A Mathematical Basis for Power-Reduction in Digital VLSI Systems

Naresh R. Shanbhag, *Member, IEEE*

Abstract—**Presented in this paper is a mathematical basis for power-reduction in VLSI systems. This basis is employed to: 1) derive lower bounds on the power dissipation in digital systems; and 2) unify existing power-reduction techniques under a common framework. The proposed basis is derived from information-theoretic arguments. In particular, a digital signal processing algorithm is viewed as a process of information transfer with an inherent information transfer rate requirement of R bits/s. Architectures implementing a given algorithm are equivalent to communication networks each with a certain capacity C (also in bits/s). The absolute lower bound on the power dissipation for any given architecture is then obtained by *minimizing* the signal power such that its channel capacity C is equal to the desired information transfer rate R. By including various implementation constraints, increasingly realistic lower bounds are calculated. The usefulness of the proposed theory is demonstrated via numerical calculations of lower bounds on power dissipation for simple static CMOS circuits. Furthermore, a common basis for some of the known power-reduction techniques such as parallel processing, pipelining and adiabatic logic is also provided.**

Index Terms—**Digital CMOS, information theory, low-power.**

I. INTRODUCTION

NUMEROUS APPLICATIONS in the area of signal processing and communications have emerged in recent years, which require implementation of highly complex algorithms. These applications include low bit-rate speech and image compression, image recognition, mass data storage, high bit-rate digital subscriber loops, asynchronous transfer mode (ATM) based local area networks (LAN) transceivers and broad-band access. The VLSI implementation of these algorithms is made difficult by their computational complexity. This problem is compounded further by the stringent requirements on the power dissipation. It is well known that reduction of power dissipation is imperative in mobile applications due to the limited life time of a battery-based energy supply. Even in nonmobile applications, reduction of power dissipation is important as it improves the reliability and reduces packaging costs. Therefore, development of power reduction techniques is currently of great interest.

An excellent review of existing power reduction techniques is provided in [6], [18], [36]. Most of the work done in the

past can be placed into the following three broad categories:

1) development of power reduction techniques;
2) development of power estimation techniques;
3) investigating the lower bounds on power dissipation.

Work done in Category 1 can be viewed as a collection of techniques to reduce power at all possible levels of the VLSI design methodology. These techniques include pipelining and parallel processing [6], [18] (architectural level), precomputation logic [1], logic minimization [25] (logic level), optimal coding [7], adiabatic computation [2], [12] (circuit level), device threshold voltage (V_t) reduction [5] (device level), and at the technological level [11] etc.

Estimating the power dissipation (Category 2) for a given architecture is another important problem that has received a lot of attention in recent years [22], [23], [26], [27], [33], [34]. In [22], an information theoretic approach to power estimation was proposed. In particular, the average value for the number of transitions occurring at the output of a Boolean function was obtained as $H(Y)/2$, where $H(Y)$ is the entropy of the output Y. An empirical measure for the average number of transitions in an implementation of a Boolean transformation was derived in [27]. This measure was found to be a function of the input and output entropies and the number of inputs and outputs. An elegant way to estimate the power dissipation of Boolean circuits from the register-transfer level (RTL) model was then presented in [27].

While a lot of effort has been directed toward the development of power reduction and estimation techniques, few results exist in Category 3. Some of the work done in Category 3 include [4], [20], [24], [37]. In [37], the lower bound on power dissipation per pole for analog circuits was presented. This bound is based upon the desired signal-to-noise ratio (SNR). Empirical lower bound estimates for digital circuits was also presented in [37], which was based upon a desired SNR, estimates of gate complexity and energy per operation. For an irreversible logical operation, a lower bound of kT (k is the Boltzmann constant and T is the absolute temperature) was proposed in [4]. The corresponding lower bound for logically reversible operation was presented to be zero. Employing thermodynamic arguments [24], the thermal noise power spectrum was determined to be kT and the minimum energy required for a logic change was shown to be lower bounded by approximately $4kT$. In [20], the *order* of the lower bound on switching energy was determined for logic circuits. This bound was derived by assuming a certain amount of switching energy per logic transition and the topology of logic gates.

In this paper, we employ an information-theoretic approach to develop a mathematical basis for power-reduction in VLSI systems. The proposed basis has two advantages: 1) it allows us to derive lower bounds on the power dissipation in digital systems; and 2) it enables us to unify existing power-reduction techniques under a common framework. In particular, we view any given digital signal processing transformation Γ as a process requiring an information transfer rate of R bits/s (bps). Here Γ could represent a Boolean truth table or even a finite-precision digital signal processing algorithm. This rate R is shown to be an inherent property of the algorithm and the input statistics and is independent of its implementation. Furthermore, we view a given architectural implementation of Γ (to be referred to as $Arch(\Gamma)$) as a network of communication channels with a capacity C. Employing the well known source-channel coding theorem from [32], we know that the network capacity $C > R$ (for error-free information transfer) and that C is a monotonically increasing function of the SNR. Hence, a minimum SNR ($\mathrm{SNR_{min}}$) is required to achieve this rate of information transfer. Finally, given the knowledge of the noise properties of the medium and $\mathrm{SNR_{min}}$, we determine the minimum signal power and hence the lower bound on power dissipation. Thus, the bounds presented in this paper are algorithm and technology dependent, with the link between the two being provided via information-theoretic concepts.

Inherent in [37] and [24] is the knowledge of the desired $\mathrm{SNR_{min}}$. In this paper, we provide a framework in which $\mathrm{SNR_{min}}$ can be computed via information theoretic results [32]. In addition, we will develop a unified mathematical basis for some of the major power reduction techniques from Category 1 via the proposed theory. In particular, we show that power reduction is achieved in all cases by reducing C toward the required minimum given by R. Hence, the theory presented in this paper, bridges the gap between work done in Category 1 and Category 3. This is achieved by employing the proposed theory to provide a unifying link between the seemingly disparate power reduction techniques in Category 1.

As mentioned before, information-theoretic approaches have been employed to develop power estimation techniques [22], [27]. Furthermore, these concepts have also been employed in the past to define a measure of *computational work* of a Boolean transformation [9], [17]. This measure of computational work is closely related to the area [8] occupied by an implementation of the Boolean function. Our work differs significantly in that we begin with an implementation-independent view of an algorithm (Γ) and then proceed to determine the lower-bounds on power dissipation for a given architecture ($Arch(\Gamma)$). Nevertheless, we will see that estimation of the output entropy is an integral part of the proposed theory. This is a difficult problem for large systems and therefore techniques proposed in [22], [27] would need to be applied. Without loss of generality, we will assume in this paper that the output entropy is known.

This paper is organized as follows. In Section II, we will review the relevant concepts from information theory. The inherent information transfer rate requirements R of a logic transformation will be defined in Section III. Next, in Section IV,
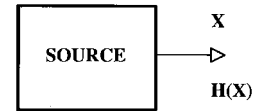

Fig. 1. A data source.

we will show how a digital architecture can be modeled as a communication network. In Section V, we present a theory for obtaining lower bounds on power dissipation for a given digital system. Applications of the proposed theory will be discussed in Section VI where we will determine the lower bounds on the power dissipation for certain digital systems and also provide a unifying link between some of the existing power reduction techniques such as parallel processing, pipelining and adiabatic logic.

II. PRELIMINARIES

In order to provide the necessary background, we will review some basic information-theoretic concepts including Shannon's joint source-channel coding theorem [32] in this section. The interested reader is referred to [10], [15], [21], [32] for a more detailed formulation of these concepts.

A. Entropy: A Measure of Information

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

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

where $P_i \stackrel{\text{def}}{=} \Pr(X = X_i)$ for $i = 0, \cdots, L-1$ and $H(X)$ is in bits. 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 (2.1)).

If the source generates symbols at a rate of f_{op} symbols per second, then the *information generation rate* G is given by

$$G = f_{\mathrm{op}} H(X). \qquad (2.2)$$

The entropy of source provides us with a lower bound on compressibility of the source. In other words, no lossless encoder can represent a source with B bits, where $B < H(X)$.

It is easy to show that

$$H(X) \leq L \qquad (2.3)$$

and the distribution (for a discrete alphabet) which achieves this upper bound is the uniform distribution. A corresponding formulation for the entropy of a continuous source also exists. However, given our interest in digital systems in this paper, (2.1)–(2.3) will suffice.
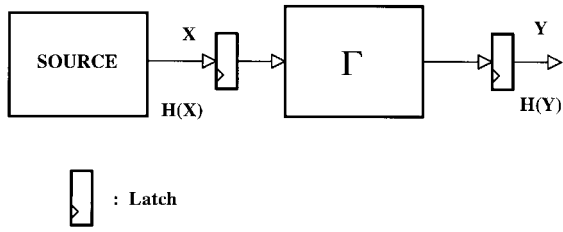
Fig. 2. A data source followed by a deterministic transformation Γ.

B. Information Transfer Rate

Assume that the output of the source in Fig. 1 is passed through a transformation $\Gamma:Z^L \to Z^M$ (see Fig. 2), where Z is the set of integers, to generate an output $Y = \Gamma(X)$. Without loss of generality, we assume that the inputs and outputs of Γ are latched. The *mutual information* $I(X;Y)$ is defined as

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) \quad (2.4)$$

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

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

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 transformation Γ to its output. Thus the *information transfer rate* R is defined as

$$R = f_{\text{op}} I(X;Y) \quad (2.6)$$

where f_{op} is the rate at which the symbols are generated by the source. Note that R is a function of the input statistics and the transformation Γ.

Consider the special case of $Y = X$ corresponding to an identity transformation. In this case $H(X|Y) = 0$ and from (2.4), $I(X;Y) = H(X)$. Thus, the transformation Γ in Fig. 2 needs to transfer $H(X)$ bits of information per symbol or $R = G$ bits/s, where G is given by (2.2). Note that all linear digital signal processing systems can be modeled as a transformation Γ and therefore have a certain information transfer rate R associated with them. We will elaborate upon this point in Section II.

C. Channel Capacity

A generic communication system is shown in Fig. 3(a). The source with entropy $H(X)$ generates the data X. The encoder may involve removal of redundancy (data compression) or addition of redundancy (error control) to the input X in order to produce the output X'. The transmit shaping block converts the encoder output X' into a form with appropriate spectral characteristics, which makes it suitable for transmission over the physical channel. In particular, the transmit shaping
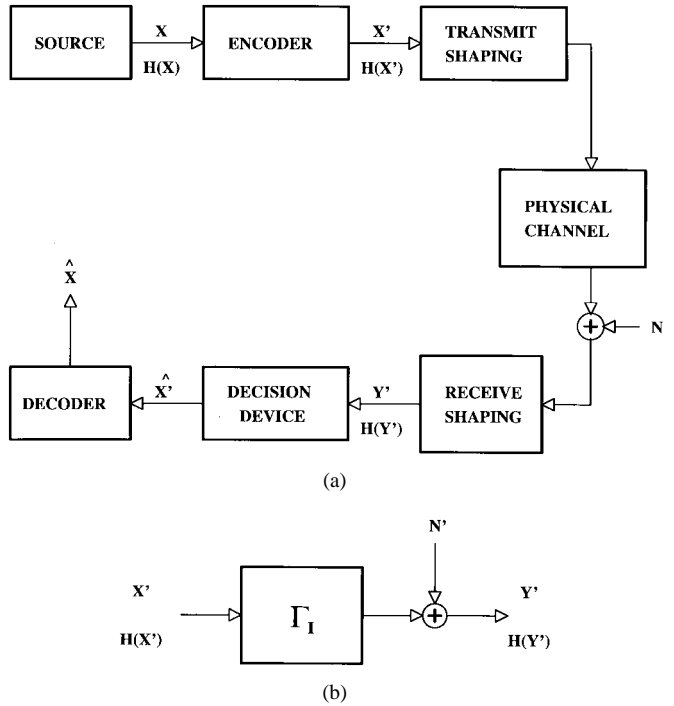

Fig. 3. A generic digital communication channel. (a) Block diagram. (b) Its alternative representation.

determines the *transmission bandwidth* W. Furthermore, the physical channel itself has an *inherent bandwidth* given by W_{ch}, which is a function of the physical characteristics of the media. The transmit shaping guarantees that $W \leq W_{ch}$ and therefore the overall system comprising of the transmit shaping, the physical channel and receive shaping has a bandwidth of W.

The channel output is typically a distorted version (amplitude distortion or phase distortion or both) of its input. The channel also has noise (N) superimposed to generate the input to the receiver. The noise N is typically assumed to be white with a Gaussian distribution but could also be a cyclostationary interferer such as near-end cross talk. The receiver front-end shapes the received signal to remove noise and channel distortion as much as possible to generate Y', where

$$Y' = X' + N' \quad (2.7)$$

and N' is the residual noise component in Y'. The ratio of the power in signal X' to that of the noise N' in (2.7) provides us with the SNR for the channel under consideration. In general, the SNR is a function of frequency.

The channel from the encoder output (X') to the receive filter output (Y') in Fig. 3(a) can be recast into the form shown in Fig. 3(b). Thus, a generic digital communication system is a noiseless identity transformation Γ_I followed by additive noise N'. From the discussion in the previous subsection (see Fig. 2), we know that the information that is transferred across a transformation Γ is given by $I(X;Y)$ in (2.4). In Fig. 3(b), the information that needs to be transferred across the identity transformation Γ_I is $H(X')$. However, due to the additive noise N', $H(Y'|X')$ is not equal to zero and

is a function of the probability distribution of X' given the probability distribution of N'. By maximizing (2.4) over all possible distributions of X', we obtain the channel capacity C.

In his seminal work [32], Shannon showed that the capacity (C) of a channel bandlimited to frequency W, whose output is of the form (2.7), is given by

$$C = \int_0^W \log_2[1 + \mathrm{SNR}(f)] \, df \qquad (2.8)$$

where C is in bps. In this case, the equivalent channel comprises of a cascade of the transmit shaping block, the physical channel and the receive shaping. From (2.8), it is clear that the capacity C depends upon the SNR and the transmission bandwidth W.

It was also shown in [32] that an information transfer rate R with probability of error approaching zero is *achievable* (via appropriate coding) as long as $R < C$. While the existence of the best codes was shown in [32], the design of practical codes which achieve the capacity is still in progress. It must be mentioned that since the publication of [32], significant progress has been made to find these optimal codes. In most cases, however, the accompanying decoding procedure is computationally intensive.

The decision device in Fig. 3(a), determines the symbol $\hat{X}'$ that was most likely to have been transmitted. Therefore, in an error-free situation $\hat{X}' = X'$. In the simplest case, the decision device could be a slicer, which makes a decision based upon the Euclidean distance of Y' from all possible transmitted symbols. Such a device is said to make decisions on a symbol-by-symbol basis. On the other hand, in *sequence detection*, decision is based upon observing a sequence of symbols. An example of such a scheme is the Viterbi detector [35]. Finally, the decoder employs an inverse of the encoder mapping to transform the detected symbols $\hat{X}'$ to detected data $\hat{X}$, where in an error-free case $\hat{X} = X$.

D. Network Information Theory

As mentioned in Section I and to be discussed further in Section IV, there is a direct correspondence between an architectural implementation of an algorithm and a communication network. Hence, there is a need for a generalized version of (2.8) (which describes the capacity of a point-to-point link) for communication networks. Computation of the capacity of such a general network and demonstrating its achievability is still an open problem [10]. However, capacity calculations for certain special cases such as parallel channels, broadcast channel etc., have been done. In the absence of such a general theory, our application of information theoretic concepts to the problem at hand would be restricted to the employment of (2.8) and the capacity formulas for the special cases. In this paper, we will employ an abstraction (see Section IV) of a digital system implemented in a noisy media, which allows us to apply (2.8) to general digital systems.

III. INFORMATION TRANSFER RATE OF A LOGIC TRANSFORMATION

In this section, we will show that any logic transformation Γ that is to be implemented, can be viewed as a process
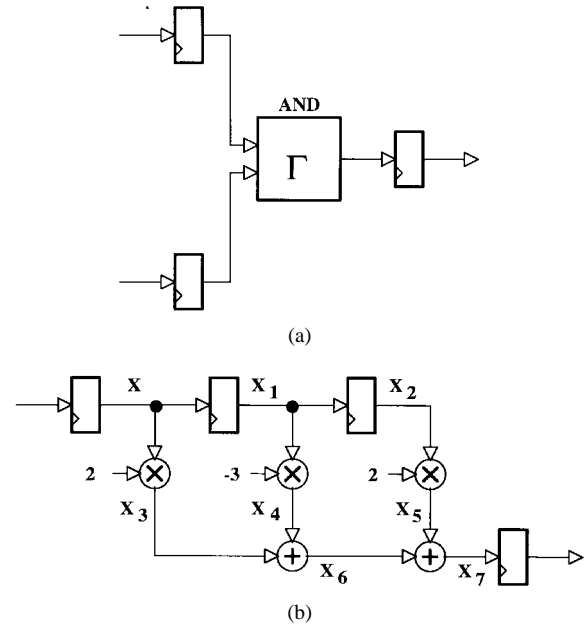

Fig. 4. Digital systems. (a) Latched AND gate. (b) An FIR filter.

of information transfer. We will then determine the inherent information transfer rate requirements R associated with such a logic transformation Γ. In all cases, we will assume that the inputs and outputs of Γ are latched synchronously. We will see that R is related to the output entropy $H(Y)$ and that it is possible to calculate the output entropy $H(Y)$ of a transformation Γ if the input probability distribution is known.

Example 1: An AND Gate: Consider a latched two-input AND gate as shown in Fig. 4(a). If we assume that both inputs are independent and equally likely to be a "1" or a "0," then the entropy of the input is 2 bits. The entropy of the AND gate output Y is given by

$$
\begin{aligned}
H(Y) &= -P(0)\log_2(P(0)) - P(1)\log_2(P(1)) \\
&= -\tfrac{3}{4}\log_2(\tfrac{3}{4}) - \tfrac{1}{4}\log_2(\tfrac{1}{4}) \\
&= 0.8112.
\end{aligned}
\qquad (3.1)
$$

Thus, the AND gate output has a smaller information content than its input.   §

A similar analysis for a 1-bit full adder in [22] determined that for a uniformly distributed input X, $H(X) = 3$ bits and $H(Y) = 1.8113$ bits. In general, we present the following result [10] for any deterministic transformation

Theorem 1: If $\Gamma: \mathbf{Z}^L \to \mathbf{Z}^M$ is a deterministic mapping, where $\mathbf{Z}$ is the set of integers, then

$$H(X) \geq H(Y)$$

where $H(X)$ and $H(Y)$ are the entropies of the input X and the output Y, respectively.

Proof: From (2.4), we obtain

$$
\begin{aligned}
H(X) - H(X|Y) &= H(Y) - H(Y|X) \\
&= H(Y)
\end{aligned}
\qquad (3.2)
$$

where we have utilized the fact that $H(Y|X) = 0$ for a deterministic mapping Γ. Furthermore, we know that the entropy

H is a nonnegative function of the probability distribution of its argument. This implies that $H(X|Y) \geq 0$ and substituting this result in (3.2) gives

$$H(X) \geq H(Y) \tag{3.3}$$

which is the desired result. ∎

From an information theoretic point of view, Theorem 1 implies that a digital logic transformation either reduces or maintains the information content of its input. This implication is consistent with our notion of a logic gate as being incapable of introducing any uncertainty (and therefore increasing the entropy) in transforming its inputs to generate the outputs.

We can also interpret Theorem 1 from a digital signal processing viewpoint. Any linear finite-precision digital filter can be represented as a transformation Γ as defined in Theorem 1. Clearly, such a filter will either remove information (if the input signal has energy in the filter stop-band) or maintain it (if all the energy in the input signal lies in the filter pass-band). In both cases, the entropy of the output cannot be greater than that of the input, which is also predicted by Theorem 1. Note that the transformation Γ can be either recursive or nonrecursive.

Corollary 1: $H(Y)$ achieves the upper bound in Theorem 1, when Γ is an injective (one-to-one) mapping.

Proof: If Γ is an injective mapping, then $H(X|Y) = 0$. Substituting this value for $H(X|Y)$ in (3.2) gives the desired result. ∎

In the case where Γ is injective, the output Y is simply a recoding of the input X. Hence, the uncertainty in the input is translated directly to the output.

Corollary 2: For the mapping Γ in Theorem 1, the mutual information is given by

$$I(X;Y) = H(Y) \tag{3.4}$$

and the information transfer rate is

$$R = f_{\mathrm{op}}H(Y). \tag{3.5}$$

Proof: The left-hand side of (3.2) equals $I(X;Y)$ [see (2.4)]. Hence the proof of (3.4). From (2.6) and (3.4), we obtain (3.5). ∎

From Corollary 2, any transformation Γ requires an information transfer of $H(Y)$ bits per symbol. Thus, Γ can be viewed as an information sieve, which takes in $H(X)$ bits/symbol and lets $H(Y)$ bits/symbol pass through it.

From (3.5), the information transfer rate R is a function of f_{op} and $H(Y)$. We also know that $H(Y)$ is dependent upon the probability distribution of X and the transformation Γ. Thus, we conclude that R is a function of the input statistics and Γ and is independent of its implementation. An alternative formulation of (3.5) is possible whereby the dependence of R on the input statistics can be eliminated. To do this we can maximize $I(X;Y)$ over all possible probability distributions of X. This will provide us with the maximum possible information transfer rate R, which now depends only upon Γ. However, without any loss of generality, we will employ (3.5) so as to be able to exploit the dependence of switching activity on input statistics.

It is instructive to consider the special case of $R = 0$ in (3.5). Clearly, R will equal zero if either f_{op} and/or $H(Y)$ is zero. If either $f_{\mathrm{op}} = 0$ (static input case) and/or $H(Y) = 0$ (known output case) then there is no need to realize the system at all in which case there will be zero power consumption. This observation is consistent with the result predicted by the proposed theory that if $R = C = 0$ then (from (2.8)) $\mathrm{SNR}(f) = 0$. In that case, the signal power and therefore the power dissipation is also zero.

Example 2: An FIR Filter: In this example, we will demonstrate the calculation of the information transfer rate R for a digital filter given by $H(z^{-1}) = 2z^{-1} - 3z^{-2} + 2z^{-3}$. While the calculation of the output entropy $H(Y)$ is independent of the architectural implementation, we will employ the direct-mapped architecture shown in Fig. 4(b) only to illustrate the procedure. In Fig. 4(b), the input has a 2-bit representation ($X(n) \in \{0, 1, 2, 3\}$) and the filter has three taps. Without loss of generality, we assume that the input $X(n)$ is independent and identically distributed (i.i.d) with a uniform distribution. In that case, the entropies $H(X_1) = H(X_2) = 2$ bits because latches have an injective input-output mapping. Similarly, we can see that the signals X_3, X_4 and X_5 also have values, which are unique for different inputs. This follows from the injective nature of multiplication and therefore $H(X_3) = H(X_4) = H(X_5) = 2$ bits. However, the mapping which generates X_6 as an arithmetic sum of X_3 and X_4 is not injective. In this case, $X_3 \in (0, 2, 4, 6)$ and $X_4 \in (0, -3, -6, -9)$. All combinations (X_3, X_4) are equiprobable with a probability of $1/16$. Furthermore, the combinations $(0, 0)$ and $(6, -6)$ result in the output being a zero. Thus, we may calculate $H(X_6)$ as shown below

$$H(X_6) = -(12)(\tfrac{1}{16})\log_2(\tfrac{1}{16}) - (2)(\tfrac{2}{16})\log_2(\tfrac{2}{16}) = 3.75. \tag{3.6}$$

In a similar fashion, the entropy of the output $H(Y) = H(X_7)$ can be obtained as follows:

$$\begin{aligned} H(X_7) = &-(8)(\tfrac{5}{64})\log_2(\tfrac{5}{64}) - (4)(\tfrac{3}{64})\log_2(\tfrac{3}{64}) \\ &-(4)(\tfrac{2}{64})\log_2(\tfrac{2}{64}) - (4)(\tfrac{1}{64})\log_2(\tfrac{1}{64}) \\ =&\, 4.128. \end{aligned} \tag{3.7}$$

Thus, we may view the FIR filter in Fig. 4(b) as a transformation Γ, which requires an information transfer of $H(X_7)$ bits per symbol. In this example, the input entropy is equal to the entropy of the vector $[X, X_1, X_2]$. This is because in the proposed model, the latches are viewed as synchronous data transceivers. Therefore, the FIR filter in Fig. 4(b), has three input transmitters and one output receiver. It can be seen that the output entropy $H(X_7)$ is less than the input entropy $H(X) + H(X_1) + H(X_2) = 6$ (input X is i.i.d), which is consistent with Theorem 1. §

Thus, all digital transformations, in particular linear finite-precision digital signal processing algorithms have an inherent information transfer rate requirement R given by (3.5). This requirement is an inherent property of the transformation and is *independent of the implementation media or the architecture*. In the next Section, we will show that an architecture, which implements Γ, is a communication network with a certain capacity C. Based on the results of this Section and from [32],

we know that C has to be greater than R for any meaningful computation to take place.

IV. DIGITAL SYSTEM AS A COMMUNICATION NETWORK

In this section, we will consider a *particular implementation* $Arch(\Gamma)$ of a *given transformation* Γ and determine parameters such as inherent channel bandwidth W_{ch} and capacity C. Clearly, there can be many different digital architectures which achieve the same functionality. For each architecture, the information transfer capacity C can be determined and would be a function of the topology and the SNR. From Shannon's source-channel coding theorem [32], this capacity C should be greater than or equal to R.

The design of a communication system usually begins with the characterization of the physical link or the channel. The goal is then to transmit a certain amount of information rate R through this channel under certain constraints such as transmit power, BER etc.. Sophisticated data/signal processing (such as line coding and forward error correction) is employed at the transmitting and more so at the receiving end in order to achieve the specified performance. In contrast, we will see that the design of a digital system is akin to the selection or development of an appropriate communication network topology with sufficient capacity C. At the present time, due to technological constraints, the line coding is restricted to binary signalling and data/signal processing involved is a simple two-level slicing.

We first provide a physical basis in order to motivate the discussion. This is done by qualitatively analyzing a single-bit two-latch system [see Fig. 5(a)] implemented in the commonly employed complementary metal-oxide semiconductor (CMOS) technology. Next, we will show that any particular architectural implementation of a general transformation Γ can be viewed as a network of communication channels with a certain inherent bandwidth W_{ch} and capacity C, which can be computed.

A. A Single-Bit Two-Latch System

Consider a digital system comprising of two latches as shown in Fig. 5(a). It can be seen that this system is an implementation of the identity transformation Γ_I [see Fig. 3(b)].

The latch [LATCH(T)] is the *transmitter*, where the encoder maps a logical "1" to a "1" and a logical "0" to a "0" (assuming that the latch output is noninverting). The *transmit shaping* is provided by the power supply waveform and the mode of switching of the output transistors in the latch. As mentioned in Section II(C), the transmit shaping determines the *transmission bandwidth* W. We can assume, without any loss of generality, that the output stage of the latch is an inverter and the power supplies are direct current (dc). Furthermore, if the transistors are modeled as ideal switches [see Fig. 5(b)], then output waveform would correspond to a *binary antipodal* signalling scheme, where a logical "1" corresponds to a positive or V_{dd} pulse and logical '0' would corresponds to a negative or V_{ss} pulse. There are situations where the power supply could be also modulated as in [13], which would then correspond to a different pulse shape.
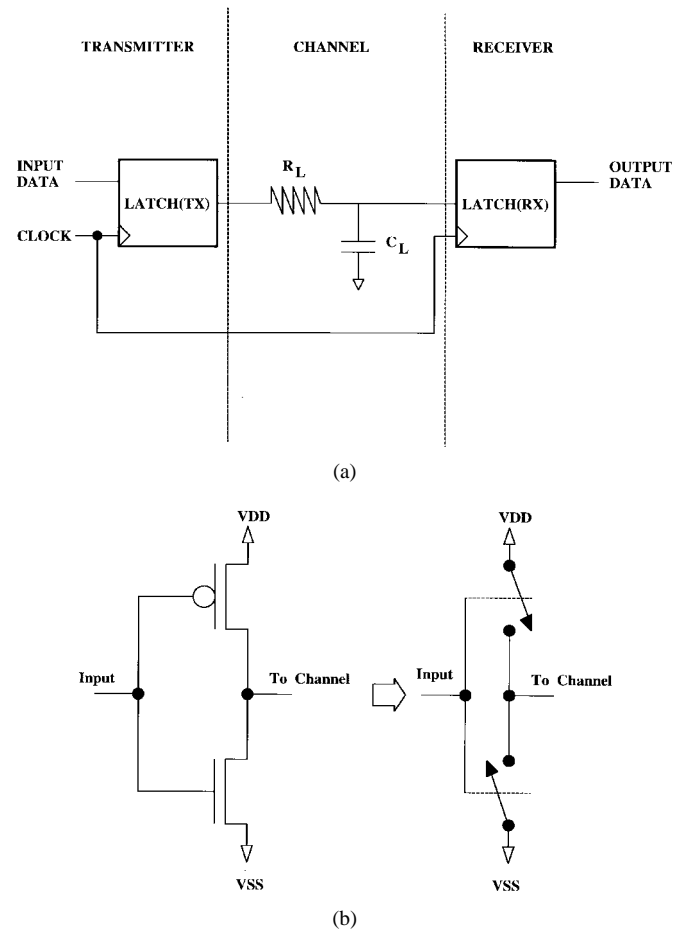

Fig. 5. A two-latch system. (a) Block diagram. (b) Switch-level representation for the single bit.

The connection between the two latches is the physical channel, which is characterized by the resistance R_L and the capacitance C_L of the link. While more accurate formulas for the delay experienced by a transmitted pulse can be employed [3], we shall ignore it without any loss of generality and for the sake of simplicity. As the transmit pulse is provided by the switching of the power supply, the resistance R_L would include the ON-resistance of the NMOS or PMOS transistor and the wire resistance. The capacitance C_L is a combination of the interconnect capacitance and the capacitance of the input stage of LATCH(R). It is clear that the 3 dB bandwidth of such a channel is given by $1/(R_L C_L)$ Hz, and hence we define the inherent bandwidth W_{ch} as

$$W_{ch} = \frac{1}{R_L C_L}. \qquad (4.1)$$

Note that in CMOS circuits, both R_L and C_L are dependent upon the circuit layout and the supply voltage V_{dd}. For simplicity and without loss of generality, we will ignore the dependence of R_L and C_L on V_{dd}.

In accordance with Fig. 3, the noise source N needs to be determined for the system in Fig. 5(a). Unlike communication systems, the total noise power in digital systems is a function of numerous factors such as signal power, temperature, semiconductor properties, frequency of operation, etc. Some of the noise sources, such as shot noise, flicker noise, thermal noise

etc. are described in [16] along with their respective power spectral densities. However, for conventional digital systems the noise is mainly due to the phenomenon of ground bounce. This noise occurs due to the inductive coupling of the power supply lines and the switching that occurs in digital systems. From [19, Ch. 12], the ground bounce voltage is given by

$$V_{gb} = \frac{4C_L V_{dd} L}{t_s^2} \qquad (4.2)$$

where L is the inductance and t_s is the time for which the switching current is flowing in the interconnect. For present day voltages and technologies, V_{gb} is of the order of a few hundred millivolts. Taking $t_s = 2R_L C_L$ and substituting this value in (4.2), we obtain

$$V_{gb} = \frac{V_{dd} L}{C_L R_L^2} = \frac{V_{dd} L W_{ch}}{R_L}. \qquad (4.3)$$

From (4.3), we notice that the noise reduces with W_{ch}. This is very much in line with what is observed in communications systems, where reduction in the bandwidth results in reduced noise power at receiver. A complete characterization of the relevant noise sources, i.e., sources which can cause a detection error, needs to be done before the application of the proposed theory. In this paper, we may assume for the sake of simplicity that the total noise can be represented as an uncorrelated source with power $\sigma_N^2 = V_{gb}^2$ over the inherent channel bandwidth W_{ch}.

The block LATCH(R) functions as the receiver in Fig. 5(a). While there is no receive shaping (see Fig. 3) in LATCH(R), it does act as a detector. The function of this receiver is to determine if the received signal is a "1" or a "0." Without loss of generality we may assume the input stage to be an inverter. In that case, the detector behaves identical to a symbol-by-symbol slicer, where the decision made by comparing the received signal with a logic threshold voltage V_{TH}. The logic threshold voltage V_{TH} for a logic gate depends upon the technology, the relative geometry of the transistors, and the temperature [39] etc.. We may assume the logic threshold voltage for an inverter to have a nominal value of $V_{dd}/2$, where V_{dd} is the supply voltage. Furthermore, analogous to the system in Fig. 3, the receiver LATCH(R) can make an error if the noise level is sufficiently high. In other words, detection error's can occur at the receiver if the SNR is small enough. For traditional digital systems, the supply voltage V_{dd} and therefore the SNR is very high. This is the reason why digital systems have excellent noise-immunity as compared to analog systems. For the same reason, traditional digital systems have a higher power dissipation than equivalent analog systems [37].

Timing recovery is an integral part of any communication system as it provides the correct sampling epoch for discrete-time processing in the receiver. From Fig. 5(a), we see that the transmitter LATCH(TX) and the receiver LATCH(RX) have the same clock. In the framework of a digital communication system, this is analogous to the case where a separate link ("clock under the table") is provided for transmitting timing information. In case of an asynchronous digital system, the use of the "handshake" protocol implies the employment of a forward and a reverse timing channel.
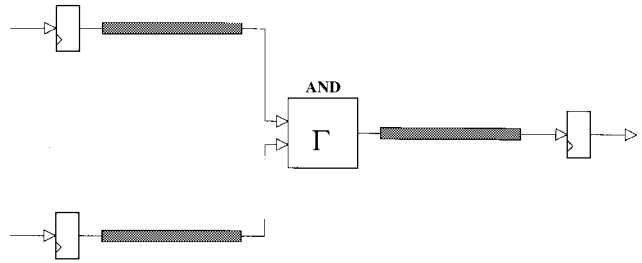

Fig. 6. Fine-grain representation of the latched AND gate system.

Note that the rate of information transfer R over the link in Fig. 5(a) is given by

$$R = f_{op} * H(X) \qquad (4.4)$$

where $H(X)$ is the entropy of the input data and f_{op} is the desired frequency of operation. Assuming that the input data is binary with ones and zeros being equally likely, we get $R = f_{op}$.

Therefore, we have now been able to relate all the parameters and components of the generic communication channel in Fig. 3 to those in the two-latch digital system. Extending the single-bit case to the multiple-bit case is straightforward and will not be considered here. Next we will consider a general digital system.

B. Digital System: Fine-Grain Representation

We will start with a *fine-grain* representation and determine the inherent channel bandwidth W_{ch} and capacity C for a given digital system. Under this representation, we view all logic gates as *data-driven transceivers* and latches as synchronous transceivers. A data-driven transceiver consists of a receiver followed by a transmitter without any timing information. In this case, the data-driven transceiver operates in a mode, where the receiver makes a new decision whenever there is a change in the received signal. The recovered data is then provided to the transmitter, which generates a new pulse whenever a new decision is made. In addition, the connections between the logic gates are viewed as bandlimited communication channels of the type described in Section IV-A.

Example 3: Latched AND Gate:

As an example, we show the latched AND gate of Fig. 4(a) in Fig. 6 employing a fine-grain representation. The latched AND gate consists of one data-driven transceiver (the AND gate itself), three synchronous transceivers (the two input latches and one output latch) and three communication channels $ch_i, i = 0, 1, 2$. In Fig. 6, the thick lines indicate physical interconnect with a certain inherent channel bandwidth $W_{ch,i} = 1/(R_{L,i} C_{L,i}), i = 0, 1, 2$. The quantities $R_{L,i}$ and $C_{L,i}$ are a function of the layout and the process parameters. However, in our search for low-power architectures and/or comparison of two different architectures, we will assume that each physical link is identical and has the same total resistance R_L and capacitance C_L. This assumption will allow us to focus upon the topology of $Arch(\Gamma)$ instead of the physical details. In that case, each channel has an inherent channel bandwidth W_{ch} given by (4.1) and the noise voltage given by

V_{gb} in (4.3). Furthermore, the capacity $C_i, i = 0$, 1, 2 of each link can be computed from (2.8) with $W \le W_{ch}$.   §

Each of the signal lines in Fig. 6 has a certain information transfer rate $R_i, i = 0$, 1, 2. Therefore, the capacity $C_i > R_i$, $i = 0$, 1, 2, and hence the minimum SNR_i for each channel can also be computed. The total power dissipation for $Arch(\Gamma)$ would then be obtained by summing over each link. In summary then, the fine-grain representation treats each link between logic blocks as a noisy bandlimited communication channel. While the latched AND gate consists of 3 communication channel, the FIR filter in Fig. 4(b) has 10 communication channels and in general it would have $4N - 2$ channels for N-taps. Thus, the complexity of a fine-grain representation increases quite rapidly with the size of the digital system. Therefore, it is of interest to develop a *coarse-grain* representation, which will be presented next.

C. Digital System: Coarse-Grain Representation

A primary goal of the *coarse-grain* representation is to be able to lump the nonideal characteristics (bandlimited noisy channels) of $Arch(\Gamma)$ at the output Y. This will allow us to represent any implementation $Arch(\Gamma)$ in an equivalent form shown in Fig. 3(b) with $\Gamma_I = \Gamma$. Under the coarse-grain representation, we have a noiseless implementation followed by a noisy bandlimited channel. In this subsection, we will determine the noise source and the bandwidth W_{ch} of the coarse-grain representation.

First, we consider the determination of the equivalent noise source. We can represent any $Arch(\Gamma)$ (with a corresponding mapping Γ') in a noisy media as shown in Fig. 7(a), where noise could have many sources including the implementation media itself. The definition of Γ' is shown in Fig. 7(b), where the input space S_X is mapped onto the output space $S_{Y'}$. The dark dots in the set S_X represent the discrete values that the input X can assume. In addition, the dark dots in the set S_Y denote the values that the output can assume if the noise power were zero. Note that for a given input X, the output could belong to a set of values whose boundary is shown in dotted lines in Fig. 7(b). The output latch acts as a receiver and can produce a noiseless output Y provided the subsets (shown by dotted lines) do not intersect. The range of values that can be assumed by the noisy output Y' is entirely a function of output SNR. If the probability density function of the noise is identical for all possible noiseless outputs Y, then we can represent the system in Fig. 7(a) as shown in Fig. 7(c) with the corresponding mapping for Γ as shown in Fig. 7(d). In this figure, all the noise in Γ' has been *referred to the output* and we now have a noiseless transformation Γ mapping the input space S_X to a noiseless output space S_Y. In this paper, we will assume that the Fig. 7(c) can be employed to represent any $Arch(\Gamma)$ in a noisy media.

All that remains in completing the coarse-grain description is a definition of the bandwidth W_{ch}. As mentioned in Sections IV-A and B, the bandwidth W_{ch} for a point-to-point link is given by the quantity $1/(R_L C_L)$ [see (4.1)], which is also equal to the reciprocal of the delay from the input to the output of the channel. Therefore, we choose to define the bandwidth of the coarse-grain representation of $Arch(\Gamma)$ as follows:

$$W_{ch} = 1/t_{cp} \tag{4.5}$$

where t_{cp} is the delay time of the critical path of the architecture. Clearly, W_{ch} is a function of the topology and the bandwidths of the individual point-to-point links in $Arch(\Gamma)$. From Fig. 7(c) and (2.4), we can determine the capacity C by maximizing the mutual information $I(X; Y)$ as follows:

$$\max[I(X; Y')] = \max[H(Y') - H(Y'|X)]$$
$$= \max[H(Y') - H(Y'|Y)]. \tag{4.6}$$

Comparing (4.6) with (2.4) and Fig. 7(c) with Fig. 3(b), we conclude that the capacity of the channel from the input X to the output Y' in Fig. 7(c) is given by (2.8) with $W = W_{ch}$ [see (4.5)] and an appropriate noise source.

Example 4: Latched AND Gate: A coarse-grain representation of the latched AND gate is shown in Fig. 8, where the input data is processed in a noiseless fashion and then sent into a noisy bandlimited channel at the output. The equivalent channel bandwidth [from (4.5)] is given by

$$W_{ch} = \frac{1}{2R_L C_L} \tag{4.7}$$

where we have utilized the fact that the critical path of the system in Fig. 6 is $t_{cp} = 2R_L C_L$. The output referred noise voltage would be a sum of the inherent noise of the output link (V_{gb}) and the additional noise due to the input. The noise in the inputs to Γ in Fig. 6 is invisible to the output unless it is sufficiently high to cause a detection error. In the latter case, we need to add voltage spikes of amplitude $V_{TH} + \epsilon$ to get the following equivalent noise source:

$$V_{gb,eq} = V_{gb} + V_{TH} + \epsilon \tag{4.8}$$

where V_{TH} is the logic threshold and $\epsilon > 0$. A probabilistic description of (4.8) can be generated by realizing the fact that the noise spike will be created only when the input makes a transition. For simplicity and without any loss of generality, we will assume throughout this paper that $V_{gb,eq} = V_{gb}$.   §

Therefore, in a coarse-grain representation of $Arch(\Gamma)$, we model all latches as transceivers. The logic between any two latches is modeled as a front-end noiseless implementation of Γ followed by a noisy bandlimited channel. Being noiseless, the front-end has infinite capacity [see (2.8)] with an infinitesimally small signal power. Hence, the front end consumes zero power. The noisy bandlimited channel will consume power and the lower bounds obtained by analyzing this channel will represent the lower bounds for $Arch(\Gamma)$. In this paper, we will consider only a coarse-grain representation due to its inherent simplicity. Furthermore, we do not lose any generality with this representation if the output referred additive noise source [see Fig. 7(c)] has the correct power spectral density and the probability distribution function.
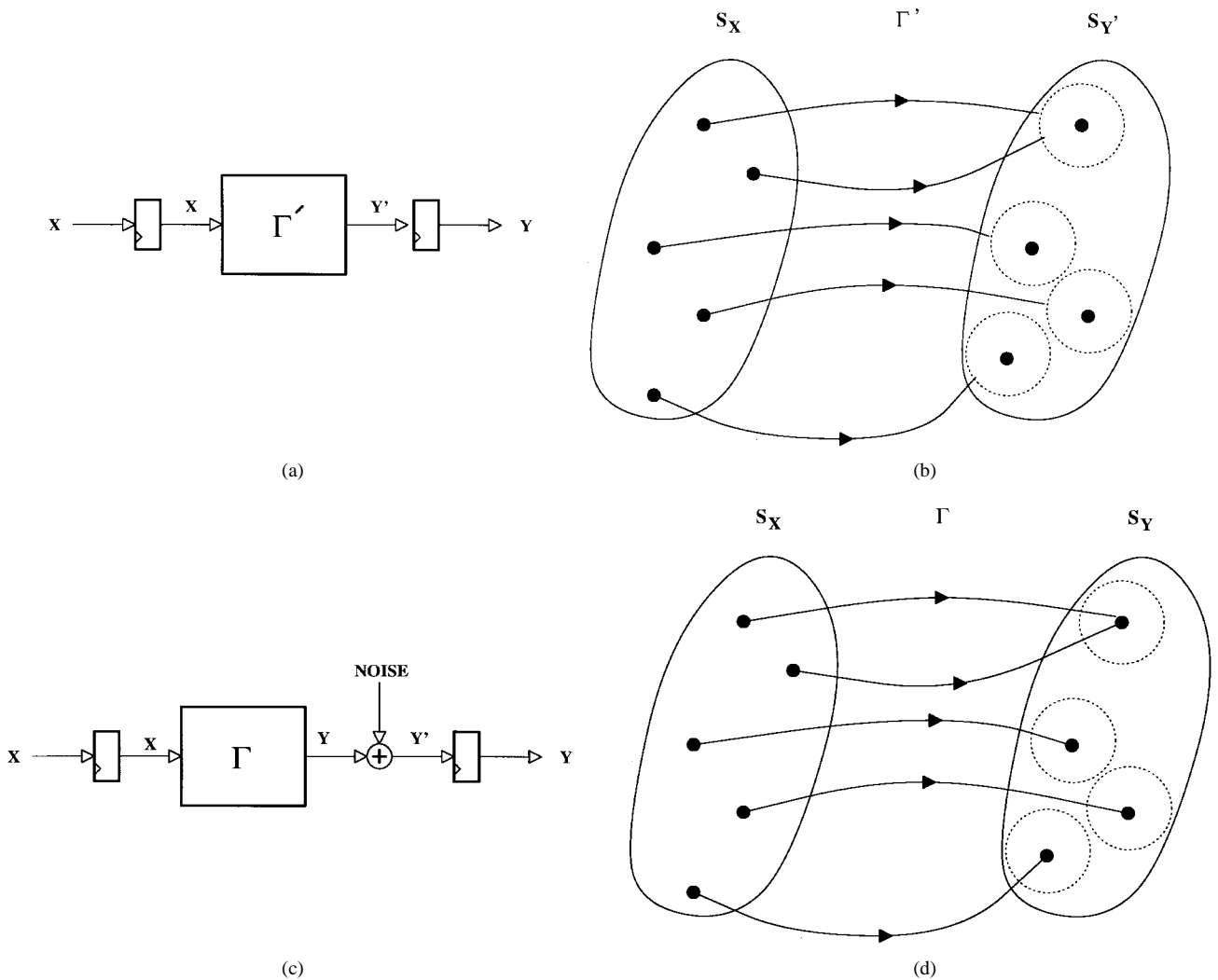
Fig. 7. Implementation of $Arch(\Gamma)$ in a noisy media. (a) Block diagram. (b) Input–output mapping for (a). (c) Representation with output referred noise. (d) Input–output mapping for (c).
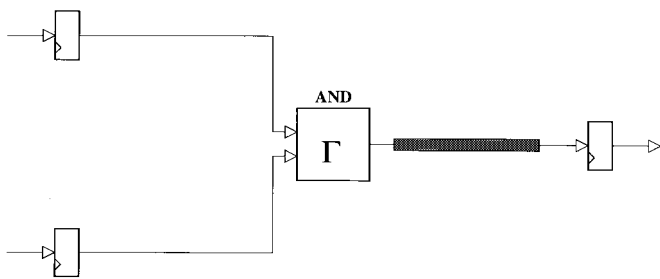

Fig. 8. Coarse-grain representation of the latched AND gate system.

V. LOWER BOUNDS ON POWER DISSIPATION

In this section, we present the main results of this paper. In particular, we employ the relationship between channel capacity and the SNR for a given channel to determine the lower bound on power dissipation in digital systems.

A. Lower Bounds

As discussed in Section III, we may view any architecture $Arch(\Gamma)$ as a communication channel of the form shown in Fig. 7(c). Furthermore, from the discussion in Section III-B, we know that the expression (2.8) applies to the channel in Fig. 7(c) and that it relates the channel capacity C to the SNR. As we know the desired rate of information transfer R [from (3.5)], which depends upon the frequency of operation f_{op} and the entropy of the noiseless channel output $H(Y)$, we can determine the minimum required SNR. Finally, we employ the relationship between signal power and power dissipation to obtain the lower bound.

Formally, we present the following result.

Theorem 2: For a given channel with the following properties:

1) bandlimited to W Hz;
2) with a noise power spectral density of $S_{NN}(f)$;
3) a desired information transfer rate R bps;
4) power dissipation at frequency f ($P_D(f)$) being related to the signal power spectrum $S_{XX}(f)$ as

$$P_D(f) = F[S_{XX}(f)] \qquad (5.1)$$

where $F(\ )$ is a linear monotonically increasing function of its argument.

The lower bound on the power dissipation for such a channel is given by

$$P_D > F\left[\int_0^W (\nu_{\min} - S_{NN}(f))^+ \, df\right] \qquad (5.2)$$

where $(arg)^+$ denotes the positive part of (arg) and $\nu_{\min}$ is a unique constant which can be obtained as a solution to the following equation:

$$R = \int_0^W \log_2\left[1 + \frac{(\nu_{\min} - S_{NN}(f))^+}{S_{NN}(f)}\right] df. \qquad (5.3)$$

The proof of Theorem 2 is provided in Appendix A. We can derive the following corollaries from Theorem 2.

Corollary 3: If the channel described in Theorem 2 has a flat signal and noise power spectra, the lower bound on power dissipation is given by

$$P_D > F[(2^{R/W} - 1)\sigma_N^2] \qquad (5.4)$$

where σ_N^2 is the integrated noise power from zero to W Hz.

Proof: From the joint source-channel coding theorem [32], we obtain

$$R < \int_0^W \log_2\left[1 + \frac{\sigma_X^2}{\sigma_N^2}\right] df \qquad (5.5a)$$

$$R < W \log_2\left[1 + \frac{\sigma_X^2}{\sigma_N^2}\right] \qquad (5.5b)$$

$$(2^{R/W} - 1)\sigma_N^2 < \sigma_X^2 \qquad (5.5c)$$

$$(2^{R/W} - 1)\sigma_N^2 < F^{-1}[P_D] \qquad (5.5d)$$

$$F[2^{R/W} - 1)\sigma_N^2] < P_D. \qquad (5.5e)$$

Hence the proof. ■

Corollary 4: The lower bound (5.2) is achieved, when the capacity of the channel C equals the desired rate of information transfer R.

Proof: This is evident from the definition of $\nu_{\min}$ in (5.3). ■

The proof of Theorem 2 and Corollary 4 is summarized in Fig. 9. We start with the minimum required information transfer rate R and the knowledge of the noise power spectrum $S_{NN}(f)$. We know from the joint source-channel coding theorem [32] that reliable transmission is possible as long as the capacity $C > R$. Hence, we compute the minimum $\mathrm{SNR}(f)$ such that $C = R$. This provides us with the minimum SNR ($\mathrm{SNR}_{\min}(f)$). Employing the knowledge of $S_{NN}(f)$ and $\mathrm{SNR}_{\min}(f)$, we can obtain the desired signal power spectral density $S_{XX}(f)$. Substituting $S_{XX}(f)$ into (5.1) and integrating over the transmission bandwidth W gives us the lower bound on power dissipation.

For a given signal processing transformation Γ, Theorem 1 allows us to calculate the information transfer rate R. Furthermore, Theorem 2 enables us to determine the lower bound on the power dissipation of a particular $Arch(\Gamma)$. Thus, the design of low-power architectures is equivalent to searching for an appropriate communication network. The best network (from a low-power point of view) would be the one which has the maximum capacity for a given SNR
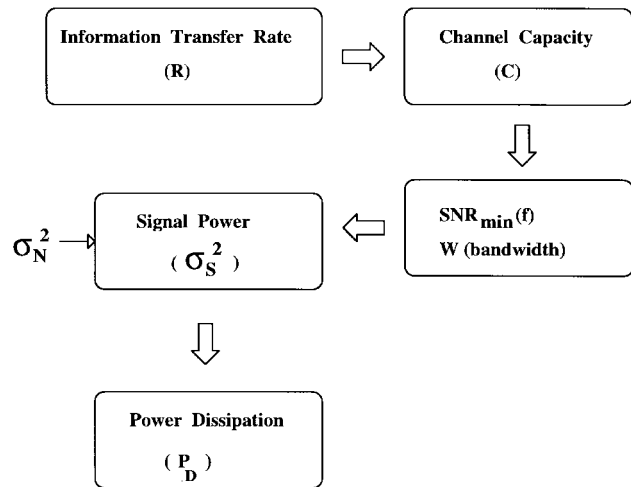

Fig. 9. The proposed theory.

and a fixed number of point-to-point links. Recall that [see Section IV-B] different architectures correspond to different communication networks. In both cases, Theorem 2 does not provide us with a technique for achieving this lower bound. This is not surprising in the light of the fact that Theorem 2 is derived from Shannon's joint source-channel coding theorem [32]. The joint source-channel coding theorem [32] provides an *achievable* (via appropriate coding) upper bound on the information carrying capacity of a given channel. However, it does not indicate the method for achieving this bound.

It is possible to compute lower bounds by including constraints from the implementation domain. For example, we may restrict the transmit levels to just two. This will imply a certain transmission bandwidth W and hence a different lower bound than the absolute bounds, which are calculated without any constraints. Therefore, a variety of lower bounds on the power dissipation can be obtained via Theorem 2. Furthermore, we may employ the bounds to compare the power reduction capabilities of different power-reduction techniques. This will be elaborated further in Section VI.

VI. APPLICATIONS

In this section, we demonstrate the use of Theorem 2 to calculate the lower bounds on power dissipation digital VLSI systems. Just for the sake of demonstration, and without any loss of generality, we assume that the noise spectrum is flat with average power $\sigma_N^2 = 10^{-2}$ V^2 over a channel bandwidth of $W_{ch} = 150$ MHz. This is a typical number for a submicron CMOS technology, where the ground bounce is approximately a few hundred millivolts. Based on the discussion in Section III-A, and assuming ideal switches, we can approximate the transmit pulse as a square-wave with an amplitude of $\pm V_{dd}/2$ volts for transmitting a "1" and "0," respectively. The signal power σ_X^2 (or the variance) is therefore equal to $V_{dd}^2/4$.

We must now define the function F for CMOS technology (see Theorem 2), which relates the signal power to the power dissipation. In order to do this, we have to consider the fact that the derivation of (2.8) (see [10] [15], [21], [32]) assumes

a maximum of $2W$ uses per second of the channel in Fig. 3. This maximum is dictated by the well-known Nyquist criterion for the transmission of noninterfering pulses over a channel bandlimited to W Hz. A question that needs to be answered is whether it is possible to have a lower power dissipation by having fewer than $2W$ uses per second and yet meet the information transfer rate of R. In Appendix B, we show that the lowest power dissipation is achieved when the channel is indeed used at the maximum possible rate of $2W$ uses per second. Hence, the function F for CMOS is defined as

$$F = C_L V_{dd}^2 W \tag{6.1}$$

where we assume $2W$ uses per second.

In this section, we will compute the lower bounds on power dissipation employing circuit, logic and architectural level techniques. Implicit in the computation is the assumption that the detection threshold voltage $V_{TH} = V_{dd}/2$. This is a commonly employed assumption for static digital CMOS circuits and can be met via proper transistor sizing.

A. Circuit Level: A Single Bit Two-Latch System

For the single-bit two latch system [see Fig. 5(a)], we have a requirement on the information transfer rate of $R = 100$ Mb/s. Assume further that $W = W_{ch} = 150$ MHz. From Theorem 2, we get

$$C > R$$
$$W \log_2(1 + \mathrm{SNR}) > R$$
$$W \log_2\left(1 + \frac{V_{dd}^2}{4\sigma_N^2}\right) > R$$
$$V_{dd} > [(2^{R/W} - 1)4\sigma_N^2]^{1/2}. \tag{6.2}$$

Substituting (6.2) into (6.1) gives us the lower bound on the power dissipation as follows

$$P_D > C_L(2^{R/W} - 1)4\sigma_N^2 W. \tag{6.3}$$

The lower bounds on the supply voltage and the power dissipation for the single bit case can be obtained by substituting $R = 100$ Mb/s, $W = 150$ MHz, $\sigma_N^2 = 10^{-2}$ V^2, and $C_L = 0.5$ pF into (6.2) and (6.3), respectively. These lower bounds are

$$V_{dd} > 0.1533 \quad \text{V} \tag{6.4a}$$
$$P_D > 0.001\,762 \quad \text{mW.} \tag{6.4b}$$

The value of V_{dd} obtained via the proposed theory is in the same range as the 20 MHz encoder-decoder circuit in [5], where $V_{dd} = 0.2$ V was employed. As mentioned before, the lower bound in (6.4) is achievable but we do not provide a technique for doing so.

B. Circuit Level: A Multibit Two-Register System

For the multibit case, the total capacity of a M-bit word-to-word link is given by

$$C = \sum_{i=0}^{M-1} W_{ch,i} \log_2[1 + \mathrm{SNR}_i] \tag{6.5}$$

where SNR_i and $W_{ch,i}$ are the SNR and the bandwidth for the ith channel, respectively. Without loss of generality, we assume that all the M channels are identical with $\mathrm{SNR}_i = \mathrm{SNR}$ and $W_{ch,i} = W$. From (6.5), we obtain

$$C = MW \log_2\left(1 + \frac{V_{dd}^2}{4\sigma_N^2}\right). \tag{6.6}$$

Employing Theorem 2 and (6.6), the minimum value of V_{dd} is obtained as follows:

$$C > R$$
$$MW \log_2\left(1 + \frac{V_{dd}^2}{4\sigma_N^2}\right) > R$$
$$V_{dd} > [(2^{R/MW} - 1)4\sigma_N^2]^{1/2}. \tag{6.7}$$

Comparing (6.7) and (6.2), we observe that the minimum supply-voltage for the multibit case is smaller than that of a serial channel. The lower bound on the power dissipation is obtained as follows:

$$P_D > C_L(2^{R/MW} - 1)4\sigma_N^2 W * M$$
$$> C_L(2^{R/MW} - 1)4\sigma_N^2 MW. \tag{6.8}$$

Thus, for $M = 4$, the lower bounds are

$$V_{dd} > 0.070 \text{ V} \tag{6.9a}$$
$$P_D > 0.001\,469\,5 \text{ mW.} \tag{6.9b}$$

A comparison of (6.9) and (6.4) shows that the fundamental lower bound on P_D for a parallel set of M channels is less that of a serial channel with the same information transfer rate requirements. Note that we are comparing the *lower bounds* on power dissipation for a serial and a parallel architecture via (6.4) and (6.9). This is quite different from the comparisons done traditionally [6], where the power reduction due to parallel processing was computed. Nonetheless, our results are consistent with [6] as it implies that a parallel architecture operating at its minimum level of power dissipation would consume less power than a corresponding serial architecture.

C. Circuit Level: Adiabatic Logic

Adiabatic computation [2], [12] and pulsed power-supply CMOS logic [13] work in a remarkably different fashion as compared to other power reduction techniques. In particular, these techniques rely upon the fact that power dissipation can be minimized by ensuring that the voltage across any resistor is kept as small as possible. This is shown in Fig. 10, where the capacitor C_L is charged up by applying a voltage source $V(t)$, whose rise time $T_r \gg R_L C_L$. Under this condition, the power dissipation can *approach* zero. A lower bound on the energy dissipation for adiabatic logic has been calculated in [2], [12] for a given value of T_r. In this subsection, we show that T_r is a function of the supply-voltage V_{dd} and the information transfer rate R. Hence, there is an upper limit for T_r and hence a lower bound on the power dissipation. In practice, R_L is the source to drain resistance of a MOSFET, which is a function of the current. However, for the sake of
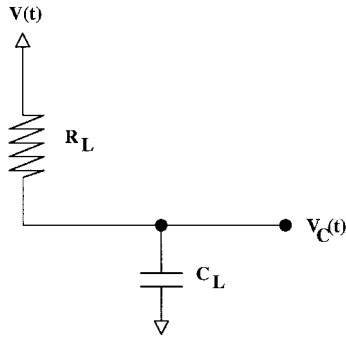
Fig. 10.   Principle of adiabatic computation.

simplicity and without any loss of generality, we will assume that R_L is a constant.

The power reduction property of adiabatic computation [2], [12] and pulsed power-supply CMOS logic [13] can be explained via the information-theoretic framework presented in this paper. By applying a time-varying voltage source $V(t)$, adiabatic computation involves reducing the bandwidth of the transmit pulse and hence the transmission bandwidth W. From (2.8), we see that reducing W also reduces the capacity C of the channel. Furthermore, it can be shown that the power dissipated in an adiabatic computation is a function of W. For a desired information transfer rate R and supply voltage V_{dd}, there is a lower limit on W and hence on the power dissipation. The limit to which the power dissipation can be reduced is computable from Theorem 2. Assuming a sinusoidal $V(t)$, we show in Appendix C that the power dissipation for adiabatic computation is given by

$$P_{D,adb} = \frac{V_{dd}^2}{2R_L}[\tan^{-1}(2\pi W R_L C_L)]^2 \qquad (6.10)$$

where W is the required transmission bandwidth as is given by

$$W_{\min,adb} = \frac{R}{\log_2\left(1 + \frac{V_{dd}^2}{2\sigma_N^2}\right)}. \qquad (6.11)$$

From (6.11), we see that for a given supply voltage V_{dd} and R, there is a lower limit on W. In fact, the relationship between V_{dd} and $W_{\min,adb}$ is in inverse proportion. A lower bound on W implies a lower bound on P_D, which is given by (6.10). Unlike other methods of power reduction, the lower bound given by (6.10) is easily approachable. This is due to the fact that we have included all implementation constraints in the derivation of (6.10).

Note that (6.10) is not really a lower bound as it is a function of the supply voltage V_{dd}. In order to compute the lower bound, we first substitute (6.11) into (6.10) to obtain

$P_{D,\min,adb}$

$$= \frac{V_{dd}^2}{2R_L}\left[\tan^{-1}\left(2\pi \frac{R}{\log_2\left(1 + \frac{V_{dd}^2}{2\sigma_N^2}\right)}R_L C_L\right)\right]^2. \qquad (6.12)$$

From (6.12), we notice that $P_{D,\min,adb}$ is a monotonically decreasing function of V_{dd} for sufficiently small V_{dd}. This is because the function $\tan^{-1}(x)$ attains a constant value as x approaches infinity. Clearly, the fundamental lower bound would be obtained if the minimum value of V_{dd} could be determined. This minimum value is obtained by recognizing the fact that the transmission bandwidth $W \leq W_{ch}$ and $W_{ch} = 1/(R_L C_L)$ Hz. Hence, the maximum value of $W_{\min,adb}$ in (6.11) is equal to $1/(R_L C_L)$. Substituting $W_{\min,adb} = 1/(R_L C_L)$ in (6.11) and solving for V_{dd}, we obtain the following lower bound on the supply voltage

$$V_{dd,\min,adb} = [(2^{RR_L C_L} - 1)2\sigma_N^2]^{1/2}. \qquad (6.13)$$

Substituting (6.13) back into (6.12) we obtain the fundamental lower bound on P_D as follows:

$$P_{D,\min,adb} = \frac{(2^{RR_L C_L} - 1)2\sigma_N^2[\tan^{-1}(2\pi)]^2}{R_L}. \qquad (6.14)$$

Clearly, from (6.14) the power dissipation for adiabatic logic will be zero only if the desired information transfer rate $R = 0$. As mentioned in Section III, R is a function of the required transformation.

D. Logic Level: A Latch-AND gate-Latch System

In this system [see Fig. 4(a)], we know that $f_{\text{op}} = 100$ Mhz and $H(X) = 2$ bits. From (3.1), we know that the information transfer rate $R = 81.2$ Mb/s. Furthermore, employing the coarse-grain representation, from (4.5) we obtain a channel bandwidth $W_{ch} = 75$ MHz and a noise power that is also reduced by a factor of half. Proceeding in exactly the same fashion as in Section VI-A, we can compute the minimum supply voltage for transmitting data at the rate of $R = 81.2$ Mb/s. The value of V_{dd}, which achieves this is given by $V_{dd} = 0.1496$ V and the lower bound on P_D is given by

$$\begin{aligned} P_D &> (0.5 * 2) * 10^{-12} * (0.1496)^2 * (150/2) * 10^6 \\ &> 0.001\,677 \text{ mW}. \end{aligned} \qquad (6.15)$$

Comparison of (6.15) and (6.4b) shows that the system in Fig. 4(a) has a smaller lower bound than the system in Fig. 5(a). Clearly, this result is possible if the output referred noise power for the AND gate is same as that for the system in Fig. 5(a) and the information transfer rate is smaller for the AND gate. This is exactly what we have assumed in this example.

A question that needs to be answered is what would the lower bound be if computed via a fine-grain approach. Ideally, both approaches should result in the same lower bound. In order to apply the fine-grain approach to the latched AND gate system in Fig. 4(a), we need to refer the noise power to the input. Clearly, due to the large gain of the AND gate, the input referred noise would be very small and hence the two input links would have a very high capacity C. This in turn implies that the power dissipation for the latched AND gate is dominated by the output link.
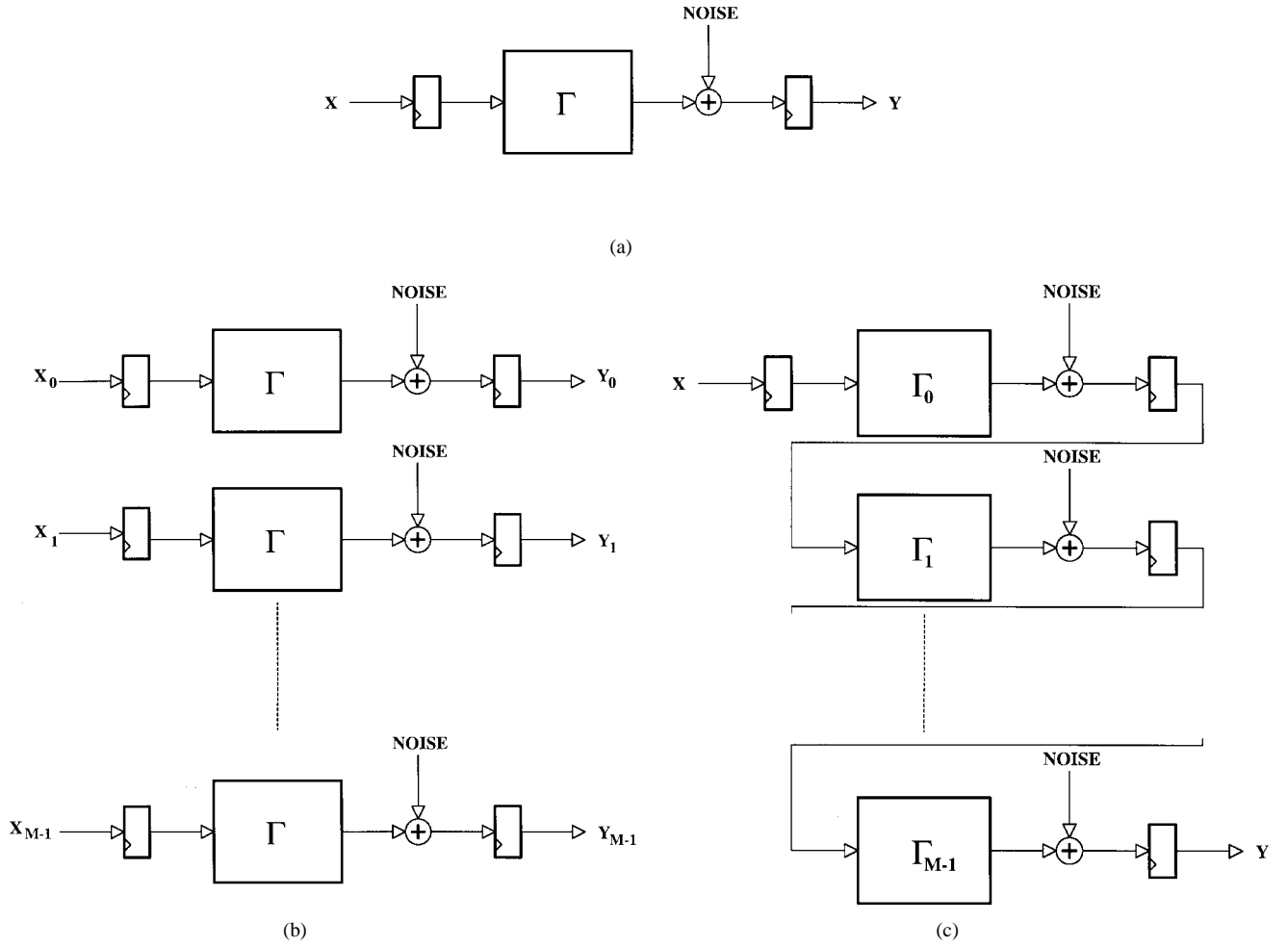
Fig. 11. Architectures. (a) Serial. (b) Parallel. (c) Pipelined architectures.

E. Architectural Level: Parallel Processing and Pipelining

Parallel processing and pipelining [6] along with power-supply voltage reduction had been proposed as a technique for reducing the power dissipation. In this subsection, we will present lower bounds for serial, parallel and pipelined architectures.

In Fig. 11, we show a serial [Fig. 11(a)], a parallel [Fig. 11(b)] and a pipelined [Fig. 11(c)] architecture. In Fig. 11(c), we have assumed that the transformation Γ in Fig. 11(a) can be written as a cascade of transformations Γ_i ($i = 0, \cdots, M - 1$). First, we will compute the channel capacity C and then *equate* it with required information transfer rate R. This will provide us with a value for $V_{dd,\min}$, where $V_{dd} > V_{dd,\min}$. Employing this value of $V_{dd,\min}$, we will then determine $P_{D,\min}$ such that $P_D > P_{D,\min}$.

Let R be the required information transfer rate (for all architectures). Furthermore, let $W = W_{ch}$ be the channel bandwidth for the serial architecture, σ_N^2 be the noise power integrated over W_{ch}, and C be the channel capacity for the serial architecture in Fig. 11(a). We will see that the lower bound for the serial architecture is identical to that of the single-bit two-latch system [see Fig. 5(a)]. This is not surprising given the fact that we have employed an abstraction of the serial architecture by defining its information transfer rate R.

The lower bound on the supply voltage for a serial architecture $V_{dd1,\min,ser}$ is calculated as follows:

$$C = R$$

$$W \log_2 \left[1 + \frac{V_{dd,\min,ser}^2}{4\sigma_N^2} \right] = R$$

$$V_{dd,\min,ser} = [(2^{R/W} - 1)4\sigma_N^2]^{1/2} \quad (6.16)$$

where $V_{dd,\min,ser}$ is now a strict lower bound on V_{dd} for a serial architecture. Substituting (6.16) into (6.1) gives us the strict lower bound on P_D for a serial architecture $P_{D,\min,ser}$ as shown below

$$P_{D,\min,ser} = C_L(2^{R/W} - 1)4\sigma_N^2 W. \quad (6.17)$$

For the parallel architecture, we assume that each channel in Fig. 11(b) is identical to the serial channel in Fig. 11(a). Proceeding in an identical fashion as before, we get

$$C = R$$

$$\sum_{i=0}^{M-1} W \log_2 \left[1 + \frac{V_{dd,\min,par}^2}{4\sigma_N^2} \right] = R$$

$$MW \log_2 \left[1 + \frac{V_{dd,\min,par}^2}{4\sigma_N^2} \right] = R$$

$$V_{dd,\min,par} = [(2^{R/MW} - 1)4\sigma_N^2]^{1/2}$$

$$(6.18)$$

The lower bound on P_D is given by

$$P_{D,\min,par} = \sum_{i=0}^{M-1} C_L(2^{R/MW} - 1)4\sigma_N^2 W$$

$$= C_L(2^{R/MW} - 1)4\sigma_N^2 MW. \qquad (6.19)$$

Comparing (6.19) and (6.17), we see that a parallel architecture is equivalent to a serial architecture which has a usable bandwidth of MW with $2W$ uses/s. In other words, a parallel architecture achieves power reduction by increasing its transmission bandwidth and reducing the supply voltage to a greater extent (than a serial architecture), while maintaining the same information transfer rate. From (6.17) and (6.19), it can be seen that the parallel architecture will always have a smaller lower bound than the serial architecture for $M > 1$. As expected, for $M = 1$, the lower bounds for the serial and parallel architectures are identical.

The pipelined architecture in Fig. 11(c) achieves power reduction in a slightly different manner than a parallel architecture. Pipelining by M-levels results in an M-fold expansion of the transmission bandwidth per stage due to the reduction in the critical path length. Assuming identical pipeline stages, the noise power per stage over a bandwidth of W is equal to σ_N^2. This implies that the noise power per stage over a bandwidth of MW is equal to $M\sigma_N^2$. Hence, the lower bound on the V_{dd} (for each stage) for the pipelined architecture can be obtained as follows:

$$C = R$$

$$MW \log_2 \left[1 + \frac{V_{dd,\min,pip}^2}{4M\sigma_N^2}\right] = R$$

$$V_{dd,\min,pip} = [(2^{R/MW} - 1)4M\sigma_N^2]^{1/2}. \qquad (6.20)$$

Note that all the stages in a pipelined architecture need to transmit R bits/s of information. Assuming a maximum usage of $2MW$ uses/s, the lower bound on the P_D is

$$P_{D,\min,pip} = C_L(2^{R/MW} - 1)4\sigma_N^2 M^2 W. \qquad (6.21)$$

Interestingly, a comparison between (6.19) and (6.21) indicates that $P_{D,\min,par}$ is lower by a factor of M as compared to $P_{D1,\min,pip}$. This is counter to the well-prevalent notion that parallelization and pipelining have identical power-reduction capabilities. However, both architectures are equivalent if the $Area \times Power$ product is considered. This is due to the fact that a M-level parallel architecture requires M-times the area of a serial architecture. On the other hand, the area requirements of a pipelined architecture is of the same order as that of a serial architecture [29], [31].

In Fig. 12, we have plotted the lower bounds on P_D for the parallel and pipelined architectures normalized with respect to that of the serial architecture for values of $R/W = 10$ and $R/W = 8$. As mentioned before, the lower bounds for the parallel architecture are always smaller than that of the serial architecture. This is clearly seen in Fig. 12. However, for the pipelined architecture, we find that the power reduction capabilities of the pipelined architecture reduces along with
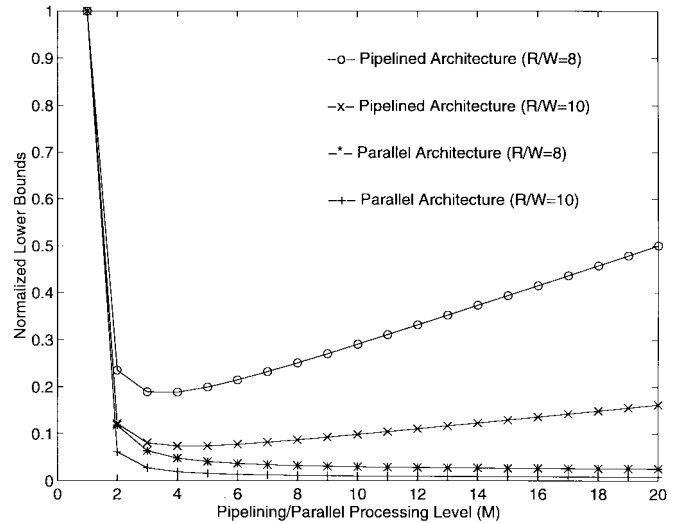

Fig. 12. Normalized lower bounds on power dissipation for parallel and pipelined architectures.

R/W. Furthermore, for a sufficiently large pipelining level M, the lower bound for the pipelined architecture can in fact be greater than that of the serial architecture. This is due to the fact that W_{ch} increases with M and therefore the noise power admitted into the receiver will also increase. This implies the need for a higher signal power to achieve a certain capacity C and therefore a higher power dissipation.

Thus, we see that the proposed theory can be applied to determine the lower bounds on the power dissipation at the circuit, logic and the architectural levels of the design hierarchy. There are numerous other power reduction techniques that can be considered for analysis. A commonly employed technique is that of clock-gating and selective power down. The proposed theory can also be applied to such techniques. In particular, note that when a certain section is being powered down, the function Γ will now change to Γ_1, which has a lower information transfer rate requirement R. The theory can now be applied to Γ_1 and calculations for Γ_1 will result in lower bounds that are smaller than those for Γ.

VII. CONCLUSIONS

A mathematical theory, based upon Shannon's joint source-channel coding theorem [32], for determining the lower-bounds on power dissipation in digital systems is presented. The proposed theory provides a fundamental theoretical basis, which unifies existing power reduction techniques and allows us to compare the lower bounds achievable by each of these techniques. The proposed theory can be employed to understand the complex interaction between various power-reduction techniques, when applied simultaneously [6]. Accurate characterization of noise sources and incorporation of implementation domain constraints in computing the lower bounds will be undertaken next. Future work also involves application of the proposed theory to other power reduction techniques not considered in this paper and possible development of new techniques. Extension to more complex digital signal processing systems and other modes of signal processing such as analog signal processing also needs to

be done. Determining the lower-bounds on the power-area product along similar lines is also a possible extension of this paper. We hope that the present work will provide a foundation for the development of advanced power-area-speed optimal VLSI signal processing systems in future.

APPENDIX A
PROOF OF THEOREM 2

There are three steps in the proof, which are as follows.

Step 1)

From the joint source-channel coding theorem [32], we know that information transfer rate R is achievable, with the probability of error approaching zero, as long as $R < C$. This fact along with (2.8) gives us

$$R < \int_0^W \log_2 \left[1 + \frac{S_{XX}(f)}{S_{NN}(f)} \right] df \qquad (A.1)$$

where $S_{XX}(f)$ and $S_{NN}(f)$ are the signal and noise power spectral densities, respectively.

Step 2)

From the water-filling argument [10] in the spectral domain, we know that the signal power spectrum which maximizes the capacity for the power constraint

$$\int_0^W S_{XX}(f) = P_X \qquad (A.2)$$

is given by

$$S_{XX}(f) = (\nu - S_{NN}(f))^+ \qquad (A.3)$$

where ν is chosen to satisfy (A.2).

Substituting (A.3) into (A.1) results in

$$R < \int_0^W \log_2 \left[1 + \frac{(\nu - S_{NN}(f))^+}{S_{NN}(f)} \right] df. \qquad (A.4)$$

It is trivial to show that

$$1 + \frac{(\nu - S_{NN}(f))^+}{S_{NN}(f)} \geq 1. \qquad (A.5)$$

Furthermore, we can also show that

$$\frac{(\nu_1 - S_{NN}(f))^+}{S_{NN}(f)} \geq \frac{(\nu_2 - S_{NN}(f))^+}{S_{NN}(f)} \qquad (A.6)$$

for any $\nu_1 \geq \nu_2$.

From the monotonically increasing property of the logarithm function and (A.5)–(A.6), there would exist a minimum value of ν say $\nu_{\min}$ such that (5.3) is satisfied. The signal power spectrum corresponding to this value of $\nu_{\min}$ can be obtained by substituting $\nu_{\min}$ into (A.3) and will be referred to as $S_{XX,\min}(f)$.

Step 3)

Substituting the value of $S_{XX,\min}(f)$ (from (A.3) with $\nu = \nu_{\min}$) into (5.1), we obtain

$$P_{D,\min} = \int_0^W F[(\nu_{\min} - S_{NN}(f))^+] \cdot df. \qquad (A.7)$$

As $F(\ )$ is a linear function therefore (A.7) can be written as

$$P_{D,\min} = F\left[\int_0^W (\nu_{\min} - S_{NN}(f))^+ df \right]. \qquad (A.8)$$

All that remains to be proved is that $P_{D,\min}$ in (A.8) is indeed the lower bound. In order to do this we assume that there exists a power spectrum $S_{XX,1}(f)$, which permits reliable information transfer rate of R, such that

$$P_{D,1} = F\left[\int_0^W S_{XX,1}(f) df \right], \qquad (A.9)$$

and

$$P_{D,1} \leq P_{D,\min}. \qquad (A.10)$$

From (A.9)–(A.10) and the fact that $F(\ )$ is a linear monotonically increasing function, we obtain

$$\int_0^W S_{XX,1}(f) df \leq \int_0^W (\nu_{\min} - S_{NN}(f))^+ df. \qquad (A.11)$$

From (5.3) and (A.11), we conclude that the channel capacity C_1 with signal power spectrum equal to $S_{XX,1}(f)$ satisfies

$$C_1 \leq R. \qquad (A.12)$$

Finally, from the joint source-channel coding theorem [32] we know that information transfer of R bps is not achievable with $S_{XX,1}(f)$ as the signal power spectral density. This contradicts our assumption. Hence, $P_D > P_{D,\min}$.

APPENDIX B
PROOF OF (6.1)

Assume that the channel bandwidth is W but it is used X times per second, where $0 > X \leq 2W$ (from the Nyquist criterion). In that case the channel capacity formula is given by

$$C = \frac{X}{2} \log_2 \left[1 + \frac{V_{dd}^2}{4\sigma_N^2} \right] \qquad (B.1)$$

where we assume that the SNR is constant over the bandwidth W.

Next we equate (B.1) to the desired information transfer rate R to obtain the lower bound on V_{dd} as follows

$$C = R$$
$$\frac{X}{2} \log_2 \left[1 + \frac{V_{dd,\min}^2}{4\sigma_N^2} \right] = R$$
$$V_{dd,\min}^2 = 4(2^{2R/X} - 1)\sigma_N^2. \qquad (B.2)$$

We can assume that each use of the channel corresponds to one signal transition. This will be true pulse-amplitude modulation

$$P_D(t) = \frac{[V_{dd}\,\sin(2\pi Wt) - V_{dd}\,\sin(2\pi Wt + \tan^{-1}(2\pi WR_LC_L))]^2}{R_L}$$
$$= \frac{V_{dd}^2}{R_L}\,\cos^2(2\pi Wt)[\tan^{-1}(2\pi WR_LC_L)]^2 \tag{C.5}$$

(PAM), where a sequence of pulses are sent over the channel. Hence, the power dissipation is given by

$$P_D = \frac{1}{2}C_L V_{dd}^2 X. \tag{B.3}$$

Substituting (B.2) into (B.3) we obtain

$$P_{D,\min,X} = 2C_L(2^{2R/X} - 1)\sigma_N^2 X. \tag{B.4}$$

By inspection of (B.4), we see that (for $X > 0$) as X increases, the *exponential term* decreases and the *linear term* increases. Thus, the right-hand side of (B.4) attains its minimum value in the interval $0 < X \leq 2W$ when X is equal to its maximum value $2W$. Substituting $X = 2W$ into (B.3), we obtain

$$P_D = C_L V_{dd}^2 W \tag{B.5}$$

which is the desired equation.

APPENDIX C
LOWER BOUND CALCULATION FOR ADIABATIC COMPUTATION

Assume that the supply voltage is given by

$$V(t) = V_{dd}\,\sin(2\pi Wt). \tag{C.1}$$

The power in $V(t)$ is known to be $V_{dd}^2/2$. Employing (2.8) with a constant $\text{SNR}(f)$, we get

$$R < W \log_2\left[1 + \frac{V_{dd}^2}{2\sigma_N^2}\right]$$
$$\frac{R}{\log_2\left[1 + \dfrac{V_{dd}^2}{2\sigma_N^2}\right]} < W \tag{C.2}$$

which is identical to (6.11).

Next, we assume that $W \gg 1/(R_LC_L)$, so that the voltage across the capacitor $V_C(t)$ is given by

$$V_C(t) = V_{dd}\,\sin(2\pi Wt + \tan^{-1}(2\pi WR_LC_L)). \tag{C.3}$$

The power dissipation in Fig. 10 is given by

$$P_D(t) = \frac{[V(t) - V_C(t)]^2}{R_L}. \tag{C.4}$$

Substituting for $V(t)$ [from (C.1)] and $V_C(t)$ [from (C.3)] into (C.4), we obtain (see (C.5) at the top of the page) where we have utilized the fact that for a small enough value of $\tan^{-1}(2\pi WR_LC_L)$,

$$\cos(\tan^{-1}(2\pi WR_LC_L)) \approx 1 \tag{C.6a}$$
$$\sin(\tan^{-1}(2\pi WR_LC_L)) \approx \tan^{-1}(2\pi WR_LC_L). \tag{C.6b}$$

The lower bound on P_D is obtained by averaging $P_D(t)$ over one period of $V(t)$ and is given by

$$P_D = W \int_0^{1/W} P_D(t)\,dt$$
$$= \frac{V_{dd}^2[\tan^{-1}(2\pi WR_LC_L)]^2}{2R_L} \tag{C.7}$$

which is the desired equation.

REFERENCES

[1] M. Alidina, J. Monterio, S. Devdas, A. Ghosh, and M. Papaefthymiou, "Precomputation-based sequential logic optimization for low-power," *IEEE Trans. VLSI Syst.*, vol. 2, pp. 426–436, Dec. 1994.
[2] W. C. Athas, L. J. Svensson, J. G. Koller, N. Tzartzanis, and E. Y.-C. Chou, "Low-power digital systems based on adiabatic switching principles," *IEEE J. Solid-State Circuits*, vol. 2, pp. 398–407, Dec. 1994.
[3] H. B. Bakoglu and J. D. Meindl, "Optimal interconnection circuits for VLSI," *IEEE Trans. Electron Devices*, vol. ED-37, pp. 903–909, May 1985.
[4] C. H. Bennett, "Logical reversibility of computation," *IBM J. Res. Develop.*, pp. 525–532, Nov. 1973.
[5] J. B. Burr and J. Schott, "A 200 mv self-test encoder/decoder using Stanford ultra-low power CMOS," in *ISSCC '94*, San Fransisco, CA, pp. 84–85.
[6] A. Chandrakasan and R. W. Brodersen, "Minimizing power consumption in digital CMOS circuits," *Proc. IEEE*, vol. 83, pp. 498–523, Apr. 1995.
[7] A. Chandrakasan, S. Sheng, and R. W. Brodersen, "Low power CMOS digital design," *IEEE J. Solid-State Circuits*, vol. 27, pp. 473–484, Apr. 1992.
[8] K.-T. Cheng and V. Agrawal, "An entropy measure for the complexity of multi-output Boolean functions," in *27th ACM/IEEE Design Automation Conf.*, Orlando, FL, June 24–28, 1990, pp. 302–305.
[9] R. W. Cook and M. J. Flynn, "Logical network cost and entropy," *IEEE Trans. Comput.*, vol. C-22, pp. 823–826, Sept. 1973.
[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[11] 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.
[12] A. G. Dickinson and J. S. Denker, "Adiabatic dynamic logic," in *Proc. Custom Integrated Circuits Conf.*, 1994.
[13] T. Gabara, "Pulsed power supply CMOS—PPS CMOS," in *1994 IEEE Symp. Low Power Electron.*, San Fransisco, CA, pp. 98–99.
[14] R. G. Gallager. *Information Theory and Reliable Communication*. New York: Wiley, 1968.
[15] R. D. Gitlin, J. F. Hayes, and S. B. Weinstien, *Data Communications Principles*. New York: Plenum, 1992.
[16] P. R. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*. New York: Wiley, 1993.
[17] L. Hellerman, "A measure of computational work," *IEEE Trans. Comput.*, vol. C-21, pp. 439–446, May 1972.
[18] M. Horowitz, T. Indermaur, and R. Gonzalez, "Low-power digital design," in *1994 IEEE Symp. Low Power Electron.*, San Diego, CA, pp. 8–11.
[19] S.-M. Kang and Y. Leblebici, *CMOS Digital Integrated Circuits: Analysis and Design*. New York: McGraw-Hill, 1995.
[20] G. Kissin, "Measuring energy consumption in VLSI circuits," Ph.D. dissertation, Dep. Comput. Science, Univ. Toronto, Canada, 1982.
[21] E. A. Lee and D. G. Messerschmitt, *Digital Communications*. Boston, MA: Kluwer Academic, 1988.

[22] D. Marculescu, R. Marculescu, and M. Pedram, "Information theoretic measures of energy consumption at register transfer level," in *ACM/IEEE Int. Symp. Low Power Design*, 1995, pp. 81–86.

[23] R. Mehra and J. Rabaey, "High level estimation and exploration," in *Proc. 1994 Int. Symp. Low Power Design*.

[24] J. D. Meindl, "Low power microelectronics: Retrospect and prospect," *Proc. IEEE*, vol. 83, pp. 619–635, Apr. 1995.

[25] J. Monteiro, S. Devdas, and A. Ghosh, "Retiming sequential circuits for low power," in *IEEE Int. Conf. Computer-Aided Design*, pp. 398–402.

[26] F. N. Najm, "A survey of power estimation techniques in VLSI circuits," *IEEE Trans. VLSI Syst.*, pp. 446–455, Dec. 1994.

[27] F. N. Najm, "Toward a high-level power estimation capability," in *ACM/IEEE Int. Symp. Low Power Design*, 1995, pp. 87–91.

[28] Y. Nakagome, K. Itoh, M. Isoda, K. Takeuchi, and M. Aoki, "Sub-1-V swing internal bus architecture for future low-power ULSI's," *IEEE J. Solid-State Circuits*, vol. 28, pp. 414–419, Apr. 1993.

[29] K. K. Parhi and D. G. Messerschmitt, "Pipeline interleaving and parallelism in recursive digital filters—Part I: Pipelining using scattered look-ahead and decomposition," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 37, pp. 1099–1117, July 1989.

[30] M. Potkonjak and J. Rabaey, "Fast implementation of recursive programs using transformations," in *Proc. ICASSP*, San Fransisco, CA, Mar. 1992, pp. V-569–572.

[31] N. R. Shanbhag and K. K. Parhi, *Pipelined Adaptive Digital Filters*. Boston, MA: Kluwer Academic, 1994.

[32] C. E. Shannon, "A mathematical theory of communications," *Bell Syst. Tech. J.*, Part I, vol. 27, pp. 379–423; Part II, pp. 623–656.

[33] 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.

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

[35] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Info. Theory*, vol. IT-13, pp. 260–269, Apr. 1967.

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

[37] ——, "Future of analog in the VLSI environment," in *Proc. ISCAS '90*, 1990, pp. 1372–1375.

[38] ——, "Micropower techniques," in *Design of Analog-Digital VLSI Circuits for Telecommunications and Signal Processing*, J. Franca and Y. Tsividis, Eds. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[39] N. Weste and K. Eshraghian, *Principles of CMOS VLSI Design: A Systems Perspective*. Reading, MA: Addison-Wesley, 1985.

Naresh R. Shanbhag (M'93) received the B.Tech. degree from the Indian Institute of Technology, New Delhi, India, in 1988, and the Ph.D. degree from the University of Minnesota, Minneapolis, in 1993, all in electrical engineering.

From July 1993 to August 1995, he was with AT&T Bell Laboratories, Murray Hill, NJ, working in the Wide-Area Networks Group, where he was responsible for the development of VLSI algorithms, architectures and implementation for high-speed data communications applications. In particular, he was the lead chip architect for AT&T's 51.84 Mb/ls s transceiver chips over twisted-pair wiring for Asynchronous Transfer Mode (ATM)-LAN and the Interactive Multimedia Television (IMTV) transmitter-receiver chip-set. In August 1995, he joined the Coordinated Science Laboratory and the Electrical and Computer Engineering Department, University of Illinois, Urbana-Champaign as an assistant professor. His research interests (see URL http:\ls \ls uivlsi.csl.uiuc.edu/~shanbhag) are in the area of VLSI architectures and algorithms for signal processing and communications. This includes the design of high-speed and \ls or low-power algorithms for speech and video processing, adaptive filtering and high-bit rate digital communications systems. In addition, he is also interested in efficient VLSI implementation methodologies for these applications. He is the coauthor of the research monograph *Pipelined Adaptive Digital Filters* (Boston, MA: Kluwer Academic, 1994).

Dr. Shanbhag received the NSF CAREER Award in 1996, the 1994 Darlington best paper award from the IEEE Circuits and Systems Society