achieved by current day systems. This indicates that substantial energy savings are possible via the use of more sophisticated coding schemes.

5 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 $24X$ below what present day systems achieve and that a $3X$ energy reduction can be achieved by employing a noise-tolerant scheme based on a simple Hamming code.

Future work needs to be directed towards multiple output functions, 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 deep submicron noise. Efficient approaches to noise in the deep micron era would require a judicious combination of noise-tolerance and noise-reduction.

References

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

[2] H. H. Chen and D. D. Ling, "Power supply noise analysis methodology for deep-submicron VLSI chip design," *1997 Design Automation Conference*, Anaheim, CA, pp. 638-643.

[3] K. L. Shepard and V. Narayanan, "Noise in deep submicron digital design," *ICCAD96*, pp. 524-531.

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

[5] A. Devgan, "Efficient coupled noise estimation for on-chip interconnects," *ICCAD97*, pp. 147-151.

[6] 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.

[7] P. E. Landman and J. M. Rabaey, "Architectural Power Analysis: The Dual Bit Type Method," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 3, no. 2, pp. 173–187, June 1995.

[8] N. R. Shanbhag and M. Goel, "Low-power adaptive filter architectures and their application to 51.84 Mb/s ATM-LAN", *IEEE Trans. on Signal Processing*, vol. 45, no. 5, pp. 1276-1290, May 1997.

[9] B. Davari, R. H. Dennard, and G. G. Shahidi, "CMOS scaling for high-performance and low-power - The next ten years," *Proceedings of the IEEE*, vol. 83, no. 4, pp. 595-606, April 1995.

[10] D. Marculescu, R. Marculescu, and M. Pedram, "Information theoretic measures for power analysis", *IEEE Trans. on CAD*, vol. 15, no. 6, pp. 599-610, June 1996.

[11] R. Mehra and J. Rabaey, "High level estimation and exploration," *Proc. 1994 International Symposium on Low Power Design*.

[12] F. N. Najm, "A survey of power estimation techniques in VLSI circuits", *IEEE Trans. on VLSI Systems*, pp. 446-455, Dec. 1994.

[13] A. Shen, A. Ghosh, S. Devdas, and K. Keutzer, "On average power dissipation and random pattern testability of CMOS combinational logic networks," *IEEE Int. Conf. on Computer-Aided Design*, pp. 402-407, 1992.

[14] C.-Y. Tsui, M. Pedram, and A. M. Despain, "Efficient estimation of dynamic power dissipation with an application," *Proc. 1993 Intl. Conf. on Computer-Aided Design*.

[15] C. H. Bennett, "Logical reversibility of computation," *IBM Journal of Research and Development*, pp. 525-532, Nov. 1973.

[16] J. D. Meindl, "Low power microelectronics: Retrospect and prospect," *Proceedings of IEEE*, vol. 83, no. 4, pp. 619-635, April 1995

[17] E. A. Vittoz, "Low-power design: Ways to approach the limits," *ISSCC '94*, pp. 14-18, San Fransisco, CA.

[18] 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.

[19] 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.

[20] 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.

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

[22] S. Kang and Y. Leblebici, *CMOS Digital Integrated Circuits*. McGraw-Hill, 1996.

[23] R. E. Blahut, *Digital Transmission of Information*. Addison-Wesley Publishing Company, Massachusetts, 1990.

[24] S. Ramprasad, N. R. Shanbhag and I. N. Hajj, "Achievable bounds on signal transition activity, " *ICCAD97*, pp. 126-129.

[25] D. Liu and C. Svensson, "Trading speed for low power by choice of supply and threshold voltages," *IEEE Journal of Solid State Circuits*, Vol. 28, No. 1, Jan. 1993.

[26] D. P. Siewiorek, *Reliable Computer Systems*. Digital Press, 1992.

[27] B. Lau, et. al., "A 2.6 GB/s multi-purpose chip-to-chip interface," *ISSCC98*, pp. 162-163.

[28] H. Tamura, et. al., "PRD-based global-mean-time signaling for high-speed chip-to-chip communications," *ISSCC98*, pp. 164-165.

[29] S. Sidiropoulos and M. Horowitz, "A 700-Mb/s/pin CMOS signaling interface using current integrating receivers," *IEEE Journal of Solid-State Circuits*, vol. 32, no. 5, pp. 681-690, May 1997.

[30] S. Lin and D. J. Costello, *Error control Coding: Fundamentals and Applications*. Prentice Hall, Inc., 1983.

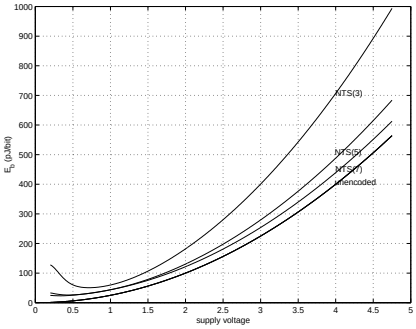

Figure 6: Plot of supply voltage vs. Energy dissipation per information bit for the traditional unencoded scheme and proposed NTS.

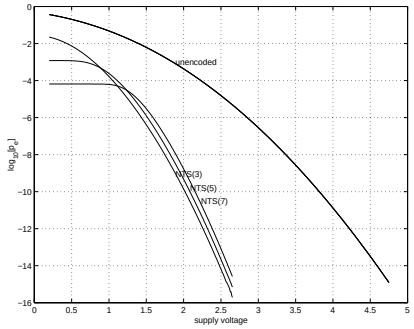

Figure 7: Plot of supply voltage vs. $log_{10}(\text{BER})$ per information bit for the traditional unencoded scheme and proposed NTS.

encoder capacitance, 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 which is given by $p_d = p_c - p_e$.

The expression for energy dissipation in the original scheme (32) includes only the dynamic component of energy dissipation. The static component of energy is negligible due to a high device threshold voltage in the buffer circuit. In the proposed scheme (33), both static and dynamic components of energy dissipation are taken into account. This is because the device threshold voltage in this case is low due to low supply voltage levels. The expression for E_{new} also involves energy dissipation in the encoder and decoder circuits where the supply voltage is kept high to maintain high noise margins.

4.3 Results

We now compare the performance and energy dissipation of the two schemes discussed above. The value of the off-state leakage current I_{sub} in the buffers is computed using (21). Note that though we have taken three sources of energy dissipation into consideration, due to the high value of the off-chip bus capacitance, the dynamic component of energy dissipation is the primary source of power dissipation for the proposed scheme.

Fig. 6 shows the plot of supply voltage vs. energy dissipation for the traditional and the proposed schemes. Note that at a given supply voltage 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. It should however be noted that at a given supply voltage the BER of the traditional scheme is much higher than that of the proposed scheme. This is indicated in fig. 7 which shows the plot of supply voltage vs. BER (bit error rate) for the original and the proposed schemes. It can be seen that at a given supply voltage the bit error rate offered by the proposed scheme is several orders of magnitude better than the one offered by the traditional scheme where only the supply voltage is varied to provide the desired level of reliability in terms of bit error rate.

Fig. 8 shows the plot of performance (in terms of bit error rate) vs. energy dissipation for the schemes considered above. Note that as expected, to achieve a specified value of bit error rate, the traditional scheme where no coding is employed consumes the maximum amount of energy. For the proposed scheme, note that as the value of m is increased, the energy dissipation required to achieve a desired level of performance decreases. This trend will however change when the complexity of the encoder and the decoder in terms

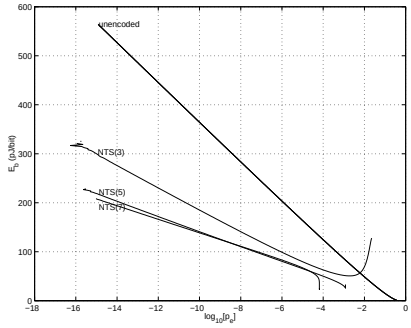

Figure 8: Plot of $log_{10}(\text{BER})$ vs. energy dissipation for the traditional unencoded scheme and proposed NTS.

of their switching capacitances becomes comparable to that of the bus.

We now compare the schemes discussed with to the lower bounds obtained in section 3 From Fig. 4, we see that the lower bound equals $E_b = 20.5$ pJ/bit. This is shown in Fig. 9 along with the values of E_b achieved by proposed noise-tolerant schemes. Also shown is 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 codes does offer about $3X$ reduction in energy dissipation *while maintaining the throughput* to achieve a $BER = 10^{-15}$. Note also that the lower bound on E_b is about $24X$ below the E_b

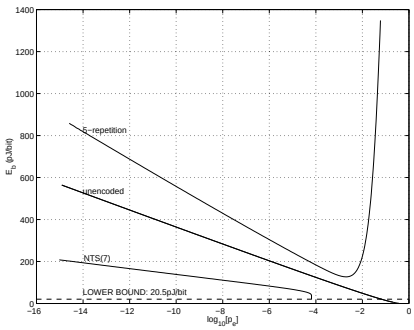

Figure 9: Comparison of lower bound and other schemes for off-chip bus data communication

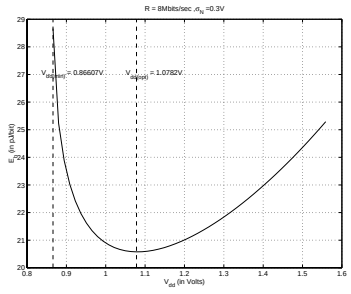

Figure 4: Plot of V_{dd} vs. E_{dyn}.

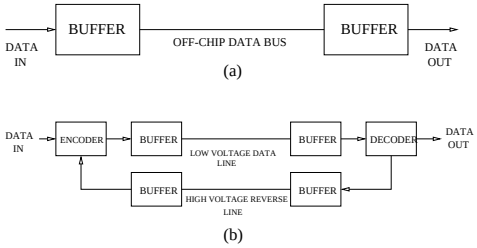

Figure 5: (a)traditional and (b)proposed schemes for chip I/O signaling

The plot of $E_b = \frac{P_{dyn}}{R}$ Vs. V_{dd} is shown in Fig. (4). First of all, note that minimum energy dissipation does not occur at the minimum possible V_{dd} which in this case is $V_{dd-min} = 0.8660V$. Increase in V_{dd} enables the circuit to be operated faster (higher f_c) and hence results in a decrease in t. Note that the decrease in t indeed offsets the increase in V_{dd} till $V_{dd} = 1.0782V$. Further increase in V_{dd} does not result in sufficient reduction in t. Hence, we see an increase in energy dissipation beyond $V_{dd} = 1.0782V$.

4 Design Example: High-Speed Chip I/O Signaling

In this section, we will demonstrate how we can get close to the lower bounds derived in the previous section for high-speed chip I/O signaling. 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 [27, 28, 29]. Thus, the problem boils down to the design of low power transmitter and receiver circuits in the presence of noise.

We make the following assumptions: 1.) $C_{bus} = 50pF$, 2.) gate capacitance $C_g = C_{bus}/5000$, 3.) $\sigma_N = 0.3V$, 4.) $R = 8$ Mb/s, 5.) $k_m = 750\mu A/V^2$, and 6.) desired bit-error rate (BER) $= 10^{-15}$. The traditional scheme (see Fig. 9) 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 3X reduction in E_b while achieving the same BER.

4.1 Noise Tolerance via Error Control

The proposed noise tolerant scheme (NTS) is shown in Fig. 6(b). The forward channel has employed a reduced voltage level V_{dd}/K_v (where V_{dd} is the supply voltage in Fig. 6(a)). The forward channel is noisy and makes errors with probability p_c given by (16. The errors due to noise are handled via error detection and error correction using retransmission. Retransmission requests are made over a reverse channel with signaling voltage V_{dd}. The reverse channel will be used infrequently if the errors are infrequent. The encoder and decoder operate at V_{dd}. 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.

Every k bits of the data stream to be transmitted is mapped onto a codeword of length n bits where $n > k$. Note that as $n > k$, the possible 2^k *message symbols* are mapped on to only a subset (with size 2^k) of the possible 2^n *codewords*. This property is made use of at the receiver end to perform error detection.

The encoded bit stream is then transmitted over the off-chip bus. The complexity of the encoder in terms of number of gates required to implement it depends on n, k and the coding scheme employed. 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 transmitted codewords. The transmitter is informed of the error using a 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 [26]) is highly inefficient in terms of energy. However, a simple Hamming code [30] results in substantial power savings.

For any positive integer $m \geq 3$, there exists a Hamming Code with 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 Hamming codes is given by

$$p_e = \left[2^m[1 + (2^m - 1)(1 - 2p_c)^{2^{m-1}}] - (1 - p_c)^{2^m - 1}\right]/k \quad (31)$$

where p_c is the error probability per bit over the bus line and is given in the present context by (16).

4.2 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 [22].

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

$$E_{old} = 0.5V_{dd}^2 C_{bus} \text{ J/bit}, \quad (32)$$

where V_{dd} is the supply voltage at which $p_c = 10^{-15}$ 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

$$E_{new} = \frac{V_{dd}^2 C_{bus}}{1 - p_d}\left[\frac{1}{2K_v^2}\frac{n}{k} + \frac{p_d}{k}\right] + \frac{V_{dd} I_{sub}}{R} + 0.5V_{dd}^2 \frac{1}{k}(C_{dec} + C_{enc})\frac{1}{1 - p_d}\text{J/bit}, \quad (33)$$

where I_{sub} is the off-state leakage current in the buffer, f_s is the input data rate in bits/sec, C_{dec} is the decoder capacitance, C_{enc} is

where the function $Q(x)$ is defined as,

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy. \quad (17)$$

It can be shown that the probability of channel error p_c increases as V_{dd} reduces. For present day voltages and technologies, σ_N is of the order of a few hundred millivolts and V_{dd} is in the range of a couple of volts. 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 1.2μ *CMOS:* Assume that $V_{dd} = 1.5V$, $\sigma_N = 0.5V$, $k = 80\mu A/V^2$ and $C_L = 50fF$. Substituting the values of V_{dd} and σ_N into (16), we obtain a value of $p_c = 0.067$. From, (13) with $p_c = 0.067$, we get $C_u = 0.6461$ bits/use. The second component f_c of the capacity equation (12) needs to be computed. In order to do this, we substitute the values of k_m, V_{dd} and C_L into (15) to get $f_c(1.5V) = 2.4 \times 10^9$ uses/sec. Therefore, the information transfer capacity of this gate is given by $C = C_u f_c = 1.55 \times 10^9$ bits/sec.§Thus, the AND gate in *Example 2* has a relatively high capacity in spite of the large noise standard deviation of $\sigma_N = 0.5V$. Employing the information-theoretic framework presented so far, we now determine the lower bounds on energy.

3.4 Lower Bound on Energy consumption

The three major sources of energy consumption in CMOS VLSI circuits [6] 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 temporary direct paths from V_{dd} to ground.

For the sake of simplicity, we assume: 1.) single output logic gates, and 2.) lumped capacitances. The expressions for the different components of power dissipation are shown below [6, 20, 21]:

$$P_{dyn} = tC_L V_{dd} f_c, \quad (18)$$

$$P_{stat} = I_{sub} V_{dd}, \quad (19)$$

$$P_{sc} = \frac{k_m}{12}(V_{dd} - 2V_t)^3 \tau f_c, \quad (20)$$

where

$$I_{sub} = K\mu C_{ox} V_t^2 e^{1.8} \left(\frac{W}{L}\right) \exp\left(\frac{-V_t}{nV_T}\right), \quad (21)$$

t is the average transition probability, C_L is the capacitance being switched, V_{dd} is the supply voltage, f_c is the channel signaling rate, τ is the 'rise-fall time' of the input signal, K is a constant dependent on gate topology ($K = 1$ for a CMOS inverter), μ is the carrier mobility, C_{ox} is the gate oxide capacitance, $\frac{W}{L}$ is the effective width-to-length ratio of the inverter, $V_T = \frac{kT}{q}$ and n is a constant with a value ranging from 1.4 to 1.5.

We employ the *energy dissipation per information bit* E_b (in joule/bit) as measure of energy efficiency given by

$$E_b = \frac{P_{tot}}{R} = \frac{P_{dyn} + P_{stat} + P_{sc}}{R}, \quad (22)$$

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. Therefore, the lower bounds on E_b can be obtained by solving the following optimization problem:

$$minimize \quad E_b = \frac{P_{tot}}{R} = \frac{P_{dyn} + P_{stat} + P_{sc}}{R} \quad (23)$$

subject to:

$$[h(t) - h(p_c)] f_c = R \quad (24)$$

$$f_c = \frac{k_m (V_{dd} - V_t)^2}{V_{dd} C_L}, \quad (25)$$

Note that the constraint (24) is obtained by replacing the information-theoretic constraint $C > R$ with a more general one: $I(X;Y) f_c \geq R$, where $I(X;Y) = [h(t) - h(p_c)] f_c$ (see (4)) for a symmetric, single output logic gate. We assume that transition signaling is used to minimize the transition activity at the output of the logic gate. It can be shown that the relationship between R, σ_N, t, C_L, V_{dd}, V_t, f_c, k, and p_c in (23), (24), and (25) is quite intricate. In the next two subsections, we employ (23) to derive two new results.

3.5 Lower Bound on Transition Activity of Noisy Logic Gates

The lower bound on transition activity derived in [24] is shown to be a special case of the bound derived in this paper.

Theorem 2: The lower bound on transition activity at the output of a symmetric, single output, noisy logic gate employing transition signaling is given by

$$t \geq h^{-1}\left[\frac{R}{f_c} + h(p_c)\right]. \quad (26)$$

Note that in the absence of noise, i.e. $p_c = 0$, and 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 (26), we obtain the lower bound on t as follows

$$t \geq h^{-1}\left[\frac{\mathcal{H}}{R_b}\right]. \quad (27)$$

which is identical to the bound in [24] derived for noiseless gates.

3.6 Lower Bound on Dynamic Energy Consumption of Noisy Logic Gates

We will now study the impact of reduction in transition activity on dynamic power dissipation. Assuming that a given circuit is operating at the lower bound given by (26), the transition activity at its output is given by,

$$t = h^{-1}\left[\frac{R}{f_c} + h(p_c)\right]. \quad (28)$$

Assuming that $V_t << V_{dd}$, from (15), we get

$$f_c = \frac{k_m V_{dd}}{C_L}. \quad (29)$$

Note that if k is fixed, V_{dd} needs to be increased in order to increase f_c to obtain a reduction in t. Substituting (28) and (29) in (18), the dynamic power dissipation is given by,

$$P_{dyn} = h^{-1}\left[\frac{RC_L}{kV_{dd}} + h(p_c)\right] V_{dd}^3 k. \quad (30)$$

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 two input AND gate from Example 1. Let the rate at which the input is being processed (rate at which the gate is clocked) be $10MHz$. For the same input statistics as in Example 1, the information transfer rate is given by

$$R = 10 \times 10^6 \times 0.8112 = 8.112 Mbits/sec. \quad (10)$$

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

2.5 Channel Capacity

The channel capacity per use C_u is obtained by maximizing (4) over all possible distributions of the channel input X. In other words,

$$C_u = \max_{\forall p(x)} \quad I(X;Y). \quad (11)$$

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

$$C = C_u f_c, \quad (12)$$

It was also shown in [19] that it is possible to achieve an information transfer rate R, (defined in (9) for a digital system [18]) for a digital system [18]) with a probability of error p_e approaching zero (via appropriate coding of the inputs) as long as $R < C$.

Example 3: For the AND gate in Example 2, assuming that the probability of error for all input combinations is 0.1 and the rate of channel use $f_c = 20MHz$, it can be shown that information transfer capacity of this gate is $10.62 Mbits/sec$. Note that as we have $C > R$, it should be possible to achieve completely reliable operation according to [19]. In practice, the probability of error can be reduced as follows. As the rate at which the circuit is operated is twice the rate at which the input arrives, for every bit of input information, we can add a parity bit that would enable us to detect and correct the errors in the output. The design of practical coding schemes that provide the best performance with least overhead is still an actively pursued area of research.

3 Information-Theoretic Framework for Noisy Gates

In this section, we first develop an information-theoretic model for a noisy digital gate and then employ it to determine the lower bounds on energy. We consider only single output functions in this paper.

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. 3(a). This diagram indicates that the probability of the output being correct is $1 - p_c$ and the probability that it is incorrect is p_c, for all inputs. Such a model is also referred to as *binary symmetric channel* (BSC) [23]. Thus, we assume that the magnitude of intermittent noise voltage is sufficient to cause logic errors. A noisy inverter and a noisy two-input AND gate can similarly be represented as shown in Fig. 3(b) and Fig. 3(c), respectively.

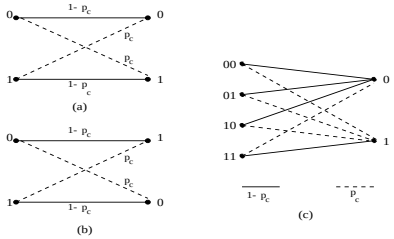

Figure 3: A channel models for:(a) binary symmetric channel, (b) an inverter, and (c) for a 2-input AND gate.

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 p_c is given by

$$C_u = 1 - h(p_c), \quad (13)$$

where $h()$ is the entropy function defined in (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:

$$p_{y,i} = \frac{1}{2}, \quad for \quad i = 0, 1. \quad (14)$$

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

For relatively high value of V_{dd} with respect to the noise voltage, $p_c = 0$, i.e., the circuit becomes error-free. In that case, (11) indicates that C_u is equal to unity and the capacity $C = f_c$ (from (12)). This is consistent with the conventional measure of capacity as being the maximum rate at which the circuit can be clocked. The result in *Theorem 1* has a dependence on the technology and the circuit style through p_c and f_c. This dependence is described next.

3.3 Characterization of f_c and p_c

We assume complementary MOS (CMOS) technology and static design style. Assuming further that the gates have been designed with balanced rise and fall times, we obtain the required expression for f_c [21] as follows.

$$f_c = \frac{k_m (V_{dd} - V_t)^2}{V_{dd} C_L}, \quad (15)$$

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 p_c 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 [3, 4, 5], in this paper we will assume that all the major sources of noise contribute a noise voltage V_N and that the gate output is in error when V_N exceeds the logic threshold voltage $V_{th} = \frac{V_{dd}}{2}$ [22]. Assuming further that V_N has a normal distribution with a variance σ_N^2, we obtain [23] p_c as,

$$p_c = Q(\frac{V_{dd}}{2\sigma_N}), \quad (16)$$

bility p_c every time it is used. The value of p_c depends upon the supply voltage V_{dd} and the variance σ_N^2 of the noise voltage V_N. This model is applicable if: 1.) the system failure modes being considered are intermittent and 2.) the system performance is quantified in terms of bit-error rates (BER) or signal-to-noise ratios (SNRs). Digital signal processing and communications systems satisfy 2.), while deep submicron technology satisfies 1.). For simplicity of exposition, we consider the noisy module to be a logic gate or a off-chip wire.

2 Information-Theoretic Preliminaries

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

2.1 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 follows

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

where $p_i \stackrel{\text{def}}{=} Pr(X = X_i)$ for $i = 0, \ldots, L-1$ 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 is its information content. In a similar fashion, a source with zero uncertainty would have zero information content and therefore its entropy would identically be equal to zero (from (1)).

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

$$h(p) = -plog_2(p) - (1-p)log_2(1-p), \tag{2}$$

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

$$h^{-1}(q) = \{p : h(p) = q, 0 \leq p \leq \frac{1}{2}\}, \tag{3}$$

where $0 \leq q \leq 1$. The function $h(p)$ is shown in Fig. 2, where we see that it achieves its maximum value of unity when $p = 0.5$. This is the same as saying that if p is the probability of observing a '1' in a binary signal then $h(p)$ is maximized if '1's and '0's are equally likely. This fact will be employed in obtaining the information transfer capacity of noisy gates.

2.2 Mutual Information and Conditional Entropy

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

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X), \tag{4}$$

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

$$H(X|Y) = -\sum_{Y \in S_Y} \sum_{X \in S_X} Pr(X,Y) log_2(Pr(X|Y)), \tag{5}$$

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

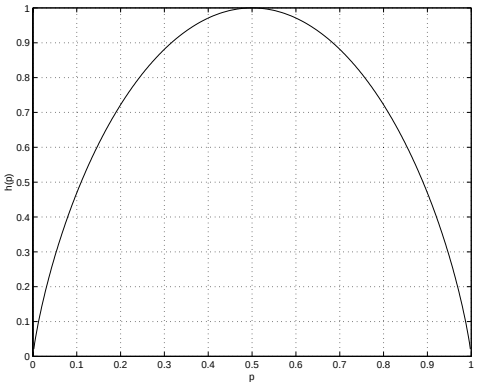

Figure 2: The entropy function $h(p)$.

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 (4) along with the fact that for a noiseless channel $H(Y|X) = 0$, provides us with the defining equation (9) for the information transfer rate R.

The following example will illustrate some of these concepts as applied to digital gates.

Example 1: An AND Gate: Consider a two-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 (1)) is given by

$$\begin{aligned} H(Y) &= -P(0)log_2(P(0)) - P(1)log_2(P(1) \\ &= -\frac{3}{4}log_2(\frac{3}{4}) - \frac{1}{4}log_2(\frac{1}{4}) \\ &= 0.8112 \text{ bits.} \end{aligned} \tag{6}$$

2.3 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

$$H(\mathcal{X}) = \lim_{n \to \infty} \frac{1}{n} H(X_1, X_2, \cdots X_n). \tag{7}$$

Note that if the stochastic process is i.i.d, we get

$$H(\mathcal{X}) = \lim_{n \to \infty} \frac{1}{n} H(X_1, X_2, \cdots X_n) = \frac{nH(X_1)}{n} = H(X_1), \tag{8}$$

which is equal to the entropy of each sample.

2.4 Information Transfer Rate

In [18], 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

$$R = f_s H(Y), \tag{9}$$

ENERGY-EFFICIENCY IN PRESENCE OF DEEP SUBMICRON NOISE[†] [*]

Rajamohana Hegde and Naresh R. Shanbhag
Coordinated Science Laboratory/ ECE Department
University of Illinois at Urbana-Champaign
Urbana, Illinois 61801
E-mail: rhegde@uivlsi.csl.uiuc.edu, shanbhag@uivlsi.csl.uiuc.edu

Abstract

Presented in this paper are: 1.) *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. A** *discrete channel model* **for noisy digital logic in deep submicron technology that captures the manifestation of circuit noise is presented. The lower bounds are derived via an information-theoretic approach whereby a VLSI architecture implemented in a certain technology is viewed as a channel with information transfer capacity C (in bits/sec). A computing application is shown to require a minimum information transfer rate R (also in bits/sec). Lower bounds are obtained by employing the information theoretic constraint $C > R$. This constraint ensures reliability of computation though in an asymptotic sense. Lower bounds on transition activity at the output of noisy logic gates are also obtained using this constraint. Past work (for noiseless bus coding) is shown to fall out as a special case. In addition, lower bounds on energy dissipation is computed by solving an optimization problem where the objective function is the energy subject to the constraint of $C > R$. A surprising result is that in a scenario where capacitive component of power dissipation dominates:** *the voltage for minimum energy is greater than the minimum voltage for reliable operation*. **For an off-chip I/O signaling example, we show that the lower bounds are a factor of 24X below present day systems and that a very simple Hamming code can reduce the energy consumption by a factor of 3X. This indicates the potential of noise tolerance (via error control coding) in achieving low energy operation in the presence of noise.**

1 Introduction

The 1997 National Roadmap for Semiconductors [1] describes 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. This is difficult to do in deep submicron 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. In other words, deep submicron technology is inherently noisy with noise comprising of ground bounce, cross-talk, process variations, charge sharing, charge leakage, etc.. This is the reason for the recent interest in deep sub-micron noise analysis [2, 3, 4, 5].

[*†] This work was supported by DARPA contract DABT63-97-C-0025 and NSF CAREER award MIP-9623737.

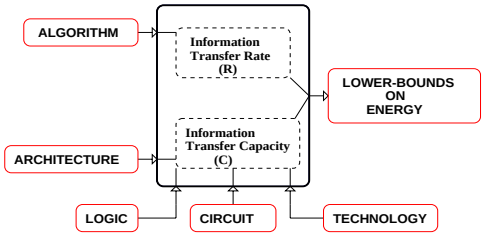

Figure 1: The information-theoretic framework for VLSI.

Designing low-power integrated circuits in the presence of deep submicron noise is a challenging problem because it requires us to explore the energy-reliability curve. It is not enough to reduce energy per se but to be able achieve sufficiently reliable operation at the same time. At present, low-power design is of great interest [6] driven mainly by the need to extend battery life per unit weight in mobile applications. Research in this area revolves around the development of *low-power design* techniques at various levels of the design hierarchy [6, 7, 8, 9], *power estimation* techniques [10, 11, 12, 13, 14], and investigating the *lower bounds on power dissipation* [15, 16, 17, 18]. However, the impact of noise has not been considered so far and in particular the following important questions 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 especially in the presence of noise?" In this paper, we provide answers to some of these questions for simple digital systems.

In this paper, in continuation of our past work [18], we present an information-theoretic framework (see Fig. 1) for determining lower bounds on energy of digital systems while ensuring reliable computation. Our main thesis shown in Fig. 1, is that computation needs to be viewed as a process of information transfer over a noisy channel. We develop a *discrete channel* model for digital systems and then develop the capacity formula to calculate the lower bounds. In this model, we assume that a module can make errors with proba-