaging effect takes place in practise making the total non-bus wire environment behaving very much like ground.

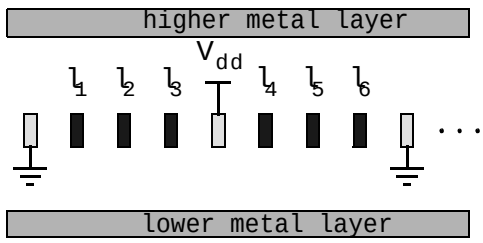

Fig. 5. Bus line fabric

Example 2. In the case of a bus with structure like that of Figure 5, capacitive coupling between non consecutive lines is very weak relatively to that between consecutive lines. An approximate model of the total capacitance network is shown in Figure 6.

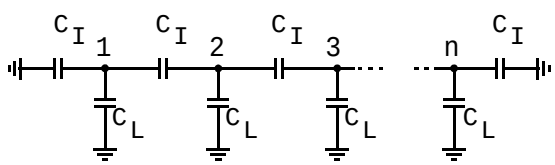

Fig. 6. Equivalent capacitive network ignoring coupling between non adjacent lines

The admitance matrix $\mathcal{A}$ is simplified as:

$$\mathcal{A} = \begin{bmatrix} 1+2\lambda & -\lambda & 0 & \cdots & 0 \\ -\lambda & 1+2\lambda & -\lambda & \cdots & 0 \\ 0 & -\lambda & 1+2\lambda & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1+2\lambda \end{bmatrix} \cdot C_L \tag{15}$$

where $\lambda = C_I/C_L$.

3. Statistical Energy Estimation and the Transition Activity Matrix

Let $l_i(1), l_i(2), \ldots$ be a sequence of bits transmitted through the ith line. If the ith line is coupled only to ground, through total capacitance C_L, the energy loss during the kth cycle, given by expression (11), is:

$$E_i(k) = \frac{1}{2}\left[l_i(k) - l_i(k-1)\right]^2 \cdot V_{dd}^2 \cdot C_L \tag{16}$$

Assuming the random bit sequence is stationary with the wide sense, the expected energy is given by:

$$\overline{E_i} = \left[R_i(0) - R_i(1)\right] \cdot V_{dd}^2 \cdot C_L \tag{17}$$

where $R_i(r) = \overline{l_i(k) \cdot l_i(k+r)}$ is the autocorrelation function of the bit sequence (over-line stands for expectation) . The quantity $T_i^a = R_i(0) - R_i(1)$ is the *transition activity* of the line. If f is the clock frequency of the bus, the expected energy and power can be written respectively as:

$$\overline{E_i} = T_i^a \cdot V_{dd}^2 \cdot C_L \ , \quad P_i = T_i^a \cdot f \cdot V_{dd}^2 \cdot C_L \tag{18}$$

This formula applies directly to a bus with decoupled lines and has been used extensively in the past. In the case of buses with coupled lines we need to generalize it. Let $L(k) = [l_1(k), l_2(k), \cdots, l_n(k)]^T,\ k = 1, 2, \ldots$ be a sequence of random bit vectors transmitted through the bus. It is of course, $V(k) = V_{dd} \cdot L(k)$. From equation (11) we get:

$$\begin{aligned} \overline{E} &= \frac{1}{2}\overline{(V(k) - V(k-1))^T \mathcal{A} (V(k) - V(k-1))} \\ &= \frac{1}{2}\overline{trace\left((V(k) - V(k-1))^T \mathcal{A} (V(k) - V(k-1))\right)} \\ &= \frac{V_{dd}^2}{2} trace\left(\mathcal{A}\,\overline{(L(k) - L(k-1)) \cdot (L(k) - L(k-1))^T}\right) \\ &= \frac{V_{dd}^2}{2} trace\left(\mathcal{A} \cdot \left[2R(0) - R(1) - R^T(1)\right]\right) \end{aligned} \tag{19}$$

where $R(r) = [R_{i,j}(r)]_{i,j=1}^{n} = \overline{L(k) \cdot L^T(k+r)}$ is the autocorrelation matrix of the vector sequence $L(k)$. We define the *transition activity matrix*, $T^a = \left[T_{i,j}^a\right]_{i,j=1}^{n}$ as:

$$T^a = R(0) - \frac{1}{2}\left(R(1) + R^T(1)\right) \tag{20}$$

Then, from expression (19), the expected energy consumption can be written as:

$$\overline{E} = V_{dd}^2 \cdot trace\left(\mathcal{A} \cdot T^a\right) \tag{21}$$

Note that the transition matrix is a generalization of the transition activity and that expression (21) is a generalization of the left expression in (18). The elements of the transition matrix are:

$$T_{i,j}^a = \overline{l_i(k) \cdot l_j(k)} - \frac{1}{2}\left(\overline{l_i(k-1) \cdot l_j(k)} + \overline{l_i(k) \cdot l_j(k-1)}\right) \tag{22}$$

Example 3. Neglecting the coupling between non-adjacent lines in the bus we get a total capacitance lumped network like that of Figure 6 and the admitance matrix $\mathcal{A}$ given by expression (15). In this case, because of the symmetry of $\mathcal{A}$, formula (21) of the expected energy becomes:

$$\overline{E} = \left\{(1+2\lambda)\sum_{i=1}^{n} T_{i,i}^a - 2\lambda \sum_{i=1}^{n-1} T_{i,i+1}^a\right\} \cdot V_{dd}^2 \cdot C_L$$

Example 4. Lets assume the setup of the previous example. If the transmitted bits are independent and uniformly distributed in $\{0,1\}$ then:

$$\overline{l_i(k) \cdot l_j(k+r)} = \begin{cases} 1/2 & \text{if } \ i = j \ \text{ and } \ r = 0 \\ 1/4 & \text{otherwise} \end{cases}$$

and the transition activity matrix is $T^a = \frac{1}{4}\mathcal{I}$, where $\mathcal{I}$ is the identity matrix. In this case the expected energy is:

$$\overline{E} = \frac{n(1+2\lambda)}{4} \cdot V_{dd}^2 \cdot C_L \tag{23}$$

Example 5. In a recent work, permutation of the data bits was proposed as an approach to reduce the expected power consumption [13]. Instead of transmitting the sequence of vectors $L(k) = [l_1(k), l_2(k), \cdots, l_n(k)]^T,\ k = 1, 2, \ldots$ we can transmit the sequence $L_\pi(k) = [l_{\pi(1)}(k), l_{\pi(2)}(k), \cdots, l_{\pi(n)}(k)]^T,\ k = 1, 2, \ldots$, where π is a permutation of the indices $1, 2, \ldots, n$. The desirable in this approach is to minimize opposite and maximize concurrent transitions in adjacent lines by choosing the appropriate permutation. A heuristic approach was presented in [13]. Using the formulation introduced in the previous sections, the problem can be written formally as follows. Let Π be the $n \times n$ permutation matrix corresponding to π. If T^a is the transition activity matrix of the original bus, then the transition activity matrix of the bus with the permuted data bits is:

$$T_\pi^a = \Pi \cdot T^a \cdot \Pi^T$$

and the expected energy consumption is given by the expression:

$$\overline{E}_\pi = V_{dd}^2 \cdot trace\left(\mathcal{A} \cdot \Pi \cdot T^a \cdot \Pi^T\right) \tag{24}$$

The desirable is to minimize expression (24) with respect to the permutation matrix Π. A lower bound of the minimum energy, $\min_\pi \overline{E}_\pi$, can be easily derived analytically if we allow Π to be a doubly stochastic matrix.

4. Energy Limits of Communication

Here we use the energy formula derived in Section to study the relation between energy consumption and information transmission. Information is carried by data or address sequences. In both cases the transmitted sequence of $n - bit$ vectors, $L(k)$, $k = 1, 2, \ldots$, is a random process with statistics that can be approximately estimated or measured [28].

Example 6. Sequences of address vectors are highly predictable; in most cycles k, it is: $L(k) = L(k-1) + 1$ [‡] We can model $L(k)$, $k = 1, 2, \ldots$ approximately as a (first order) Markov chain [29]. Then for $x, y = 0, 1, \ldots, 2^n - 1$ it is:

$$\Pr(L(k) = y | L(k-1) = x) = p_{x,y}$$

with $p_{x,x+1}$ very close to 1. In some cases we can simplify the model of the process even further by assuming that for some small δ it is: $p_{x,x+1} = 1 - (2^n - 1)\delta$ and $p_{x,y} = \delta$ if $y \neq x + 1$.

Now, a measure of the information rate through the bus is needed. We can use the *entropy rate* $\mathcal{H}(L)$ of the random process $L(k)$, $k = 1, 2, \ldots$, defined as [30]:

$$\mathcal{H}(L) = \lim_{m \to \infty} \frac{H\left(L(1), L(2), \ldots, L(m)\right)}{m} \tag{25}$$

where $H\left(L(1), L(2), \ldots, L(m)\right)$ is the entropy of the partial sequence $L(1)$,$L(2)$,..., $L(m)$, that is:

$$H\left(L(1), L(2), \ldots, L(m)\right) = -\sum_{L(1),L(2),\ldots,L(m)} \Pr\left(L(1), L(2), \ldots, L(m)\right) \cdot \log_2\left(\Pr\left(L(1), L(2), \ldots, L(m)\right)\right) \tag{26}$$

The n "physical" bits transmitted during a clock cycle may not be statistically independent, neither to each other nor to the previous bits transmitted. In the sense of the Shannon-McMillan-Breiman theorem, $\mathcal{H}(L)$ equals the expected number of bits needed to express the information content of the vector $L(k)$ [§30].

[‡] For notational purposes, the numbers $0, 1, \ldots, 2^n - 1$ are identified with their binary expansions. This is done throughout the paper.

[§] When for example $L(k)$ is stationary and ergodic.

Example 7. For the Markov address process of the previous example we have:

$$\begin{aligned} H\left(L(1), L(2), \ldots, L(m)\right) &= H(L(1)) + H(L(2)|L(1)) + \ldots + H(L(m)|L(m-1)) \\ &= H(L(1)) + (m-1)H(L(2)|L(1)) \end{aligned}$$

where $H(\cdot|\cdot)$ is the *conditional entropy function.* In the example, $L(k)$ is uniformly distributed in $0, 1, \ldots, 2^n - 1$ and therefore it is:

$$\begin{aligned} H\left(L(2)|L(1)\right) &= -\frac{1}{2^n}\sum_{x,y} p_{x,y} \cdot \log_2\left(p_{x,y}\right) \\ &= -(2^n-1)\cdot\delta\cdot\log_2(\delta) - (1-(2^n-1)\delta)\cdot\log_2(1-(2^n-1)\delta) \end{aligned}$$

So, the entropy rate of the address process is given by:

$$\mathcal{H}(L) = -(2^n-1)\cdot\delta\cdot\log_2(\delta) - (1-(2^n-1)\delta)\cdot\log_2(1-(2^n-1)\delta) \tag{27}$$

Example 8. Now lets consider a bus with the capacitive structure of Figure 6, $n = 8$, $\lambda = 5$, $V_{dd} = 1V$ and $C_L = 100fF$ [¶] The transition energy is given by equation (11) where the admitance matrix $\mathcal{A}$ is given by expression (15). Suppose the bus carries sequences of address vectors that can be modeled by a Markov process like that of example 6 with $\delta = 1/2^{10}$. From expression (27) we get that the process L carries $\mathcal{H}(L) = 2.8$ *bits* (per cycle or per transmission or per bus transition) in average. In addition, the expected energy cost per cycle is:

$$\overline{E} = \sum_{X,Y} \frac{V_{dd}^2 \cdot C_L}{2}\left(Y-X\right)^T \mathcal{A}\left(Y-X\right)\cdot \Pr(L(k-1)=X, L(k)=Y) \tag{28}$$

where X, Y take all $n-bit$ vector values. Evaluating the above expression (with $\Pr(X,Y) = \delta/2^8$ if $Y \neq X+1$ and $\Pr(X, X+1) = [1-(2^8-1)\delta]/2^8$) we get $\overline{E} = 1.37pJ$. So it is:

$$\overline{E}/\mathcal{H}(L) = 1.37pJ/2.8\ bits = 0.49pJ/bit \tag{29}$$

In average, $0.49pJ$ must be dissipated per bit of information that is transmitted through the bus.

The discussion in the example 8 above motivates the following formal definitions.

Definition 1. *Suppose a random process L, of $n-bit$ vectors $L(k)$, $k = 1, 2, \ldots$ is transmitted through a bus. The expected energy per information bit of the process is* [‖]

$$E_b(L) = \overline{E(L(k-1), L(k))}/\mathcal{H}(L) \tag{30}$$

[¶]These are realistic numbers for modern deep sub-micron buses.

[‖]To avoid technicalities we assume that process L is stationary. The definitions and results of this section can be directly expanded to more general classes of processes.

It is important to notice the dependance between the expected energy and the entropy rate as well as to clarify the difference between the expected energy and the expected energy per information bit.

Example 9. Lets examine a trivial bus with only one line which carries a sequence of bits $b(1), b(2), \ldots$. Furthermore, suppose that for some (large) number m and for every $r = 0, 1, 2, \ldots$, exactly one of the bits $b(r2^m), b(r2^m+1), \ldots, b(r2^m+(2^m-1))$ is 1. It is clear that we can encode m bits of information bits into 2^m bits that are physically transmitted. In this case, the entropy rate $\mathcal{H}(b)$, of the sequence $b(k)$, is $m/2^m$ *bits* (per cycle). The expected energy per cycle is $\overline{E} \cong \frac{1}{2^m} \cdot C \cdot V_{dd}^2$ approximately. Letting m become large, $\overline{E}$ becomes arbitrarily small! and the energy per information bit, $\overline{E}/\mathcal{H}(b) = \frac{1}{m} \cdot C \cdot V_{dd}^2$ becomes arbitrarily small as well! Unfortunately, the entropy rate $\mathcal{H}(b)$ tends to zero too.

Example 10. Lets alter the encoding in the previous example. Lets assume that for every r, m, we allow exactly one $k \in \{0, 1, \ldots, 2^m - 1\}$ such that $b(r2^m + k) = b(r2^m + k + 1)$. All other consecutive (physical) bits have complementary binary values. Again, we can encode m information bits into 2^m physical bits. Letting m became large we have that $\overline{E} \to \frac{1}{2} C V_{dd}^2$, $\mathcal{H}(b) \to 0$ and $\overline{E}/\mathcal{H}(b) \to \infty$.

Another important parameter is the size of the bus, that is "the bandwidth of the communication channel". The following example shows that information rate cannot be considered independently of the size of the bus.

Example 11. Lets consider the *one hot encoding* scheme where m bits are encoded into 2^m bus lines simply by having exactly one line carrying a 1 and the rest of them carrying zeros. The energy loss per cycle is bounded above by a constant e, i.e. $\overline{E} < e$. Letting m become large we have that the energy per bit, which is less than e/m, can become arbitrarily small. At the same time, information rate, that is equal to m, becomes arbitrarily large. It looks like a win-win situation but unfortunately it is not. We need to consider the size of the bus, 2^m lines, that grows exponentially with m. Therefore, the energy per information bit must be considered with respect to the information rate and the size of the bus.

Definition 2. *Let L be a random process, of $n-$ bit vectors $L(k)$, $k = 1, 2, \ldots$, that is transmitted through a bus with n lines. The utilization α, $0 \leq \alpha \leq 1$, of the bus by the process is:*

$$\alpha = \mathcal{H}(L)/n \tag{31}$$

The utilization α is the percentage of the "bandwidth" of the bus that is occupied by the transmission of process L. In example 11, the utilization approaches zero as the parameter m tends to infinity.

Since it is always desirable to transmit information at low energy cost, the interesting question to answer is: *What is the minimum energy required to transmit a bit of information through the bus?*

Definition 3. *The minimum required energy per information bit transmitted, through a bus with n lines that is utilized by a factor α, is:*

$$E_b^*(\alpha) = \min_{L\,:\,\mathcal{H}(L)/n=\alpha} E_b(L) \tag{32}$$

The following theorem provides us with the limit of communication energy. The knowledge of the energy cost $E(x,y)$, of the transition from a state x to a state y, is required. Equation (11) may be used.

Theorem 1. *When the bus is utilized by a factor α, $0 \le \alpha \le 1$, the minimum energy per information bit is:***

$$E_b^*(\alpha) = \ln(2) \cdot \left(\gamma - \frac{1}{\frac{\partial}{\partial\gamma} \ln\left(\ln\left(\mu(\gamma)\right)\right)} \right)^{-1} \tag{33}$$

where γ is the positive solution of the equation:

$$\alpha = -\frac{1}{n\,\ln(2)}\,\gamma^2 \frac{\partial}{\partial\gamma}\left(\frac{\ln\left(\mu(\gamma)\right)}{\gamma}\right) \tag{34}$$

and $\mu(\gamma)$ is the maximal eigenvalue of the matrix:

$$W(\gamma) = \left[e^{-\gamma\,E(x,y)} \right]_{x,y=0}^{2^n-1} \tag{35}$$

Furthermore, the minimum is attained by a stationary, ergodic, Markov process having transition probabilities:

$$\Pr\left(y|x\right) = \frac{1}{\mu(\gamma)}\,\frac{g_y}{g_x}\,e^{-\gamma\,E(x,y)}$$

where $g = (g_x)_x$ is the right eigenvector of matrix $W(\gamma)$ corresponding to $\mu(\gamma)$.

Proof: The proof of the theorem is sketched in Appendix B. Details can be found in [31].

The maximum utilization of the bus, $\alpha = 1$, corresponds to the case where the bits of the random vector process, $L(k) = [l_1(k), l_2(k), \cdots, l_n(k)]^T,\ k = 1, 2, \ldots,$ are independent and uniformly distributed in $\{0, 1\}$. The rate of this process is of course n bits per cycle and the expected energy per cycle is:

$$\overline{E}_u = \frac{1}{2^{2n}} \sum_{x,y=0}^{2^n-1} E(X,Y)$$

For a bus with the capacitive structure of Figure 6, the expected energy per cycle is given by (23).

**Although expression (33) holds for the case of energy function (11), for general cost functions there is one exception: if there are some c, θ_x, $x = 0, 1, \ldots, 2^n - 1$ such that $E(x,y) = c + \theta_x - \theta_y$, for every x, y, then it is $E_b^*(\alpha) = c/(\alpha\, n)$.

Example 12. Consider again a bus with the capacitive structure of Figure 6. The energy function of the bus is given by expression (11) with the admitance matrix $\mathcal{A}$ given by (15). In Figures 7 we see the normalized minimum energy per information bit, $E_b^*(\alpha)/\overline{E}_u$, as a function of the bus utilization, α. The three graphs correspond to the cases: $n = 2, 4$ and 8 with $\lambda = 5$. Energy increases rapidly around $\alpha = 0$ and $\alpha = 1$ and it is zero at $\alpha = 0$. This confirms example 11, that is, the energy per information bit can be arbitrarily low for sufficiently small rate. Note that this would not be true if the energy expression (11) included leakage or other terms that would contribute to energy consumption when *there is no transition in the bus.*

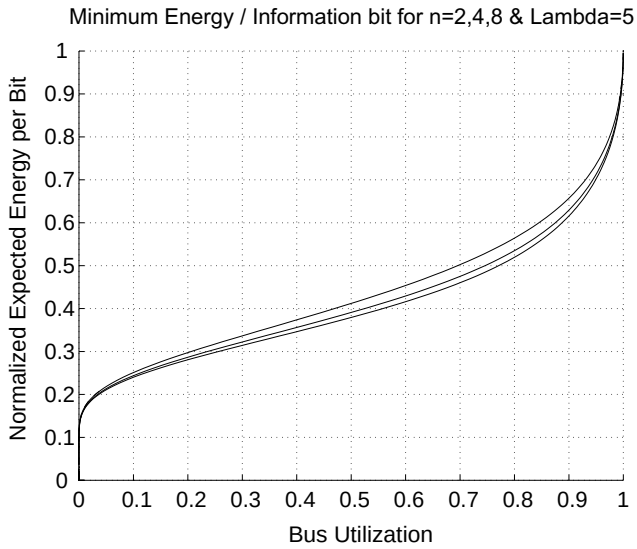

Fig. 7. Communication energy limits for a family of buses

5. Energy Efficiency of Communication

Theorem 1 provides us with an energy efficiency measure of information transmission. We can estimate or measure the entropy rate of data or instruction sequences [28], estimate or measure the energy consumption and finally compare this amount of energy with the minimum possible energy at the same rate that is given by the theorem.

Definition 4. *Consider a bus with n lines and a random process, of bit vectors $L(k)$, $k = 1, 2, \ldots$, transmitted through the bus. The communication efficiency of process L is:*

$$\eta(L) = \frac{E_b^*(\mathcal{H}(L)/n)}{E_b(L)} \tag{36}$$

The meaning of communication efficiency is the following: The power consumption, when transmitting process L, is (definition 4.1):

$$P_L = f \cdot \mathcal{H}(L) \cdot E_b(L),$$

where f if the clocking frequency of the bus. Now let L^* be an energy-optimal process of the same rate, i.e., $E_b(L^*) = E_b^*(\mathcal{H}(L)/n)$ and $\mathcal{H}(L^*) = \mathcal{H}(L)$. Let L^* be transmitted and the bus be clocked at a frequency f^*. Then the corresponding power consumption is:

$$\begin{aligned} P_{L^*} &= f^* \cdot \mathcal{H}(L) \cdot E_b(L^*) \\ &= f^* \cdot \mathcal{H}(L) \cdot \eta(L) \cdot E_b(L) \end{aligned}$$

The two transmissions result in the same power consumption, $P_{L^*} = P_L$, if $f = \eta(L) \cdot f^*$. Therefore, if we fix the power level, process L^* achieves $1/\eta(L)$ times higher communication rate than that process L achieves.

Example 13. Consider the setup of example 8. The entropy rate of the process is 2.8 *bits* per cycle, the expected energy per information bit is $0.49pJ/bit$ and the bus has $n = 8$ lines. For utilization $\alpha = 2.8/8 = 0.35$ the minimum energy per information bit is: $E_b^*(0.35) = 0.91 * V_{dd}^2 \cdot C_L/bit = 0.091pJ/bit$. Therefore, the communication efficiency of the process in the example is: $\eta(L) = 0.091/0.49 = 0.19$. So, for a given energy amount, the amount of information that can be transmitted is $1/0.19 = 5.26$ times larger than that carried by process L.

6. Conclusions

The expected energy consumption and the limits of communication energy were derived using an analytical energy model appropriate for buses with coupled lines. The transition activity matrix was introduced to generalize the transition activities of individual lines and the problem of optimal permutation of bus lines was analytically formulated. The communication energy efficiency factor was introduced to measure the redundancy in energy consumption.

7. Appendix A

Defining the $n \times n$ diagonal matrix $C^r = diag(C_1^r, C_2^r, \ldots, C_n^r)$ we can write equation (4) in vector form as:

$$I(L,t) = C^r \frac{\partial V(L,t)}{\partial t} \tag{37}$$

Integration of equation (1) over $0 \leq x \leq L$ gives:

$$I(0,t) = I(L,t) + \int_0^L A(x) \frac{\partial V(x,t)}{\partial t} dx \tag{38}$$

Integrating (38) over $0 \leq t \leq T$ and using (37) we get:

$$\begin{aligned}
\int_0^T I(0,t) &= \int_0^T C^r \frac{\partial V(L,t)}{\partial t} dt + \int_0^T \left(\int_0^L A(x) \frac{\partial V(x,t)}{\partial t} dx \right) dt \\
&= C^r (V^f - V^i) + \int_0^L \left(A(x) \int_0^T \frac{\partial V(x,t)}{\partial t} dt \right) dx \\
&= C^r (V^f - V^i) + \left(\int_0^L A(x)\, dx \right) (V^f - V^i) \qquad (39)
\end{aligned}$$

From equations (3) and (7) we have:

$$\begin{aligned}
E^{Vdd} = \sum_{i=1}^n E_i^{Vdd} &= \sum_{i=1}^n V_i^f \int_0^T \left\{ I_i(0,t) + C_i^d \frac{\partial V_i(0,t)}{\partial t} \right\} dt \\
&= \sum_{i=1}^n V_i^f \int_0^T I_i(0,t) dt + \sum_{i=1}^n V_i^f C_i^d \left(V_i^f - V_i^i \right)
\end{aligned}$$

Defining the $n \times n$ diagonal matrix $C^d = diag(C_1^d, C_2^d, \ldots, C_n^d)$ we can write:

$$E^{Vdd} = (V^f)^T \int_0^T I(0,t) dt + (V^f)^T C^d \left(V^f - V^i \right)$$

Finally, replacing (39) in the equation above we get:

$$E^{Vdd} = (V^f)^T \left[\int_0^L A(x)\, dx + C^d + C^r \right] (V^f - V^i)$$

Equations (9) and (10) imply that $\mathcal{A} = \int_0^L A(x)\, dx + C^d + C^r$, therefore it is:

$$E^{Vdd} = (V^f)^T \mathcal{A} \left(V^f - V^i \right)$$

8. Appendix B

Here we sketch the proof of Theorem 1. For simplicity we assume that the energy function E is symmetric, this it is true for expression (11). Observe that given a stationary processes L, with transition probability matrix $P = [p_{i,j}]_{i,j=0}^{2^n-1}$, we can define a stationary Markov process L^M that has transition probability matrix and stationary probability vector q equal to those of L. It can be verified directly that $qP = q$, that processes L^M and L have the same expected energy, and that L^M has entropy rate greater or equal to that of L. Therefore the minimum in expression (32) is achieved within the subset of Markov processes.

A Markov process can be defined by either the pair (P, q) or the joint probability matrix $\Pi = [\pi_{i,j}]_{i,j=0}^{2^n-1}$, with $\pi_{i,j} = q_i\, p_{i,j}$. Moreover, we can express the entropy

rate as:

$$\mathcal{H}(\Pi) = -\sum_{i,j} \pi_{i,j} \log_2(\frac{\pi_{i,j}}{\sum_k \pi_{i,k}}) \tag{40}$$

It is of course: $\mathcal{H}(\Pi) = \mathcal{H}(L^M)$. When transmitting process L^M, the expected energy (per cycle), as a function of Π, is:

$$\overline{E}(\Pi) = \sum_{i,j} \pi_{i,j}\, E(i,j) \tag{41}$$

With the above setup the minimization problem (32) is written as:

$$E_b^*(\alpha) = \frac{1}{an} \cdot \min_{\Pi\,:\,\mathcal{H}(\Pi)=\alpha\, n} \overline{E}(\Pi) \tag{42}$$

Since Π is a joint probability matrix of a stationary process, it must satisfy the following constrains:

$$\pi_{i,j} \geq 0\ , \quad \sum_{i,j} \pi_{i,j} = 1\ , \quad \sum_j \pi_{i,j} = \sum_k \pi_{k,i} \tag{43}$$

Ignoring the factor $1/(\alpha n)$ and the change of logarithmic basis the Lagrangian of the problem is:

$$\begin{aligned} \mathcal{L}(\Pi, v, \lambda, \mu) &= v\left(\sum_{i,j} \pi_{i,j} - 1\right) + \sum_{i,j} \lambda_i\,(\pi_{i,j} - \pi_{j,i}) - \\ &- \theta\left(\sum_{i,j} \pi_{i,j}\, \ln\left(\frac{\pi_{i,j}}{\sum_k \pi_{i,k}}\right) - \alpha n\right) + \sum_{i,j} \pi_{i,j}\, E(i,j) \end{aligned} \tag{44}$$

and the partial derivative of the Lagrangian with respect to the variable $\pi_{i,j}$ is :

$$\frac{\partial \mathcal{L}}{\partial \pi_{i,j}} = v + \lambda_i - \lambda_j - \theta\, \ln\left(\frac{\pi_{i,j}}{\sum_k \pi_{i,k}}\right) + E(i,j) \tag{45}$$

We want to solve the system of equations $\frac{\partial \mathcal{L}}{\partial \pi_{i,j}} = 0$ and to find the stationary points. First note that $\theta = 0$ implies: $E(i,j) = -v - \lambda_i + \lambda_j$ which has been excluded (see footnote of theorem 1) and it is incompatible with the expression (11). Suppose now that $\theta \neq 0$. Equations $\frac{\partial \mathcal{L}}{\partial \pi_{i,j}} = 0$, $i,j = 0, 1, \ldots, 2^n - 1$ imply:

$$\frac{\pi_{i,j}}{\sum_k \pi_{i,k}} = e^{\frac{v + \lambda_i - \lambda_j + E(i,j)}{\mu}} \tag{46}$$

Parameters v, λ_i and μ are real, so applying the transformation : $f = e^{v/\mu}$, $g_i = e^{-\lambda_i/\mu}$ and $\gamma = -1/\mu$, we have $f > 0$ and $g_i > 0$ for every i. Moreover, expression (46) becomes:

$$\frac{\pi_{i,j}}{\sum_k \pi_{i,k}} = f\,\frac{g_j}{g_i} e^{-\gamma E(i,j)} \tag{47}$$

Summing both sides of (47) over j we get $g_i = f \sum_j e^{-\gamma E(i,j)} g_j$ which written in matrix form gives: $g = fWg$ with $g = (g_0, g_1, \ldots, g_{2^n-1})^T$ and $W = [e^{-\gamma E(i,j)}]_{i,j}$. The matrix fW is always positive and therefore g must be its unique (up to a factor) positive eigenvector [32]. Furthermore, f must be the inverse of the maximal eigenvalue μ of W. The symmetry of the energy cost function E gives: $g' = f\, g'W$ or more explicitly :

$$g_j = f \sum_i g_i e^{-\gamma E(i,j)} \tag{48}$$

We define the probability vector $q = (g_0^2, g_1^2, \ldots, g_{2^n-1}^2)/\|g\|^2$ and matrix $P = [p_{i,j}]_{i,j=0}^{2^n-1}$ with $p_{i,j} = \frac{\pi_{i,j}}{\sum_k \pi_{i,k}}$. It can be verified that P is a stochastic matrix and that:

$$\begin{aligned}
\sum_i q_i p_{i,j} &= \sum_i \frac{g_i^2}{\|g\|^2} f \frac{g_j}{g_i} e^{-\gamma E(i,j)} \\
&= \frac{g_j}{\|g\|^2} f \sum_i g_i e^{-\gamma E(i,j)} \\
&= \frac{g_j^2}{\|g\|^2} = q_j
\end{aligned}$$

So q is a left eigenvector of P. Finally, it is easy to show that $\pi_{i,j} = q_i p_{i,j}$, i.e.:

$$\pi_{i,j} = \frac{g_i\, g_j}{\|g\|^2} f e^{-\gamma E(i,j)} \tag{49}$$

Also note that $\pi_{i,j} = \pi_{j,i}$. From (40), (47) and (49) we have:

$$\begin{aligned}
\ln(2) \cdot \mathcal{H}(\Pi) &= -\sum_{i,j} \pi_{i,j} \ln(\frac{\pi_{i,j}}{\sum_k \pi_{i,k}}) \\
&= -\sum_{i,j} \pi_{i,j} \ln(\frac{1}{\mu} \cdot \frac{g_j}{g_i} e^{-\gamma E(i,j)}) \\
&= -\sum_{i,j} \pi_{i,j} \left\{-\ln(\mu) + \ln(g_j) - \ln(g_i) - \gamma E(i,j)\right\} \\
&= \sum_{i,j} \pi_{i,j} \left\{\ln(\mu) + \gamma E(i,j)\right\} \\
&= \ln(\mu) + \gamma \sum_{i,j} \pi_{i,j}\, E(i,j) \\
&= \ln(\mu) + \frac{\gamma}{\mu} \sum_{i,j} \frac{g_i\, g_j}{\|g\|^2} e^{-\gamma E(i,j)}\, E(i,j) \\
&= \ln(\mu) - \frac{\gamma}{\mu} \cdot \frac{\partial \mu}{\partial \gamma}
\end{aligned} \tag{50}$$

where in the last step we used a result from perturbation theory [32]. Manipulating (50) we get expression (34). Similarly, starting from (41) we have:

$$\begin{aligned}
\overline{E}(\Pi) &= \sum_{i,j} \pi_{i,j}\, E(i,j) \\
&= \frac{1}{\mu} \sum_{i,j} \frac{g_i\, g_j}{\|g\|^2}\, e^{-\gamma E(i,j)}\, E(i,j) \\
&= -\frac{1}{\mu} \cdot \frac{d\mu}{d\gamma} \\
&= -\frac{\partial}{\partial \gamma} \ln\left(\mu(\gamma)\right)) \qquad (51)
\end{aligned}$$

The minimal expected energy per information bit is $\overline{E}/(\alpha n)$ which after some manipulation gives expression (33).

Acknowledgments

The authors acknowledge support from the MARCO Focus Research Center on Interconnect funded at MIT through a subcontract from Gatech. The program is supported by MARCO and DARPA. Paul Sotiriadis has been partially supported by the Alexander S. Onassis Public Benefit Foundation, the Greek Section of Scholarships and Research.

Also, the authors would like to thank professor Andreas Andreou for his helpful suggestions on the manuscript.

9. References

1. A. Chandrakasan, R. Brodersen , *Low Power Digital Cmos Design* , Kluwer Academic Publishers 1995.
2. M. Pedram, "Power minimization in IC design: Principles and applications", *ACM Transactions on Design Automation Electronic Systems, Vol. 1, No. 1, pp. 1-54, Jan. 1996.*
3. F. Najm, "A survey of power estimation techniques in VLSI circuits", *IEEE Transactions on VLSI Systems, Vol. 2, pp. 446-455, Dec. 1994.*
4. A. Ghosh, S. Devadas, K. Keutzer, J. White, "Estimation of average switching activity in combinatorial and sequential circuits", *Proceedings of ACM/IEEE Design Automation Conference, pp. 253-259, June 1992.*
5. R. Marculescu, D. Marculescu, M. Pedram, "Efficient power estimation for highly correlated input streams", *Proceedings of ACM/IEEE Design Automation Conference, pp. 628-634, June 1995.*
6. S. Borkar, "Low power design challenges for the decade", *Asia and South Pacific Design Automation Conference 2001, pp. 293-296.*
7. T. Sakurai, "Design Challenges for $0.1\mu m$ and Beyond", *Asia and South Pacific Design Automation Conference 2000, pp. 553-558.*
8. K. Yamashita, S. Odanaka, "Interconnect Scaling Scenario Using a Chip Level Interconnect Model", *IEEE transactions on electron devices Vol. 47, No. 1, Jan. 2000.*

9. N. Menezes, L. Pillegi, "*Analyzing on-chips Interconnect effects*", Design of High Performance Microprocessor Circuits, A. Chandrakasan, W.J. Bowhill, F. Fox, Eds., IEEE Editions 2001, Chapter 16.
10. S. Das, W. Smith, C. Paul, "Modeling of data bus structures using numerical methods", *International Symposium on Electromagnetic Compatibility, 1993, pp. 409-414.*
11. C. Cheng, J. Lillis, S. Lin, N. Chang, *Interconnect Analysis and Synthesis*, John Wiley and Sons, 2000.
12. L. Pileggi, "Coping with RC(L) interconnect design headaches", *IEEE/ACM International Conference on CAD 1995, pp. 246-253.*
13. L. Macchiarulo, E. Macii, M. Poncino, "Low-Energy Encoding for Deep-Submicron Address Buses", *IEEE/ACM International Symposium on Low Power Electronics and Design 2001, pp. 176-181.*
14. S. Ramprasad, N. Shanbhag, I. Hajj, "Information-Theoretic Bounds on Average Signal Transition Activity", *IEEE Trans. on VLSI, Vol. 7, No. 3, Sept. 1999.*
15. 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.*
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", *Proceedings of the Great Lakes Symposium, VLSI, pp. 77-82, 1997.*
17. P. Sotiriadis and A. Chandrakasan, "Low Power Bus Coding Techniques Considering Inter-wire Capacitances", *Proceedings of the Custom Integrated Circuits Conference, May 2000.*
18. P. Sotiriadis, A. Chandrakasan, "Coding for Energy Reduction in Deep Sub-Micron Technologies, *The Transition Pattern Coding Approach", To appear in the IEEE Transactions on Circuits and Systems.*
19. J. Henkel, H. Lekatsas, "A2BC: Adaptive Address Bus Coding for Low Power Deep Sub-Micron Design, *Design Automation Conference, pp. 744-749, 2001.*
20. T. Sakurai, "Closed-form expressions for interconnection delay, coupling, and crosstalk in VLSIs", *IEEE Transactions on Electron Devices, Vol. 40, Issue 1, Jan. 1993, pp. 118-124.*
21. K. Tang, E. Friedman, "Peak noise prediction in loosely coupled interconnect [VLSI circuits]", *International Symposium on Circuit and Systems 1999, Vol.1, pp. 541-544.*
22. T. Sakurai, S. Kobayashi, M. Noda, "Simple expressions for interconnection delay, coupling and crosstalk in VLSI", *International Symposium on Circuit and Systems 1991, Vol. 4, pp. 2375-2378.*
23. M. Becer, I. Hajj, "An analytical model for delay and crosstalk estimation in interconnects under general switching conditions", *IEEE International Conference on Electronics, Circuits and Systems, 2000, vol.2, pp. 831-834.*
24. J. Davis, J. Meindl, "Compact Distributed RLC Interconnect Models Part II: Coupled Line Transient Expressions and Peak Crosstalk in Multilevel Networks", *IEEE Transactions on Electron Devices, Vol. 47, No. 11, Nov. 2000.*
25. S. Das, W. Smith, C. Paul, "Modeling of data bus structures using numerical methods", *International Symposium on Electromagnetic Compatibility, 1993, pp. 409-414.*
26. C. Cheng, J. Lillis, S. Lin, N. Chang, *Interconnect Analysis and Synthesis*, John Wiley and Sons, 2000.
27. J. Rabaey, *Digital Integrated circuits*, Prentice Hall 1996.
28. R. Marculescu, D. Marculescu, M. Pedram, " Probabilistic Modeling of Dependencies During Switching Activity Analysis", *IEEE Transactions Computer-Aided Design of Integrated Circuits and Systems*, vol. 17, pp. 73-83, June 1998.

29. R. Gallager, *Discrete Stochastic Processes*, Kluwer Academic Publishers, 1998.
30. T. Cover, J. Thomas, *Elements of Information Theory*, John Wiley and Sons, 1991.
31. P. Sotiriadis, *Interconnect Modeling and Optimization in Deep Sub-Micron Technologies*, Ph.D. Thesis, Massachusetts Institute of Technology, May 2002.
32. R. Horn, C.Johnson, *Matrix Analysis*, Cambridge University Press 1994.