Toward Achieving Energy Efficiency in Presence of Deep Submicron Noise

Rajamohana Hegde, *Student Member, IEEE,* and Naresh R. Shanbhag, *Senior Member, IEEE*

***Abstract*—Presented in this paper are 1) information-theoretic *lower bounds* on energy consumption of noisy digital gates and 2) the concept of *noise tolerance* via coding for achieving energy efficiency in the presence of noise. In particular, lower bounds on a) circuit speed f_c and supply voltage V_{dd}; b) transition activity t in presence of noise; c) dynamic energy dissipation; and d) total (dynamic and static) energy dissipation are derived. A surprising result is that in a scenario where dynamic component of power dissipation dominates, *the supply voltage for minimum energy operation* ($V_{dd,\,opt}$) *is greater than the minimum supply voltage* ($V_{dd,\,\min}$) *for reliable operation*. We then propose noise tolerance via coding to approach the lower bounds on energy dissipation. We show that the lower bounds on energy for an off-chip I/O signaling example are a factor of 24× below present day systems. A very simple Hamming code can reduce the energy consumption by a factor of 3×, while Reed–Muller (RM) codes give a 4×reduction in energy dissipation.**

***Index Terms*—Coding, energy dissipation, gate capacity, lower bound, noise, noise-tolerant computing.**

I. INTRODUCTION

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 circuits. The 1999 International Technology Roadmap for Semiconductors [1] has identified the ability to continue affordable scaling as one of the grand challenges. The ubiquity of digital systems is also partly due to the reason that, unlike analog circuits that are noise sensitive, digital circuits are inherently immune to noise due to their nonlinear voltage transfer characteristics. However, noise immunity becomes difficult to achieve in deep submicron (DSM) era due to reduced feature sizes, smaller supply voltages (smaller noise margins), and higher density. These features render DSM technology inherently noisy with noise comprising ground bounce, IR drops [2], capacitive and inductive cross-talk [3], charge sharing, charge leakage, process variations, etc. An awareness of this fact has resulted in the recent interest in deep submicron noise analysis [2]–[5].

Energy-efficient VLSI circuit design is of interest [6] given the proliferation of mobile computing devices, the need to reduce packaging cost, and the desire to extend operational life of VLSI systems by improving reliability. Designing low-power integrated circuits in the presence of deep submicron noise is a challenging problem because it compels us to address the issues of energy reduction and reliable operation in a unified manner. At present, research in the area of low-power design revolves around the development of *low-power design* techniques at various levels of the design hierarchy [6]–[8], *power estimation* techniques [9]–[11], and investigating the *lower bounds on power dissipation* [12]–[16]. However, the impact of noise on energy reduction has not been considered so far and, in particular, the following important questions still remain unanswered. 1) "What is the lower-bound on power dissipation?" 2) "How far are we from these bounds?" and 3) "How do we approach the lower bounds systematically in the presence of noise? i.e., how do we design noise-tolerant low-power VLSI systems?"

The existence of lower bounds on energy dissipation at various levels of the system design hierarchy was proposed by Meindl in [14]. Although thermal noise was shown to be the limiting factor for energy reduction at the devices level, noise sources such as ground bounce and cross-talk were not considered. In [15], the lower bound on power dissipation per pole for analog circuits and empirical lower bound estimates for digital circuits was presented. These bounds were estimated from the desired signal-to-noise ratio (SNR). The bounds in [14] and [15] are derived under the assumption that to compute reliably, one requires reliable/noise-free elements. This assumption is too conservative, especially if energy efficiency is desired and a finite (but small) probability of error at the output is acceptable.

The problem of realizing reliable Boolean functions using noisy logic gates using hardware redundancy was first addressed by Von Neumann [17]. The construction proposed in [17] was to build reliable Boolean functions by interleaving computational layers with error correcting layers to keep the probability of error of the overall network under control. Error correction was done using triple modular redundancy and majority voting. It was shown that a given logic function with arbitrarily high reliability can be realized by the proposed construction using noisy logic gates with probability of error $\epsilon < 0.0073$. In [18], it was shown that the depth of a reliable Boolean function implemented with noisy logic gates must be higher than that of the realization of the same with noiseless gates. However, unlike Von Neumann's work, it was not shown how the bound on the depth can be realized. In [19], it was shown that when a given logic function is implemented with 3-input logic gates, it is possible to compute reliably if $\epsilon < 1/6$. It was also shown that, for the 3-input case, reliable computation is not possible if $\epsilon \geq 1/6$ improving on the bound in [18].

In this paper, we extend our past work [16], where the central thesis (shown in Fig. 1) is that computation needs to be viewed

Manuscript received June 8, 1998; revised September 15, 1999. This work was supported by NSF CAREER award MIP 96-23737 and DARPA Contract DABT63-97-C-0025.

The authors are with the Coordinated Science Laboratory/ECE Department, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: rhegde@uivlsi.csl.uiuc.edu; shanbhag@uivlsi.csl.uiuc.edu).

Publisher Item Identifier S 1063-8210(00)04345-6.

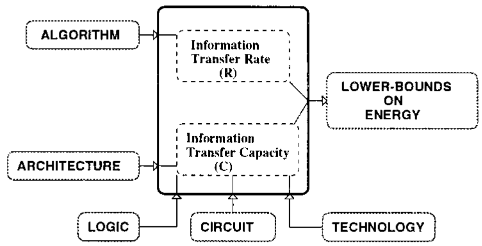


Fig. 1. Information-theoretic framework for VLSI.

as a process of information transfer over a noisy channel. In [16], we have shown that any system function with input X and output Y has a minimum *information transfer rate* requirement of R bits/s. Any implementation of the algorithm is viewed as a communication channel/network with an information transfer capacity C (also in bits/s). The capacity C is a function of the speed, SNR, and the architecture. For reliable information transfer (or system operation), we need $C > R$ [20]. The condition $C > R$ ensures that it is possible to transfer information at a rate of R bits/s with the probability of error (in the information bits) approaching zero, i.e., perfect reliability. In the past, we have applied [16] to the framework in Fig. 1 to provide a common basis to power reduction techniques at various levels of the design hierarchy such as pipelining, parallel processing, adiabatic logic, etc. In particular, we have shown that all power reduction techniques tend to bring C close to R.

In this paper, we propose a *discrete channel* model for digital modules and then develop the capacity formula to calculate the lower bounds. In this model, we assume that every time a module is used, it can make an error with a certain probability ϵ. The value of ϵ depends on the supply voltage V_{dd} and the variance σ_N^2 of the noise voltage V_N. While we consider the module to be a logic gate and an off-chip wire, it could potentially be a complete digital system. We then use the discrete channel model to show how the absolute lower bound on energy dissipation can be obtained by only using the information-theoretic constraint $C > R$ to ensure reliability.

While the information-theoretic framework provides us with the lower bounds, it does not indicate how to achieve them. In addition to deriving the lower bounds, we also propose the concept of noise tolerance to approach the lower bound on energy dissipation. Improving the reliability of noisy logic gates by employing error correcting codes was considered by Elias in [21] with the conclusion that an arbitrarily high level of reliability can be achieved only when the computational rate approaches zero. This argument was strengthened in [22], where it was shown that, except for symmetric gates such as XOR and NOT, one cannot do better than repetition coding to improve the reliability of noisy logic gates. However, by relaxing one of the conditions imposed on the encoders and decoders, it was shown in [23] that Reed–Muller codes can be employed to achieve high levels of reliability without an abrupt drop in computational rate. Coding schemes have been successfully used to provide fault tolerance for computing applications, without considering energy efficiency. Arithmetic codes [24] have been proposed to improve the reliability of arithmetic units such as adders and multipliers against faults, and can also be used for noise tolerance. In contrast, in this paper we employ error-control codes to provide the desired level of reliability while explicitly reducing the overall energy dissipation thereby approaching the lower bound.

In summary, the primary contributions of this paper are twofold: 1) obtaining the lower bounds on energy consumption of noisy logic gates via an information-theoretic framework and 2) achieving energy efficiency via noise-tolerant coding in the presence of deep submicron noise. The rest of this paper is organized as follows. In Section II, we provide the necessary information-theoretic preliminaries. In Section III, we develop an information-theoretic view of VLSI systems for computation followed by our main result in Section IV, where we show how absolute lower bounds on energy consumption can be derived. In Section V, for off-chip I/O signaling, we show how the lower bounds derived can be approached using the concept of noise tolerance via error-control coding schemes.

II. Preliminaries

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

A. Entropy

Consider a discrete source generating symbols X from the set $S_X = X_0, X_1, \cdots X_{L-1}$ 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 follows:

$$H(X) = -\sum_{i=0}^{L-1} p_i \log_2(p_i) \tag{2.1}$$

where $p_i \stackrel{\text{def}}{=} \Pr(X = X_i)$ for $i = 0, \cdots, L-1$ and $H(X)$ is in bits. Note that we have $L = 1$ for a single bit line, while an m-bit bus has $L = 2^m - 1$.

This definition of the measure of information implies that the greater the *uncertainty* in the source output, the higher its information content.

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

$$h(p) = -p \log_2(p) - (1-p) \log_2(1-p) \tag{2.2}$$

where $0 \leq p \leq 1$. Similarly, the *inverse entropy function* $h^{-1}(q)$ is defined as

$$h^{-1}(q) = \{p\colon h(p) = q,\ 0 \leq p \leq \tfrac{1}{2}\} \tag{2.3}$$

where $0 \leq q \leq 1$. The function $h(p)$ is shown in Fig. 2, where it can be seen that it achieves its maximum value of unity when $p = 0.5$.

B. Mutual Information and Conditional Entropy

The *mutual information* $I(X;Y)$ is defined as

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) \tag{2.4}$$

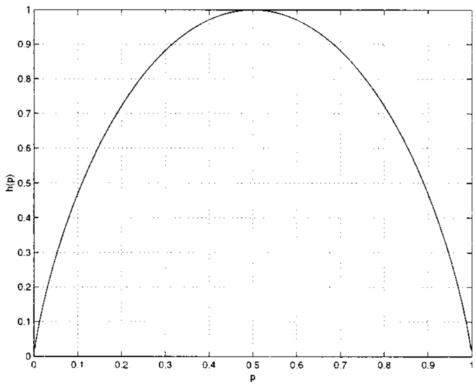


Fig. 2. Entropy function $h(p)$.

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

$$H(X|Y) = -\sum_{Y \in S_Y} \sum_{X \in S_X} \Pr(X, Y) \log_2(\Pr(X|Y)) \quad (2.5)$$

where the set $S_X = \{X_0, X_1, \cdots, X_{L-1}\}$ and $S_Y = \{Y_0, Y_1, \cdots, Y_{M-1}\}$.

The conditional entropy $H(X|Y)$ can be interpreted as the *residual uncertainty* in X given the knowledge of Y. In a similar fashion, the mutual information $I(X;Y)$ can be viewed as the *reduction in uncertainty* in X due to the knowledge of Y. This reduction in uncertainty [by an amount $I(X;Y)$] in X is due to the information transferred from the input of the channel to its output *per use* of the channel. The definition of mutual information in (2.4), along with the fact that for a noiseless channel $H(Y|X) = 0$, provides us with the defining equation (2.9) for the information transfer rate R. The following example will illustrate some of these concepts as applied to digital gates.

Example 1: AND Gate. Consider a 2-input AND gate operating at 100 MHz where both inputs are independent and identically distributed (i.i.d) with the probability of a “1” on each being equal to 0.5. In that case, the entropy of the input (taken either as a single 2-bit source with $L = 3$ or as two single-bit sources) is 2 bits. The entropy of the AND gate output Y [from (2.1)] is given by

$$\begin{aligned} H(Y) &= -P(0)\log_2(P(0)) - P(1)\log_2(P(1) \\ &= -\tfrac{3}{4}\log_2\left(\tfrac{3}{4}\right) - \tfrac{1}{4}\log_2\left(\tfrac{1}{4}\right) \\ &= 0.8112 \text{ bits.} \end{aligned} \quad (2.6)$$

C. Entropy Rate

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

$$H(\mathcal{X}) = \lim_{n \to \infty} \frac{1}{n} H(X_1, X_2, \cdots X_n). \quad (2.7)$$

Note that if the stochastic process is independent and identically distributed (i.i.d), we get

$$H(\mathcal{X}) = \lim_{n \to \infty} \frac{1}{n} H(X_1, X_2, \cdots X_n) = \frac{nH(X_1)}{n} = H(X_1) \quad (2.8)$$

which is equal to the entropy per sample.

D. Information Transfer Rate

In [16], 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

$$R = f_s H(Y) \quad (2.9)$$

where $H(Y)$ is the entropy of the output Y and f_s is the rate at which the input symbols are being generated. This information transfer rate R is *implementation-independent*.

Example 2: Consider the 2-input AND gate from Example 1. Let the rate at which the input is being generated f_s =10 MHz. For the same input statistics as in Example 1, the information transfer rate is given by

$$R = 10 \times 10^6 \times 0.8112 = 8.112 \text{ Mbits/s.} \quad (2.10)$$

Note that the information transfer rate is dependent on both the rate at which the data is being generated and the probability distribution of the input.

E. Channel Capacity

The channel capacity per use C_u is obtained by maximizing (2.4) over all possible distributions of the channel input X. In other words [20]

$$C_u = \max_{\forall\, p(x)} \quad I(X;Y). \quad (2.11)$$

Multiplying C_u with the rate at which the channel is used f_c (in Hz), we obtain

$$C = C_u f_c. \quad (2.12)$$

Example 3: For the AND gate in Example 2, assuming that the probability of error for all input combinations is $\epsilon = 0.1$ and switching speed $f_c = 20$ MHz, the capacity of this gate is given by

$$\begin{aligned} C &= \max_{\forall\, p(x)} \quad I(X;Y) f_c \\ &= [1 - h(\epsilon)] f_c, \text{ (from Th.1, Sec. III-B)} \\ &= [1 - h(0.1)] 20 \times 10^6 \text{ bits/s} \\ &= 10.62 \text{ Mbits/s} \end{aligned} \quad (2.13)$$

where the input distribution is such that the output distribution is uniform, resulting in $p_y = 0$ and, therefore, $h(p_y) = 1$.

Note that f_c, the speed at which the circuit is operated, is different from f_s, the rate at which the source generates information. As we have $C(10.62 \text{ Mbits/s}) > R(8.112 \text{ Mbits/s})$, it should be possible to achieve an arbitrarily high level of reliability [20]. In practice, the probability of error can be reduced by adding redundant bits to enable error detection and correction at the output.

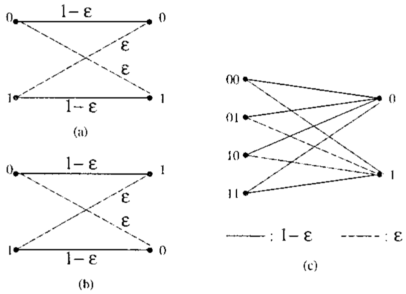


Fig. 3. Discrete channel models for (a) a noninverting buffer, (b) an inverter, and (c) a 2-input AND gate.

It was shown in [20] that it is possible to achieve an information transfer rate R, (defined in (2.9) for a digital system [16]) with a probability of error δ approaching zero (via appropriate coding of the inputs) as long as $R < C$ and the information source is an ergodic and stationary process. We assume that the input sources considered here are also ergodic and stationary.

III. Information-Theoretic Framework for Noisy VLSI Circuits

In this section, we present an information-theoretic framework for noisy digital systems. We then employ this framework to calculate the lower bounds on power dissipation for single-output gates.

A. Discrete Channel Model for Noisy Gates

A discrete channel model for a noisy noninverting buffer is represented by a *trellis*, as indicated in Fig. 3(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. Such a model is also referred to as a *binary symmetric channel* (BSC) [20]. Thus, we assume that the magnitude and duration of intermittent noise voltage is sufficient to cause logic errors. A noisy inverter and a noisy 2-input AND gate can similarly be represented as shown in Fig. 3(b) and (c), respectively. Note that logic errors can be permanent due to delay faults induced by technology variations, such as those in V_t. In that case, for the input combinations that excite such faults, we obtain $\epsilon = 1$.

B. Information Transfer Capacity of Noisy Gates

The following theorem quantifies the information transfer capacity per use C_u of a single-output noisy gate.

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

$$C_u = 1 - h(\epsilon) \tag{3.1}$$

where $h(\,)$ is the entropy function defined in (2.2) and C_u is in bits per use of the channel. Furthermore, the output distribution that achieves this capacity is the uniform distribution given by

$$p_{y,\,i} = \tfrac{1}{2}, \qquad \text{for } i = 0,\, 1 \tag{3.2}$$

where $p_{y,\,i}$ is the probability of observing the ith output combination.

Proof: The information transfer capacity of an n-input, 1-output gate is obtained from (2.11) and (2.4) as follows:

$$C_u = \max_{\forall\, p(x)} \quad [H(Y) - H(Y|X)]. \tag{3.3}$$

As the gate is symmetric, with error probability ϵ, we have $H(Y|X) = h(\epsilon)$. As Y is a binary random variable, we have $\max\, H(Y) = 1$, which can be achieved by choosing an appropriate input distribution. Substituting in (3.3), we get (3.1). ■

Note that Theorem 1 indicates that when $\epsilon = 0$ (i.e., a noiseless gate), the information transfer capacity per use C_u is maximum and equal to 1 bit per use. This is same as saying that the 2-input, 1-output AND gate discussed in Example 1, Section II can transfer one bit of information for every use. This is possible provided the gate is noiseless and that the output distribution is uniform. This, in turn, implies that the input combination "11" should occur with a probability of 0.5 and the rest of the input combinations should occur with a probability of 0.5. Furthermore, this AND gate has a capacity of $C_u = 0$ if it makes an error half of the time, i.e., $\epsilon = 0.5$. An interesting observation to be made at this point is that $C_u > 0$ if the AND gate makes errors more than half of the time, i.e., $\epsilon > 0.5$. It will be shown later that for a relatively high value of V_{dd} with respect to the noise, the value of $\epsilon = 0$, i.e., the circuit becomes error-free. In that case, from (3.1), we see that C_u is equal to unity and the capacity $C = f_c$ [from (2.12)]. This is consistent with the conventional measure of capacity as being the maximum rate at which the circuit can be clocked. Most modern day digital systems operate in this region, which is also the primary reason for the high power dissipation of such systems.

Theorem 1 seems to suggest that the information transfer capacity C is independent of technology and circuit implementation style. However, this is not the case, because both ϵ and f_c depend upon the technology and the circuit style. In particular, we note that f_c and noise are a function of the supply voltage V_{dd}, transistor transconductance k_m, and the device threshold voltage V_t, as described next.

C. 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) and the circuit design style is static with dual NMOS and PMOS transistor networks. Assuming further that the gates have been designed with balanced rise and fall times, the maximum signaling rate f_c at a supply voltage V_{dd} equals the reciprocal of the average propagation delay and is approximately given by [25]

$$f_c = \frac{k_m(V_{dd} - V_t)^2}{V_{dd}C_L} \tag{3.4}$$

where

- k_m transconductance of the NMOS/PMOS transistor;
- V_{dd} supply voltage;
- V_t NMOS/PMOS transistor threshold voltage;
- C_L 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 ongoing work [3], [4], in this paper we

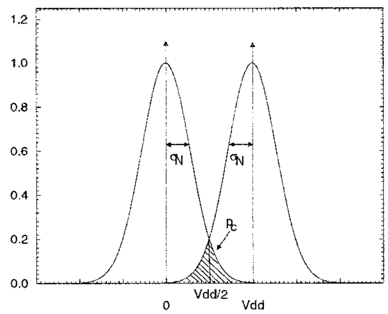


Fig. 4. Function $Q(x)$ and its relation to p_c.

will assume that all the major sources of noise contribute a noise voltage V_N at the gate output. Therefore, the gate output is in error when V_N exceeds the gate decision threshold voltage V_{th} [25] defined as

$$V_{th} = \frac{V_{dd} - |V_{t,p}| + V_{t,n}}{2} \tag{3.5}$$

where $V_{t,p}$ and $V_{t,n}$ are the threshold voltages of the PMOS and NMOS transistors, respectively.

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 equals the shaded overlap area in Fig. 4, and is given by

$$\epsilon = Q\left(\frac{V_{dd}}{2\sigma_N}\right) \tag{3.6}$$

where the function $Q(x)$ is the Gaussian pulse (see Fig. 4) and is defined as

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy. \tag{3.7}$$

Note that (3.6) provides us with the necessary dependence of ϵ on the supply voltage V_{dd}. Fig. 4 indicates that as V_{dd} reduces the two curves approach each other, thereby increasing the overlap and, hence, the probability of channel error ϵ. The Gaussian assumption on noise distribution has been shown to be valid for off-chip I/O signaling [26].

We summarize all the assumptions made in deriving (3.4) and (3.6).

1) CMOS technology with fully static design style.
2) The low-to-high and high-to-low propagation delays are equal. For small $V_{t,n}$ and $V_{t,p}$, this implies the transconductance of the equivalent NMOS and PMOS transistors are equal.
3) The threshold voltages $V_{t,n} = -V_{t,p}$ so that with equal NMOS and PMOS transconductance, we have $V_{th} = V_{dd}/2$.
4) The variance σ_N^2 of the noise voltage V_N is independent of V_{dd} and has a normal distribution with zero mean. The noise pulse has a time duration greater than the delay of the gate.

Note that if any of the assumptions are violated, then these changes can be incorporated into the expressions for ϵ [in (3.6)] and f_c [in (3.4)]. We now consider an example, which illustrates the application of the concepts presented so far.

Example 4: Capacity of a 2-Input AND Gate in $1.2\ \mu$ *CMOS.* Assume that $V_{dd} = 1.5$ V, $V_t = 0.5$ V, $\sigma_N = 0.5$ V, $k_m = 80\ \mu$A/V^2 and $C_L = 50$ fF. Substituting the values of V_{dd} and σ_N into (3.6), we obtain a value of $\epsilon = 0.067$. From (3.1), with $\epsilon = 0.067$, we get $C_u = 0.6461$ bits/use. The second component f_c of the capacity (2.12) needs to be computed. To do this, we substitute the values of k_m, V_{dd}, V_t and C_L into (3.4) to get $f_c = 1.07 \times 10^9$ uses/s. Therefore, the information transfer capacity of this gate is given by $C = C_u f_c = 6.84 \times 10^8$ bits/s.

It can be seen that the AND gate in Example 2 has a relatively high capacity in spite of the large noise standard deviation of $\sigma_N = 0.5$ V. Employing the information-theoretic framework presented so far, we now determine the lower bounds on energy in Section IV.

IV. Lower Bound on Energy Dissipation for Noisy Gates

In this section, we employ the condition for reliable operation $C > R$ as a constraint while minimizing energy dissipation to derive the desired lower bounds. We first formulate the problem of determining the lower bounds as a constrained optimization problem in Section IV-A. We then determine the lower bound on speed of operation of the circuit and the supply voltage in Section IV-B. In Section IV-C, we derive the lower bound on transition activity of a noisy gate. Sections IV-D and IV-E discuss the lower bounds on dynamic energy dissipation and total (dynamic and static) energy dissipation, respectively.

A. Energy Minimization Problem

The three major sources of power consumption in CMOS VLSI circuits are 1) dynamic power dissipation (P_{dyn}), due to capacitive switching; 2) static power dissipation (P_{stat}), due to leakage and subthreshold currents; and 3) short-circuit power dissipation (P_{sc}), due to direct path currents caused due to nonzero rise and/or fall times of inputs. For the sake of simplicity, we will only consider single-output logic gates to derive the lower bound on energy consumption. We will also assume that all the capacitances of the gate are lumped at the output, though this assumption results in some loss of accuracy. The average dynamic power dissipation of a CMOS inverter is given by [25]

$$P_{dyn} = 0.5 t C_L V_{dd}^2 f_c \tag{4.1}$$

where

t	average transition probability;
C_L	capacitance being switched;
V_{dd}	supply voltage;
f_c	rate at which the gate is clocked.

The average static power dissipation is given by

$$P_{stat} = I_{sub} V_{dd} \tag{4.2}$$

where I_{sub} is the average standby current. For a single-output logic gate, I_{sub} is given by [27]

$$I_{sub} = K \mu C_{ox} V_t^2 e^{1.8} \left(\frac{W}{L}\right) \exp\left(\frac{-V_t}{n_c V_T}\right) \tag{4.3}$$

where K is a constant dependent on gate topology ($K = 1$ for a CMOS inverter); μ is the carrier mobility; C_{ox} is the oxide capacitance; W/L is the effective width to length ratio of the PMOS/NMOS networks; $V_T = kT/q$; and n_c is a constant. For a typical deep submicron technology, the value of n_c ranges from 1.4 to 1.5. The average short-circuit power dissipation is given by [6]

$$P_{sc} = t\frac{K_m}{12}(V_{dd} - 2V_t)^3 \tau f_c \tag{4.4}$$

where τ is the "rise–fall time" of the input signal. We assume that the gate is operated at the maximum possible rate given by (3.4).

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

$$E_b = \frac{P_{tot}}{R} = \frac{P_{dyn} + P_{stat} + P_{sc}}{R} \tag{4.5}$$

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. R is a fixed quantity and, hence, minimizing P_{tot} also minimizes E_b. Therefore, our goal is to minimize (4.5) 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 (2.4). For the special case of a symmetric, single-output logic gate, we get

$$I(X;Y)f_c = [H(Y) - H(Y|X)]f_c = [h(p_y) - h(\epsilon)]f_c \tag{4.6}$$

where p_y is the probability of observing a "1" at the output. If the input distribution is such that $p_y = 0.5$, then we obtain the result of Theorem 1 and the gate would be operating at capacity. For the rest of this paper, we assume that transition signaling (i.e., a "1" is represented with a transition and a "0" is represented with no transition) is employed at the output of the logic gate, due to which we obtain $p_y = t$.

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 $[h(t) - h(\epsilon)]f_c = R$.

Proof: Let the energy dissipation be minimum when $I(X;Y)f_{c1} = R_1$, where $R_1 > R$, and f_{c1} is given by (3.4). Now, since $I(X;Y)f_{c1} > R$, we can reduce f_{c1} to f_c such that $I(X;Y)f_c = R$, thus satisfying the reliability constraint. Note that since P_{dyn} and P_{sc} are directly proportional to f_c, signaling at f_c reduces both P_{dyn} and P_{sc}. Keeping the signaling rate at f_c, we can now increase V_t such that the circuit again operates at its maximum possible speed. Note that increasing V_t reduces both P_{stat} and P_{sc}. Hence, minimum energy dissipation cannot occur when $I(X;Y)f_c > R$, which contradicts our initial assumption. Therefore, energy dissipation is minimum when $I(X;Y)f_c = R$. ∎

Using (4.5) and Lemma 1, we now formulate the following optimization problem to obtain the absolute lower bound on energy consumption as follows:

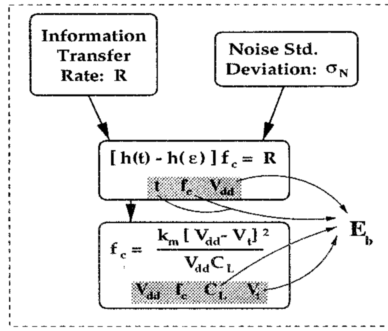


Fig. 5. Relationship between the various parameters involved in the optimization problem.

Minimize

$$E_b = \frac{P_{tot}}{R} = \frac{P_{dyn} + P_{stat} + P_{sc}}{R} \tag{4.7}$$

subject to

$$[h(t) - h(\epsilon)]f_c = R \tag{4.8}$$

$$f_c = \frac{k_m(V_{dd} - V_t)^2}{V_{dd}C_L}. \tag{4.9}$$

There is an intricate relationship between R, σ_N, t, C_L, V_{dd}, V_t, f_c, k_m, t, and ϵ in (4.7)–(4.9). In Fig. 5, this relationship is illustrated as follows.

- Consider the constraint in (4.8). For a given R, as V_{dd} is increased, f_c increases and, hence, a lower value of t can satisfy (4.8). If the increase in V_{dd} is offset by the decrease in t, there will be a net reduction in dynamic energy consumption despite an increase in supply voltage.
- For a fixed V_{dd}, an increase in f_c can be achieved in several ways. From the constraint in (4.9), it can be seen that f_c could be increased by reducing V_t, increasing k_m, and reducing C_L individually or in some combination. Any reduction in V_t will result in an exponential increase in the static energy consumption and an increase in C_L will lead to a linear increase in dynamic power consumption. These tradeoffs have been studied in [28].

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. The problem stated in (4.7)–(4.9) is employed in subsequent sections to derive the desired lower bounds. We first employ (4.8) to derive the lower bound on supply voltage, circuit speed, and transition activity at the output of a noisy logic gate. This result is then used to derive the lower bound on dynamic energy dissipation. Finally, we provide a lower bound on total energy dissipation by proposing a solution to (4.7)–(4.9).

B. Lower Bound on f_c and V_{dd}

To maintain reliability in information transfer, we need to meet the constraint $C > R$. Employing (4.8), for a single-output symmetric logic gate, we obtain the condition

$$[h(t) - h(\epsilon)]f_c \geq R. \tag{4.10}$$

In order that LHS of (4.10) 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

$$f_{c,\,\min} = \frac{R}{1-h(\epsilon)}. \tag{4.11}$$

Proof: From (4.10), since the maximum value of $h(t) = 1$, the minimum frequency at which the circuit needs to operate to maintain reliability is given by (4.11). ■

From Theorem 1, note that the denominator in (4.11) 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

$$\frac{(V_{dd,\,\min} - V_t)^2}{V_{dd,\,\min}} = \frac{RC_L}{k_m[1-h(\epsilon)]}. \tag{4.12}$$

Proof: By substituting (4.9) in (4.11), with $V_{dd} = V_{dd,\,\min}$, we get (4.12). ■

If the effect of V_t can be neglected, i.e., if $V_t = 0$, then we obtain from (4.12)

$$V_{dd,\,\min} = \frac{RC_L}{k_m[1-h(\epsilon)]}. \tag{4.13}$$

If $f_c = f_{c,\,\min}$, $V_{dd} = V_{dd,\,\min}$ then $t = 0.5$. 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.

C. Lower Bound on Transition Activity of Noisy Logic Gates

From (4.1), we see that dynamic energy dissipation is proportional to the transition activity t. Also, from an information-theoretic viewpoint, it seems that an information bearing signal should have a lower bound on the transition activity t that is proportional to its information content. This lower bound is derived in Theorem 4 below.

Theorem 4: The lower bound on transition activity at the output of a symmetric, single-output, noisy logic gate employing transition signaling is given by

$$t \geq h^{-1}\left[\frac{R}{f_c} + h(\epsilon)\right]. \tag{4.14}$$

Proof: Solving for t in (4.10), provides us with the lower bound in (4.14). ■

From (4.14), we note that reduction in t can be obtained by increasing f_c. As both f_c and t contribute to dynamic power dissipation [see (4.1)], net savings in dynamic power dissipation will be achieved only if the reduction in t offsets the increase in f_c.

In the absence of noise, i.e., $\epsilon = 0$, substituting $R = \mathcal{H} f_s$ bits/s and $f_c = R_b f_s$ (R_b is the number of code bits assigned per symbol) into (4.14), we obtain the lower bound on t for the noiseless case as follows:

$$t \geq h^{-1}\left[\frac{\mathcal{H}}{R_b}\right]. \tag{4.15}$$

Equation (4.15) is the bound in [29] derived for a noiseless bus.

```
Start:  Input: R, σ_N and K >> 1;
        compute V_dd,min and f_c,min using (4.13) and (4.11),
                respectively;
        t = 0.5; f_c = f_c,min; V_dd = V_dd,min; V_t = 0;
        δV_dd = V_dd,min/K; δt = t/K;
        compute E_b using (4.18); E_bv(1) = E_b;
        i = 1;
Vloop: repeat
                i = i + 1; V_dd,opt = V_dd;
                V_dd = V_dd + δV_dd;
                compute t_min using (4.14); t = t_min;
                compute E_b using (4.18); E_bt(1) = E_b; j = 1;
tloop:          repeat
                        j = j + 1; t_opt = t; t = t + δt;
                        if t ≥ 0.5
                          exit
                        V_t,opt = V_t;
                        compute V_t using (4.21);
                        compute E_b using (4.18); E_bt(j) = E_b;
                until (E_bt(j) > E_bt(j - 1));
                E_bv(i) = E_bt(j - 1);
        until (E_bv(i) > E_bv(i - 1));
        E_b,min = E_bv(i - 1);
        report E_b,min, V_dd,opt, V_t,opt, t_opt, V_dd,min
        compute f_c,min and f_c,opt using
                V_dd,min, V_dd,opt, V_t,opt and t_opt.
```

Fig. 6. Plot of V_{dd} versus E_{dyn}.

D. Lower Bound on Dynamic Energy Dissipation

We will now determine the lower bound on dynamic power dissipation of noisy gates. Assuming that $V_t \ll V_{dd}$, from (3.4), we get

$$f_c = \frac{k_m V_{dd}}{C_L}. \tag{4.16}$$

Substituting (4.14) and (4.16) in (4.1), we obtain the dynamic power dissipation as follows:

$$P_{dyn} = h^{-1}\left[\frac{RC_L}{k_m V_{dd}} + h(\epsilon)\right] V_{dd}^3 k_m. \tag{4.17}$$

The plot of $E_b = P_{dyn}/R$ versus V_{dd} is shown in Fig. 6. First, note that the minimum possible supply voltage V_{dd} at which reliable operation is guaranteed is $V_{dd,\,\min} = 0.8660$ V. Recall that, at $V_{dd} = V_{dd,\,\min}$ [from (4.13)], we have $f_c = f_{c,\,\min}$ [from (4.11)] and $t = 0.5$. Increasing V_{dd} enables the circuit to be operated faster (higher f_c) and, hence, permits us to decrease t. Note that the decrease in t indeed offsets the increase in V_{dd} till $V_{dd} = 1.0782$ V. This is primarily due to the fact that a small reduction in $h(t)$ results in a large reduction in t around the region $t = 0.5$. A further increase in V_{dd} does not result in a sufficient reduction in t. Hence, we see an increase in energy dissipation beyond $V_{dd} = 1.0782$ V. The minimum energy is found to be $E_{b,\,\min} = 20.5$ pJ/bit.

E. Lower Bound on Static and Dynamic Energy Dissipation

We now consider the problem of jointly optimizing static and dynamic components of energy dissipation. As the supply voltage is reduced with technology scaling, it is also required to reduce the threshold voltage to maintain throughput. Reduction in threshold voltage causes an increase in the static power dissipation. In addition to lower supply voltages, the switching capacitance in deep submicron circuits are also smaller and, hence, dynamic and static components of power dissipation are of the same order [27]. 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:

Minimize

$$E_b = \frac{K_{sa} \exp\left(\frac{-V_t}{K_{sb}}\right) V_{dd} + tV_{dd}^2 C_L f_c}{R} \tag{4.18}$$

subject to:

$$[h(t) - h(\epsilon)]f_c = R \tag{4.19}$$

$$f_c = \frac{k_m (V_{dd} - V_t)^2}{V_{dd} C_L} \tag{4.20}$$

where $K_{sa} = \mu_0 C_{ox} V_t^2 e^{1.8} (W/L)$ and $K_{sb} = nV_T$. The problem stated in (4.18)–(4.20) is a constrained optimization problem, which can be solved using the Lagrangian method. This, however, leads to a set of nonlinear equations that would have to be solved using a numerical method. Instead, the solution to this problem is found using a two-dimensional search procedure as follows.

1) For the given values of R and σ_N, with $V_t = 0$, we first obtain $V_{dd,\min}$ given by (4.13). Recall that at this value of V_{dd}, we have $t = 0.5$. With $V_{dd} = V_{dd,\min}$, $V_t = 0$, $f_c = f_{c,\min}$ and $t = 0.5$, the value of E_b is found from (4.18).
2) The value of V_{dd} is now increased in sufficiently small steps. For each V_{dd}:
   - 2.1) from (4.14), the minimum possible value of t is found;
   - 2.2) we now increase t in small steps till we see an increase in total energy dissipation at this value of V_{dd} due to an increase in dynamic component of energy.
     - 2.2.1) For each value of t, we compute V_t using

$$V_t = V_{dd} - \sqrt{\frac{RC_L V_{dd}}{k_m [h(t) - h(\epsilon)]}}. \tag{4.21}$$

Note that (4.21) is obtained by substituting (4.20) in (4.19) and then solving for V_t.

     - 2.2.2) With values of V_{dd}, t and V_t, the value of E_b is now computed from (4.18). If this value of E_b is lower than the previous value, then go to 2.2, otherwise go to 2.

This search algorithm is illustrated as a pseudo-code in Fig. 7.

The region over which the minimum is searched is shown in Fig. 8. The search begins (step 1 in the search algorithm) at the point P, at which we have $V_{dd} = V_{dd,\min}$, $t = 0.5$, $f_c = f_{c,\min}$ and $V_t = 0$. With an increase in V_{dd}, the value of t that satisfies (4.8) reduces. This is denoted by the point Q in Fig. 8. This corresponds to step 2 in the search algorithm. Also, at Q, the circuit is operating at a speed greater than $f_{c,\min}$. The value of t is now incremented in small steps. Note that this leads to a reduction in f_c [from (3.4)] and, hence, an increase in V_t to satisfy (4.9). The transition density t can be increased till we get $t = 0.5$. This is denoted by R in Fig. 8. This procedure is repeated until the value of V_{dd} is high enough to make the dynamic component of energy dissipation dominant.

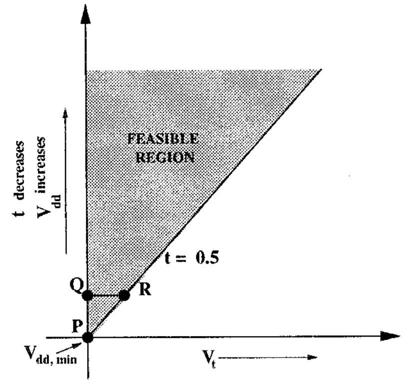


Fig. 7. Algorithm to obtain the lower bound on total energy dissipation.

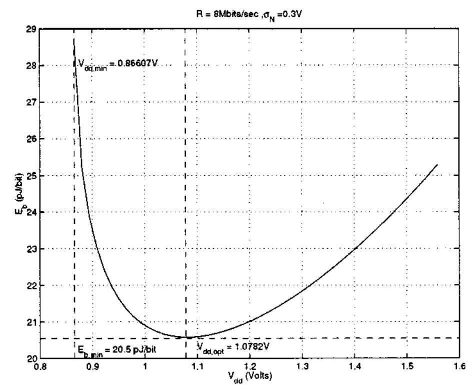


Fig. 8. Feasible region for the optimization scheme.

The optimum supply and threshold voltage values for a single-output gate to obtain an information transfer rate of 8 Mbits/s with $\sigma_N = 0.3$ V is found in Fig. 9. We assume that $k_m = 25 \times 10^{-6}$ A/V^2 and $C_L = 50$ fF. A relatively small value of load capacitance C_L is chosen so that the static power component P_{stat} is made comparable to P_{dyn}. By following the search procedure described above, we obtained $V_{dd,opt} = 0.4430$ V, $V_{t,opt} = 0.2561$ V, and $f_{c,opt} = 3.9457 \times 10^7$ Hz and the lower bound on energy is $5E_{b,\min} = 2.0416 \times 10^{-14}$ J/bit. The plot of E_b versus V_t for several values of V_{dd} is shown in Fig. 9. Each curve corresponds to a fixed value of $V_{dd} > V_{dd,\min}$ and to each iteration of step 2 in the search algorithm. It can be seen that at each V_{dd}, initially, due to low t and low V_t, the static component of power is dominant. The decrease in static power due to an increase in V_t compensates for the increase in dynamic power due to the increase in t, resulting in a net decrease in total power consumption. At higher values of V_t, the energy rises again due to the increase in

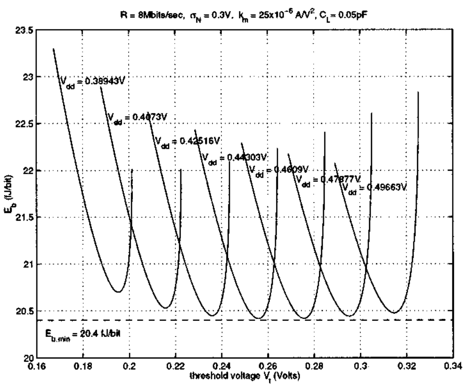


Fig. 9. Optimum supply and threshold voltage for minimum energy dissipation.

dynamic power dissipation because of higher values of t. The value of t at the beginning of iteration for a given V_{dd} decreases with the increase in V_{dd}. However, the search for a fixed value of V_{dd} always terminates at t = 0.5, at which V_t is maximum for that iteration.

We would like to point out that the lower bound we have derived in this section differs from those obtained in [28] in two significant ways. First, the minimum energy dissipation derived in [28] does not take into consideration the effect of noise in the circuit at low supply voltages. The effects of noise at lower supply voltages will dictate the circuit performance and, hence, limit the levels by which we can scale the supply voltage. Secondly, we assume a constant throughput R, which automatically imposes a throughput restriction on the circuit being designed. Significantly higher reductions in lower bounds can be obtained by trading speed for power as done in [28] and [30], where the effect of threshold voltage variation on the yield due to energy minimization has been taken into consideration.

V. Design of 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 one of great importance due to the high-data rates (0.5–2.6 Gb/s), low voltage levels (0.7–0.8 V), and noisy board environment [26]. Thus, the problem boils down to the design of low-power transmitter and receiver circuits in the presence of noise. To use the concept of noise tolerance to design complex energy efficient digital systems, effective error control techniques that are inexpensive from an energy point of view are necessary. This requires exploration of existing, as well as new, reliability improvement techniques from an energy viewpoint. However, the off-chip wire example considered here is similar to the binary symmetric channel extensively studied in the area of communications. Many error control schemes are already available to improve the reliability of this channel.

We make the following assumptions: 1) C_{bus} = 50 pF; 2) gate capacitance $C_g = C_{bus}/$ 5000; 3) σ_N = 0.3 V; 4) R = 8 Mb/s; 5) k_m = 750 μ A/V^2; and 6) desired bit-error rate (*BER*) = 10^{-14}. The traditional scheme [see Fig. 10(a)], wherein the desired level of noise immunity is achieved by appropriately choosing the supply voltage to suppress the noise,

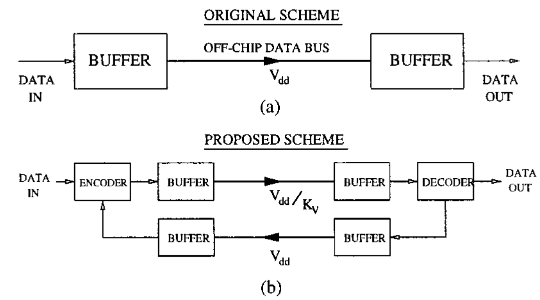


Fig. 10. (a) Traditional and (b) proposed schemes for chip I/O signaling.

requires a supply voltage of V_{dd} = 4.8 V to achieve the desired *BER* with E_b = 565 pJ/bit. Next, we propose a noise-tolerant scheme [see Fig. 10(b)] that achieves 3–4× reduction in E_b while achieving the same *BER*.

A. Noise Tolerance via Error Control

The proposed noise-tolerant scheme (NTS) is shown in Fig. 10(b). The forward channel has employed a reduced voltage level V_{dd}/K_v [where V_{dd} is the supply voltage in Fig. 10(a)], where $K_v \geq 1$ is a constant. The forward channel is noisy and makes errors with probability ϵ given by (3.6). The errors due to noise are handled via error detection using structured codes and error correction using retransmission. Retransmission requests are made over a reverse channel with signaling voltage V_{dd}. Due to the supply voltage being high, the reverse channel will be virtually noiseless. The reverse channel will be used infrequently if the errors are infrequent. The encoder and decoder operate at V_{dd} and, hence, are assumed to be noise free. Hence, energy savings would accrue only if the reduction in the supply voltage in the forward channel is able to offset the overhead due to the encoder, decoder, and the reverse channel. Note that two sets of pins are used in this scheme, compared with one in the original scheme. It may be possible to employ more sophisticated bidirectional signaling schemes as in [31], so that only one set of pins can be used in the proposed scheme.

The data to be transmitted over the low-supply line is encoded as follows. Every k bit of the data stream to be transmitted is mapped onto a code word of length n bits, where $n > k$. Note that as $n > k$, the possible 2^k *message symbols* are mapped onto only a subset (with size 2^k) of the possible 2^n *code words*. At the receiving end, the received bit stream is first decoded using a decoder. The decoder declares an error if the received string of n bits is not one of the predetermined code words. The transmitter is informed of the error using the reliable high-V_{dd} reverse channel. On receipt of an error signal from the receiver, the transmitter retransmits the message symbol.

The simplest possible coding scheme that can be employed is *n-repetition code*, where the same information bit is transmitted n times and majority logic is used at the receiving end for decoding the message bit. We show that this approach (commonly employed in fault-tolerant computing [32]) is highly inefficient in terms of energy. However, a simple Hamming code [33] results in substantial power savings.

For any positive integer $m \geq 3$, a Hamming code exists with the following parameters:

- code length: $n = 2^m - 1$;
- information symbol length: $k = 2^m - m - 1$;
- no. of parity bits: $n - k = m$.

The *BER* for an (n, k) linear code is given by [33]

$$\text{BER} = \frac{1}{n} \sum_{i=1}^{n} i A_i \epsilon^i (1-\epsilon)^{n-i} \tag{5.1}$$

where ϵ is the error probability per bit over the bus line and is given in the present context by (3.6), A_i is the number of code words with weight i in the code.

B. Energy Savings

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 that are sized to minimize delay. In the case of the conventional system [see Fig. 10(a)], the energy dissipated per information bit transmitted is given by

$$E_{b,\,old} = 0.5 V_{dd}^2 C_{bus} \tag{5.2}$$

where $V_{dd} = 4.8$ V is the supply voltage at which $\epsilon = 10^{-14}$ per bit and C_{bus} is the bus capacitance. It is assumed that the signal has a transition activity of 0.5. It can be shown (see the Appendix) that the energy dissipation for the proposed scheme [in Fig. 10(b)] is given by

$$E_{b,\,new} = \frac{V_{dd}^2 C_{bus}}{1-p_d}\left[\frac{1}{2K_v^2}\,\frac{n}{k} + \frac{p_d}{k}\right] + \frac{V_{dd} I_{sub}}{R} + 0.5 V_{dd}^2 \cdot \left((n-k)\left(2C_{and} + \frac{2k-1}{k}\,C_{xor}\right)\right)\frac{1}{1-p_d} \tag{5.3}$$

where

- I_{sub} off-state leakage current in the buffer;
- f_s input data rate in bits/s;
- C_{and} capacitance of an AND gate;
- C_{xor} capacitance of an OR gate;
- K_v factor by which the supply voltage for the forward channel in the proposed scheme is scaled down;
- p_d probability of error detection, which is given by $p_d = 1 - (p_c + p_e)$

where $p_c = (1-\epsilon)^n$ and for Hamming codes [33],

$$p_e = 2^m[1 + (2^m - 1)(1-2\epsilon)^{2^{m-1}}] - (1-\epsilon)^{2^m - 1}. \tag{5.4}$$

In the proposed scheme (5.3), both static and dynamic components of energy dissipation are taken into account as the device threshold voltage in this case is kept low to maintain throughput at low supply voltage levels. The expression for $E_{b,\,new}$ also involves energy dissipation in the encoder and decoder circuits where the supply voltage is kept high to maintain high noise margins. While we have taken three sources of energy dissipation into consideration, due to the high value of the off-chip bus capacitance, it is the dynamic component of energy dissipation that is the primary source of power dissipation for the proposed scheme.

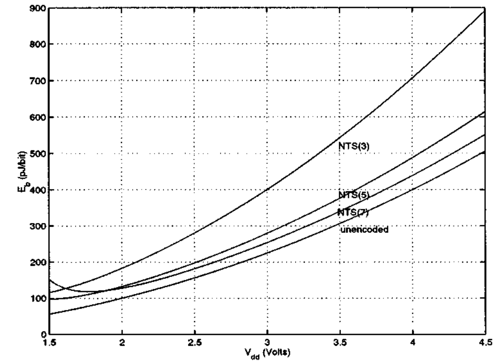


Fig. 11. Plot of V_{dd} versus E_b for the traditional scheme and proposed NTS.

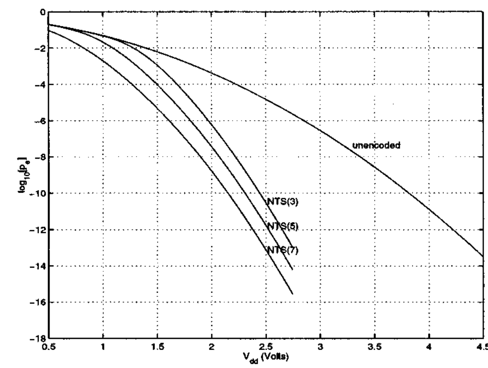


Fig. 12. Plot of V_{dd} versus $\log_{10}$ (BER) per information bit for the traditional unencoded scheme and proposed NTS.

C. Results

We now compare the performance and energy dissipation of the two schemes discussed above. Fig. 11 shows the plot of V_{dd} versus energy dissipation for the traditional and proposed schemes. Note, for the same V_{dd}, the energy dissipated in the traditional scheme is the least. For the proposed scheme, as we increase the value of m, the number of parity bits in the coding scheme, the ratio n/k decreases and, hence, energy consumption also decreases. Fig. 12 shows the plot of supply voltage versus bit error rate (BER) for the original and the proposed schemes. It can be seen that for the same V_{dd}, the BER offered by the proposed scheme is several orders of magnitude better than the one offered by the traditional scheme—especially in the range (1.5–4.8 V). Fig. 13 shows the plot of BER versus energy dissipation for the schemes considered above. Note that as expected, to achieve a specified value of BER, the traditional scheme consumes the maximum amount of energy. For the proposed scheme, the energy dissipation required to achieve a desired level of performance decreases as m is increased.

We now compare the energy dissipation of proposed schemes with the lower bounds obtained in Section III. From Fig. 6, we see that the lower bound equals $E_b = 20.5$ pJ/bit. This is shown in Fig. 14 along with the values of E_b achieved by proposed noise-tolerant schemes. Also shown is the energy efficiency of a n-repetition code with $n = 5$. Note that repetition code does not offer any energy savings as compared to the original scheme. However, a simple linear code such as the Hamming code does offer about 3× reduction in energy dissipation *while maintaining the throughput* to achieve a BER $= 10^{-14}$.

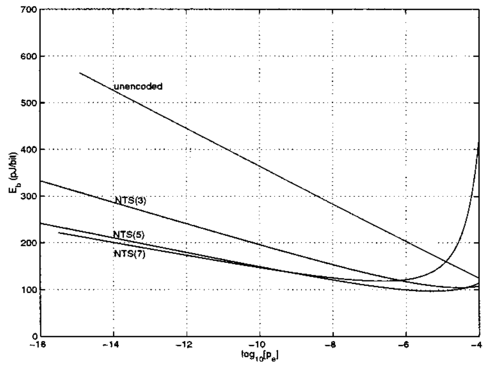


Fig. 13. Plot of $\log_{10}$ (BER) versus E_b for the traditional scheme and proposed NTS.

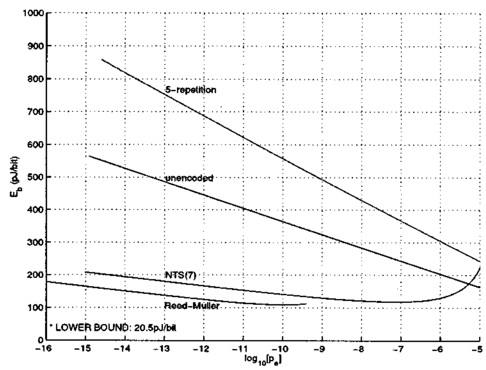


Fig. 14. Comparison of lower bound and other schemes for off-chip bus data communication.

In addition to Hamming codes, the plot of BER versus E_b for Reed–Muller (RM) codes [34] is shown in Fig. 14. Note that RM codes lead to about 4× reduction in energy dissipation at the target performance of BER $= 10^{-14}$. The improvement performance is primarily due to better error detection capability of RM codes over Hamming codes. Note, also, that the lower bound on E_b is about 24× below the E_b achieved by current day systems for the target BER $= 10^{-14}$. This indicates that substantial energy savings are possible via the use of more sophisticated coding schemes.

VI. 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. For an off-chip signaling, we have shown that the lower bounds are a factor of 24× below what present day systems achieve and that a 4× energy reduction can be achieved by employing a noise-tolerant scheme based on simple linear codes.

Future work needs to be directed toward deriving the lower bounds for multiple output functions, obtaining comprehensive noise models, and developing noise-tolerant schemes that are applicable at various levels of design hierarchy. In [35], *algorithmic noise tolerance* is applied at the system level to obtain energy-efficient implementation of DSP systems. The issue of noise tolerance at the circuit level has been recently addressed in [36], [37], where energy-efficient and noise-tolerant dynamic circuit styles have been proposed. Efficient approaches to noise in the deep micron era would require a judicious combination of noise tolerance and noise reduction at all levels of the VLSI design hierarchy.

Appendix

In this Appendix, we derive (5.3) that expresses the energy dissipation per bit for the proposed noise-tolerant scheme. The total energy dissipation per information bit is given by

$$E_{tot} = E_{dyn} + E_{stat} + E_c \tag{A.1}$$

where $E_{dyn} = E_{dyn,\,f} + E_{dyn,\,r}$ is the dynamic component of energy dissipation; ($E_{dyn,\,f}$ is the energy dissipated over the forward channel; $E_{dyn,\,r}$ is the energy dissipated over the reverse channel); E_{stat} is the static component of energy dissipation; and E_c is the energy dissipated in the encoder and the decoder.

Every time an error is detected with probability p_d, the receiver sends a retransmission request over the reliable high supply voltage (V_{dd}), reverse channel. Let N code words be transferred from the transmitting end over the noisy low supply voltage (V_{dd}/K_v) forward channel. The total number of transmissions (including retransmissions) over the forward channel required to transfer N code words is given by

$$\begin{aligned} N_t &= N[1 + p_d + p_d^2 + p_d^3 + \cdots], \\ &= \frac{N}{1 - p_d}. \end{aligned} \tag{A.2}$$

Hence, the dynamic energy dissipation over the forward channel for N code words transferred is obtained from (A.2) and (4.1) as follows:

$$E_{dyn,\,fw} = \frac{N}{1 - p_d} \frac{1}{2} \frac{V_{dd}^2}{K_v^2} C_{bus} n \tag{A.3}$$

where n is the number of bits in a code word and V_{dd}/K_v is the supply voltage over the forward channel. The number of retransmission requests made over the reverse channel while transferring N code words is given by

$$N_r = \frac{N p_d}{1 - p_d}. \tag{A.4}$$

Hence, the dynamic energy dissipation over the reverse channel for N code words transmitted is given by

$$E_{dyn,\,rw} = \frac{N p_d}{1 - p_d} V_{dd}^2 C_{bus} \tag{A.5}$$

where the transition activity is 1 due to the assumption that a retransmission request is made using transition signaling. The total dynamic energy dissipation over the two off-chip lines per information bit is given by

$$E_{dyn} = \frac{E_{dyn,\,fw} + E_{dyn,\,rw}}{Nk} \tag{A.6}$$

where k is the number of information bits transmitted per code word (recall that the coding scheme maps k information bits to an n bit code word). Substituting (A.3) and (A.5) in (A.6) we get

$$E_{dyn} = \frac{1}{1 - p_d} \left[\frac{1}{2} \frac{n}{k} \frac{1}{K_v^2} + p_d \frac{1}{k} \right] V_{dd}^2 C_{bus}. \tag{A.7}$$

The static energy dissipation per bit of information transferred is obtained from (4.2) as $E_{stat} = V_{dd}I_{sub}/R$.

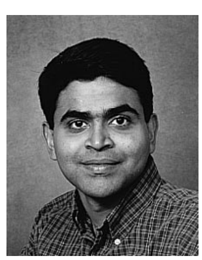
We now derive an expression for the energy dissipation in the encoder and the decoder. In the error control scheme described in this paper, we have employed Hamming codes to perform error detection. Hamming codes are linear codes which can be defined by a generator matrix and parity-check matrix in systematic form [33]. For a linear code in systematic form, the generation of the n-bit code word corresponding to a k-bit message symbol involves $k(n-k)$ AND operations and $(k-1)(n-k)$ XOR operations. The decoding of an n-bit received code word corresponding to a k-bit message symbol involves $k(n-k)$ AND operations and $k(n-k)$ XOR operations. Therefore, the total switching capacitance of the encoder and decoder is $(n-k)(2C_{and} + (2k-1/k)C_{xor})$, where C_{and} is the switching capacitance of an AND gate and C_{xor} is the switching capacitance of an XOR gate. The dynamic energy [see (4.1)] dissipated in the encoder and decoder per information bit is given by

$$E_{enc-dec} = 0.5V_{dd}^2 \left((n-k)\left(2C_{and} + \frac{2k-1}{k}C_{xor}\right)\right) \cdot \frac{1}{1-p_d}. \quad \text{(A.8)}$$

Note that when $n = k$, no coding takes place and, hence, $E_{enc-dec} = 0$ as predicted by (A.8). Also, the overhead due to retransmissions has been taken into account. Substituting (A.7) and (A.8) in (A.1), we obtain the expression for the total energy dissipation for the proposed noise-tolerant scheme as given by (5.3).

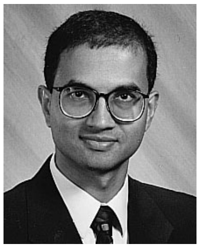
REFERENCES

[1] *The 1999 International Technology Roadmap for Semiconductors* [Online] Available: http://www.sematech.org.
[2] H. H. Chen and D. D. Ling, "Power supply noise analysis methodology for deep-submicron VLSI chip design," in *1997 Design Automation Conf.*, Anaheim, CA, pp. 638–643.
[3] K. L. Shepard and V. Narayanan, "Noise in deep submicron digital design," in *ICCAD96*, pp. 524–531.
[4] A. Devgan, "Efficient coupled noise estimation for on-chip interconnects," in *ICCAD97*, pp. 147–151.
[5] L. Gal, "On-chip cross talk—the new signal integrity challenge," in *IEEE Custom Integrated Circuits Conf.*, 1995, pp. 251–254.
[6] A. Chandrakasan and R. W. Brodersen, "Minimizing power consumption in digital CMOS circuits," *Proc. IEEE*, vol. 83, pp. 498–523, Apr. 1995.
[7] P. E. Landman and J. M. Rabaey, "Architectural power analysis: The dual bit type method," *IEEE Trans. VLSI Syst.*, vol. 3, pp. 173–187, June 1995.
[8] B. Davari, R. H. Dennard, and G. G. Shahidi, "CMOS scaling for high-performance and low-power—The next ten years," *Proc. IEEE*, vol. 83, pp. 595–606, Apr. 1995.
[9] 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.
[10] F. N. Najm, "A survey of power estimation techniques in VLSI circuits," *IEEE Trans. VLSI Syst.*, pp. 446–455, Dec. 1994.
[11] A. Shen, A. Ghosh, S. Devdas, and K. Keutzer, "On average power dissipation and random pattern testability of CMOS combinational logic networks," in *IEEE Int. Conf. Computer-Aided Design*, 1992, pp. 402–407.
[12] C. H. Bennett, "Logical reversibility of computation," *IBM J. Res. Develop.*, pp. 525–532, Nov. 1973.
[13] R. Landauer, "Dissipation and noise immunity in computation and communication," *Nature*, pp. 779–784, Oct. 1988.
[14] J. D. Meindl, "Low power microelectronics: Retrospect and prospect," *Proc. IEEE*, vol. 83, pp. 619–635, Apr. 1995.
[15] E. A. Vittoz, "Low-power design: Ways to approach the limits," in *ISSCC '94*, San Fransisco, CA, pp. 14–18.
[16] 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.
[17] J. Von Neumann, "Probabilistic logics and the synthesis of reliable organizms from unreliable components," in *Automata Studies*, C. E. Shannon and J. McCarthy, Eds. Princeton, NJ: Princeton Univ. Press, 1956, pp. 43–98.
[18] N. Pippenger, "Reliable computation by formulas in the presence of noise," *IEEE Trans. Inform. Theory*, vol. 34, pp. 194–197, Mar. 1988.
[19] B. Hajek and T. Weller, "On the maximum tolerable noise for reliable computation by formulas," *IEEE Trans. Inform. Theory*, Mar. 1991.
[20] C. E. Shannon, "A mathematical theory of communications," *Bell Syst. Tech. J.*, pt. I/II, vol. 27, p. 379–423, 623–656, 1948.
[21] P. Elias, "Computation in the presence of noise," *IBM J. Res. Develop.*, vol. 2, pp. 346–353, Oct. 1958.
[22] W. W. Peterson and M. O. Rabin, "On codes for checking logical operations," *IBM J. Res. Develop.*, vol. 3, pp. 163–168, Apr. 1959.
[23] D. K. Pradhan and S. M. Reddy, "Error-control techniques for logical processors," *IEEE Trans. Comput.*, vol. C-21, Dec. 1972.
[24] T. R. N. Rao and E. Fujiwara, *Error-Control Coding for Computer Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
[25] J. M. Rabaey, *Digital Integrated Circuits: A Design Perspective*. Englewood Cliffs, NJ: Prentice-Hall, 1996.
[26] C.-K. K. Yang *et al.*, "A 0.5-μm CMOS 4.0Gb/s serial link transceiver with data recovery using oversampling," *IEEE J. Solid-State Circuits*, vol. 33, pp. 713–722, May 1998.
[27] R. X. Gu and M. I. Elmasry, " Power dissipation analysis and optimization of deep submicron CMOS digital circuits," *IEEE J. Solid-State Circuits*, vol. 31, May 1996.
[28] D. Liu and C. Svensson, "Trading speed for low power by choice of supply and threshold voltages," *IEEE J. Solid State Circuits*, vol. 28, Jan. 1993.
[29] S. Ramprasad, N. R. Shanbhag, and I. N. Hajj, "Achievable bounds on signal transition activity," in *ICCAD97*, pp. 126–129.
[30] R. Gonzalez *et al.*, "Supply and threshold voltage scaling for low power CMOS," *IEEE J. Solid-State Circuits*, vol. 32, Aug. 1997.
[31] C. Kim *et al.*, "A 640MB/s bi-directional data-strobed, double-data-rate SDRAM with a 40mW DLL circuit for a 256MB memory," in *1998 IEEE Int. Solid-State Circuits Conf.*, Feb. 1998.
[32] D. P. Siewiorek, *Reliable Computer Systems*. Bedford, MA: Digital, 1992.
[33] S. Lin and D. J. Costello, *Error Control Coding: Fundamentals and Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1983.
[34] R. Hegde and N. R. Shanbhag, "Lower bounds on energy dissipation and noise-tolerance for deep submicron VLSI," in *1998 IEEE Int. Symp. Circuits Syst.*, Orlando, FL.
[35] ——, "Energy-efficient signal processing via algorithmic noise-tolerance," in *1999 IEEE Int. Symp. Low Power Design*, San Diego, CA.
[36] L. Wang and N. R. Shanbhag, "Noise-tolerant dynamic circuit design," in *Proc. IEEE Int. Symp. Circuits Syst.*, Orlando, FL, May/June 1999.
[37] G. Balamurugan and N. R. Shanbhag, "Energy-efficient dynamic circuit design in presence of cross-talk noise," in *1998 IEEE Int. Symp. Low Power Design*, San Diego, CA.



Rajamohana Hegde (S'95) received the B.E. degree in electrical and electronics engineering from Manipal Institute of Technology, Manipal, India, in 1992, and the M.S. degree in electrical engineering from Wright State University, Dayton, OH, in 1996. He is currently pursuing the Ph.D. degree in electrical engineering from the University of Illinois at Urbana-Champaign.

From May 1999 to August 1999, he was a summer intern at the Circuits Research Laboratory of Microcomputer Research Laboratory, Intel Corporation, Hillsboro, OR, where his work involved high-speed and noise-tolerant datapath design. His current research focus is on exploration of performance limits of VLSI systems and development of algorithmic noise-tolerance techniques for DSP and communication systems to achieve reliable low-energy operation. His research interests are in the area of VLSI design of DSP and communications systems for low-power and design of noise-tolerance techniques for deep submicron VLSI systems.



Naresh R. Shanbhag (M'88–S'88–M'93–SM'98) received the B. Tech. degree from the Indian Institute of Technology, New Delhi, India, in 1988, the M.S. degree from Wright State University, Dayton, OH, in 1990, and the Ph.D. degree from the University of Minnesota, Minneapolis, in 1993, all in electrical engineering.

From July 1993 to August 1995, he worked at AT&T Bell Laboratories, Murray Hill, NJ, in the Wide-Area Networks Group, where he was responsible for development of VLSI algorithms, architectures, and implementation of high-speed data communications transceivers. In particular, he was the Lead Chip Architect for AT&T's 51.84 Mb/s transceiver chips over twisted-pair wiring for asynchronous transfer mode (ATM)-LAN and broadband access chip-sets. In August 1995, he joined the Coordinated Science Laboratory and the Electrical and Computer Engineering Department, University of Illinois at Urbana-Champaign, as an Assistant Professor. His research interests are in the design and integrated circuit implementation of low-power/high-performance signal processing and communications systems. He has published more than 70 journal/conference articles/book chapters and holds three U.S. patents on these topics. He is also a co-author of the research monograph *Pipelined Adaptive Digital Filters* (Norwell, MA: Kluwer, 1994).

Dr. Shanbhag received the 1999 IEEE Leon K. Kirchmayer Best Paper Award, the 1999 Xerox Faculty Award, the National Science Foundation CAREER Award in 1996, and the 1994 Darlington best paper award from the IEEE Circuits and Systems Society. Since July 1997, he has been a Distinguished Lecturer for the IEEE Circuits and Systems Society. He is currently serving as an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS: PART II. He is a Member of the Design and Implementation of Signal Processing Systems (DISPS) Technical Committee of the IEEE Signal Processing Society, the VLSI Systems and Applications and VLSI in Communications Technical Committees of the IEEE Circuits and Systems Society. He has served on the technical program committees of various conferences, including the 1998 and 1999 International Symposium on Low-Power Electronics and Design, the 1999 IEEE Workshop on Signal Processing Systems, and the 1999 IEEE International Symposium on Circuits and Systems.