Maximum Achievable Energy Reduction Using Coding with Applications to Deep Sub-Micron Buses

Paul P. Sotiriadis
pps@mit.edu

Anantha Chandrakasan
anantha@mtl.mit.edu

Vahid Tarokh
vahid@mit.edu

Department of EECS, Massachusetts Institute of Technology, Cambridge, MA 02139

Abstract: In this work we ask: How much can we reduce the power consumption in deep-sub-micron buses using coding techniques? We answer the question in two steps. First we give the minimum energy per information bit required for communicating through deep-sub-micron buses. Then, we show that the minimum energy is asymptotically achievable using coding. In addition, a simple differential coding scheme is proposed that achieves most of the possible energy reduction. The methodology used here also applies to more general communication and computation models.

Introduction

During the last few years researchers have examined devices in VLSI circuits from an information and communication theory perspective [1],[2],[3],[4]. A lot of exciting new results have been produced including the relation between information rate and transition activity in binary sequences [3]. The transition activity T_a of a circuit node (or line) is a useful power measure when the node (or line) is *de-coupled* from any other active node in the circuit. If this is true, then the transition activity is translated into power consumption equal to $P = T_a C_L V_{dd}^2/2$, where C_L is the capacitance between the node (line) and ground. As technology scales down not many nodes, neither lines can be regarded isolated or shielded any more. Coupling between nodes (or distributed lines) implies that power dissipation depends also on their *cross-activities* (see [7] for an extensive discussion) and therefore the simple power formula of the transition activity that has been used extensively in the past is not valid here. In many cases like deep-sub-micron buses, the coupling between lines is much stronger than the coupling between individual lines and ground. The purpose of this work is to answer the fundamental questions:

1) What is the minimum energy required per information bit for communicating through deep sub-micron buses? 2) Is the minimum energy achievable using coding?

To study the relation between power dissipation and information rate in deep-sub-micron buses, we introduce a new mathematical framework based on the notion of noiseless finite-state (NLFS) channel (def. 1). Buses are special cases of symmetric NLFS channels. In Theorem 1 the fundamental lower bound of the energy/bit is given. Theorem 2 shows that the bound is asymptotically achievable using coding. Finally, a class of simple differential coding schemes is presented. The schemes have low encoding and decoding complexity and provide significant energy reduction when used with DSM buses. The details of the proofs along with extensive discussion and generalizations can be found in [5].

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 is partially supported by the Alexander S. Onassis Public Benefit Foundation, the Greek Section of Scholarships and Research.

1. Information Transmission and Energy Cost in NLFS Channels; Definitions

Consider a single isolated line in a digital circuit carrying a sequence of binary values $x(1), x(2), x(3), \ldots$. The energy dissipated at time k by charging and discharging the parasitic capacitance between the line and ground depends only on the two values $x(k-1)$ and $x(k)$, i.e. the transition of the line (and of course the capacitance and V_{dd}). This property can be generalized:

Definition 1: A *noiseless finite-state* (NLFS) channel is a discrete communication device of some alphabet A and such that the energy cost E for transmitting $x(k+1)$ depends only on $x(k+1)$ and $x(k)$, that is, the energy is a function of the transition, $E = E(x(k) \to x(k+1))$.

The cost of transmitting the sequence $x(1), x(2), \ldots, x(m)$ assuming the channel starts with $x(0)$ is,

$$E(x(0), x(1), \ldots, x(m)) = \sum_{k=0}^{m-1} E(x(k) \to x(k+1)) \qquad (1)$$

Example 1: For the case of a bus with n *isolated* lines each having parasitic capacitance C_L to ground, the energy dissipated during the transition from $w = (w_1, \ldots, w_n)$ to $z = (z_1, \ldots, z_n)$ is:

$$E(w \to z) = \left(\sum_{i=1}^{n} w_i \oplus z_i \right) \frac{V_{dd}^2 C_L}{2} \qquad (2)$$

Example 2: A deep sub-micron (DSM) technology bus can be approximately modeled as in Figure 1 [6]. The boundary capacitances are due to fringing effects and the energy dissipated during the transition from w to z is given by [7]:

$$E(w \to z) = (z-w)^T \cdot C \cdot (z-w)\ V_{dd}^2/2 \qquad (3)$$

(the formula holds for the case of distributed wires with inductive coupling as well). The matrix C is given by expression (4) and $\lambda = C_I/C_L$.

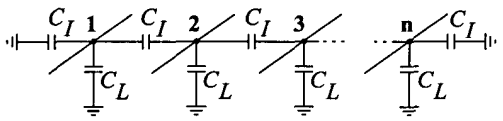


Figure 1: Shielded DSM bus.

$$C = \begin{bmatrix} 1+2\lambda & -\lambda & 0 & \dots & 0 \\ -\lambda & 1+2\lambda & -\lambda & \vdots & 0 \\ 0 & -\lambda & & \vdots & \vdots \\ \vdots & \vdots & \vdots & 1+2\lambda & -\lambda \\ 0 & 0 & \dots & -\lambda & 1+2\lambda \end{bmatrix} \cdot C_L \qquad (4)$$

Consider a NLFS channel with a given transition energy function E and alphabet A, carrying a stochastic process $X = \{X(k)\}_k$.

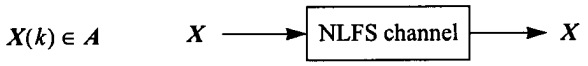


Figure 2: Noiseless finite-state channel.

To quantify the information rate through the channel, we pose the minimal assumption that the process is *stationary*[1]. Then, the expected number of transmitted information-bits (i-bits) per channel use equals the binary entropy rate of the process:

$$H(X) = \lim_{m \to \infty} \frac{H(X(1), X(2), \dots, X(m))}{m}$$

(in i-bits/channel use). The channel has alphabet A and so the maximum bit rate is the capacity of the channel, $H_m = \log_2|A|$. Since the random process X carries only $H(X)$ i-bits per channel use we can say that X utilizes the bus by a factor of

$$\rho(X) = H(X)/H_m$$

We call $\rho(X)$ the *utilization* of the channel by the random process X. It is of course $0 \le \rho(X) \le 1$.

Example 3: A bus with n lines can carry $H_m = n$ bits at a time. Suppose that $n = 2$ and X is the stationary Markov information source of Figure 3.

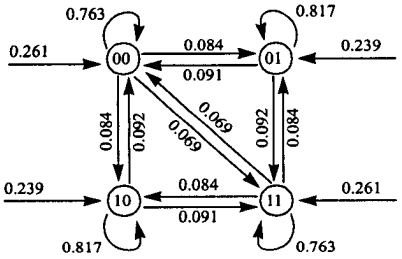


Figure 3: A stationary Markov information source.

The process X has entropy rate $H(X) = 1.02$ and therefore it utilizes the bus by the factor $\rho(X) = 1.02/2 = 0.501$.

Regardless of the particular structure of a NLFS channel the information transmission is always associated with certain energy consumption. The stationarity[1] of X allows the following definitions:

1. The stationarity is stronger than what we need. All the results can be directly extended for cyclostationary or other more general classes of random processes.

Definitions: The *average energy per channel use* of the stationary stochastic process X is:

$$E_{av}(X) = \lim_{m \to \infty} \frac{1}{m} \sum_{k=0}^{m-1} \overline{E(X(k) \to X(k+1))} \qquad (5)$$

where the expectation (over-line) is taken with respect to the random variables $X(k)$ and $X(k+1)$. Since $H(X)$ i-bits per channel use are transmitted in average we define the *average energy per information bit* of the process X to be:

$$E_b(X) = E_{av}(X)/H(X)$$

Example 4: Consider the DSM bus of Figure 1 with $n = 2$ and $\lambda = 5$. The energy costs of the 16 transitions can be calculated using (3). They are shown in Figure 4 in multiples of $V_{dd}^2 C_L$. (the costs are symmetric).

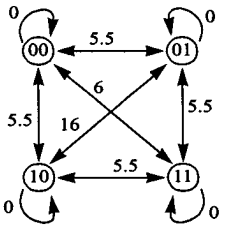


Figure 4: Transition costs of a DSM bus with $n = 2$ and $\lambda = 5$.

Suppose now that the Markov process of the example 3 is transmitted through this bus. From the diagrams of Figure 3 and Figure 4 we can easily calculate the average energy per channel use to be $E_{av}(X) = 1.18 \cdot V_{dd}^2 C_L$. The entropy rate of the process is $H(X) = 1.02$ (i-bits per bus use) and therefore it has energy per transmitted i-bit equal to $E_b(X) = 1.15 \cdot V_{dd}^2 C_L$ (Joule per bit). This must be compared to the energy per i-bit when the bus is completely utilized. This is exactly the case that the uniform *i.i.d.* process U is transmitted. Again, from expression (3) we calculate $E_{av}(U) = 5.5 \cdot V_{dd}^2 C_L$. Since two i-bits are transmitted at each bus use we also have that $E_b(U) = 2.75 \cdot V_{dd}^2 C_L$ (Joules per bit). The table below summarizes the entropy rate (in i-bits per channel use), the utilization of the bus, the average energy per channel use (in $V_{dd}^2 C_L$ units) and the average energy per transmitted i-bit (in $V_{dd}^2 C_L$ units) of the two processes X and U.

	H	ρ	E_{av}	E_b
X:	*1.02*	*0.501*	*1.18*	*1.15*
U:	*2*	*1*	*5.5*	*2.75*

Table 1: Information rate & energy rates of the processes X, U

2. MINIMUM ENERGY PER I-BIT TRANSMITTED THROUGH A NLFS CHANNEL

In example 4 the process X carries data with an energy cost per i-bit less than the half of that of U. On the other hand, the bit rate of X is about the half of the bit rate of U. In most communication channels, like buses, data must be transmitted at a given rate, or equivalently the channel must be utilized by a given factor. It is natural to ask the following question:

What is the minimum energy required per i-bit transmitted through a NLFS channel when the channel is utilized by a factor α *?*

Mathematically the problem is to find $E_b^*(\alpha)$ with $0 < \alpha \leq 1$:

$$E_b^*(\alpha) = \min_{\rho(X) = \alpha} E_b(X) \tag{6}$$

with X being stationary (or a more general process). The answer is given by Theorem 1. The details of the proof can be found in [5].

Theorem 1: *For a NLFS channel utilized by a factor* $\alpha \in (0, 1]$ *the* ***minimum energy per i-bit*** *is given by equation* (7)[1] *with* γ *being the unique non-negative number satisfying* (8):

$$E_b^*(\alpha) = \frac{1/\log_2(e)}{\gamma - \dfrac{1}{\dfrac{\partial}{\partial\gamma}\ln(\ln(\mu(\gamma)))}} \tag{7}$$

$$\alpha = -\frac{\log_2(e)}{H_{max}}\gamma^2 \frac{\partial}{\partial\gamma}\left[\frac{\ln(\mu(\gamma))}{\gamma}\right] \tag{8}$$

The number $\mu(\gamma)$ *is the maximal positive eigenvalue of the matrix:* $W(\gamma) = [e^{-\gamma E(w \to z)}]_{w,z \in A}$.

Example 5: The minimum possible energy per i-bit transmitted through the bus of example 4 when the bus is utilized by a factor $\alpha = 0.501$ is indeed $1.15 \cdot V_{dd}^2 C_L$. One of the stochastic processes achieving this bound is the one generated by the information source of Figure 3.

In the following example we demonstrate the result of Theorem 1 for the cases of some DSM buses.

Example 6: In Figure 5 we see the minimum possible energy per i-bit communicated through DSM buses as a function of their utilization. The buses have $n = 2, 4, 8$, $\lambda = 5$ and structure as in Figure 1. For total utilization, $\alpha = 1$, the minimum possible energy per i-bit is $2.75 \cdot V_{dd}^2 C_L$. This is exactly the expected energy per i-bit when i.i.d. and uniform data is transmitted. For utilization slightly less than one there is a steep drop in the energy/i-bit. The slope is infinite at $\alpha = 1$. Thus, insignificant sacrifice of bit rate results to a significant reduction in energy per i-bit (see Theorem 2 for the achievability of $E_b^*(\alpha)$). The energy/i-bit is cut to half for utilizations 0.683 , 0.734 and 0.765 when $n = 2, 4$ and 8 respectively. When the utilization approaches zero, the energy/i-bit approaches zero as well. Therefore, by reducing the bit rate appropriately, ***data can be transmitted at as low energy/i-bit as is desired.***

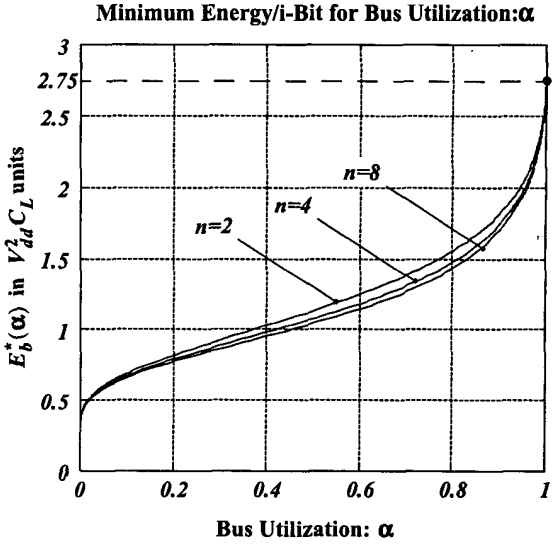


Figure 5: Minimum achievable energy per transmitted i-bit through DSM buses with $n = 2, 4, 8$, $\lambda = 5$ and utilization α .

Some of the aforementioned energy properties of buses hold for the general case of NLFS channels that have non-trivial transition costs i.e. - when information cannot be transmitted at no cost.

2.1 Buses with isolated lines - The Entropy Bound

In the case of buses with isolated lines, i.e. $\lambda = 0$, that are coupled to ground through capacitances of the same size, C_L , the result of Theorem 1 simplifies to: $E_b^*(\alpha) = \dfrac{2h^{-1}(\alpha)}{\alpha} \cdot \dfrac{V_{dd}^2 C_L}{4}$.

The function h^{-1}: $[0, 1] \to [0, 1/2]$ is the inverse of the binary entropy function h when restricted to the domain $[0, 1/2]$. This special case was also established in [3]. We call this bound, the *Entropy Bound* and we note that it is independent of the number of lines. This is in contrast to the case that there is coupling between the lines as shown in Figure 5 for DSM buses.

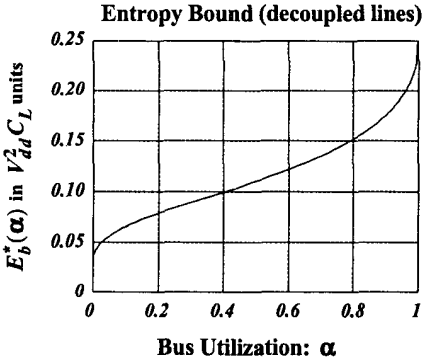


Figure 6: The entropy bound on the energy/i-bit for decoupled lines as a function of the bus utilization α .

1. There is one exception: if there are some c, η_z, $z \in A$ s.t. $E(w \to z) = c + \eta_w - \eta_z$ then $E_b^*(\alpha) = c/(\alpha \cdot H_{max})$

3. Asymptotic Achievability of the Minimum Energy/I-Bit using Coding

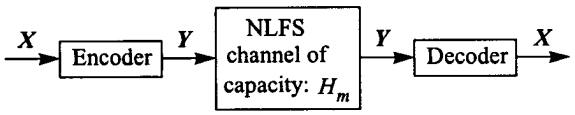
Consider the transmission of a process X of entropy rate $H(X)$ through a NLFS channel that has capacity $H_m > H(X)$. The process utilizes the channel by the factor $\rho(X) = H(X)/H_m$, therefore Theorem 1 implies that in average, an amount of energy equal or greater to $E_b^*(\rho(X))$ is needed per transmitted i-bit. We ask:

Is it possible to encode the random sequence X before transmitting it, so that the average energy required per i-bit is as close to the minimum, $E_b^*(\rho(X))$, as desired?

Theorem 2 answers the above question *affirmatively* for the case that X is stationary and ergodic. The details of the proof are available in [5].

Theorem 2: *A stationary ergodic process X of entropy rate $H(X)$ can be encoded and transmitted through a NLFS channel of capacity $H_m > H(X)$ at an average energy cost per i-bit arbitrarily close to $E_b^*(H(X)/H_m)$.*

The situation is illustrated in the Figure below:



$$H(Y) = H(X) \qquad E_b(Y) \cong E_b^*(H(X)/H_m)$$

Figure 7: Optimal energy reduction coding for NLFS channel.

Theorem 2 shows that the limit $E_b^*(H(X)/H_m)$ is asymptotically achievable using coding. For a practical design though, the power consumption of the encoder and decoder must be taken into account as well.

4. Differential Low-Weight Coding

A class of practical, simple differential coding schemes that can be used to reduce the power consumption in DSM buses is shown below.

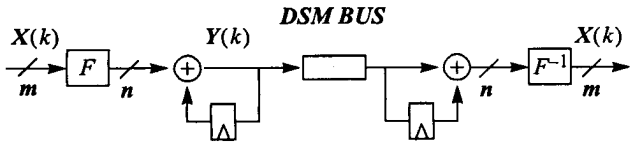


Figure 8: Simple differential coding for deep-sub-micron buses asymptotically achieving the entropy bound.

The scheme encodes the m input bits, vector $X(k)$, into the n bits of vector $Y(k)$ that is transmitted through the (extended) bus, $(n > m)$. The function F is chosen so that the 2^m input vectors are mapped in a 1 – 1 fashion to 2^m vectors of length n that have minimum possible Hamming weights. Low Hamming weight codewords have been used in different contexts in [8] and [9]. The two sequences X and Y are related as,

$$Y(k) = Y(k-1) \oplus F(X(k)) \tag{9}$$

It can be shown that for i.i.d. uniform input data X, the energy per i-bit transmitted through the bus is ([5]):

$$E_b(Y) \geq \frac{2h^{-1}(m/n)}{m/n} \cdot E_b(X)$$

where $E_b(X) = (1+2\lambda)V_{dd}^2 C_L/4$ is the energy per bit when the data X is transmitted uncoded. The equality holds asymptotically when $n \to \infty$ and m/n approaches a utilization α. Figure 9 shows the energy reduction using the scheme of Figure 8 for different values of n and bus utilizations m/n.

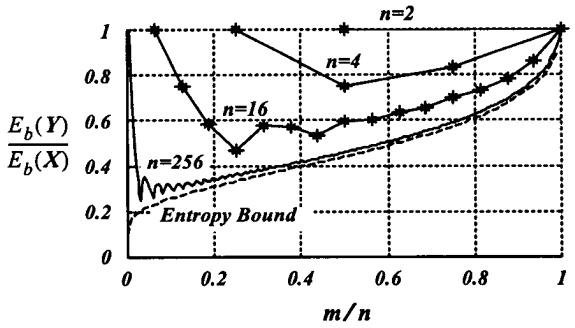


Figure 9: Energy reduction by differential low weight coding.

5. Conclusions

In this work we gave the *minimum possible energy per i-bit required for communicating through noiseless finite-state channels* at a given rate. We showed that this energy per i-bit is asymptotically achievable using coding. The general results were used to characterize the energy/i-bit vs. bit rate behavior of deep submicron buses. A simple differential coding scheme that achieves significant energy reduction was proposed.

References

[1] D. Marculescu, R. Marculescu, and M. Pedram, "Information theoretic measures for power analysis," *IEEE Trans. Computer-Aided Design*, vol. 15, pp. 599-610, June 1996.

[2] N. R. Shanbhag, ``A mathematical basis for power reduction in digital VLSI systems", *IEEE Trans. Circuits Syst. II*, Vol. 44, pp. 935-951, Nov. 1997.

[3] S. Ramprasad, N. R. Shanbhag, and I. N. Hajj, "Information-Theoretic Bounds on Average Signal Transition Activity," *IEEE Trans. on VLSI*, Vol. 7, No. 3, Sept. 1999.

[4] M. Stan, W. Burleson, "Low-power encodings for global communication in CMOS VLSI", *IEEE Transactions on VLSI Systems*, pp. 444-455, Vol. 5, No. 4.

[5] P. Sotiriadis, V. Tarokh, A. Chandrakasan, "Information Theoretic Treatise of Energy Reduction in Deep Submicron Computation Modules", *Submitted to the IEEE Transactions on Information Theory.*

[6] D. Sylvester, Chenming Wu, "Analytical modeling and characterization of deep-submicrometer interconnect", *Proc. IEEE, Vol. 89 Issue: 5, pp. 634 -664, May 2001.*

[7] P. Sotiriadis and A. Chandrakasan, "A Bus Energy Model for Deep Sub-micron Technology", *To appear in the IEEE Transactions on VLSI Systems.*

[8] J. Tabor, "Noise Reduction Using Low Weight and Constant Weight Coding Techniques", *Master thesis, Dept. of EECS, MIT, May 1990.*

[9] M. Stan, W. Burleson, "Limited-Weight codes for low-power I/O", *Int. Workshop on Low-Power Design, Napa, CA, 1994.*