LOWER BOUNDS ON ENERGY DISSIPATION AND NOISE-TOLERANCE FOR DEEP SUBMICRON VLSI†

Rajamohana Hegde and Naresh R. Shanbhag

Coordinated Science Laboratory/ECE Department
University of Illinois at Urbana-Champaign
Urbana, Illinois 61801
email: [rhegde, shanbhag]@uivlsi.csl.uiuc.edu

ABSTRACT

In this paper, we obtain the lower bounds on total energy dissipation of deep submicron (DSM) VLSI circuits via an information-theoretic framework. This framework enables the derivation of lower bounds under the constraint of reliable operation in presence of DSM noise. We employ noise-tolerance via coding to approach the lower bounds on energy dissipation. It is shown that the lower bounds on energy for an off-chip I/O signaling example are a factor of 24X below present day systems. It is also shown that reduction in energy consumption by a factor of 4X can be obtained by employing Reed-Muller codes to achieve the desired bit-error-rate (BER) in presence of DSM noise.

1. INTRODUCTION

Energy-efficient VLSI circuit design is of great interest[1, 2] given the proliferation of mobile computing devices, the need to reduce packaging cost and the desire to improve reliability and extend operational life of VLSI systems. The ability to scale CMOS technology has been one of the prominent reasons for its widely successful use in building low cost and increasingly complex digital VLSI systems. In order to sustain this growth, it is essential to be able to scale in a cost effective manner. The 1997 National Roadmap for Semiconductors [3] has identified the ability to continue affordable scaling as one of the Grand Challenges. For future technologies to be affordable, it is essential that high yields be obtained without putting stringent requirements on the manufacturing tolerances. However, noise immunity becomes difficult to achieve in DSM technology due to reduced feature sizes, smaller supply voltages (smaller noise margins), faster transistors, slow and noisy interconnect, and increasing density due to the trend towards building systems-on-a chip. An awareness of this fact has resulted in the recent interest in DSM noise analysis [4].

A framework to jointly address the issues of noise, energy dissipation and reliability was presented in [5]. It was shown that any system function with input X and output Y has a minimum *information transfer rate* requirement of R bits/sec. Any implementation of the algorithm is viewed as a communication channel/network with an information transfer capacity C (also in bits/sec). The capacity C is a function of W (speed), and signal-to-noise ratio (SNR). For reliable information transfer (or system operation), we need $C > R$ [7]. In particular, it shown that all power reduction techniques tend to bring C as close as possible to R. In [6], this framework was extended to digital gates and the lower bounds on transition activity at the output of logic gates and dynamic energy dissipation were derived. It was also shown that, using Hamming codes, substantial energy reduction can be obtained while meeting the desired reliability specifications for off-chip signaling.

In this paper, we employ the information-theoretic framework [5, 6] to obtain lower bounds on total (both static and dynamic) energy dissipation. In addition, we demonstrate the viability of the noise-tolerance approach proposed in [6] by obtaining about 4X reduction in energy reduction by employing Reed-Muller codes for data transmission over off-chip bus. The rest of this paper is organized as follows. In section 2, we provide the necessary information-theoretic preliminaries. In section 3 we develop an information-theoretic view of VLSI systems for computation followed by our main result in section 4 where we show how absolute lower bounds on total energy consumption can be derived. In section 5, for off-chip I/O signaling, we show how the lower bounds derived can be approached using the concept of noise-tolerance via Reed-Muller Codes.

2. PRELIMINARIES

In this section, we describe information-theoretic preliminaries such as *entropy*, *mutual information*, *conditional entropy* and *channel capacity*.

Entropy: Consider a discrete source generating symbols X from the set $S_X = X_0, X_1, \ldots X_{L-1}$ according to a probability distribution function $p(x)$. A measure of the information content of this source is given by its *entropy* $H(X)$, which is defined as

$$H(X) = -\sum_{i=0}^{L-1} p_i log_2(p_i). \quad (1)$$

We define a related *entropy function* $h(p)$ as follows:

$$h(p) = -plog_2(p) - (1-p)log_2(1-p), 0 \leq p \leq 1. \quad (2)$$

Mutual Information and Conditional Entropy: The *mutual information* $I(X;Y)$ is defined as

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X). \quad (3)$$

$H(X|Y)$ is the *conditional entropy* of X conditioned on Y.

Entropy Rate: Just as *entropy* is the average number of bits required to describe the outcome of a random experiment, *entropy*

† THIS WORK WAS SUPPORTED BY DARPA CONTRACT DABT63-97-C-0025 AND NSF CAREER AWARD MIP-9623737.

0-7803-5471-0/99/$10.00©1999 IEEE

rate is the average number of bits per symbol required to describe a random process. Formally, the *entropy rate* of a stochastic process $\{X_i\}$ is defined by

$$H(\mathcal{X}) = \lim_{n\to\infty} \frac{1}{n} H(X_1, X_2, \cdots X_n). \quad (4)$$

Information Transfer Rate: In [5], we have shown that any system function with input X and output Y has a minimum *information transfer rate* requirement of R bits/s given by $R = f_s H(Y)$, where $H(Y)$ is the entropy of the output Y and f_s is the rate at which the input symbols are being generated. The information transfer rate R is *implementation-independent.*

Channel Capacity: The channel capacity per use C_u is obtained by maximizing (3) over all possible distributions of the channel input X. In other words [7],

$$C_u = \max_{\forall p(x)} \quad I(X;Y). \quad (5)$$

Multiplying C_u with the rate at which the channel is used f_c (in Hz), we obtain, the *information transfer capacity* $C = C_u f_c$. Note that the information transfer capacity is *implementation dependent.*

3. THE INFORMATION-THEORETIC FRAMEWORK

In this section, an information-theoretic framework for noisy digital systems is presented. This framework is then employed to calculate the lower bounds on energy dissipation for such gates.

3.1. Discrete Channel Model for Noisy Gates

A discrete channel model for a noisy gate is represented by a trellis as indicated in Fig. 1(a). This diagram indicates that the probability of the output being correct is $1-\epsilon$ and the probability that it is incorrect is ϵ, for all inputs. A noisy inverter and a noisy two-input AND gate can similarly be represented as shown in Fig. 1(b) and Fig. 1(c), respectively.

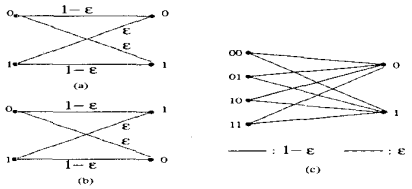

Figure 1: Discrete channel models for:(a) binary symmetric channel, (b) an inverter, and (c) for a 2-input AND gate.

3.2. Information Transfer Capacity of Noisy Gates

While it may seem that information can never be reliably transferred over such a channel, information theory says otherwise. For such gates, we present the following theorem, which quantifies the information transfer capacity per use C_u:

Theorem 1: *The information transfer capacity per use C_u of an n-input, 1-output symmetric gate that makes an error with probability ϵ is given by $C_u = 1 - h(\epsilon)$, where $h()$ is the entropy function defined in (2) and C_u is in bits per use of the channel.*

It has been shown in [5] (for continuous alphabet channels) that minimum energy operation is achieved when $C \approx R$. A similar conclusion was arrived at in [6] paper for discrete channels.

3.3. Characterization of f_c and ϵ

For the remainder of this section and the paper, we will assume that the technology is complementary MOS (CMOS) . Assuming further that the gates have been designed with balanced rise and fall times, the maximum signaling rate at a supply voltage V_{dd} (f_c) equals the reciprocal of the average propagation delay and is approximately given by [9],

$$f_c = \frac{k_m (V_{dd} - V_t)^2}{V_{dd} C_L}, \quad (6)$$

where k_m is the transconductance of the NMOS/PMOS transistor, V_{dd} is the supply voltage, V_t is the NMOS/PMOS transistor threshold voltage, and C_L is the load capacitance.

Characterization of ϵ is difficult as it requires the knowledge of various noise sources and their dependence upon the supply voltage. As this is an on-going work [4], in this paper we will assume that the gate output is in error when V_N exceeds the gate decision threshold voltage [9] $V_{th} = (V_{dd} - |V_{t,p}| + V_{t,n})/2$, where $V_{t,p}$ and $V_{t,n}$ are the threshold voltages of the PMOS and NMOS transistors, respectively. Without loss of generality, we assume that $V_{t,p} = -V_{t,n}$ so that $V_{th} = V_{dd}/2$.

Assuming that a signaling waveform has a certain noise voltage V_N added on to it and V_N has a normal distribution with a variance σ_N^2. It can be shown that the probability of channel error ϵ is given by, $\epsilon = Q(\frac{V_{dd}}{2\sigma_N})$, where the function $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$. The Gaussian assumption on noise distribution has been shown to be a valid for off-chip I/O signaling [10].

Note that none of the assumptions listed above are mandatory for the applications of the results in this paper. If any of the assumptions are violated then these changes can be incorporated into the expressions for ϵ and f_c (in (6)).

4. LOWER BOUND ON ENERGY DISSIPATION FOR NOISY GATES

In this section, the condition for reliable operation $C > R$ is employed as a constraint that needs to be satisfied while minimizing energy dissipation.

4.1. The Energy Minimization Problem

The three major sources of power consumption in CMOS VLSI circuits are: 1.) dynamic power (P_{dyn}) – due to capacitive switching, 2.) static power (P_{stat}) – due to sub-threshold currents and 3.) short circuit power (P_{sc}) – due to direct path currents caused due to non-zero rise and/or fall times of inputs. For the sake of simplicity, we will consider single output logic gates to derive the lower bound on energy consumption. The average dynamic power dissipation of a single output CMOS gate is given by [9],

$$P_{dyn} = (t/2) C_L V_{dd}^2 f_c, \quad (7)$$

where t is the average transition activity, C_L is the capacitance being switched, V_{dd} is the supply voltage and f_c is the rate at which the gate is clocked. The average static power dissipation is given by $P_{stat} = I_{sub} V_{dd}$, where I_{sub} is the average stand-by current. We assume that the gate is operated at the maximum possible rate given by (6). We will employ the *energy per information bit* (E_b) as the measure to compute the minimum energy requirements, where E_b (in joule/bit) is given by

$$E_b = \frac{P_{tot}}{R} = \frac{P_{dyn} + P_{stat} + P_{sc}}{R}, \quad (8)$$

and R is the information transfer rate. Note that E_b is the energy required to transfer one bit of information over the channel, which in our case is the logic gate. Our goal is to minimize (8) subject to the information-theoretic constraint $C > R$. We employ the fact that a more general condition for reliable operation is $I(X;Y)f_c \geq R$, where $I(X;Y)$ is defined in (3). For the special case of a symmetric, single output logic gate, we get

$$I(X;Y)f_c = [H(Y) - H(Y|X)]f_c = [h(p_y) - h(\epsilon)]f_c, \quad (9)$$

where p_y is the probability of observing a '1' at the output.

Lemma 1: *The energy dissipation is minimum when $I(X;Y)f_c = R$. For a symmetric, single output logic gate, energy dissipation is minimum when $[h(t) - h(\epsilon)]f_c = R$.*

Using (8) and *Lemma 1*, we now formulate the following optimization problem to obtain the absolute lower bound on energy consumption as follows:

$$minimize \quad E_b = \frac{P_{tot}}{R} = \frac{P_{dyn} + P_{stat} + P_{sc}}{R} \quad (10)$$

$$\text{subject to:} \quad [h(t) - h(\epsilon)]f_c = R \quad (11)$$

$$f_c = \frac{k_m(V_{dd} - V_t)^2}{V_{dd}C_L} \quad (12)$$

In this paper, we assume that the transistor transconductance k_m is fixed. Hence, the lower bound is now a function of t, V_{dd} and V_t. We first employ (11) to derive the lower bound on supply voltage and circuit speed. We then provide a lower bound on total energy dissipation by proposing a solution to (10)–(12).

4.2. Lower Bound on f_c and V_{dd}

In order to maintain reliability in information transfer, we need to meet the constraint that the information transfer capability of a VLSI system be greater than the information transfer requirement of the given application. Employing (11), for a single output symmetric logic gate, we obtain the following condition.

$$[h(t) - h(\epsilon)]\, f_c \geq R. \quad (13)$$

where f_c is the rate at which the gate is operated, t is the transition activity, ϵ is the probability that an error occurs which is dependent on V_{dd} and R is the desired information transfer rate. In order that LHS of (13) is positive, we need to have $t \geq \epsilon$.

Theorem 2: *The lower bound on f_c for reliable operation of a symmetric, single output, noisy logic gate is given by*

$$f_{c,min} = \frac{R}{1 - h(\epsilon)}. \quad (14)$$

From *Theorem 1*, the denominator in (14) is the capacity C_u. Hence, when $f_c = f_{c,min}$, the gate needs to operate at its capacity.

Theorem 3: *The minimum value of V_{dd} for reliable operation of a symmetric, single output, noisy logic gate, denoted by $V_{dd,min}$ satisfies the equation*

$$\frac{(V_{dd,min} - V_t)^2}{V_{dd,min}} = \frac{RC_L}{k_m\,[1 - h(\epsilon)]}. \quad (15)$$

If the effect of V_t can be neglected, i.e, if $V_t \approx 0$, then we obtain from (15), $V_{dd,min} = \frac{RC_L}{k_m[1-h(\epsilon)]}$. If $f_c = f_{c,min}$, $V_{dd} = V_{dd,min}$, we get $h(t) = 1$ and hence $t = 0.5$. However, this condition may not result in minimum energy dissipation. An increase in V_{dd} leads to an increase in f_c and hence a decrease in t. The decrease in t can offset the increase in V_{dd} and f_c resulting in a net reduction in dynamic energy consumption as we see in the next section.

4.3. Lower Bound on Static and Dynamic Energy Dissipation

We now consider the problem of jointly optimizing static and dynamic components of energy dissipation. We assume that the inputs to the circuit have zero rise and fall times and hence the short circuit power dissipation is zero. We also assume that the parameters k_m and C_L are fixed. The objective then is to find the optimum values of V_{dd} and V_t such that the sum of static and dynamic power consumption is minimized. The problem is stated as follows:

$$\text{minimize } E_b = \frac{K_{sa}\exp\left(\frac{-V_t}{K_{sb}}\right)V_{dd} + tV_{dd}^2C_Lf_c}{R}, \quad (16)$$

subject to (11) and (12). The problem stated in (16) is a constrained optimization problem which can be solved using the Lagrangian Method. This, however, leads to a set of nonlinear equations which would have to be solved using a numerical method. Instead, the solution to this problem is found using a two dimensional search procedure. The optimum supply and threshold voltage values for a single-output gate to obtain an information transfer rate of 8Mbits/sec with $\sigma_N = 0.3V$ is found in Fig. 2. We assume that $k_m = 25 \times 10^{-6}A/V^2$ and $C_L = 50fF$. By following a two dimensional search procedure described above, we obtained $V_{dd,opt} = 0.4430V$, $V_{t,opt} = 0.2561V$ and $f_{c,opt} = 3.9457 \times 10^7 Hz$ and the lower bound on energy is $E_{b,min} = 2.0416 \times 10^{-14}J/bit$. The plot of E_b vs. V_t for several values of V_{dd} is shown in Fig. 2. Each curve corresponds to a fixed value of $V_{dd} > V_{dd,min}$.

5. HIGH-SPEED LOW POWER CHIP I/O SIGNALING

In this section, we propose the use of noise-tolerance to approach the lower bounds for high-speed chip I/O signaling schemes. This is conceptually a simple problem but of great importance due to the high-data rates (0.5 Gb/s-2.6 Gb/s), low voltage levels (0.7V-0.8V) and noisy board environment [10]. We make the following assumptions: 1.) $C_{bus} = 50pF$, 2.) gate capacitance $C_g = C_{bus}/5000$, 3.) $\sigma_N = 0.3V$, 4.) $R = 8Mb/s$, 5.) $k_m = 750\mu A/V^2$, and 6.) desired bit-error rate (BER) = 10^{-14}. The traditional scheme (see Fig. 9), wherein the desired level of noise immunity is achieved by appropriately choosing the supply voltage to suppress the noise, requires a supply voltage of $V_{dd} = 4.8V$ to achieve the desired BER with $E_b = 565pJ/bit$. Next, we propose a noise-tolerant scheme that achieves a $4X$ reduction in E_b while achieving the same BER.

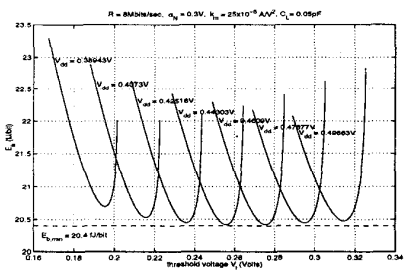

Figure 2: Optimum supply and threshold voltage for minimum energy dissipation

5.1. Noise-Tolerance via Error Control

The noise-tolerant scheme proposed in [6] is shown in Fig. 3(b). The forward channel has employed a reduced voltage level V_{dd}/K_v (where V_{dd} is the supply voltage in Fig. 3(a) where $K_v \geq 1$ is a constant. The forward channel is noisy and makes errors with probability ϵ. The errors due to noise are handled via error detection using first order Reed-Muller codes [11] as opposed to Hamming codes in [6] and error correction using retransmission.

5.2. Energy Savings

In order to compute the power dissipation, the capacitance of the bus-line is modeled as a lumped capacitance C_{bus} at the output of the transmitter buffer. We assume that buffers in both the transmitter and the receiver are a tapered series of inverters which are sized to minimize delay [9].

In case of the conventional system, the energy dissipated per information bit transmitted is given by

$$E_{b,old} = 0.5V_{dd}^2 C_{bus} \quad J/bit, \tag{17}$$

where $V_{dd} = 4.8V$ is the supply voltage at which $\epsilon = 10^{-14}$ per bit. It is assumed that the signal has a transition activity of 0.5. It can be shown that the energy dissipation for the proposed scheme is given by

$$\begin{aligned} E_{b,new} &= \frac{V_{dd}^2 C_{bus}}{1-p_d}\left[\frac{1}{2K_v^2}\frac{n}{k} + \frac{p_d}{k}\right] + \frac{V_{dd} I_{sub}}{R} \\ &\quad + 0.5V_{dd}^2 C_{ED}\frac{1}{1-p_d} \quad J/bit, \end{aligned} \tag{18}$$

where I_{sub} is the off-state leakage current in the buffer, f_s is the input data rate in bits/sec, C_{ED} is the capacitance of the encoder and decoder, K_v is the factor by which the supply voltage for the forward channel in the proposed scheme is scaled down, and p_d is the probability of error detection. The number of bits in the message symbol and the codeword are denoted by k and n respectively.

5.3. Results

We now compare the performance and energy dissipation of the two schemes discussed above. Fig. 4 shows the plot of BER vs. energy dissipation for the schemes considered above. Note that as expected, to achieve a specified value of bit error rate, the traditional scheme consumes the maximum amount of energy. By ignoring the static power dissipation, the lower bound for this case is $E_b = 20.5pJ/bit$. Note that RM codes offer about $4X$ reduction in energy dissipation *while maintaining the throughput* to achieve a $BER = 10^{-14}$ and the lower bound on E_b is about $24X$ below that achieved by current day systems.

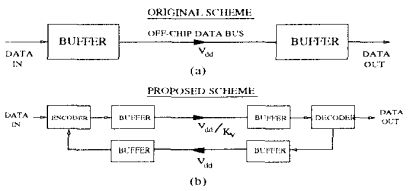

Figure 3: (a) traditional and (b) proposed schemes for chip I/O signaling

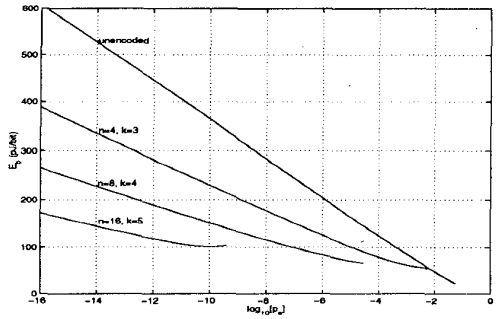

Figure 4: Plot of log_{10}(BER) vs. E_b for the traditional scheme and proposed NTS.

6. CONCLUSIONS

The main conclusions of this paper are: 1.) noise-tolerance is an attractive technique for achieving low energy operation in presence of noise, and 2.) lower bounds on energy can be derived via information-theoretic concepts. Future work needs to be directed towards comprehensive noise models, efficient noise-tolerant schemes for on-chip logic and digital signal processing filters so that energy efficiency can be achieved in the presence of DSM noise. Efficient approaches to noise in the DSM era would require a judicious combination of noise-tolerance and noise-reduction.

7. REFERENCES

[1] A. Chandrakasan and R. W. Brodersen, "Minimizing power consumption in digital CMOS circuits," *Proceedings of the IEEE*, vol. 83, no. 4, pp. 498-523, April 1995.

[2] M. Horowitz, T. Indermaur and R. Gonzalez, "Low-power digital design," *1994 IEEE Symposium on Low Power Electronics* pp. 8-11, San Diego, Oct. 10-12.

[3] The 1997 National Semiconductor Technology Roadmap, *URL: http://www.sematech.org*.

[4] K. L. Shepard, et. al., "Global Harmony: Coupled noise analysis for full-chip RC interconnect networks," *ICCAD97*, pp. 139-146.

[5] N. R. Shanbhag, "A mathematical basis for power-reduction in digital VLSI systems", *IEEE Trans. on Circuits and Systems, Part II*, vol. 44, no. 11, pp. 935-951, Nov. 1997.

[6] R. Hegde and N. R. shanbhag, "Energy efficiency in presence of deep submicron noise", *ICCAD98*, San Francisco, CA.

[7] C. E. Shannon, "A mathematical theory of communications," *Bell System Technical Journal*, vol. 27, Part I, pp. 379-423, Part II, pp. 623-656, 1948.

[8] R. X. Gu and M. I. Elmasry, " Power dissipation analysis and optimization of deep submicron CMOS digital circuits." *IEEE Journal on Solid State Circuits*, vol. 31, No. 5, May 1996.

[9] J. M. Rabaey, *Digital Integrated Circuits: A Design Perspective*. Prentice-Hall, New Jersey, 1996.

[10] C. -K. K. Yang et. al., "A 0.5-μm CMOS 4.0Gb/s serial link transceiver with data recovery using oversampling," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 5, pp. 713-722, May 1998.

[11] S. B. Wicker, *Error control systems for digital communication and storage,* Prentice Hall, New Jersey, 1995.