Information Theoretic Measures for Power Analysis[1]

Diana Marculescu, Radu Marculescu, Massoud Pedram

Department of Electrical Engineering - Systems

University of Southern California, Los Angeles, CA 90089

Abstract - *This paper considers the problem of estimating the power consumption at logic and register transfer levels of design from an information theoretical point of view. In particular, it is demonstrated that the average switching activity in the circuit can be calculated using either entropy or informational energy averages. For control circuits and random logic, the output entropy (informational energy) per bit is calculated as a function of the input entropy (informational energy) per bit and an implementation dependent information scaling factor. For data-path circuits, the output entropy (informational energy) is calculated from the input entropy (informational energy) using a compositional technique which has linear complexity in terms of the circuit size. Finally, from these input and output value, the entropy (informational energy) per circuit line is calculated and used as an estimate for the average switching activity. The proposed switching activity estimation technique does not require simulation and is thus extremely fast, yet produces sufficiently accurate estimates.*

I. Introduction

Modern design tools have changed the entire design process of digital systems. As a result, most of the systems are conceived and designed today at the behavioral/logic level, with little or no knowledge of the final gate level implementation and the layout style. In particular, designers are becoming more and more interested in register-transfer level (RTL) modules (adders, multipliers, registers, multiplexers) and strategies to put them together in order to build complex digital systems. Power minimization in digital systems is not an exception to this trend. Having as soon as possible in the design cycle an estimate of power consumption, can save significant redesign efforts or even completely change the entire design architecture.

A. Basic Issues and Prior Work

Gate-level power estimation techniques can be divided into two general categories: simulative and non-simulative [1][2]. Simulative techniques have their roots in direct simulation and sampling techniques [3]-[5]. The main advantages of these techniques are that existing simulators can be used for estimation purposes and

---

[1]This research was supported in part by Advanced Research Project Agency under Contract F33615-95-C-1627 and by Semiconductor Research Corporation under contract 94-DJ-559.

issues such as hazard generation and propagation, reconvergent fanout induced correlations are automatically taken into consideration. The disadvantage is their strong input pattern dependence and long running times (in sampling techniques this is needed for achieving high levels of accuracy). Non-simulative techniques are based on probabilistic or stochastic techniques [6]-[9]. Their main advantages are higher speed and lower dependence on the input patterns. Generally speaking, however, they tend to be less accurate due to the simplified models used and the approximations made in order to achieve higher efficiency.

Higher levels of abstraction have also been considered, but here many problems are still pending a satisfactory solution. At this level, consistency is more important than accuracy, that is, relative (as opposed to absolute) evaluation of different designs is often sufficient. Most of the high level prediction tools combine deterministic analysis with profiling and simulation in order to address data dependencies. Important statistics include the number of instructions of a given type, the number of bus, register and memory accesses, and the number of I/O operations executed within a given period [10][11]. Analytic modeling efforts have been described in [12] where a parametric power model was developed for macromodules. However, the trade-off between flexibility and accuracy is still a challenging task as major difficulties persist due to the lack of precise information and the conceptual complexity which characterizes this level of design abstraction.

Power dissipation in CMOS circuits comes from three sources: leakage currents which include the reverse-biased p-n junction and subthreshold currents, short-circuit currents which flow due to the DC path between the supply rails during output transitions and capacitive switching currents which are responsible for charging and discharging of capacitive loads during logic transitions. In well-designed circuits with relatively high threshold voltages, the first two sources are small compared to the last one. Therefore, to estimate the total power consumption of a module (in a gate level implementation), we may only account for the capacitive switching currents, yet achieve sufficient levels of accuracy [13]:

$$P_{avg} = \frac{f_{clk}}{2} \cdot V_{DD}^{2} \cdot \sum_{n} \left(C_n \cdot sw_n \right) \tag{1}$$

where f_{clk} is the clock frequency, V_{DD} is the supply voltage, C_n and sw_n are the capacitive load and the average switching activity of gate n, respectively (the summation is performed over all gates in the circuit). As we can see, in this formulation the average switching activity per node is a key factor, and therefore its correct calculation is essential for accurate power estimation. Note that for the same implementation of a module, different input sequences may give rise to different switching activities at the circuit inputs and at the outputs of internal gates and consequently completely different power values.

The problem of power estimation at the RT-level is different from that at the logic level: whilst at gate level it is desirable to determine the switching activity at each node (gate) in the circuit (Fig.1(a)), for RT-level designs an average estimate per module is satisfactory (Fig.1(b)). In other words, some accuracy may be sacrificed in order to obtain an acceptable power estimate early in the design cycle at a significantly lower computational cost.

Fig.1 comes here

In the data-flow graph considered in Fig.1(b), the total power consumption may be estimated as: $P_{total} = P_{wires} + P_{modules}$. Usually, the interconnect power consumption is either estimated separately or included in the power consumption of the modules, therefore we can write: $P_{total} \approx \sum_{m_j \in M} P_{m_j} \propto \sum_{m_j \in M} (C_{m_j} \cdot SW_{m_j})$ where the summation is performed over the set of modules M used in the data-flow graph, and $C_{m_{jj}}$, SW_{m_j} stand for the capacitance loading and the average switching activity of module m_j, respectively. Basically, what we propose is to characterize the average switching activity of a module (SW_{m_j}) through the average switching activity for a *typical signal line* in that module (sw_{avg}). More formally, for a generic module m_j having n internal lines (each characterized by its capacitance and switching activity values c_i and sw_i, respectively), we have:

$$P_{m_j} \propto \sum_{i=1}^{n} (c_i \cdot sw_i) \approx sw_{avg} \cdot \sum_{i=1}^{n} c_i = sw_{avg} \cdot C_{m_j} \tag{2}$$

We assume that module capacitances C_{m_j} are either estimated or taken from a library, therefore we concentrate on estimating the average switching activity per module. This is a quite different strategy compared to the previous work. The only other proposed method for power estimation at RT-level is simulative in nature, requires pre-characterization of the modules and may be summarized as follows: first, RT-level simulation is performed to obtain the average switching activity at the inputs of the modules and then, this switching activity is used to "modulate" a switched capacitance value (product of the switching activity and the physical capacitance) which is pre-computed and stored for each module in the library to obtain the power dissipation estimate [14]. Compared to this methodology, the distinctive feature of the present approach, is that *it does not require simulation*; its predictions are based only on the characteristics of the input sequence and some knowledge about the function and/or structure of the circuit (see Section IV for details).

B. Overview and Organization of the Paper

In this paper, we address the problem of power estimation at the RT-level from an information theoretical point of view [15]. Traditionally, entropy has been considered a useful measure for solving problems of area estimation, timing analysis [17][18] and testing [19][20]. We propose two new measures for estimating the power consumption of each module based on *entropy* and *informational energy*. Our entropy/informational energy-based measures simply provide an approximation for the functional activity in the circuit without having to necessarily simulate the circuit. With some further simplifications, simple closed form expressions are derived and their value in practical applications is explored.

We should point out that although this paper targets RT-level and behavioral design, it also presents, as a by-product, a technique applicable to logic level designs. This is a first step in building a unified framework for power analysis from gate level to behavioral level.

The paper is organized as follows. Sections II and III introduce the main concepts and the motivation behind our model. In Section IV we present some practical considerations and in Section V, we give the results obtained by analyzing a common data-path and benchmark circuits. Finally, we conclude by summarizing our main ideas.

II. THEORETICAL FRAMEWORK

A. An Entropy-Based Approach

Let $A_1, A_2, ..., A_n$ be a complete set of events which may occur with the probabilities $p_1, p_2, ..., p_n$ that is:

$$p_k \geq 0 \qquad k = 1, 2, \ldots, n \qquad \sum_{k=1}^{n} p_k = 1 \tag{3}$$

Let us consider an experiment where the outcome is unknown in the beginning; such an experiment exposes a probability finite field $\mathcal{A}_n$, completely characterized by the discrete probability distribution $p_1, p_2, ..., p_n$. In order to quantify the content of information revealed by the outcome of such an experiment, Shannon introduced the concept of entropy [21].

Definition 1: (Entropy)

Entropy of a finite field $\mathcal{A}_n$ (denoted by $H(\mathcal{A}_n)$) is given by:

$$H(\mathcal{A}_n) = H(p_1, p_2, \ldots, p_n) = -\sum_{k=1}^{n} p_k \log p_k \tag{4}$$

Entropy satisfies some basic properties detailed in [21]. We plot in Fig.2 the entropy of a Boolean variable as a function of its signal probability, that is $H(\mathcal{A}_2) = -p \cdot \log p - (1-p) \cdot \log(1-p)$.

Fig.2 comes here

First, we should note that log $(p_k) \le 0$ since $0 < p_k \le 1$ and so $H(\mathcal{A}_n) \ge 0$. Thus, the entropy can never be negative. Second, let $p_1 = 1$, $p_2 = ... = p_n = 0$. By convention, $p_k \log(p_k) = 0$ when $p_k = 0$ and hence, in this case, $H(\mathcal{A}_n) = 0$. Conversely, $H(\mathcal{A}_n) = 0$ implies that $p_k \log(p_k) = 0$ for all k, so that p_k is either 0 or 1. But only one p_k can be unity since their sum must be 1. Hence, entropy is zero if and only if there is complete certainty.

Definition 2: (Conditional Entropy)

Conditional entropy of some finite field $\mathcal{A}_n$ with probabilities $\{p_i\}_{1 \le i \le n}$ with respect to $\mathcal{B}_m$ (with probabilities $\{q_i\}_{1 \le i \le m}$) is defined as:

$$H(\mathcal{A}_n | \mathcal{B}_m) = -\sum_{j=1}^{n} \sum_{k=1}^{m} q_k \cdot p_{kj} \cdot \log(p_{kj}) \tag{5}$$

where p_{kj} is the *conditional probability* of events A_j and B_k ($p_{kj} = prob\ (A_j | B_k)$). In other words, conditional entropy refers to the uncertainty left about $\mathcal{A}_n$ when $\mathcal{B}_m$ is known.

Definition 3: (Joint Entropy)

Given two finite fields $\mathcal{A}_n$ and $\mathcal{B}_m$, their joint entropy is defined as:

$$H(\mathcal{A}_n \times \mathcal{B}_m) = -\sum_{j=1}^{n} \sum_{k=1}^{m} q_k \cdot p_{kj} \cdot \log(q_k \cdot p_{kj}) \tag{6}$$

Based on these two concepts, one can find the information *shared* by two complete sets of events:

$$I(\mathcal{A}_n; \mathcal{B}_m) = H(\mathcal{A}_n) + H(\mathcal{B}_m) - H(\mathcal{A}_n \times \mathcal{B}_m)$$

which is called *mutual information* (or *transinformation*). Moreover, by using the above definitions, one can show that:

$$I(\mathcal{A}_n; \mathcal{B}_m) = H(\mathcal{A}_n) - H(\mathcal{A}_n | \mathcal{B}_m) \tag{7}$$

The Venn diagram for these relations is shown in Fig.3.

Fig.3 comes here

The concept of entropy is equally applicable to partitioned sets of events. More precisely, given a partitioning $\prod = \{A_1, A_2, ..., A_n\}$ on the set of events, the entropy of this partitioning is:

$$H(\Pi) = -\sum_{i=1}^{n} p(A_i) \log p(A_i) \tag{8}$$

where $p(A_i)$ is the probability of class A_i in partition $\prod$.

Example 1: The truth table for a randomly excited 1-bit full adder is given below:

Fig.4 comes here

where x_i, y_i are the inputs, c_i is carry-in, s_i is the sum bit and c_{i+1} is carry-out. The output space is partitioned in four classes as $\prod = \{A_1, A_2, A_3, A_4\} = \{10, 01, 11, 00\}$, where $p(A_1) = p(A_2) = 3/8$, $p(A_3) = p(A_4) = 1/8$. Applying (8) we obtain $H(\prod) = 1.8113$. We observe that within a class there is no activity on the outputs; this means that output transitions may occur only when one has to cross class boundaries in different time steps. If the output sequence is purely random, then exactly H bits are needed to represent the output sequence; therefore the average number of transitions per word (or average switching activity per word) will be $H/2$. In any other nonrandom arrangement, for a minimum length encoding scheme, the average number of transitions per word will be $\leq H/2$, so in practice, $H/2$ can serve as a conservative upper bound on the number of transitions per word. In our example, we find an average switching value approximately equal to 0.905 which matches fairly well the exact value 1 deduced from the above table. Such a measure was suggested initially by Hellerman to quantify the *computational work* of simple processes [16].

More formally, if a signal x is modelled as a lag-one Markov chain with conditional probabilities p_{00}, p_{01}, p_{10}, p_{11} and signal probabilities p_0 and p_1 as in Fig.5,

Fig.5 comes here

then we have the following result:

Theorem 1: $h(x^+|x^-) \geq 2 \cdot sw(x) \cdot (p_{00} + p_{11})$, where $h(x^+|x^-)$ is the conditional entropy between two successive time steps and $sw(x)$ is the switching activity of line x.

Proof:

According to Definition 2, $h(x^+|x^-) = -p_1 (p_{10} \log p_{10} + p_{11} \log p_{11}) - p_0 (p_{00} \log p_{00} + p_{01} \log p_{01})$. The signal probabilities can be expressed in terms of conditional probabilities as: $p_1 = \frac{p_{01}}{p_{01} + p_{10}}$ and $p_0 = \frac{p_{10}}{p_{01} + p_{10}}$, respectively [9]. Using the well-known identity $-\ln(1-a) = \sum_{k=1}^{\infty} \frac{a^k}{k}$ for $0 < a < 1$, we obtain:

$$h(x^+|x^-) = \frac{1}{2 \cdot \ln(2)} \cdot \frac{2 \cdot p_{01} \cdot p_{10}}{p_{01} + p_{10}} \cdot \sum_{k=1}^{\infty} \left[\frac{\left(p_{00}^{k-1} + p_{01}^{k-1}\right) \cdot p_{00} + \left(p_{10}^{k-1} + p_{11}^{k-1}\right) \cdot p_{11}}{k} \right].$$

We note that $\frac{2 \cdot p_{01} \cdot p_{10}}{p_{01} + p_{10}}$ is exactly the switching activity of line x. The above identity is general, but it is hard to use in practice in this form; therefore, we try to bound the summation. To begin with, we note that $x \cdot f(x) \geq x \cdot min\,[f(x)] \quad for\;\; x \geq 0$ and that $a^{k-1} + b^{k-1}$ (for $a + b = 1$) is minimized when $a = b = 0.5$.

Thus, $h\,(x^+|x^-) \geq \frac{1}{2 \cdot \ln\,(2)} \cdot sw\,(x) \cdot \sum_{k=1}^{\infty} \frac{(p_{00} + p_{11}) \cdot \frac{1}{2^{k-2}}}{k}$ or, using again the above identity, we get:

$h\,(x^+|x^-) \geq 2 \cdot sw\,(x) \cdot (p_{00} + p_{11})$.

■

We have thus obtained an upper bound for the switching activity of signal x when it is modelled as a lag-one Markov chain. To obtain an upper bound useful in practice, we use the following result:

Corollary 1: Under the *temporal independence* assumption, we have that $sw\,(x) \leq h\,(x)\,/\,2$.

Proof:

If we assume temporal independence, that is, $p_1 = p_{01} = p_{11}$, $p_0 = p_{10} = p_{00}$ and $h\,(x^+|x^-) = h\,(x)$, then we have: $p_{00} + p_{11} = 1$ and hence the relationship between $sw\,(x)$ and $h\,(x)$ is exactly the one based on intuition: $sw\,(x) \leq h\,(x)\,/\,2$.

■

To evaluate H, one can use basic results from information theory concerning transmission of information through a module. More precisely, for a module with input X and output Y, we have $I\,(X;Y) = H\,(X) - H(X|Y)$ and by symmetry $I\,(X;Y) = H\,(Y) - H(Y|X)$ due to the commutativity property of mutual information. When input X is known, no uncertainty is left about the output Y and thus $H(Y|X)$ is zero. Therefore, the information transmitted through a module can be expressed as $H\,(Y) = H\,(X) - H(X|Y)$ which represents the amount of information provided about X by Y. For instance, in Fig.4, $I\,(X;\,Y) = H\,(Y) = 3 - 1.1887 = 1.8113$; thus, informally at least, the observation of the output of the module provides 1.8113 bits of information about the input, on average. However, in real examples, this type of characterization becomes very expensive when the input/output relation is not a one-to-one mapping. This usually requires a large number of computations; for instance, the exact calculation of the output entropy of an n-bit adder, would require the knowledge of joint input/output probabilities and a double summation with 2^{2n} terms (as in (5)). As a consequence, in order to analyze large designs, we target a *compositional approach* where the basic modules are already characterized in terms of transinformation and what is left to find, is only a *propagation mechanism* among them (details are given in Section IV).

As we have seen, an appropriate measure for the average switching activity of each net in the circuit is its entropy value. Basically, what we need is a mapping $\xi \rightarrow \xi'$ from the actual set ξ of the nets in the target circuit (each having a possibly distinct switching activity value) to a virtual set ξ', which contains the same collection of wires, but this time each net has the same value of switching activity. More formally, $\xi \rightarrow \xi'$ is a mapping such that the following conditions are satisfied:

$$\begin{cases} |\xi| = |\xi'| \\ sw(x_i) = sw(x_j) \quad \text{for any } x_i, x_j \in \xi' \end{cases} \tag{9}$$

Bearing in mind this, one can express the total number of transitions per step as:

$$SW(\xi) = SW(\xi') \leq n \cdot \frac{h(\xi')}{2} \tag{10}$$

where n stands for the presumed cardinality of ξ' and $h(\xi')$ represents the average entropy per bit of any net in ξ'. To clarify these ideas, let us consider the simple circuit in Fig.6.

Fig.6 comes here

In this example, we feed the circuit with a 3-bit random sequence and tabulate in the right side, the logic values obtained in the entire circuit by logic simulation. We have that $\xi = \{x, c, y, a, b, z\}$ with the switching profile (in number of transitions) {4, 4, 4, 2, 2, 2}, respectively. Doing a quick calculation, we get $SW(\xi)$ = 2.25 transitions per step. On the other hand, $\xi' = \{x, c, y, a, b, z\}$ with the average entropy $h(\xi') = (3 * 1 + 2 * 0.811 + 0.954) / 6 = 0.9295$ characterizing each signal in the set; using relation (10) we get an expected value $SW(\xi') = 2.73$ which is greater than $SW(\xi)$, but sufficiently close to it.

Unfortunately, in large circuits, it is expensive to compute $h(\xi')$ as the complete set of events characterizing a circuit is exponential in the number of nodes. To avoid the brute force approach, that is exhaustive logic simulation, we make some simplifying assumptions which will be detailed in Section III.A.

B. An Informational Energy-Based Approach

Assuming that we have a complete set of events $\mathcal{A}_n$ (as in (3)), we may regard the probability p_k as the *information* associated with the individual event A_k. Hence, as in the case of entropy, we may define an average measure for the set $\mathcal{A}_n$:

Definition 4: (Informational Energy)

The global information of the finite field $\mathcal{A}_n$ (denoted by $E(\mathcal{A}_n)$) may be expressed by its *informational energy* as (also called the *Gini Function*[1]):

$$E(\mathcal{A}_n) = \sum_{j=1}^{n} p_j^2 \tag{11}$$

We plot in Fig.7 the informational energy of a Boolean variable as a function of its signal probability, that is $E(\mathcal{A}_2) = p^2 + (1-p)^2$.

Fig.7 comes here

Due to its simplicity, the informational energy was used mainly in statistics (not necessarily in conjunction with Shannon's entropy) as a characteristic of a distribution (discrete as is the case here, or continuous in general). However, without a precise theory, its use was rare until its usefulness was proved in [22].

The value of the informational energy is always upper-bounded by 1. This is because $\sum_{j=1}^{n} p_j^2 \le \left(\sum_{j=1}^{n} p_j\right)^2 = 1$ with equality iff one of the events has probability 1 and the rest of them have the probability 0. Thus, $E(\mathcal{A}_n) = 1$ iff the experiment provides the same determinate and unique result. The basic properties satisfied by the informational energy can be found in [23][24]. We also give the following definition:

Definition 5: (Conditional Informational Energy)

Conditional informational energy of some finite field $\mathcal{A}_n$ with probabilities $\{p_i\}_{1 \le i \le n}$ with respect to $\mathcal{B}_m$ (with probabilities $\{q_i\}_{1 \le i \le m}$) is defined as:

$$E(\mathcal{A}_n | \mathcal{B}_m) = \sum_{j=1}^{n} \sum_{k=1}^{m} q_k \cdot p_{kj}^2 \tag{12}$$

where p_{kj} is the conditional probability of events A_j and B_k.

Based on this measure, one can find a relationship between the switching activity and the informational energy. Let $e(x)$ denote the informational energy of a single bit x. Considering x modelled as a lag-one Markov chain with conditional probabilities p_{00}, p_{01}, p_{10}, p_{11} and signal probabilities p_0 and p_1 as in Fig.5, we can give the following result:

Theorem 2: $e(x^+|x^-) = 1 - sw(x) \cdot (p_{00} + p_{11})$ where $e(x^+|x^-)$ is the conditional informational energy between two successive time steps.

---

[1]This was first used by Corrado Gini in a study from "Atti del R. Ist. Veneta di Scienze," *Lettere ed Arti*, 1917, 1918, v. LXXVII

Proof:

From Definition 5, $e\,(x^+|x^-) = p_1\,(p_{10}^2 + p_{11}^2) + p_0\,(p_{00}^2 + p_{01}^2)$. Since the switching activity of line x is $\frac{2 \cdot p_{01} \cdot p_{10}}{p_{01} + p_{10}}$ and $p_1 = \frac{p_{01}}{p_{01} + p_{10}}$, $p_0 = \frac{p_{10}}{p_{01} + p_{10}}$ one can write the following relationship between conditional informational energy and switching activity:

$$e\left(x^+ \,\middle|\, x^-\right) = \frac{p_{01} \cdot \left(p_{10}^2 + p_{11}^2\right) + p_{10} \cdot \left(p_{00}^2 + p_{01}^2\right)}{p_{01} + p_{10}} =$$

$$= \frac{p_{01} \cdot (1 - 2 \cdot p_{10} \cdot p_{11}) + p_{10} \cdot (1 - 2 \cdot p_{00} \cdot p_{01})}{p_{01} + p_{10}} = 1 - sw\,(x) \cdot (p_{00} + p_{11})$$

■

In the most general case, p_{00} and p_{11} can take any values and thus even if we have an exact relation between energy and switching activity (i.e. $sw\,(x) = (1 - e\,(x^+|x^-)) \,/\, (p_{00} + p_{11})$) we cannot bound the sum from the denominator. However, under the temporal independence assumption, we have an *exact* relation since $p_{00} + p_{11} = 1$ and thus:

Corollary 2:

$$sw\,(x) = 2p\,(x)\,(1 - p\,(x)) = 1 - e\,(x)\,. \tag{13}$$

■

For instance, returning to the 1-bit full-adder example in Fig.4, we find $E_{input} = 0.125$ and $E_{output} = 0.875$. Thus, on average, the output exposes an informational energy of 0.437 and, based on (13), a switching activity of 0.563 (compared to the exact value of 0.5). Thus, informational energy (along with entropy) seems to be a reliable candidate for estimation of energy consumption.

Once again we consider the virtual mapping $\xi \rightarrow \xi'$, where each net in ξ' is characterized by the same amount of average informational energy $e\,(\xi')$. Based on (9), the expected switching activity per step in the whole circuit $SW\,(\xi')$, can be expressed as:

$$SW\,(\xi) = SW\,(\xi') = n \cdot (1 - e\,(\xi')) \tag{14}$$

where the cardinality of ξ' is assumed to be n.

Considering the simple case in Fig.6, we get $e\,(\xi') = (3 * 0.5 + 2 * 0.625 + 0.531) / 6 = 0.546$ and therefore, $SW\,(\xi') = 2.71$ which matches well the actual value (2.25). However, in real circuits, direct computation of $e\,(\xi')$ is very costly; to develop a practical approach, we need further simplifications as will be shown subsequently.

C. Quantitative Evaluations

In order to derive a consistent model for energy consumption at the RT-level, we first have to abstract somehow the information which is present at the gate level. Thus, a simplified model is adopted as a starting point.

Let us consider some combinational block realized on n levels as a leaf-DAG[1] of 2-input NAND gates (a similar analysis can be carried out for 2-input NOR gates and the final result is the same). We assume that inverters may appear only at primary inputs/outputs of the circuit; we do not include these inverters in the level assignment step. One can express the signal probability of any net at level $j+1$ as a function of the signal probability at level j by:

$$p_{j+1} = 1 - p_j^2 \qquad \forall j = 0, \ldots, n-1 \tag{15}$$

Similarly, the signal probability of any net at level $j+2$ is given by:

$$p_{j+2} = 1 - \left(1 - p_j^2\right)^2 \qquad \forall j = 0, \ldots, n-2 \tag{16}$$

The average entropy per net at level j is given by:

$$h_j = -p_j \cdot \log p_j - (1 - p_j) \cdot \log(1 - p_j) \tag{17}$$

Using the corresponding average entropy per net at level $j+2$, the parametrized relationship between h_j and h_{j+2} can be approximated by $h_{j+2} \approx \frac{h_j}{2}$ when j is sufficiently large (values greater than 6). Hence we get expressions for entropy per bit at even/odd levels of the circuit: $h_{2j} \approx \frac{h_0}{2^j}$ and $h_{2j+1} \approx \frac{h_1}{2^j}$, where h_0, h_1 are entropies per bit at the primary inputs and first level, respectively. To get a closed form expression, we may further assume that h_1 may be estimated in terms of h_0 as $h_1 \approx \frac{h_0}{\sqrt{2}}$ (in fact, the exact entropy decrease for a NAND gate excited by pseudorandom inputs is 0.8113, but for uniformity, we use $\frac{1}{\sqrt{2}} = 0.707$). Thus, for a 2-input NAND gate leaf-DAG, the entropy per bit at level j may be approximated as:

$$h_j \approx \frac{h_0}{2^{j/2}} \tag{18}$$

This may be further generalized for the case of f-input NAND gate leaf-DAGs, observing that increasing the

[1] In a leaf-DAG, only the leaf nodes have multiple fanouts.

fanin from 2 to f, produces a decrease in the number of levels by log *(f)*. Hence, for a fanin of f, (18) becomes:

$$h_j \approx \frac{h_0}{2^{(j \cdot \log f)/2}} = \frac{h_0}{f^{j/2}} \tag{19}$$

Definition 6: We call $\frac{1}{\sqrt{f}}$ the *information scaling factor (ISF);* it characterizes each logic component (gate, module or circuit).

We will see how relation (19) is affected by the circuit structure and functionality in general. In any case, this provides a starting point for estimating the total entropy at each level in the circuit. In general, the total entropy over all levels N in the circuit is given as:

$$H_{total} = \sum_{j=0}^{N} H_j = \sum_{j=0}^{N} n_j \cdot h_j \tag{20}$$

where H_j is the total entropy at level j and n_j is the number of nodes on level j.

All these considerations can be easily extended for the case of informational energy. Considering the same assumptions as in previous section and using relation (15), the informational energy per net at level j may be expressed as:

$$e_j = p_j^2 + (1 - p_j)^2 \tag{21}$$

Applying (15) for level j+2 and substituting in (21), we get the following parameterized dependency between the informational energies at levels $j + 2$ and j:

$$\begin{aligned} e_{j+2} &= \left(1 - \left(1 - p_j^2\right)^2\right)^2 + \left(1 - p_j^2\right)^4 \\ e_j &= p_j^2 + (1 - p_j)^2 \end{aligned} \tag{22}$$

Using a similar approach as in the case of entropy, we get the following expression for the average informational energy per bit at level j in a circuit with fanin f:

$$e_j = 1 - \frac{1 - e_0}{f^{j/2}} \tag{23}$$

From here, an estimate can be drawn for the total energy at level j, and thus for the total energy over all the levels of the circuit:

$$E_{total} = \sum_{j=0}^{N} E_j = \sum_{j=0}^{N} n_j \cdot e_j \tag{24}$$

where E_j is the total energy at level j and again n_j is the number of nodes on level j.

III. Information Modeling

A. Theoretical Results

As we have seen, an estimate of the average switching activity for a module can be obtained from the total entropy (informational energy) over all levels of the circuit. An exact technique would be too expensive to use in practice; at the same time, since we are dealing with RT-level designs, the internal structure may be unknown. Therefore, to manage the complexity, we will use the following simplifying assumptions:

A_1. *Uniform Network Structure:* Nodes are uniformly distributed over the levels of the circuit.

In other words, we assume the same number of nodes on each level of the circuit. In addition, all the gates on each level are assumed to get their inputs from the previous level. This will significantly simplify our task in obtaining closed-form formulae for average switching activity per module (see Section IV.A for the effect of other common network structures when assumption A_1 is relaxed).

As we have seen, for a leaf-DAG containing 2-input NAND/NOR gates, there is a simple relationship between the values of entropy (informational energy) on adjacent levels in the circuit. Unfortunately, in practice this circuit structure is too restrictive since typical logic circuits exhibit a large fanout, not only at the primary inputs, but also at internal nodes. In addition, logic circuits contain a mixture of gates: while NAND (AND), NOR (OR) gates are *entropy decreasing*, XOR gates and inverters are *entropy preserving* gates. More precisely, the output entropy of XORs and inverters is 1 when they are randomly excited and therefore their information scaling factor is 1. This behavior is still described by (19) for $f = 1$ (similar considerations apply to informational energy). In general, any generic "gate" having equal-sized ON and OFF sets, exposes the same *entropy preserving* characteristic. For instance, given two boolean functions $f(a, b, c)$ and $g(a, b, c)$ with ON-sets of cardinality 4 and 2, respectively, then under random inputs $H_f = 1$ (with an *ISF* of 1), and $H_g = 0.8113$ (with an *ISF* of 0.8113).

A_2. *Uniform Information Variation:* The entropy and informational energy per bit at level j are estimated as:

$$h_j = \frac{h_0}{f_{eff}^{j/2}} \quad \text{and} \quad e_j = 1 - \frac{1 - e_0}{f_{eff}^{j/2}}, \quad j = 0, 1, ..., N. \tag{25}$$

Differently stated, we assume that each "generic gate" in a given circuit is characterized by an *effective information scaling factor* (ISF_{eff}) whose value depends on both structure and functionality of gate:

$ISF_{eff} = \frac{1}{n} \cdot \sum_{i=1}^{n} ISF_i$ where n is the total number of gates in the circuit. Under assumptions A_1 and A_2, we may state the following:

Proposition 1: The average entropy (informational energy) per bit in an N-level circuit, may be estimated as:

$$h_{avg} = h_{in} \cdot \frac{1 - \left(\frac{h_{out}}{h_{in}}\right)^{\frac{N+1}{N}}}{(N+1) \cdot \left(1 - \left(\frac{h_{out}}{h_{in}}\right)^{\frac{1}{N}}\right)}$$

$$e_{avg} = 1 - (1 - e_{in}) \cdot \frac{1 - \left(\frac{1 - e_{out}}{1 - e_{in}}\right)^{\frac{N+1}{N}}}{(N+1) \cdot \left(1 - \left(\frac{1 - e_{out}}{1 - e_{in}}\right)^{\frac{1}{N}}\right)} \tag{26}$$

where h_{in} (e_{in}), h_{out} (e_{out}) are the average input and output entropies (energies) per bit[1].

Proposition 1 gives us an estimate of the average entropy/informational energy in a circuit with N levels. The factor f_{eff} is "hidden" in the relationship between N, h_{in} (e_{in}) and h_{out} (e_{out}) since the outputs are assumed to be on level N:

$$h_{out} = h_N = \frac{h_0}{f_{eff}^{N/2}} = \frac{h_{in}}{f_{eff}^{N/2}}$$

$$e_{out} = e_N = 1 - \frac{1 - e_0}{f_{eff}^{N/2}} = 1 - \frac{1 - e_{in}}{f_{eff}^{N/2}} \tag{27}$$

The above equations show that the loss of information per bit from one level to the next decreases with the number of levels. For circuits with a large logical depth, one can obtain from (26) simpler equations by letting N approach infinity. This is also useful in cases where information about the logic depth of the circuit is not readily available. We therefore make the following assumption:

A_3. *Asymptotic Network Depth:* The number of levels N is large enough to be considered infinity ($N \to \infty$). Using this assumption, we get the following:

Corollary 3: For sufficiently large N, the average entropy and informational energy per bit in the circuit are given by:

$$h_{avg} = \frac{h_{in} - h_{out}}{\ln \frac{h_{in}}{h_{out}}} \qquad e_{avg} = 1 - \frac{e_{in} - e_{out}}{\ln \frac{1 - e_{out}}{1 - e_{in}}} \tag{28}$$

[1] Proofs that are omitted here can be found in [24].

Note: In the above derivations, trivial cases such as a value of zero for the input or output entropy and a value of one for the input or output energy are excluded.

What we have obtained so far are simple formulae for estimating the average entropy (informational energy) per bit, and from these, the average switching activity over all the nets in the module. The main difficulty in practice is to estimate the actual output entropy h_{out} (or informational energy e_{out}) since the information usually available at this level of abstraction is not detailed.

B. The Influence of Structure and Functionality

All logic gates belonging to a given module can be characterized by an effective factor f_{eff} which captures information about the circuit structure and functionality. How can we model a general circuit for entropy/energy based evaluations? One can consider relations (26) and (27), where the information scaling factor reflects not only the structure, but also the fraction of information preserving gates.

Example 2: Let us consider for instance, circuit C17 given below.

Fig.8 comes here

To levelize it properly (every wire that connects the output of a gate at level i to the input of a gate at level $i+2$ must go through some buffer gate at level $i+1$), we added three "dummy" components x, y, z. Logically, x, y, z function as buffers but informationally, they are entropy preserving elements. Considering the nodes uniformly distributed throughout the circuit (according to assumption A_1), the average number of nets per level is 4.25. Applying random vectors at the circuit inputs, the exact value of the entropy per bit at the output is obtained as $h_{out} = 0.44$. The effective scaling factor can be calculated as a weighted sum over all the gates in the circuit; thus the corresponding f_{eff} is 1.55 (there are 3 entropy preserving and 6 entropy decreasing gates). From (27) we get an estimate for the output bit entropy ($j = 3$) as $h_{out} = 0.51$ which is reasonably close to the exact value. Based on the input and output entropy, we may get an estimate for the average entropy per bit and thus for the switching activity. The average switching activity for a generic net in the circuit is $sw_{avg(sim)}$ = 0.437 (from simulation) and based on (26), we get $sw_{avg(est)}$ = 0.382 which is very good compared to simulation. A similar analysis can be performed for the informational energy.

IV. Practical Considerations

A. Using Structural Information

These considerations are equally applicable to data-path operators with known internal structure as well as to control circuits represented either at gate or RT-level.

If some structural information is available (such as the number of internal nodes, the number of logic levels), the average entropy (informational energy) may be evaluated using the actual values of f_{eff}, N and the distribution of nodes on each level. In all cases, the output entropy (informational energy) is the same, computed as in (27). The average entropy (or informational energy) for the whole module depends on the actual distribution of nodes in the circuit. In practice, some common distributions are:

a) Uniform distribution (this case was treated in detail in Section III).

b) Linear distribution (e.g. circuit C17 in Fig.8):

In this case, using a similar approach as the one in Section III, we found the following result (valid for $N \to \infty$) for a generic n input, m output module:

$$h_{avg} = \frac{2 \cdot n \cdot h_{in}}{(n+m) \cdot \ln\left(\frac{h_{in}}{h_{out}}\right)} \cdot \left(1 - \frac{m}{n} \cdot \frac{h_{out}}{h_{in}} - \frac{\left(1 - \frac{m}{n}\right) \cdot \left(1 - \frac{h_{out}}{h_{in}}\right)}{\ln\left(\frac{h_{in}}{h_{out}}\right)}\right) \tag{29}$$

c) Exponential distribution (e.g. a balanced tree circuit with 8 inputs):

In this case, we have for $N \to \infty$:

$$h_{avg} = \frac{h_{in} \cdot \ln\left(\frac{n}{m}\right)}{1 - \frac{m}{n}} \cdot \frac{1 - \frac{m \cdot h_{out}}{n \cdot h_{in}}}{\ln\left(\frac{n \cdot h_{in}}{m \cdot h_{out}}\right)} \tag{30}$$

<u>Note:</u> The main advantage of equations (29), (30)[1] is that they allow an estimate of average entropy (and therefore of average switching activity) of modules or gate-level circuits, without resorting to logic simulation or probabilistic techniques like those presented in [3]-[9].

Similar derivations apply for informational energy. We can see that when $n = m$, we get the same results as in Section III (see equation (28)).

B. Using Functional Information

For common data-path operators, can be efficiently estimated based on the "compositional technique" introduced in [20]. There, the Information Transmission Coefficient (*ITC*) is defined as the fraction of information that is transmitted through a function; it may be computed by taking the ratio of the entropy on the outputs of a function and the entropy on the inputs of that function. For convenience, we call *ITC*s "Entropy

[1]More general expressions for (29), (30) (i.e. when N is not approaching infinity) can be found in [24].

Transmission Coefficients” (*HTC*s) and we characterize them as follows:

$$HTC^{comp} = \frac{W_{out}}{W_{in}} \cdot \frac{h_{out}}{h_{in}} \tag{31}$$

where W_{in} (W_{out}) is the number of bits on the input (output) and h_{in} (h_{out}) is the input (output) average bit entropy.

Therefore, the output entropy h_{out} is estimated based solely on the RT-level description of the circuit through a postorder traversal of the circuit. The main advantage of such an approach, is that *it needs only a high-level view of the design* in order to derive useful information.

A similar technique can be introduced to compute the output informational energy as follows.

Definition 7 (Energy Transmission Coefficient):

The fraction of informational energy transmitted through a function called “Energy Transmission Coefficient” (*ETC*) is defined as the ratio of the output and input informational energy.

In Table 1 we give the values of the *ETC* coefficients for the same data-path operators considered in [20].

Table 1 comes here

Similar to (31), we may evaluate ETC_{out} as a function of *ETC* for the component of interest and the *ETC* values for all inputs:

$$ETC^{comp} = \frac{W_{out}}{W_{in}} \cdot \frac{e_{out}}{e_{in}} \tag{32}$$

where e_{in} (e_{out}) is the input (output) average bit informational energy.

Common data-path circuits (e.g. arithmetic operators) have the *scalability property*, that is, their *HTC/ETC* values do not depend significantly on the data-path width. Unfortunately, there are many other circuits (e.g. control circuits) which cannot be treated in this manner. In those cases, relations (28), (29) and (30) have to be used in conjunction with some information about the circuit structure in order to get reliable estimates for average switching activity.

C. HTC and ETC Variations with the Input Statistics

As presented in [20], Thearling and Abraham’s compositional techniques is only an approximation because it does not consider any dependency which may arise in practical examples. In reality, every module may be embedded in a larger design and therefore its inputs may become dependent due to the structural dependencies (namely, the reconvergent fan-out) in the preceding stages of the logic. As a consequence, the values given in

[20] or here in Table 1 (which correspond to the case of pseudorandom inputs) result in large errors as we process a circuit with internal reconvergent fan-out. To be accurate we need a more detailed analysis as will be described in the following.

Without loss of generality, we restrict ourselves to the case of 8- and 16-bit adders and multipliers (denoted below by *add* and *mul*, respectively) and for each of them, we consider two scenarios:

- Each module is fed by biased input generators, that is input entropy (informational energy) per bit varies between 0 and 1 (respectively 0.5 and 1); each such module is separately analyzed.
- Modules are included in a large design with reconvergent fanout branches, so that inputs of the modules cannot be considered independent. Details of this experiment are reported in [24]. As shown there, the dependence of *HTC*s and *ETC*s on the input statistics can be described empirically by the following simple relations:

$$HTC^{add} \approx HTC_0^{add} \cdot (2 - h_{in}) \qquad HTC^{mul} \approx HTC_0^{mul} \cdot 2^{h_{in} - 1}$$
$$ETC^{add} \approx ETC_0^{add} \qquad ETC^{mul} \approx ETC_0^{mul} \tag{33}$$

where the 0-subscripted values correspond to the pseudorandom case (reported in [20] and here in Table 1). These equations can be easily used to adjust the *HTC/ETC* coefficients in order to analyze large designs more accurately. Differently stated, using equations (33) we loose less information when traversing the circuit from one level to next because we account for structural dependencies.

In any case, our proposed framework is also open to simulation (the *zero-knowledge scenario*); this may provide accurate values for output entropy (informational energy) values, but with a much higher computational cost. In practice, we thus have the following options to analyze a logic- or RT-level design:

Fig.9 comes here

In general, using structural information can provide more accurate results either based on entropy or informational energy measures; supporting evidence for this claim is given in [24]. On the other hand, evaluations based on functional information require less information about the circuit and therefore may be more appealing in practice as they provide an estimate of power consumption earlier in the design cycle. The structural approach is thus more appropriate to be used when a gate-level description is available (and therefore detailed information can be extracted) whilst the functional approach (using the compositional technique) is more suitable for RT/behavioral level descriptions.

V. EXPERIMENTAL RESULTS

In order to asses the accuracy of the proposed model, two experiments were performed: one involving individual modules (ISCAS'85 benchmarks and common data-path components) and the other involving a collection of data-path modules specified by a data flow graph.

A. Validation of the Structural Approach (Gate-Level Descriptions)

The experimental setup consisted of a pseudorandom input generator feeding the modules under consideration. The values of the entropy and informational energy for the circuit inputs were extracted from the input sequence, while the corresponding values at the circuit outputs and the average values of entropy or informational energy were estimated as in Section III (using structural information). These average values were then used to estimate the average switching activity per node; the latter, when weighted by an average module capacitance, is a good indicator of power/energy consumption.

We report in Table 2 our results on benchmark circuits and common data-path operators. $Power_{sim}$ and sw_{sim} are the exact values of power and average switching activity obtained through logic simulation under SIS ($Power_{sim} = \frac{1}{2} \cdot V_{DD}^2 \cdot f_{clk} \cdot \sum_{i=1}^{k} sw_{sim}^i \cdot C_{gate}^i$ and $sw_{sim} = \frac{1}{k} \cdot \sum_{i=1}^{k} sw_{sim}^i$, where k is the number of primary inputs and gates in the circuit). In the third column, C_{module} stands for the module capacitance $C_{module} = \sum_{i=1}^{k} C_{gate}^i$. We also report for comparison under the $Power_{approx}$ column the value of power obtained using the approximation in (2). The error introduced by this approximation (under pseudorandom input data) is on average 4.56%[1]. In the next four columns we report our results for average switching activity and power calculated as in (2) for both entropy- and informational energy-based approaches.

Table 2 comes here

As we can easily see, the *average percentage error* (over all the circuits) is 11.38% (7.71%) for entropy (informational energy)-based evaluations of average switching activity, whilst for total power estimation it is 7.03% (4.42%) for entropy (informational energy)-based approaches. These results were found to be consistent for different input signal probabilities. For comparison, in Fig.10, we present the percentage error variation obtained for a pseudorandom and a biased input sequence. In the latter case, the average percentage error was

[1]The percentage error was calculated as $\left|\frac{value_{sim} - value_{est}}{value_{sim}}\right| \cdot 100$.

11.17% (15.91%) for entropy (informational energy) power estimations. All results were generated in less than two seconds of CPU time on a SPARC 20 workstation with 64 MBytes of memory.

Fig.10 comes here

B. Validation of the Functional Approach (Data-Flow Graph Descriptions)

In Fig.11 we consider a complete data-path represented by the data-flow graph of the differential equation solver given in [25]. All the primary inputs were considered as having 8 bits and the output entropy of the entire module was estimated with the compositional technique based on *HTC*s or *ETC*s. The *HTC*/*ETC* values for the multipliers and adders were calculated according to (33). Using the entropy based approach, the average switching activity was estimated as 0.1805, whilst when using the informational energy, the average switching activity was 0.1683. These estimations should be compared with the exact value of 0.1734 obtained by behavioral simulation.

Fig.11 comes here

Our technique can also be applied to analyze different implementations of the same design in order to trade-off power for area, speed or testability. Suppose we select from the data-path in Fig.11 only the part which computes u_1. In Fig.12 we give two possible implementations of the selected part of the data-path. One is the same as above and the other is obtained using common algebraic techniques such as factorization and common subexpression elimination. All the primary inputs were assumed to be random (except inputs '3' and '*dx*' which are constants). Each adder or multiplier is labelled with its average switching activity value *SW*. In the first case, applying the compositional technique based on entropy, we obtain an average switching activity of 0.186, whilst using informational energy this value is 0.197. For the second implementation, the corresponding values are 0.242 from entropy and 0.282 from informational energy which show an average increase in switching activity of 30%. However, considering the module capacitance of adders and multipliers (as given in Table 2), we actually get a total switched capacitance of 331.15 for the first design and 268.88 for the second one (using entropy-based estimations). This means a decrease of 19%, and thus, the second design seems to be a better choice as far as power consumption is concerned.

Fig.12 comes here

C. Limitations of the Present Approach

The simplifying assumptions considered in Sections II and III were aimed at making this approach predictive in nature. However, they introduce as a side effect some inherent inaccuracies. Generally speaking, simple network structure, uniform information variation and asymptotic network depth are reasonable hypotheses which perform well on average, but may not be satisfied for some circuits.

On the other side, we do not consider temporal correlations at the modules inputs and in some cases this could compromise the accuracy of the predictions, mostly for switching activity estimates. Our preliminary experiments show that the temporal effects are not as important at RT-level as they are at circuit or gate-level but undoubtfully, the overall quality of the approach would benefit from incorporating them into analysis.

Moreover, using the functional information through *HTC/ETC*s coefficients as suggested in Section IV, is applicable only to modules that perform a single well-defined function; multifunctional modules like complex ALU's would require a more elaborate analysis than the one presented here.

Lastly, our model is intended only for zero-delay model; a general delay model cannot be supported in a straightforward manner by this information theoretic approach.

VI. CONCLUSION

In this paper, we presented and information theoretic approach for power estimation at the RT and gate levels of design abstraction. Noting that entropy characterizes the uncertainty of a sequence of applied input vectors and is thus intuitively related to the input switching activity, we have mathematically shown that (under temporal independence assumption) the average switching activity of a signal line is upper-bounded by one-half of its entropy. We then presented two techniques for calculating the entropy at the circuit outputs from the input entropies.

The first technique, which is more applicable when some information about the circuit implementation is provided, calculates the output entropy using an effective information scaling factor (derived from the number and type of logic gates used) and the number of logic levels in the circuit. This technique requires parsing the circuit netlist to generate the appropriate parameters but otherwise, relies on closed form expressions for power estimation from that point on.

The second techniques, which is more applicable when only functional/data-flow information about the circuit is given, calculates the output entropies using a compositional technique. It has linear complexity in the size of the high-level specification of the circuit and is based on a precharacterization of library modules in terms of their entropy transmission coefficients.

Having obtained the output entropies, the average entropy per signal line in the circuit is calculated using closed form expressions derived for three different types of gate distribution per logic level (uniform, linear and exponential). Finally, the average entropies were used to estimate the average switching activity characterizing the module under consideration.

Similar techniques, were developed and presented for informational energy, which is yet another information theoretic measure related to the switching activity in the circuit.

Results obtained for common benchmarks show that the proposed power estimation techniques are extremely fast (few CPU seconds), yet sufficiently accurate (10% relative error on average) to be of use in practical cases.

REFERENCES

[1] F. N. Najm, 'A Survey of Power Estimation Techniques in VLSI circuits', *IEEE Transactions on VLSI Systems,* vol.2, no.4, pp. 446-455, Dec.1994.

[2] M. Pedram, 'Power Minimization in IC Design: Principles and Applications', in *ACM Transactions on Design Automation of Electronic Systems,* vol.1, no.1, pp.1-54, Jan.1996

[3] F.N. Najm, 'A Monte Carlo Approach for Power Estimation', *IEEE Transactions on VLSI Systems*, vol.1, no.1, pp. 63-71, Mar. 1993.

[4] B.J.George, D.Gossain, S.C.Tyler, M.G.Wloka, and G.K.Yeap, 'Power Analysis and Characterization for Semi-Custom Design', in *Proc. Intl. Workshop on Low Power Design,* pp.215-218, April 1994.

[5] C.X.Huang, B.Zhang, A.-C.Deng, and B.Swirski, 'The Design and Implementation of PowerMill', in *Proc. Intl. Workshop on Low Power Design,* pp.105-110, April 1995

[6] A. Ghosh, S. Devadas, K. Keutzer, and J. White, 'Estimation of Average Switching Activity in Combinational and Sequential Circuits', in *Proc. ACM/IEEE Design Automation Conference,* pp. 253-259, June 1992.

[7] F. N. Najm, 'Transition Density: A New Measure of Activity in Digital Circuits', *IEEE Transactions on CAD*, vol. 12, no. 2, pp. 310-323, Feb. 1993.

[8] C.-Y. Tsui, M. Pedram, and A. M. Despain, 'Efficient Estimation of Dynamic Power Dissipation with an Application', in *Proc. IEEE/ACM Intl. Conference on Computer Aided Design,* pp. 224-228, Nov. 1993.

[9] R. Marculescu, D. Marculescu, and M. Pedram, 'Efficient Power Estimation for Highly Correlated Input Streams', in *Proc. ACM/IEEE Design Automation Conference*, pp. 628-634, June 1995.

[10] A. Chandrakasan, et. al, 'HYPER-LP: A System for Power Minimization Using Architectural Transformation', in *Proc. IEEE/ACM Intl. Conference on Computer Aided Design,* pp. 300-303, Nov.1992.

[11] R. Mehra, J. Rabaey, 'High Level Estimation and Exploration', in *Proc. Intl. Workshop on Low Power Design,* pp.197- 202, April 1994.

[12] P. Landman, J. Rabaey, 'Power Estimation for High Level Synthesis', in *Proc. European Design Automation Conferenc*e, pp. 361-366, Feb.1993.

[13] N. Weste and K Eshraghian, 'Principles of CMOS VLSI Design: A Systems Perspective', 2nd Edition, Addison-Wesley Publishing Co., 1994.

[14] P. Landman and J. Rabaey, 'Black-Box Capacitance Models for Architectural Power Analysis', in *Proc.Intl. Workshop on Low Power Design*, pp. 165-170, April 1994.

[15] L. Brillouin, 'Science and Information Theory', Academic Press, New York, 1962.

[16] L. Hellerman, 'A Measure of Computational Work', in *IEEE Trans. on Computers*, vol. C-21, No.5, pp. 439-446, 1972.

[17] R.W. Cook, M.J. Flynn, 'Logical Network Cost and Entropy', in *IEEE Trans. on Computers*, vol. C-22, No. 9, pp. 823-826, Sept. 1973.

[18] K.-T. Cheng, V. Agrawal, 'An Entropy Measure for the Complexity of Multi-Output Boolean Functions', in *Proc. ACM/IEEE Design Automation Conference,* pp. 302-305, June 1990.

[19] V. Agrawal, 'An Information Theoretic Approach to Digital Fault Testing', in *IEEE Trans. on Computers*, vol. C-30, No. 8, pp. 582-587, Aug. 1981.

[20] K. Thearling, J. Abraham, 'An Easily Computed Functional Level Testability Measure', in *Proc. Intl. Test Conference,* pp. 381-390, 1989.

[21] A. Papoulis, ‘Probability, Random Variables, and Stochastic Processes’, McGraw-Hill Co., 1984.

[22] O. Onicescu, ‘Theorie de l’information. Energie informationelle’, Paris, C.R. Acad.Sci., Tome 263, 1966.

[23] O. Onicescu, V. Stefanescu, ‘Elements of Informational Statistics with Applications’, in Romanian, Bucharest 1979.

[24] D. Marculescu, R. Marculescu, and M. Pedram, ‘RT-Level Power Analysis Using Information Theoretic Measures’, USC Technical Report CENG 95-25, Dec. 1995

[25] P. Paulin, and J. Knight, ‘Force-directed Scheduling for the Behavioral Synthesis of ASICs’, *IEEE Transactions on CAD*, vol. CAD-8, No. 6, pp. 661-679, July 1989.