# Information Theoretic Measures of Energy Consumption at Register Transfer Level* 

Diana Marculescu, Radu Marculescu, Massoud Pedram<br>Department of Electrical Engineering - Systems<br>University of Southern California, Los Angeles, CA 90089


#### Abstract

The problem of estimating the energy consumption at register transfer level is addressed from an information theoretical point of view. It is shown that the average switching activity can be predicted without simulation using either entropy or informational energy averages. Consequently, two new measures relying on these concepts are developed. The accuracy of these models is investigated using common benchmarks and the results are promising.


## 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.
Energy minimization in digital systems is not an exception to this trend. Having as soon as possible in the design cycle an estimate for power consumption, may save significant redesign efforts or even completely change the entire design architecture. Circuit and gate level power estimation techniques have been extensively explored [1]-[5]. Higher levels of abstraction have been also considered [6]-[8], but here many problems are still pending a satisfactory solution; the main difficulties arise from the lack of precise information and the conceptual complexity which characterizes these levels.
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 in the circuit, for RT-level designs an average estimate per module is satisfactory. In other words, some accuracy may be sacrificed in order to obtain efficiency for power estimation.
In this paper, we address the problem of power estimation at RT-level from an information theoretical point of view [9]. Traditionally, entropy was considered a

[^0]useful measure for solving problems of area estimation [10]-[12], timing analysis [12] or even testing [13],[14]. Assuming that the internal energy consumption of modules is the dominant contributor to the energy consumption at RT-level, we propose two new measures based on entropy and informational energy for estimating the energy consumption of each module. With some further simplifications, simple closed form expressions are derived and their value in practical applications is explored. A distinctive feature of the present approach is that it does not require logic simulation and is based only on the characteristics of the input sequence and some knowledge about the composition of the circuit.

The paper is organized as follows. Sections 2 and 3 introduce the main concepts and the motivation behind our model. In Section 4 we present some practical considerations and finally, in Section 5, we give the results obtained by analyzing a common data-path. We conclude by summarizing our main ideas and indicating possible extensions of the present work.

## II. An Entropy Based Approach

## A. The Concept and Its Use

Let $\mathrm{A}_{1}, \mathrm{~A}_{2}, \ldots, \mathrm{~A}_{\mathrm{n}}$ be a complete set of events which may occur with the probabilities $\mathrm{p}_{1}, \mathrm{p}_{2}, \ldots, \mathrm{p}_{\mathrm{n}}$ that is:
$\mathrm{p}_{\mathrm{k}} \geq 0,(\forall \mathrm{k}), \mathrm{k}=1,2, \ldots, \mathrm{n} \quad \sum_{\mathrm{k}=1}^{\mathrm{n}} \mathrm{p}_{\mathrm{k}}=1$
Let's 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 $\mathrm{p}_{1}, \mathrm{p}_{2}, \ldots, \mathrm{p}_{\mathrm{n}}$. In order to quantify the content of information revealed by the outcome of such an experiment, Shannon introduced the concept of entropy [15].
Definition 1: The entropy of the finite field $\mathcal{A}_{n}$ (denoted by $\mathrm{H}\left(\mathcal{A}_{n}\right)$ ), is given by:
$\mathrm{H}\left(\mathscr{A}_{n}\right)=\mathrm{H}\left(\mathrm{p}_{1}, \mathrm{p}_{2}, \ldots, \mathrm{p}_{\mathrm{n}}\right)=-\sum_{\mathrm{k}=1}^{\mathrm{n}} \mathrm{p}_{\mathrm{k}} \log \mathrm{p}_{\mathrm{k}}$
where $\log$ stands for $\log _{2}$. The entropy is nonnegative: $\mathrm{H}\left(\mathrm{p}_{1}, \mathrm{p}_{2}, \ldots, \mathrm{p}_{\mathrm{n}}\right) \geq 0$ and if some $\mathrm{p}_{\mathrm{i}}=1$ and $\mathrm{p}_{\mathrm{k}}=0(\mathrm{k} \neq \mathrm{i}, \mathrm{k}$ $=1,2, \ldots, n)$, then $H\left(p_{1}, p_{2}, \ldots, p_{n}\right)=0$.

The entropy in (1) is a measure of the uncertainty contained in $\mathcal{A}_{n}$ in the sense that the larger the entropy, the less certain the outcome of a random experiment in $\mathcal{A}_{n}$. Entropy is also a measure of the information
contained in each event $A_{i}$; this information is given by $\mathrm{I}\left(\mathrm{A}_{\mathrm{i}}\right)=-\log \left(\mathrm{p}_{\mathrm{i}}\right)$. Therefore the entropy is the weighted average information of all the events in the finite field $\mathcal{A}_{n}$.
Note: Information and uncertainty have the same quantitative measure (entropy), but different meanings. As information increases, the uncertainty decreases and vice versa. Indeed the uncertainty is equal to the amount of information which is needed to make the outcome of an experiment known.

Entropy satisfies the following basic properties [15].
Property 1: Entropy is maximized when all events are equiprobable, that is:
$H\left(p_{1}, p_{2}, \ldots, p_{n}\right) \leq H\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right)$
Property 2: If $\mathcal{A}_{n}$ and $\mathcal{B}_{m}$ are any two independent finite fields, $\mathcal{A}_{n} \times \mathcal{B}_{m}$ a new finite field with $n m$ events, then
$\mathrm{H}\left(\mathcal{A}_{n} \times \mathcal{B}_{m}\right)=\mathrm{H}\left(\mathscr{A}_{n}\right)+\mathrm{H}\left(\mathcal{B}_{n}\right)$
Shannon's entropy is equally applicable to partitioned sets of events. More precisely, given a partitioning $\Pi=$ $\left\{\mathrm{A}_{1}, \mathrm{~A}_{2}, \ldots, \mathrm{~A}_{\mathrm{n}}\right\}$ on the set of events $\Omega$, the entropy of this partitioning is:
$\mathrm{H}(\Pi)=-\sum_{i=1}^{\mathrm{n}} \mathrm{p}\left(\mathrm{A}_{\mathrm{i}}\right) \log \mathrm{p}\left(\mathrm{A}_{\mathrm{i}}\right)$
where $p\left(A_{i}\right)$ is the probability of class $A_{i}$ in partition $\Pi$.
Example 1: Truth table for a randomly excited 1-bit full adder is given bellow:

| $c_{i}$ | $x_{i}$ | $y_{i}$ | $s_{i} c_{i+1}$ |  |
| :---: | :---: | :---: | :---: | :---: |
| 0 | 0 | 1 | 1 | 0 |
| 1 | 0 | 0 | 1 | 0 |
| 1 | 1 | 0 | 0 | 1 |
| 1 | 1 | 1 | 1 | 1 |
| 0 | 1 | 1 | 0 | 1 |
| 1 | 0 | 1 | 0 | 1 |
| 0 | 1 | 0 | 1 | 0 |
| 0 | 0 | 0 | 0 | 0 |

where $\mathrm{x}_{\mathrm{i}}, \mathrm{y}_{\mathrm{i}}$ are the inputs, $\mathrm{c}_{\mathrm{i}}$ is carry- $\mathrm{in}, \mathrm{s}_{\mathrm{i}}$ is the sum bit and $c_{i+1}$ is carry-out. The output space is partitioned in four classes as $\Pi=\left\{\mathrm{A}_{1}, \mathrm{~A}_{2}, \mathrm{~A}_{3}, \mathrm{~A}_{4}\right\}=\{10,01,11,00\}$, where $\mathrm{p}\left(\mathrm{A}_{1}\right)=\mathrm{p}\left(\mathrm{A}_{2}\right)=3 / 8, \mathrm{p}\left(\mathrm{A}_{3}\right)=\mathrm{p}\left(\mathrm{A}_{4}\right)=1 / 8$; applying (4) we obtain $\mathrm{H}(\Pi)=1.8113$. 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 a purely random one, then exactly H bits will suffice to represent the output sequence; therefore the average number of transitions per word (or average switching activity per word) would be $\mathrm{H} / 2$. In any other nonrandom arrangement, for a minimum length encoding scheme, the average number of transitions per word would be $\leq \mathrm{H} / 2$, so in practice, this value 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 [10]; it was subsequently explored using symbolic conditional expressions for the output probabilities [16].

As we have seen, an appropriate measure for the average switching activity of each net in the circuit is its entropy value. For any set $\xi$ of nets from the circuit, we have
$\mathrm{SW}(\xi) \leq \frac{\mathrm{H}(\xi)}{2}$
where H is the entropy of the set $\xi$ and SW represents the average number of transitions of this set (in what follows, upper case letters refer to total values and lower case letters to individual values). Therefore considering $\xi$ as the set of nets in the circuit, we may say that the total number of transitions per step in the entire circuit is upper bounded by half of the total entropy of the system. It is unfortunately expensive to compute $\mathrm{H}(\xi)$ since the complete set of events characterizing the circuit is exponential in the number of nodes in the circuit. Alternatively, we can use an approximation considering some distance k (given in the number of gates on any path $\tau$ ) beyond which any two signals may be assumed independent. Partitioning the entire set of nets into groups of independent signals, one can express the entropy of the entire circuit as the sum of the partial entropies. In real cases (where the signals may be highly correlated due to reconvergent fanout), this value will be a conservative upper bound on the actual value of entropy.

## B. A Quantitative Evaluation

Consider now 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 $\mathrm{j}+1$ as a function of the signal probability at the level j by:
$\mathrm{p}_{\mathrm{j}+1}=1-\mathrm{p}_{\mathrm{j}}^{2} \quad \forall \mathrm{j}=0, \ldots, \mathrm{n}-1$
Similarly, the signal probability of any net at level $\mathrm{j}+2$ is given by:
$\mathrm{p}_{\mathrm{j}+2}=1-\left(1-\mathrm{p}_{\mathrm{j}}^{2}\right)^{2} \quad \forall \mathrm{j}=0, \ldots, \mathrm{n}-2$
The average entropy per net at level $j$ is given by:
$h_{j}=-p_{j} \cdot \log p_{j}-\left(1-p_{j}\right) \cdot \log \left(1-p_{j}\right)$
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 $h_{j+2} \approx \frac{h_{j}}{2}$ and hence we get expressions for entropy per bit at even/odd levels of the circuit: $h_{2 j} \approx \frac{h_{0}}{2^{j}}$ and $h_{2 j+1} \approx \frac{h_{1}}{2^{j}}$, 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

[^1]$h_{1} \approx \frac{h_{0}}{\sqrt{2}}$ (in fact, the exact decrease is 0.811 , but for uniformity, we chose this way). Thus, for a 2-input NAND gate leaf-DAG, the entropy per bit at level j may be approximated as:
$\mathrm{h}_{\mathrm{j}} \approx \frac{\mathrm{h}_{0}}{2^{\mathrm{j} / 2}}$
This may be further generalized for the case of f-input NAND gate leaf-DAGs, observing that increasing the fanin from 2 to $f$, produces an decrease in the number of levels by logf. Hence, for a fanin of $f,(10)$ becomes:
$h_{j} \approx \frac{h_{0}}{2^{(j \cdot \log f) / 2}}=\frac{h_{0}}{f^{j / 2}}$
We call $\frac{1}{\sqrt{\mathrm{f}}}$ information scaling factor; it characterizes each logic component (gate, module or circuit). We will see how this relation 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 in the circuit would thus be:
$H_{\text {total }}=\sum_{j=0}^{N} H_{j}$
where $\mathrm{H}_{\mathrm{j}}$ is the total entropy at level j .

## III. An Informational Energy Based Approach

## A. The Concept and Its Use

Definition 2: The global information of the finite field $\mathcal{A}_{n}$ (denoted by $\mathrm{E}\left(\mathcal{A}_{n}\right)$ ) may be expressed by its informational energy as follows ${ }^{2}$ :
$\mathrm{E}\left(\mathscr{A}_{n}\right)=\sum_{\mathrm{j}=1}^{\mathrm{n}} \mathrm{p}_{\mathrm{j}}^{2}$
For instance, using the truth table in Example 1 we find $E_{\text {input }}=0.125$ and $E_{\text {output }}=0.875$.

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 Onicescu proved its usefulness [17]. We give in the following few basic properties satisfied by the informational energy [18].

Property 3: The informational energy becomes $1 / n$ when all events are equiprobable and 1 when one of the events in $\mathcal{A}_{n}$ is certain:
$\frac{1}{\mathrm{n}} \leq \mathrm{E}\left(\mathcal{A}_{n}\right) \leq 1$
Property 4: If the uniformity (or the uncertainty) of the system increases, then its informational energy decreases. Property 5: If $\mathcal{A}_{n}$ and $\mathcal{B}_{m}$ are two independent finite

[^2]fields, then $\mathrm{E}\left(\mathcal{A}_{n} \times \mathcal{B}_{m}\right)=\mathrm{E}\left(\mathcal{A}_{n}\right) \mathrm{E}\left(\mathcal{B}_{m}\right)$.
Based on this measure, one can find a relationship between the switching activity and informational energy. Let $\mathrm{e}(\mathrm{x})$ denote the informational energy of a single bit x . For this bit, under temporal independence assumption, we have an exact relation:
$\operatorname{sw}(x)=2 p(x)(1-p(x))=1-e(x)$
Since the average switching activity over all nodes in the circuit corresponds to the arithmetic mean rather than geometric mean, we will use for the average informational energy for a set $\xi$ the following:
$\mathrm{e}=\frac{E_{\text {total }}}{|\xi|}$
where $\mathrm{E}_{\text {total }}$ represents the additive informational energy of the entire circuit (even if it is multiplicative by its nature).

However, direct computation of $\mathrm{E}_{\text {total }}$ is very costly; considering once again the partitioning strategy as in the entropy case and inductively applying Property 5, we may end up with a practical approach as will be shown subsequently.

## B. A Quantitative Evaluation

We consider the same assumptions as in Section 2.2. Using relation (6), the informational energy per net at level j may be expressed as:
$e_{j}=p_{j}^{2}+\left(1-p_{j}\right)^{2}$
Applying (6) for level $j+2$ and substituting in (18), we get the following parameterized dependency between the informational energies at levels $\mathrm{j}+2$ and j :
$\mathrm{p}_{\mathrm{j}+2}=1-\left(1-\mathrm{p}_{\mathrm{j}}^{2}\right)^{2}$
$e_{j+2}=\left(1-\left(1-p_{j}^{2}\right)^{2}\right)^{2}+\left(1-p_{j}^{2}\right)^{4}$
$e_{j}=p_{j}^{2}+\left(1-p_{j}\right)^{2}$
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 $f$ :
$\mathrm{e}_{\mathrm{j}}=1-\frac{1-e_{0}}{\mathrm{f}^{\mathrm{j} / 2}}$
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:

$$
\begin{equation*}
E_{\text {total }}=\sum_{j=0}^{N} E_{j} \tag{21}
\end{equation*}
$$

## IV. Practical Considerations

In practice, tree and leaf-DAG structures are too restrictive to be considered alone: random logic circuits exhibit a large fanout not only at the primary inputs, but also at internal nodes. This is one reason to reconsider the above relations to make them applicable in practice. In addition, logic circuits contain a mixture of gates: while NAND (AND), NOR (OR) are entropy decreasing, XORs and inverters are entropy preserving gates. More precisely, the output entropy of XORs is 1 when they are
randomly excited and therefore their information scaling factor is 1 . Their behavior is still described by (11) for $\mathrm{f}=1$ (similar considerations apply to informational energy). In general, any generic "gate" having almost equal-sized ON and OFF sets, exposes the same almost entropy preserving characteristic. In short, we may say that structural and functional aspects are important. At RT-level both of them have to be abstracted and used in an implicit manner to compensate the lack of explicit information which characterize high-level representations.

To illustrate these issues, we considered a subset of ISCAS'85 circuits and some common data-path circuits mapped with lib2.genlib library (Table 1) and look-up tables (LUTs) with 3 inputs (Table 2); our objective was to asses the structure/function influences over constrained or unconstrained mappings. In these tables, "Nodes'" represents the total number of nodes in the circuit (including inputs), 'Levels"' stands for number of levels in the circuit, ' p ', is the fraction of conservative gates and $h_{\text {avg }}, e_{\text {avg }}, s w_{\text {avg }}$ are the simulated values for average entropy, informational energy and switching activity, respectively.

Table 1: Mapped Using lib2.genlib

| Circuit | Nodes | Levels | p | $\mathrm{h}_{\text {avg }}$ | $\mathrm{e}_{\text {avg }}$ | $\mathrm{sw}_{\text {avg }}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| C432 | 234 | 31 | 0.35 | 0.7391 | 0.6558 | 0.4565 |
| C880 | 419 | 45 | 0.32 | 0.7861 | 0.6092 | 0.4531 |
| add4 | 29 | 10 | 0.40 | 0.8508 | 0.5785 | 0.4187 |
| mul2 | 40 | 12 | 0.62 | 0.4870 | 0.7588 | 0.2253 |

Table 2: Mapped Using 3-input LUT

| Circuit | Nodes | Levels | p | $\mathrm{h}_{\text {avg }}$ | $\mathrm{e}_{\text {avg }}$ | $\mathrm{sw}_{\text {avg }}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| C432 | 158 | 21 | 0.15 | 0.7643 | 0.6399 | 0.4523 |
| C880 | 252 | 19 | 0.25 | 0.8273 | 0.6032 | 0.4617 |
| add4 | 17 | 5 | 1 | 0.9689 | 0.5193 | 0.4936 |
| mul2 | 51 | 7 | 0.7 | 0.5309 | 0.7290 | 0.4764 |

As we can see, the actual structure of the circuits changes the average entropy/informational energy per bit and ultimately, the average switching activity per bit. In addition, there seems to exist a 'natural'" tendency for different functions to be mapped with a larger or smaller number of entropy/informational energy preserving gates.

How can we model a general circuit for entropy/ energy based evaluations ? One can consider relations (11) and (20), but there the information scaling factor should reflect not only the average fanin in the circuit, but also the fraction of preserving gates. The corresponding effective value $f_{\text {eff }}$ will be smaller than the average fanin $f$ due to actual gate composition of the circuit. Let us consider for instance, circuit C17 given below:


To levelize it properly, we added three "dummy" components $\mathrm{x}, \mathrm{y}, \mathrm{z}$. Logically, $\mathrm{x}, \mathrm{y}, \mathrm{z}$ function as buffers but informationally, they are entropy preserving elements. Considering the nodes uniformly distributed throughout the circuit, the average number of nets per level is $(6+5+4+2) / 4=4.25$. Applying random vectors at the circuit inputs, the entropy per bit at the output becomes $h_{\text {out }}=0.44$. The effective scaling factor can be calculated as a weighted sum over all the gates in the circuit; thus the corresponding $\mathrm{f}_{\text {eff }}$ is: $\left(\frac{9}{3 \cdot 1+6 \cdot \frac{1}{\sqrt{2}}}\right)^{2}=1.55$ (there are 3 entropy preserving and 6 entropy decreasing gates). From (11) we get an estimate for the output bit entropy $(\mathrm{j}=3)$ as $\mathrm{h}_{\text {out }}=0.51$ which is reasonably close to the exact value.

In what follows, we consider the following simplifying assumptions:
A $_{1}$. Uniform distribution: Nodes are uniformly distributed over the levels of the circuit.
$\mathbf{A}_{\mathbf{2}}$. Uniform decrease: The entropy and informational energy per bit at level $j$ are estimated in terms of $f_{\text {eff }}$ as:
$h_{j}=\frac{h_{0}}{f_{\text {eff }}^{j / 2}}$ and $e_{j}=1-\frac{1-e_{0}}{f_{\text {eff }}^{j / 2}}, j=0,1, \ldots, N$.
Under these assumptions, we may state the following:
Proposition 1. The average entropy/energy per bit in an Nlevel circuit, may be estimated as:

$$
\begin{align*}
& \mathrm{h}_{\mathrm{avg}}=\mathrm{h}_{\mathrm{in}} \cdot \frac{1-\left(\frac{\mathrm{h}_{\text {out }}}{\mathrm{h}_{\text {in }}}\right)^{\frac{\mathrm{N}+1}{\mathrm{~N}}}}{(\mathrm{~N}+1) \cdot\left(1-\left(\frac{\mathrm{h}_{\text {out }}}{\mathrm{h}_{\text {in }}}\right)^{\frac{1}{\mathrm{~N}}}\right)} \\
& \mathrm{e}_{\mathrm{avg}}=1-\left(1-\mathrm{e}_{\text {in }}\right) \cdot \frac{1-\left(\frac{1-\mathrm{e}_{\text {out }}}{1-\mathrm{e}_{\text {in }}}\right)^{\frac{\mathrm{N}+1}{\mathrm{~N}}}}{(\mathrm{~N}+1) \cdot\left(1-\left(\frac{1-\mathrm{e}_{\text {out }}}{1-\mathrm{e}_{\text {in }}}\right)^{\frac{1}{\mathrm{~N}}}\right)} \tag{22}
\end{align*}
$$

where $\mathrm{h}_{\text {in }} / \mathrm{e}_{\text {in }}, \mathrm{h}_{\text {out }} / \mathrm{e}_{\text {out }}$ are the average input and output entropies/energies per bit.

This gives us an estimate of the average entropy/ energy in a circuit with $N$ levels. The factor $f_{\text {eff }}$ is "hidden" in the relationship between $\mathrm{N}, \mathrm{h}_{\text {in }}\left(\mathrm{e}_{\text {in }}\right)$ and $\mathrm{h}_{\text {out }}\left(\mathrm{e}_{\text {out }}\right)$ since the outputs are considered to be on level N :
$h_{\text {out }}=h_{N}=\frac{h_{0}}{f_{\text {eff }}^{N / 2}}=\frac{h_{\text {in }}}{f_{\text {eff }}^{N / 2}}$
$\mathrm{e}_{\text {out }}=\mathrm{e}_{\mathrm{N}}=1-\frac{1-\mathrm{e}_{0}}{\mathrm{f}_{\text {eff }}^{\mathrm{N} / 2}}=1-\frac{1-\mathrm{e}_{\mathrm{in}}}{\mathrm{f}_{\mathrm{eff}}^{\mathrm{N} / 2}}$
The greater the number of levels is, the smaller the value for $f_{\text {eff }}$ will be, which somehow suggests that the loss of information per bit from one level to another decreases with the number of levels. However, the usefulness of
these formulas is limited, since in general at RT-level we know very little about the internal structure of the circuit. To compensate this lack of information, we add a new assumption valid for circuits with high logical depth:
$\mathbf{A}_{3}$. Asymptotic values: For sufficiently large N , the average entropy and informational energy per bit in the circuit are given by:

$$
\begin{equation*}
h_{\text {avg }}=\frac{h_{\text {in }}-h_{\text {out }}}{\ln \frac{h_{\text {in }}}{h_{\text {out }}}} \quad e_{\text {avg }}=1-\frac{e_{\text {in }}-e_{\text {out }}}{\ln \frac{1-e_{\text {out }}}{1-e_{\text {in }}}} \tag{24}
\end{equation*}
$$

Note: In the above assumption, trivial cases such as zero input or output entropy and one input or output energy are excluded.

The main difficulty in practice is to estimate the actual output entropy $\mathrm{H}_{\text {out }}$ (which is approximately m times greater than $h_{\text {out }}$, where $m$ is the number of outputs), since the information usually available at this level of abstraction is not detailed at all. Some considerations about the estimation of $\mathrm{H}_{\text {out }}$ can be made:
Case 1. Common data-path operators may allow a quick estimation of $\mathrm{H}_{\text {out }}$ based on the "composition technique" introduced in [14]. 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. We give below the ITC values for 8 -bit common datapath operators taken from [14].

| Operator | ITC | Operator | ITC |
| :--- | :---: | :---: | :---: |
| Addition | 0.500 | Negation | 1.000 |
| Subtraction | 0.500 | And, Or | 0.406 |
| Multiplication | 0.461 | $<,>$ | 0.063 |
| Divide by 2 | 0.875 | Multiplexer | 0.471 |

Using the following relationship between the ITC on the output signals and the ITCs on the input signals for a particular component, we may estimate the ITCs values throughout the circuit [14]:
ITC $_{\text {out }}=$ ITC $_{\text {comp }} \cdot \sum_{i=1}^{n} \frac{w_{i}}{W} \cdot$ ITC $_{i}$
where $\mathrm{ITC}_{\text {comp }}$ is the ITC for the component of interest, $\mathrm{ITC}_{\mathrm{i}}$ is the ITC value for input $\mathrm{i}, \mathrm{n}$ is the number of input signal paths for the component, $\mathrm{w}_{\mathrm{i}}$ is the data-path width for input $i$, and $W=\sum_{i=1}^{n} w_{i}$. In particular, the output entropy $\mathrm{H}_{\text {out }}$ may be estimated relying solely on the RTlevel description of the circuit.
Case 2. Control circuits need a more elaborate treatment:
a) Control part with unknown internal structure can be mapped using k-input LUTs. From these abstract mappings, we may estimate a set of values for $\mathrm{f}_{\text {eff }}$ and N . After that, the average switching activity is deduced as the mean over the estimated values obtained in all these mappings.
b) Alternatively, when the structural information is available, one can estimate the output signal probability
with an incremental approach as in [5]. These values can be just plugged into (23) to obtain an estimate for average output entropy/informational energy.
c) If neither structural nor functional information is available (a zero-knowledge scenario), we may rely entirely on the characteristics of the input stream and get a rough estimate for average switching values with
$\mathrm{h}_{\mathrm{avg}}=\frac{\mathrm{H}_{\text {total }}}{\mathrm{N}} \approx \frac{\mathrm{h}_{0}+\mathrm{h}_{1}}{2} \quad \mathrm{e}_{\mathrm{avg}}=\frac{\mathrm{E}_{\text {total }}}{\mathrm{N}} \approx \frac{\mathrm{e}_{0}+\mathrm{e}_{1}}{2}$
The values of the entropy and informational energy are extracted at the primary inputs, while the corresponding values at level 1 can be estimated using the recurrence equations (6), (8) and (18).
In any case, this framework is also open to logic simulation; this may provide accurate values for output entropy/informational energy values but with a much higher computational cost.

## V. EXPERIMENTAL RESULTS

The whole point about the proposed models is to assess the accuracy of relations (23) and (24). Their main advantage is the fact that they do not need any simulation to predict an average energy consumption for the target circuit.
Two types of experiments were performed: one involving individual modules (ISCAS' 85 bechmarks and common data-path operators) and another one using allocated data-paths. Due to its imprecise meaning at RTlevel, the delay was not considered as a parameter.
The experimental setup consisted of a pseudo-random input generator feeding the modules under consideration. The values of the entropy and informational energy were extracted at the primary inputs, while the corresponding values at the outputs were estimated as in Section 4. Once obtained, the average values of entropy or informational energy were directly used to estimate the average switching activity per node; this value weighted by an average node capacitance is a good indicator of power/ energy consumption.
We report in Tables 3 our results on benchmark circuits and common data-path operators. We note that the results are promising in all cases, as the max error is 0.094 (0.0951) for entropy (energy)-based estimations.

Table 3: Data-Path and ISCAS Circuits

|  | Simulated values |  |  | Estimated values |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Circuit | $\mathrm{h}_{\text {avg }}$ | $\mathrm{e}_{\text {avg }}$ | $\mathrm{sw}_{\text {avg }}$ | $\mathrm{h}_{\text {avg }}$ | $\mathrm{sw}_{\text {avg }}$ | $\mathrm{e}_{\text {avg }}$ | $\mathrm{sw}_{\text {avg }}$ |
| add8 | 0.8494 | 0.5792 | 0.4282 | 0.9394 | 0.4697 | 0.5218 | 0.4782 |
| add16 | 0.8487 | 0.5754 | 0.4369 | 0.9252 | 0.4626 | 0.5268 | 0.4732 |
| add32 | 0.9201 | 0.5486 | 0.4017 | 0.9914 | 0.4957 | 0.5037 | 0.4968 |
| mu18 | 0.6857 | 0.6615 | 0.3163 | 0.6778 | 0.3389 | 0.6365 | 0.3635 |
| mul16 | 0.7564 | 0.6952 | 0.3854 | 0.6974 | 0.3487 | 0.6265 | 0.3754 |
| mul32 | 0.7086 | 0.6857 | 0.3759 | 0.6918 | 0.3459 | 0.6316 | 0.3684 |
| C432 | 0.8391 | 0.5558 | 0.4565 | 0.9048 | 0.4524 | 0.5516 | 0.4484 |
| C880 | 0.8861 | 0.5092 | 0.4531 | 0.8236 | 0.4118 | 0.5953 | 0.4047 |
| C1908 | 0.9272 | 0.4650 | 0.4519 | 0.9928 | 0.4964 | 0.5049 | 0.4951 |

In Fig. 3 we considered a complete data-path
represented by the data-flow graph of the differential equation example [19]; 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 ITCs. The ITC value for the multipliers was taken 0.461 and the ITCs for adders and subtracters were 0.5 (we indicated in Fig. 3 the values of ITCs in the entire circuit). Based on these values, $\mathrm{h}_{\text {out }}$ was estimated as 0.1508 compared to 0.1666 obtained by exact simulation. The overall errors in estimating the average switching activities using entropy and informational energy were 0.0228 and 0.0202 , respectively.


Fig. 3: A Data-Path Example

## VI. Conclusions

In this paper, we have proposed two new measures to estimate the energy consumption in RT-level digital circuits. Their foundation relies on statistical modelling of the circuit behavior and seems to successfully overcome the lack of information which characterizes this level of abstraction. The accuracy of the model is investigated using common benchmarks for pseudorandom input sequences. As shown, the average switching activity may be predicted without simulation using either entropy or informational energy averages; the error in prediction is generally small enough to be satisfactory in practice. Future work will be devoted to extend the compositional technique in Section 4, to also handle the informational energy. In addition, it would be desirable to make the compositional technique more sensitive to different styles of implementing data-path/control hardware.

## REFERENCES

[1] M.Pedram, J.Rabaey 'Tutorial on low power design techniques', Design Automation Conference, June 1994.
[2] F. N. Najm, R. Burch, P. Yang, and I. Hajj, 'Probabilistic Simulation for Reliability Analysis of CMOS VLSI Circuits’, in IEEE Trans. on CAD, vol. CAD-9, April 1990.
[3] C.-Y. Tsui, M. Pedram, and A. M. Despain, 'Efficient Estimation of Dynamic Power Dissipation with an Application', in Proc. 1993 Intl. Conference on Computer Aided Design.
[4] P. Schneider, U. Schlichtmann, and K. Antreich, 'A New Power Estimation Technique with Application to Decomposition of Boolean Functions for Low Power', in Proc. 1994 European Design Automation Conference.
[5] R. Marculescu, D. Marculescu, and M. Pedram, 'Switching Activity Analysis Considering Spatiotemporal Correlations’, in Proc. 1994 Intl. Conference on Computer Aided Design.
[6] A. Chandrakasan, et. al, 'HYPER-LP: A System for Power Minimization Using Architectural Transformation', in Proc. 1992 Intl. Conference on Computer Aided Design.
[7] R. Mehra, J. Rabaey, 'High Level Estimation and Exploration', in Proc. 1994 Intl. Workshop on Low Power Design.
[8] P. Landman, J. Rabaey, 'Black-Box Capacitance Models for Architectural Power Analysis’, in Proc. 1994 Intl. Workshop on Low Power Design.
[9] L. Brillouin, 'Science and Information Theory', Academic Press, New York, 1962.
[10] L. Hellerman, 'A Measure of Computational Work', in IEEE Trans. on Computers, vol. C-21, No.5, 1972.
[11] R.W. Cook, M.J. Flynn, 'Logical Network Cost and Entropy', in IEEE Trans. on Computers, vol. C-22, No. 9, 1973.
[12] K.-T. Cheng, V. Agrawal, 'An Entropy Measure for the Complexity of Multi-Output Boolean Functions', in Proc. 1990 Design Automation Conference, 1990.
[13] V. Agrawal, 'An Information Theoretic Approach to Digital Fault Testing', in IEEE Trans. on Computers, Aug. 1980.
[14] K. Thearling, J. Abraham, 'An Easily Computed Functional Level Testability Measure’, in Proc. 1989 Intl. Test Conference.
[15] A. Papoulis, 'Probability, Random Variables, and Stochastic Processes', McGraw-Hill Co., 1984.
[16] R. Marculescu, 'A Probabilistic Approach for Power Estimation Under Pseudorandom Input Sequences', Class Project, Fall 1993, Univ. of Southern California.
[17] O. Onicescu, 'Theorie de l'information. Energie informationelle', Paris, C.R. Acad.Sci., Tome 263, 1966
[18] O. Onicescu, V. Stefanescu, 'Elements of Informational Statistics with Applications', in Romanian, Bucharest 1979
[19] P.G. Paulin, et. al, 'HAL: A Multiparadigm Approach to Automated Data-Path Synthesis', in Proc. 1986 Design Automation Conference, 1986


[^0]:    *This research was supported in part by the National Science Foundation's Young Investigator Award under contract no. MIP-9457392.

[^1]:    ${ }^{1}$ In a leaf-DAG, only the leaf nodes have multiple fanouts.

[^2]:    ${ }^{2}$ This was firstly used in statistics by C.Gini in 1917.

