Theoretical Bounds for Switching Activity Analysis in Finite-State Machines

Diana Marculescu
Department of Electrical Engineering
University of Maryland
College Park, MD 20742

Radu Marculescu
Department of Electrical and Computer Engineering
University of Minnesota
Minneapolis, MN 55455

Massoud Pedram
Department of Electrical Engineering - Systems
University of Southern California
Los Angeles, CA 90089

Abstract - *The objective of this paper is to provide lower and upper bounds for the switching activity on the state lines in Finite State Machines (FSMs). Using a Markov chain model for the behavior of the states of the FSM, we derive theoretical bounds for the average Hamming distance on the state lines which are valid irrespective of the state encoding used in the final implementation. Such lower and upper bounds, in addition to providing a target for any state assignment algorithm, can also be used as parameters in a high-level model of power, and thus provide an early indication about the performance limits of the target FSM. Experimental results obtained for the mcnc'91 benchmark suite show that our bounds are tighter than the bounds reported previously by other researchers and can be effectively used in a high-level power estimation framework.*

1. Introduction

Since power consumption has become a critical issue in the development of digital systems, tools that can control the power budget during the various phases of the design process are in high demand [1]. Given an initial specification of the behavior of the system, several synthesis/ optimization steps are required to generate a final, efficient implementation. Synthesis systems typically take a hardware description language (HDL) model of a design as the initial input. The synthesis path is usually composed of several steps that can be summarized as follows: *high-level synthesis*, *state assignment* of the symbolic states of the FSM describing the control part, *logic synthesis* and *library binding*.

This paper targets the very first step in this synthesis flow where the FSM characterizing the control part of the high-level representation is typically described in the form of a State Transition Graph (STG) and each state is represented in a symbolic form. Such an investigation is motivated by the fact that, despite the significant efforts invested in the research of low-power systems, little has been done from the perspective of theoretical aspects involved in the design and application of such systems. More precisely, most of the work done so far has targeted only algorithms for state assignment [2-4], re-encoding [5] or synthesis [6] for low-power. In a complementary effort, the problem of bus encoding for low-power has been tackled by several researchers [7-10]. However, little has been done in the area of finding good theoretical bounds for the average switching activity in FSMs or on buses. Preliminary results in this direction have been recently reported in [11, 12].

The particular subject this paper addresses is related to the derivation of *theoretical bounds* for FSM switching activity. More precisely, our objective is to find *lower* and *upper bounds* on the switching activity of the state lines, directly from an abstract STG description, far before the state assignment is actually made. Such lower and upper bounds, in addition to providing a target for any state assignment algorithm, can also be used as parameters in a high-level model of power, and thus provide an early indication about the performance limits of the target FSM. To be more specific, consider the typical representation of a standard FSM in Fig.1a. In Fig.1b, we represent a portion of the STG which describes the behavior of the FSM.

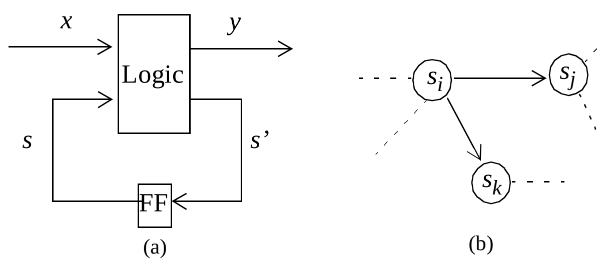

Fig.1: The FSM model

Depending on the actual encoding used in the final implementation of the FSM, the state lines *s* may switch more or less frequently. The amount of switching on the state lines is best characterized by the *average Hamming distance* between the codes assigned to consecutive states. What we aim in this paper is to provide lower and upper bounds for the average Hamming distance on the state lines *s* which are valid regardless of the state encoding used in the final implementation. These bounds can be later combined with the average Hamming distance extracted from the statistics of the vector sequence at the primary inputs *x* and primary outputs *y*, and used to derive performance limits in terms of total power consumption in the target FSM. The main advantage of doing so lies in the independence of the results from the actual implementation which gives more flexibility to the designer.

A similar issue was addressed by Tyagi in [11]. In that paper, the author introduces two lower bounds on the average Hamming distance per transition which emphasize the qualitative dependence of the average Hamming distance on the number of bits used for encoding. An interesting "by-product" of these lower bounds is a greedy state assignment algorithm. In [12], lower and upper bounds for the average switching activity on an information channel (typically, a bus) are provided. However, their applicability in the case of FSMs is severely limited by the fact that they are achievable merely through *compression* techniques rather than *encoding* techniques (which is the case for FSMs).

In this paper we improve over Tyagi's work by providing not only tighter lower bounds, but also upper bounds for the average Hamming distance. The new lower and upper bounds we propose in this paper are easy to understand and are based on either the *informational energy* or the *topology* of the FSM. Together with a technology-independent measure for the circuit complexity, they can be used to generate performance limits in terms of total power consumption.

In this paper, we target only lower and upper bounds for the switching activity, although in order to be useful for high-level power analysis, they have to be used in conjunction with an estimator for the area (complexity) of the target circuit. This is very important from a practical point of view, since it provides an estimate for the range where total power values lie, early in the design cycle. As pointed out in [16], for control circuits, one can use a power model that depends on the switching activities on the primary inputs, primary outputs and state lines:

$$P \propto C_0 \cdot \alpha_I \cdot N_I \cdot N_M + C_1 \cdot \alpha_O \cdot N_O \cdot N_M$$

where N_I and N_O denote the number of external input plus state lines and external output plus state lines for the FSM, C_0 and C_1 are regression coefficients which are empirically derived from low-level simulation of previously designed standard cell controllers, α_I and α_O denote the switching activities on the external input plus state lines and external output plus state lines, and finally N_M denotes the number of minterms in an optimized cover of the FSM and is related to the area of the two-level implementation of the circuit. Thus, having computed lower and upper bounds on the total power consumption of the sequential circuit, we can derive lower and upper bounds on the total power values.

Other efforts in the direction of estimating the area (complexity) of a circuit have been presented in [17, 18], but the problem is far from being solved. In [17], a measure of the circuit area is deduced by using the average number of literals per essential prime implicant. A multiple-output circuit is transformed into a single-output equivalent circuit and its area is estimated using the results in [19] for single-output boolean functions. In [18], as an estimate for the area of the target circuit, the authors propose a measure which is linear in the number of nodes in Binary Decision Diagram (BDD) associated with the circuit. However, since we want to keep the independence from the actual implementation of the circuit, none of these approaches is directly applicable in our case because the state encoding is unknown. Thus, we need an *encoding-independent* measure for the complexity of the circuit. Our proposed measure is based on the size of the *minimal symbolic cover* of the FSM [20].

The paper is organized as follows. In Sections 2 and 3, we present the lower and upper bounds, the complexity of their calculation and discuss their relationship with other possible bounds. The applicability of the lower and upper bounds in a high-level power model is discussed in Section 4 and the experimental results are presented in Section 5. Section 6 concludes the paper and summarizes our main contribution.

2. Informational lower and upper bounds on FSM switching

In this section we present the theoretical framework for determining lower and upper bounds on FSM switching activity. Formally, the problem to be solved can be formulated as follows:

"Given the behavior of the input data and the behavioral description of a controller in the form of the STG associated with the FSM, find lower and upper bounds on the average Hamming distance of the state lines, for a fixed length encoding of the states."

Behavior of the input data refers to the actual steady state and conditional probabilities for the primary inputs, assuming that the inputs are obtained using a Markov generator. In the following, we assume that the state lines of the FSM are modeled as a lag-one Markov chain[1] characterized by the stochastic matrix $Q = (q_{ij})_{1 \le i,j \le n}$, where n is the number reachable of states of the FSM and $q_{ij} = p\,(s = j \mid s' = i)$ is the conditional probability of the FSM being in state i given that it was previously in state j. These probabilities, along with the steady state probability vector $p = (p_i)_{1 \le i \le n}$, can be found using standard techniques for probabilistic analysis of FSMs as proposed in [13] (applicable only for uncorrelated input streams) or [14] (applicable for any high-order Markov source for the primary inputs). Alternatively, having the STG of the FSM, a very fast functional simulation can be performed for a typical input data stream and the set of reachable states can be extracted along with the characteristics of the underlying Markov chain. From here on, we will refer to lag-one Markov chains as simply Markov chains.

Definition 1. (Average Hamming distance) Given a fixed length encoding of the states of a Markov chain, the average Hamming distance of the chain is given by:

$$\tilde{d} = \sum_{1 \le i, j \le n} p_i \cdot q_{ij} \cdot d_{ij}$$

where $Q = (q_{ij})_{1 \le i,j \le n}$ is the matrix characterizing the chain, $p = (p_i)_{1 \le i \le n}$ is the steady state probability vector, and d_{ij} is the Hamming distance between the codes assigned to states i and j.

Example 1: Consider the Markov chain in Fig.2, where each edge from state s_i to state s_j is labeled with the conditional probability q_{ij}. The steady state probability vector is given by p = (1/

1.While in general the Markov chain corresponding to the state lines may be characterized by a high-order Markov chain, for the purpose of estimating the average Hamming distance, modeling the states as lag-one Markov chains is sufficient.

3 1/6 1/6 1/6 1/6). Assuming a random encoding of the states such that *code* (s_1) = ‘000’, *code* (s_2) = ‘001’, *code* (s_3) = ‘110’, *code* (s_4) = ‘010’, *code* (s_5) = ‘101’, then the average Hamming distance of the chain is given by:

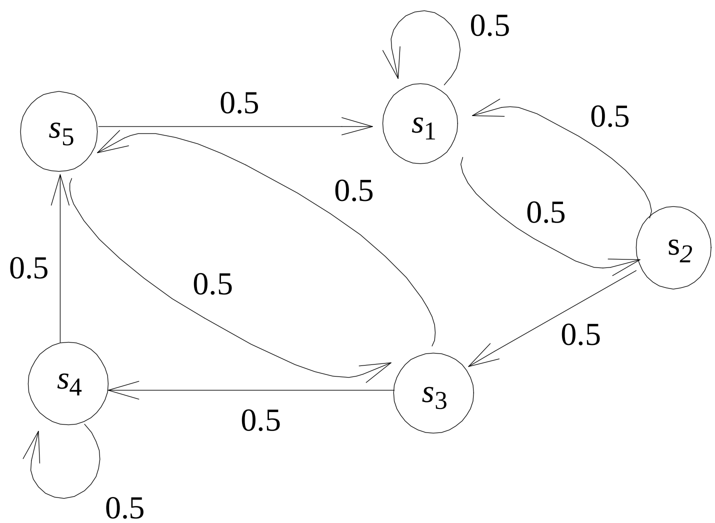

Fig.2: An example of a Markov chain

$$\tilde{d} = \frac{1}{3} \cdot 0.5 \cdot 1 + \frac{1}{6} \cdot (0.5 \cdot 1 + 0.5 \cdot 3) + \frac{1}{6} \cdot (0.5 \cdot 2 + 0.5 \cdot 1) + \frac{1}{6} \cdot 0.5 \cdot 3 + \frac{1}{6} \cdot (0.5 \cdot 2 + 0.5 \cdot 2)$$

$$= 1.33 \ \ \text{transitions/step}$$

This means that, for this particular encoding, an average number of 1.33 transitions is obtained when the graph in Fig.2 is traversed in a random manner. On the other hand, if we assume a Gray-type encoding given by: *code* (s_1) = ‘000’, *code* (s_2) = ‘001’, *code* (s_3) = ‘011’, *code* (s_4) = ‘111’, *code* (s_5) = ‘110’, the average number of transitions changes to:

$$\tilde{d} = \frac{1}{3} \cdot 0.5 \cdot 1 + \frac{1}{6} \cdot (0.5 \cdot 1 + 0.5 \cdot 1) + \frac{1}{6} \cdot (0.5 \cdot 1 + 0.5 \cdot 2) + \frac{1}{6} \cdot 0.5 \cdot 1 + \frac{1}{6} \cdot (0.5 \cdot 2 + 0.5 \cdot 2)$$

$$= 1 \ \ \text{transition/step}$$

Definition 2. (Average Hamming distance from state *i*) The average Hamming distance from state *i* is defined as:

$$\tilde{d}_i = \sum_{j=1}^{n} q_{ij} \cdot d_{ij} .$$

Example 2: For the Markov chain in Fig.2, for the first encoding we have:

$\tilde{d}_1 = 0.5 \cdot 1 = 0.5$, $\tilde{d}_2 = 0.5 \cdot 1 + 0.5 \cdot 3 = 2$, $\tilde{d}_3 = 0.5 \cdot 2 + 0.5 \cdot 1 = 1.5$,

$\tilde{d}_4 = 0.5 \cdot 3 = 1.5$, $\tilde{d}_5 = 0.5 \cdot 2 + 0.5 \cdot 2 = 2$,

whereas for the second encoding we have:

$\tilde{d}_1 = 0.5 \cdot 1 = 0.5$, $\tilde{d}_2 = 0.5 \cdot 1 + 0.5 \cdot 1 = 1$, $\tilde{d}_3 = 0.5 \cdot 1 + 0.5 \cdot 2 = 1.5$,

$\tilde{d}_4 = 0.5 \cdot 1 = 0.5$, $\tilde{d}_5 = 0.5 \cdot 2 + 0.5 \cdot 2 = 2$.

Note 1: Given a fixed length encoding of the states, there is a simple relationship between the average Hamming distance of a Markov chain and the average Hamming distance from a fixed state of the chain: $\tilde{d} = \sum_{i=1}^{n} p_i \cdot \tilde{d}_i$. Thus, if we find lower and upper bounds for $\tilde{d}_i$, we will also obtain lower and upper bounds for $\tilde{d}$.

To begin with, we give the following useful result:

Lemma 1. If $\{a_i\}_{1 \le i \le n}$ and $\{b_i\}_{1 \le i \le n}$ are two sets of positive real numbers, then the following holds:

$$\sum_{i=1}^{n} a_i \cdot b_i \le \sqrt{\sum_{i=1}^{n} a_i^2 \cdot \sum_{i=1}^{n} b_i^2}$$

with equality if and only if $a_i b_j = a_j b_i$ for any $i \neq j$.

Proof: It results from the inequality $\sum_{i=1}^{n} a_i^2 \cdot \sum_{i=1}^{n} b_i^2 - \left(\sum_{i=1}^{n} a_i \cdot b_i\right)^2 = \sum_{i \neq j} (a_i \cdot b_j - a_j \cdot b_i)^2 \ge 0$ with equality if and only if $a_i b_j = a_j b_i$ for any $i \neq j$.■

Using the above result, we present lower and upper bounds based on the information-theoretic concept of *informational energy* [15]:

Definition 3. (Informational energy) Given a discrete stochastic process characterized by steady state probabilities $\{p_i\}_{1 \le i \le n}$, the *informational energy* is defined as $E = \sum_{i=1}^{n} p_i^2$.

The above concept can be extended to the more general case of Markov chains as follows:

Definition 4. (Informational energy associated to a state) Given a Markov chain characterized by the stochastic matrix $Q = (q_{ij})_{1 \le i,j \le n}$, the *informational energy for state i* is defined as $E_i = \sum_{i=1}^{n} q_{ij}^2$.

Lemma 2. For a fixed k-bit encoding, the average Hamming distance from state i satisfies the following double inequality:

$$k \cdot \sum_{j \neq i} q_{ij} - \sqrt{(E_i - q_{ii}^2) \cdot \sum_{j \neq i} (k - d_{ij})^2} \leq \tilde{d}_i \leq \sqrt{(E_i - q_{ii}^2) \cdot \sum_{j \neq i} d_{ij}^2} \qquad (1)$$

with equality if and only if $q_{il}d_{ij} = q_{ij}d_{il}$ (for the upper bound) and $q_{il}(k - d_{ij}) = q_{ij}(k - d_{il})$ (for the lower bound) for any $l \neq j$, $l \neq i$, and $j \neq i$. $E_i = \sum_j q_{ij}^2$ represents the informational energy associated to state i.

Proof: It results directly from Lemma 1 applied to the sets $\{d_{ij}\}_{j \neq i}$ and $\{q_{ij}\}_{j \neq i}$ for a fixed i. ■

Lemma 2 basically breaks the bounds of the average Hamming distance from a given state into two terms: one characterized by the informational energy (that is, by the topology and parameters of the underlying Markov chain) and the other characterized by the actual Hamming distances given by the encoding. Since the informational energy is encoding-independent, we need to find lower and upper bounds on the encoding-dependent sums that appear in (1).

Lemma 3. Let t_i be the number of outgoing edges (excluding the self edges) from state i. For a fixed k-bit encoding, the following inequalities hold:

$$\sum_{j \neq i} (k - d_{ij})^2 \leq \sum_{l=1}^{m_i} (k - l)^2 \cdot \binom{k}{l} \text{ and } \sum_{j \neq i} d_{ij}^2 \leq \sum_{l=1}^{n_i} (k - l + 1)^2 \cdot \binom{k}{l-1}$$

where m_i, n_i are chosen such that $\sum_{l=1}^{m_i - 1} \binom{k}{l} < t_i \leq \sum_{l=1}^{m_i} \binom{k}{l}$ and $\sum_{l=1}^{n_i - 1} \binom{k}{l-1} < t_i \leq \sum_{l=1}^{n_i} \binom{k}{l-1}$.

Proof: For a fixed k-bit encoding and a given state, there is a number of $\binom{k}{1}$ codes at distance 1, a number of $\binom{k}{2}$ codes at distance 2, etc.[1]. Thus, assuming that there are t_i outgoing transitions from state i, to obtain an upper bound for the two sums, we need to assign the Hamming distances in decreasing (increasing) order until all edges are exhausted. Hence, the above inequalities are satisfied.■

Lemmas 2 and 3 can be further combined to compute encoding-independent lower and upper bounds for the average Hamming distance of a Markov chain:

Theorem 1. For a fixed k-bit encoding, the average Hamming distance of a Markov chain satisfies

1.The same observation has been used in [11].

$$\sum_i p_i \cdot \tilde{d}_{i,\,min} \le \tilde{d} \le \sum_i p_i \cdot \tilde{d}_{i,\,max}$$

where $\tilde{d}_{i,\,min} = k \cdot \sum_{j \neq i} q_{ij} - \sqrt{(E_i - q_{ii}^2) \cdot \sum_{l=1}^{m_i} (k-l)^2 \cdot \binom{k}{l}}$,

$\tilde{d}_{i,\,max} = \sqrt{(E_i - q_{ii}^2) \cdot \sum_{l=1}^{n_i} (k-l+1)^2 \cdot \binom{k}{l-1}}$, and the notation is the same as in Lemma 3. The complexity of computing these bounds is $O(t)$ where $t = \sum_{i=1}^{n} t_i$ is the total number of transitions (excluding self transitions) and n is the number of states in the underlying Markov chain.

Proof: It results immediately from Lemma 2, Lemma 3 and Definition 2.■

Example 3: For the Markov chain in Fig.2, we have $E_1 - q_{11}^2 = 0.25$, $E_2 - q_{22}^2 = 0.5$, $E_3 - q_{33}^2 = 0.5$, $E_4 - q_4^2 = 0.25$, and $E_5 - q_{55}^2 = 0.5$. Thus, assuming that we target a $k = 3$ bit encoding, the bounds for the average Hamming distance from each state are:

$$0.5 = 3 \cdot 0.5 - \sqrt{0.25 \cdot (3-1)^2} \le \tilde{d}_1 \le \sqrt{0.25 \cdot 3^2} = 1.5,$$

$$1 = 3 \cdot 1 - \sqrt{0.5 \cdot [(3-1)^2 + (3-1)^2]} \le \tilde{d}_2 \le \sqrt{0.5 \cdot (3^2 + 2^2)} = 2.55,$$

$$1 = 3 \cdot 1 - \sqrt{0.5 \cdot [(3-1)^2 + (3-1)^2]} \le \tilde{d}_3 \le \sqrt{0.5 \cdot (3^2 + 2^2)} = 2.55,$$

$$0.5 = 3 \cdot 0.5 - \sqrt{0.25 \cdot (3-1)^2} \le \tilde{d}_4 \le \sqrt{0.25 \cdot 3^2} = 1.5,$$

$$1 = 3 \cdot 1 - \sqrt{0.5 \cdot [(3-1)^2 + (3-1)^2]} \le \tilde{d}_5 \le \sqrt{0.5 \cdot (3^2 + 2^2)} = 2.55.$$

Based on these, we get the following bounds for the average Hamming distance of the chain:

$$0.75 = \frac{1}{3} \cdot 0.5 + \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 0.5 + \frac{1}{6} \cdot 1 \le \tilde{d}$$
$$\le \frac{1}{3} \cdot 1.5 + \frac{1}{6} \cdot 2.55 + \frac{1}{6} \cdot 2.55 + \frac{1}{6} \cdot 1.5 + \frac{1}{6} \cdot 2.55 = 2.03$$

Note 2: The bounds are achieved exactly if conditions in Lemma 2 are met for every state i and the topology of the Markov chain permits the assignment of Hamming distances as in Lemma 3.

The bounds presented in this section are easy to compute and give an interesting insight into the relationship between the informational energy associated with a Markov chain and the possible values for the average Hamming distance for a fixed length encoding. However, as we shall see later in this paper, these bounds are not tight enough. We present in the next section an alternative way to derive tighter bounds for the average Hamming distance.

3. Combinatorial lower and upper bounds on FSM switching

In this section, we prove formally that the bounds we are about to present are indeed tighter than the information theoretic bounds presented in the previous section and, in addition, better than the *simple bound* $\sum_{i \neq j} p_i \cdot q_{ij}$ used in [11]. The notation is the same as in the previous section. We first give the following result:

Lemma 4. Assuming that $\{q_{ij}\}_{1 \le j \le n}$ is sorted in non-increasing order, for a fixed k-bit encoding, the average Hamming distance from state i satisfies the following double inequality:

$$\sum_{l=1}^{m_i} l \cdot \sum_{j=a_l}^{a_{l+1}-1} q_{ij} \le \tilde{d}_i \le \sum_{l=1}^{n_i} (k-l+1) \cdot \sum_{j=b_l}^{b_{l+1}-1} q_{ij} \qquad (2)$$

where $a_l = 1 + \sum_{m=1}^{l-1} \binom{k}{m}$, $b_l = 1 + \sum_{m=1}^{l-1} \binom{k}{m-1}$, and m_i, n_i are as in Lemma 3.

Proof: Because $\{q_{ij}\}_{1 \le j \le n}$ is sorted in non-increasing order, by assigning the codes in increasing (decreasing) order of their Hamming distances, we obtain a lower (upper) bound for the average Hamming distance from state i. Indeed, knowing that there are exactly $\binom{k}{l}$ codes at distance l, assigning in a greedy fashion the lowest (highest) Hamming distance to the edges with the highest probability, we get that:

$$q_{i1} \cdot 1 + \dots + q_{i\binom{k}{1}} \cdot 1 + q_{i,\binom{k}{1}+1} \cdot 2 + \dots + q_{i,\binom{k}{1}+\binom{k}{2}} \cdot 2 + \dots \le \tilde{d}_i = \sum_{j \neq i} q_{ij} \cdot d_{ij} \text{ and}$$

$$q_{i1} \cdot k + q_{i,\binom{k}{0}+1} \cdot (k-1) + \dots + q_{i,\binom{k}{1}+\binom{k}{2}} \cdot (k-1) + \dots \ge \tilde{d}_i = \sum_{j \neq i} q_{ij} \cdot d_{ij}.$$

We define the series $a_l = 1 + \sum_{m=1}^{l-1} \binom{k}{m}$, $b_l = 1 + \sum_{m=1}^{l-1} \binom{k}{m-1}$, and take m_i, n_i such that $a_{m_i - 1} \le t_i \le a_{m_i}$ and $b_{n_i - 1} \le t_i \le b_{n_i}$. In other words, for the first inequality, if the outgoing transitions are sorted according to their probabilities, then the first a_1 edges are assigned a distance of 1, the next a_2 - a_1 are assigned a distance of 2 etc., until all t_i edges are exhausted. A similar rationale is used for the second inequality. Thus, the lower and upper bounds in (2) are satisfied. ■

Note 3: The bounds given in (2) are tight, that is, for a fixed state *i*, we can *always* find an encoding which achieves these bounds. For instance, for state s_2 in Fig.2, assuming that *code* (s_2) = '001' is fixed, then it is sufficient to consider *code* (s_1) = '000' and *code* (s_3) = '011' to achieve the lower bound of $\tilde{d}_2 = 0.5 \cdot 1 + 0.5 \cdot 1 = 1$ transition/step. On the other hand, if we consider the same code for s_2, *code* (s_1) = '110' and *code* (s_3) = '111', we achieve the upper bound of $\tilde{d}_2 = 0.5 \cdot 3 + 0.5 \cdot 2 = 2.5$ transitions/step.

Based on the above lemma, we give the following result which holds for any *k*-bit encoding of the states of a Markov chain:

Theorem 2. For a fixed *k*-bit encoding, the average Hamming distance of a Markov chain satisfies the following:

$$\sum_{i=1}^{n} p_i \cdot \sum_{l=1}^{m_i} l \cdot \sum_{j=a_l}^{a_{l+1}-1} q_{ij} \le \tilde{d} \le \sum_{i=1}^{n} p_i \cdot \sum_{l=1}^{n_i} (k-l+1) \cdot \sum_{j=b_l}^{b_{l+1}-1} q_{ij} \qquad (3)$$

where the notation is the same as in Lemma 4. The complexity of computing these bounds is $O(t \cdot \log n)$, where $t = \sum_{i=1}^{n} t_i$ is the total number of transitions (excluding self transitions) and *n* is the number of states in the underlying Markov chain.

Proof: Since $\tilde{d} = \sum_{i=1}^{n} p_i \cdot \tilde{d}_i$ and thus, using the result in Lemma 4 we get exactly the above claim.

As far as the complexity is concerned, we note that the critical step in computing the above bounds is sorting the outgoing transitions according to their probability. For a state with t_i outgoing transitions, this takes O(t_i log t_i). Since $t_i \le n$, doing this for all the states in the Markov chain cannot take more than O(*t* log *n*). ■

Example 4: For the Markov chain in Fig.2, assuming a 3-bit encoding, the inequalities in Lemma 4 and Theorem 2 may be written as follows:

$0.5 = 1 \cdot 0.5 \le \tilde{d}_1 \le 3 \cdot 0.5 = 1.5$, $1 = 1 \cdot 0.5 + 1 \cdot 0.5 \le \tilde{d}_2 \le 3 \cdot 0.5 + 2 \cdot 0.5 = 2.5$,

$1 = 1 \cdot 0.5 + 1 \cdot 0.5 \le \tilde{d}_3 \le 3 \cdot 0.5 + 2 \cdot 0.5 = 2.5$, $0.5 = 1 \cdot 0.5 \le \tilde{d}_4 \le 3 \cdot 0.5 = 1.5$,

$1 = 1 \cdot 0.5 + 1 \cdot 0.5 \le \tilde{d}_5 \le 3 \cdot 0.5 + 2 \cdot 0.5 = 2.5$ and hence, we get the bounds for $\tilde{d}$:

$$0.75 = \frac{1}{3} \cdot 0.5 + \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 0.5 + \frac{1}{6} \cdot 1 \le \tilde{d}$$
$$\le \frac{1}{3} \cdot 1.5 + \frac{1}{6} \cdot 2.5 + \frac{1}{6} \cdot 2.5 + \frac{1}{6} \cdot 1.5 + \frac{1}{6} \cdot 2.5 = 2$$

Note 4: The bounds in Lemma 4 are always achievable if each state i is considered in isolation. However, the bounds in Theorem 2 may not be achievable for the whole graph due to the constraints which result from the specific structure of the Markov chain. For example, for the Markov chain in Fig.2, if *code* (s_3) = ‘011’, to achieve the lower bound for state s_3, we should have *code* (s_4) = ‘111’ and *code* (s_5) = ‘010’. But this encoding does *not* achieve the lower bound for state s_4 which requires s_5 to be at a Hamming distance of 1.

Corollary 1. The bounds in Theorem 2 are always tighter than the bounds from Theorem 1.

Proof: Using Lemma 1, we have the following inequalities for a fixed state i:

$$\sum_{l=1}^{n_i} (k-l+1) \cdot \sum_{j=b_l}^{b_{l+1}-1} q_{ij}$$

$$= q_{i1} \cdot k + q_{i,\binom{k}{0}+1} \cdot (k-1) + \ldots + q_{i,\binom{k}{1}+\binom{k}{2}} \cdot (k-1) + \ldots$$

$$\le \sqrt{(E_i - q_{ii}^2) \cdot \sum_{l=1}^{n_i} (k-l+1)^2 \cdot \binom{k}{l-1}}$$

and

$$\sum_{l=1}^{m_i} l \cdot \sum_{j=a_l}^{a_{l+1}-1} q_{ij} = k \cdot \sum_{j \ne i} q_{ij} - \sum_{l=1}^{m_i} (k-l) \cdot \sum_{j=a_l}^{a_{l+1}-1} q_{ij} = k \cdot \sum_{j \ne i} q_{ij}$$

$$-\left(q_{i1} \cdot 1 + \ldots + q_{i\binom{k}{1}} \cdot 1 + q_{i,\binom{k}{1}+1} \cdot 2 + \ldots + q_{i,\binom{k}{1}+\binom{k}{2}} \cdot 2 + \ldots\right)$$

$$\ge k \cdot \sum_{j \ne i} q_{ij} - \sqrt{(E_i - q_{ii}^2) \cdot \sum_{l=1}^{m_i} (k-l)^2 \cdot \binom{k}{l}}$$

Thus, the bounds in Theorem 2 are tighter than the bounds given in Theorem 1.■

As we can see, we can trade-off the quality of the bounds versus the time spent for their computation. The informational energy-based bounds are easier to compute (in about O(t) time), but they are not as tight as the bounds given in Theorem 2 (computed in O($t\log n$) time).

Note 5: The lower bound derived in Theorem 2 is always better than the *simple bound* $\sum_{i \neq j} p_i \cdot q_{ij}$ used in [11], which is not achievable unless each transition has a Hamming distance of 1. Indeed, we have the following:

Corollary 2. The simple lower bound $\sum_{i \neq j} p_i \cdot q_{ij}$ and the lower bound in Theorem 2 satisfy:

$$\sum_{i \neq j} p_i \cdot q_{ij} \le \sum_{i=1}^{n} p_i \cdot \sum_{l=1}^{m_i} l \cdot \sum_{j=a_l}^{a_{l+1}-1} q_{ij} \le \tilde{d}$$

that is, the lower bound in Theorem 2 is always tighter than the simple lower bound $\sum_{i \neq j} p_i \cdot q_{ij}$.

Proof: It is sufficient to observe that $\sum_{l=1}^{m_i} l \cdot \sum_{j=a_l}^{a_{l+1}-1} q_{ij} \ge \sum_{l=1}^{m_i} 1 \cdot \sum_{j=a_l}^{a_{l+1}-1} q_{ij} = \sum_{j \neq i} q_{ij}$ and thus the above claim is satisfied. ■

Differently stated, although our lower bound may not be achievable, it is still tighter than the simple bound. As we shall see in the experimental part, in most of the cases, the global and local lower bounds given in [11] are worse (i.e. *looser*) than the simple bound, and hence, are worse than our lower bound.

4. An application of switching bounds in power estimation of sequential circuits

In this section, we will show how our theoretical bounds can be used in a high-level power estimation framework for sequential circuits. We assume that the behavior of the target circuit is specified as a STG and the state encoding (and therefore the final implementation) has not been determined yet.

Having computed lower and upper bounds for the switching activity, in order to be useful for high-level power analysis, they have to be used in conjunction with an estimator for the area (complexity) of the target circuit. We note that, since we want to keep the independence from the actual implementation of the circuit, the approaches described in the introduction [16-19] are not directly applicable in our case because the state encoding is unknown. Thus, we need an *encoding-independent* measure for the complexity of the circuit. For this reason, in this paper we use the following *piecewise linear* model for the power consumption of a sequential circuit:

$$P = (\alpha \cdot SW_i + \beta \cdot SW_s + \gamma \cdot SW_o) \cdot \#SPI \cdot V_{dd}^2 \cdot f \qquad (4)$$

where SW_i, SW_s, SW_o are the switching activities (or, equivalently, the average Hamming distances) of the primary inputs, state lines and primary outputs, respectively. The term #*SPI* represents the minimum number of symbolic prime implicants needed in a two-level implementation of the sequential circuit. The terms V_{dd} and f represent the voltage supply and the clock frequency, respectively.

Since we do not know the final state assignment and basically need only a rough measure for the complexity of the target circuit, we use the minimum number of prime implicants from a symbolic cover of the FSM (#*SPI*) [20, 21]. To obtain this number, given a FSM, we first assign one-hot codes to all the states. Then symbolic minimization is applied on the one-hot coded machine using *multi-valued logic minimization*. The result is a symbolic cover of the FSM. Each element of the symbolic cover is a *symbolic prime implicant*, that is a 4-tuple (x, S, S', y) where S is the set of states which transit to the same next state S' and assert the same output y when the input combination is x. The set of symbolic prime implicants has also been used in [3] for finding an optimal state assignment targeting low-power design of sequential circuits.

The number of symbolic prime implicants in the minimal symbolic cover (given by #*SPI*) is an *upper bound* on the size of the minimal boolean cover of the target FSM [20] and can be used in conjunction with the switching activity information to derive power values. In general, the coefficients α, β, γ of the piecewise linear model in (4) depend on #*SPI*, SW_i, SW_s, and SW_o. To find these coefficients, we resort to least mean square regression techniques [22].

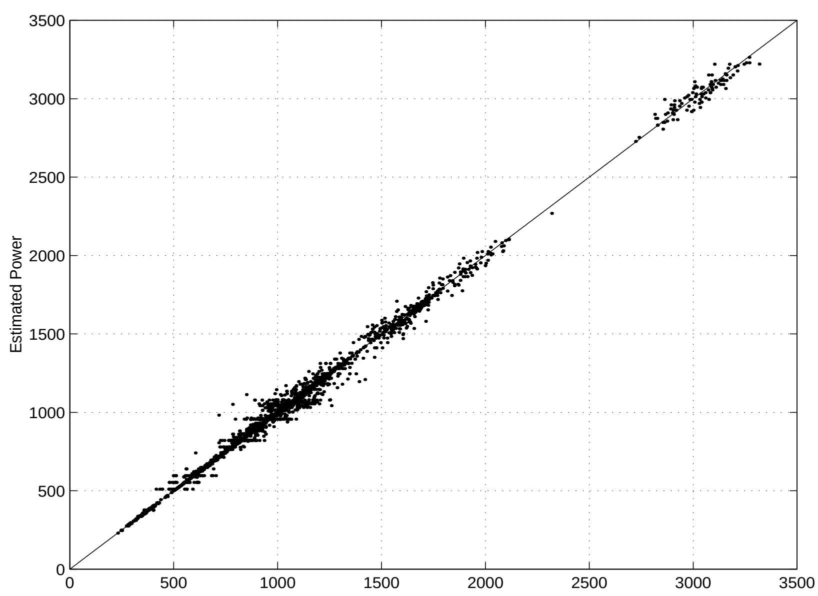

Fig.3: Estimated vs. exact power for *mcnc'91* benchmarks

We present in Fig.3 a comparison between the exact total power consumption and the total power consumption obtained using the model in (4) for 100 random implementations of each benchmark circuit from the *mcnc'91* suite. As we can see, there is a very good match between the exact and estimated values of total power and, on average, the error is 2.36% (with a maximum of 36.44%).

It should be pointed out that, typically, the coefficient corresponding to the state lines (β) is *one order of magnitude larger* than the other two coefficients. As a consequence, it is expected to have a very strong dependence between the total power values and the switching activity (or average Hamming distance) on the state lines. From this perspective, it is clearly important to have tight lower and upper bounds for the average Hamming distance of the state lines.

5. Experimental results

In this section we provide our experimental results for a subset of *mcnc'91* benchmark suite. In particular, we are interested in assessing the effectiveness of the proposed lower and upper bounds in FSM switching activity analysis and how these bounds can be used in a high-level power estimation framework. To this end, we target three sets of experiments:

a) The first set of experiments shows a comparison between different theoretical bounds for average Hamming distance (Table 1). To do this, based on the STG of each circuit, we extract the values for the conditional and steady state probabilities for the underlying Markov chain. Then, we compute our values for the minimum and maximum switching activity considering a minimal length encoding (i.e., $\log n$ bits where n is the total number of states reachable). We give in Table 1 the values for the lower and upper bounds based on informational energy (columns 2, 3) and the tighter bounds computed as in Section 3 (columns 4, 5). For comparison, we also provide the simple lower bound (column 6) and the local and global lower bounds computed as in [11] (columns 7, 8).

As we can see in Table 1, our best lower bound (column 4) is larger (that is, *tighter*) than the simple bound (column 5) which assumes that every edge is assigned a Hamming distance of 1. We also note that in some cases, the lower bound based on informational energy (column 2) is also better than the simple lower bound. Moreover, the lower bounds computed as in [1] are in all cases lower (that is, *looser*) than our proposed lower bounds and also the simple lower bound. The reason is that the entropy factor in local and global approaches from [11] captures the dynamic information in an STG, whereas K[1], roughly speaking, captures a worst-case "static" measure. When the two differ a lot (structure of the static graph and the dynamic use of it), negative values are possible. Generally, this is when the base switching is likely to do better [23].

1. K has been defined in [11] as $\sum_{1 \le i, j \le n} \frac{1}{2^{d_{ij}}}$.

Table 1: Comparison of different theoretical bounds for the average Hamming distance

Circuit	LB (inf)	UB (inf)	LB (comb)	UB (comb)	LB (simple)	LB (local) [11]	LB (global) [11]
bbara	0.1854	0.8003	0.2228	0.7851	0.2228	-0.3601	-0.9743
bbsse	0.7336	1.2667	0.7402	1.2441	0.7008	0.4584	0.0921
bbtas	0.3986	1.2983	0.4418	1.2799	0.4418	0.0030	-0.1276
beecount	0.7143	1.3089	0.7143	1.2857	0.7143	0.2905	0.0421
cse	0.2768	0.7719	0.2834	0.6613	0.2519	-0.5166	-1.4496
dk14	1.0000	2.7425	1.0000	2.7142	1.0000	0.1106	0.0669
dk16	0.9524	4.1454	0.9624	4.1238	0.9624	0.4127	0.2323
dk17	0.7777	2.2234	0.8428	2.1853	0.8408	0.2595	-0.3570
dk27	1.0000	2.5926	1.0000	2.5497	1.0000	0.1749	0.0413
dk512	1.0000	3.5356	1.0000	3.5027	1.0000	0.1666	-0.0552
ex1	0.5901	2.8842	0.7651	2.7727	0.7651	0.1432	-0.3522
ex2	0.7161	4.7164	1.0000	4.5528	1.0000	0.1586	-1.3630
ex3	0.8507	3.5243	1.0000	3.4604	1.0000	0.2931	-0.2876
ex4	0.8394	3.6825	0.9048	3.5872	0.9048	0.0133	-0.1497
ex5	0.7017	1.8275	0.7547	1.7981	0.7119	0.3516	-0.3465
ex7	0.8173	2.4667	0.9467	2.4210	0.9467	0.2887	-0.1884
keyb	0.3723	1.2939	0.4488	1.2441	0.4488	-0.1087	-0.9561
kirkman	0.7355	2.8026	0.7533	2.7922	0.7533	0.1550	-1.2543
mark1	0.7706	2.1092	0.7742	2.0968	0.7742	0.1830	-0.2352
mc	0.5714	1.1429	0.5714	1.1429	0.5714	0.0350	-0.1496
planet	0.7971	6.0511	0.9999	5.8889	0.9999	-0.0282	-0.5198
s1	0.1716	4.2261	0.7961	3.7175	0.7961	0.0420	-0.3045
sand	0.4375	2.2845	0.4947	2.2433	0.4718	0.1297	-0.4265
sse	0.7336	1.2667	0.7402	1.2441	0.7008	0.4584	0.0921
tav	1.0000	2.0000	1.0000	2.0000	1.0000	0.0000	0.0000
tbk	0.5056	1.6028	0.5238	1.5873	0.4603	-0.1148	-0.8332
train11	0.5860	2.1115	0.6029	2.0931	0.6029	0.3096	-0.4764
train4	0.5246	0.9494	0.5296	0.9357	0.5296	0.3779	0.1558

b) Second, to see how these bounds change when the number of bits used for encoding varies, we chose a typical example (circuit *tbk*) and computed the lower and upper bounds for different encoding lengths (Fig.4). For this set of experiments, the bounds were computed as in Section 3.

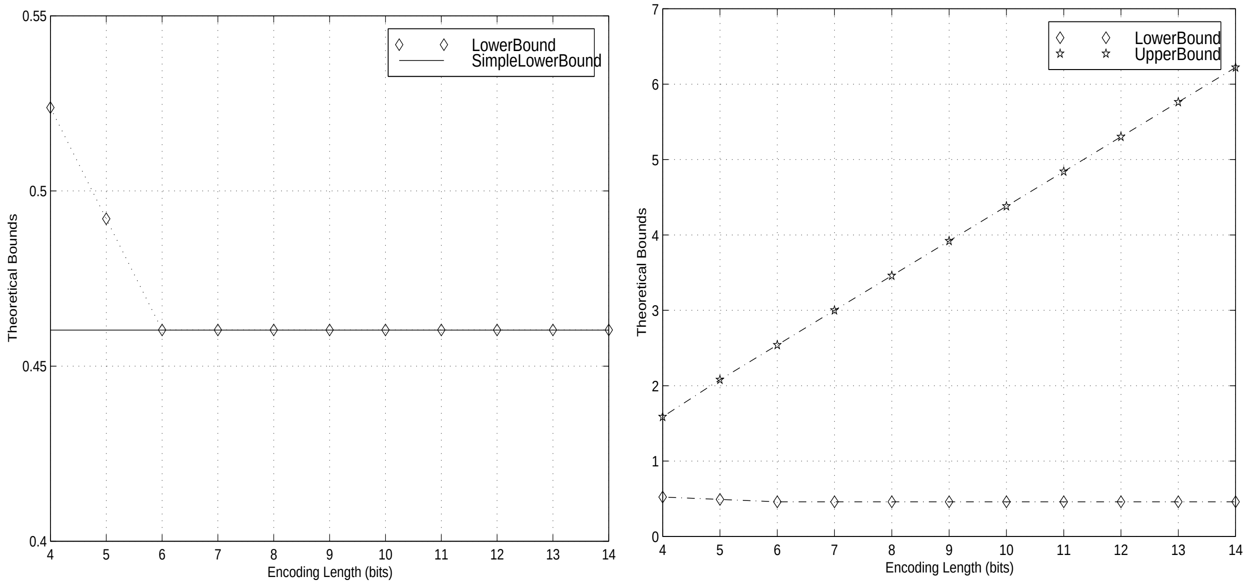

Fig.4: A typical case (circuit *tbk*)

As we can see in the first graph, the lower bound reaches the simple lower bound value after the encoding length reaches 6 bits, while the upper bound increases linearly with the encoding length (second graph). Although these bounds may not be achievable, we can draw the conclusion that increasing the number of bits used for encoding beyond some limit will bring only marginal reductions in the lower bound for switching activity on the state lines (and thus in the minimum achievable total power consumption).

c) The third set of experiments illustrates how the theoretical results from Section 3 can be applied in a high-level power estimation environment (Table 2). For the same set of circuits that was used in Section 4, to derive the power estimation model, we compute the number of symbolic prime implicants of a minimal cover (*#SPI)* using multiple-valued logic minimization. In addition, using the STG information and the structure of the underlying Markov chain, we compute, as in Section 3, the lower and upper bounds of the switching activity on the state lines. Then, based on the model proposed in Section 4 and using the theoretical bounds from Section 3 (columns 2, 3), we derived lower and upper bounds for the total power consumption of each circuit. We also provide the minimum and maximum power values obtained over 100 implementations with random state encodings which use the same number of bits (columns 4, 5). The power values were obtained using an in-house gate level power simulator developed under SIS.

Table 2: Bounds for total power values (μW @ 20MHz and V_{dd} = 5V)

Circuit	LB (comb)	UB (comb)	LB (rand)	UB (rand)
bbara	260.61	485.03	305.68	425.26
bbsse	795.19	1477.70	824.85	1372.41
bbtas	230.50	555.78	231.39	470.48
beecount	488.91	575.16	415.73	702.62
cse	579.38	821.84	604.71	782.47
dk14	717.21	965.27	721.53	943.10
dk16	557.88	1924.70	1133.90	1348.66
dk17	447.99	991.21	558.18	879.20
dk27	436.46	1024.40	561.07	895.71
dk512	500.83	1483.90	737.14	1156.16
ex1	1208.00	2722.60	1572.66	2319.51
ex2	534.61	2037.00	823.13	1621.94
ex3	508.51	1499.30	789.77	1205.71
ex4	713.14	1801.80	852.72	1466.09
ex5	375.67	786.18	457.85	679.09
ex7	488.24	1476.00	695.79	1262.40
keyb	597.78	1351.50	595.75	1309.30
kirkman	719.02	1667.80	786.89	1453.06
mark1	748.94	1825.70	812.32	1662.99
mc	284.09	469.45	281.79	401.79
planet	1893.20	4577.80	2723.32	3318.19
s1	1282.40	1853.80	1390.52	1713.35
sand	1290.90	2188.70	1555.95	1776.51
sse	795.18	1477.70	824.85	1372.41
tav	525.23	945.23	525.23	735.23
tbk	951.47	962.90	796.02	1089.03
train11	315.78	894.36	488.74	667.98
train4	237.97	362.74	247.50	347.81

As we can see, in most cases, the maximum and minimum power values over all random implementations is within the bounds we proposed. There are, however, 4 instances in which the bounds are violated and these correspond to the cases when the power model given in equation (4) of Section 4 gives the highest errors for estimated power values. For these cases, the bounds for total power values are violated by 12.38% on average.

6. Conclusion

In this paper, we have presented lower and upper bounds for the average switching activity on the state lines in FSMs. As the main theoretical contribution, we improve over the previous work by providing a tighter lower bound and a new upper bound for the average Hamming distance between the states of the target FSM. Our theoretical bounds are *encoding-independent,* and therefore can be used in a high-level power estimation framework to provide an early indication about the performance limits of the target FSM. Preliminary results are encouraging and show the effectiveness of using these bounds for estimating the range for total power values.

References

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

[2] K, Roy and S.C. Prasad, ‘Syclop: Synthesis of CMOS Logic for Low Power Application,’ in *Proc. Intl. Conf on Computer Design*, pp. 464-467, Oct. 1992.

[3] C.Y. Tsui, M. Pedram, C. Chen, and A.M. Despain, ‘Low Power State Assignment Targeting Two- and Multi-level Logic Implementations,’ in *Proc. ACM/IEEE Intl. Conf. on Computer-Aided Design,* pp. 82-87, Nov. 1994.

[4] L. Benini and G. De Micheli, ‘State Assignment for Low Power Dissipation’, in *IEEE Journal of Solid State Circuits*, vol.30, no.3, 1995.

[5] G. Hachtel, M. Hermida, A. Pardo, M. Poncino and F. Somenzi, ‘Reencoding sequential circuits to reduce power dissipation,’ in *Proc. ACM/IEEE Intl. Conf. on Computer-Aided Design*, pp. 70-73, Nov. 1994.

[6] B. Kumthekar, I.H. Moon, and F. Somenzi, ‘A Symbolic Algorithm for Low-Power Sequential Synthesis,’ in *Proc. Intl. Symposium on Low-Power Electronics and Design*, Aug. 1997.

[7] M.R. Stan and W.P. Burleson, ‘Two-Dimensional Codes for Low-Power,’ in *Proc. Intl. Symposium on Low-Power Electronics and Design*, pp. 335-340, Aug. 1996.

[8] M.R. Stan and W.P. Burleson, ‘Bus-Invert Coding for Low-Power I/O,’ in *IEEE Trans. on VLSI Systems*, vol.3, no.1, March 1995.

[9] M.R. Stan and W.P. Burleson, ‘Limited-Weight Codes for Low-Power I/O,’ in *Proc. Intl. Workshop on Low Power Design*, pp. 209-214, April 1994.

[10] C.L. Su, C.Y. Tsui, and A.M. Despain, ‘Saving Power in the Control Path of Embedded Processors,’ in *IEEE Design and Test of Computers*, vol.11, no.4, Dec. 1994.

[11] A. Tyagi, ‘Entropic Bounds on FSM Switching,’ in *IEEE Trans. on VLSI Systems*, vol.5, no.4, Dec. 1997.

[12] S. Ramprasad, N.R. Shanbhag, and I.N. Hajj, ‘Achievable Bounds on Signal Transition Activity,’ in *Proc. ACM/IEEE Intl. Conf. on Computer-Aided Design*, Nov. 1997.

[13] G. Hachtel, E. Macii, A. Pardo, and F. Somenzi, ‘Markovian Analysis of Large Finite State Machines,’ in *IEEE Trans. on CAD of Integrated Circuits*, vol.15, no.12, Dec. 1996.

[14] D. Marculescu, R. Marculescu, and M. Pedram, ‘Trace-Driven Steady-State Probability Estimation with Application to Power Estimation,’ to appear in *Proc. Design Automation and Test in Europe Conf.*, Feb. 1998.

[15] D. Marculescu, R. Marculescu, and M. Pedram, ‘Information Theoretic Measures for Power Analysis,’ in *IEEE Trans. on CAD of Integrated Circuits*, vol.15, no.6, June 1996.

[16] P.E. Landman and J.M. Rabaey, ‘Activity-Sensitive Architectural Power Analysis for the Control Path,’ in *Proc. Intl. Symposium on Low-Power Design*, pp. 93-98, April 1995.

[17[M. Nemani and F.N. Najm, 'High-Level Area and Power Estimation for VLSI Circuits,' in *Proc. ACM/ IEEE Intl. Conf. on Computer-Aided Design*, pp. 114-119, Nov. 1997.

[18] F. Ferrandi, F. Fummi, E. Macii, M. Poncino, and D. Sciuto, 'Power Estimation of Behavioral VHDL Descriptions,' in to appear in *Proc. Design Automation and Test in Europe Conf.*, Feb. 1998.

[19] M. Nemani and F.N. Najm, 'High-Level Area Prediction for Power Estimation,' in *Proc*. *Custom Integrated Circuits conference*, pp. 483-486, May 1997.

[20] G. De Micheli, R.K. Brayton, and A. Sangiovanni-Vincentelli, 'Optimal State Assignment for Finite State Machines,' in *IEEE Trans*. *on CAD of Integrated Circuits*, vol.4, no.7, July 1985.

[21] T. Villa and A. Sangiovanni-Vincentelli, 'NOVA: State Assignment of Finite State Machines for Optimal Two-Level Logic Implementation,' in *IEEE Trans. on CAD of Integrated Circuits*, vol.9, no.9, Sept. 1990.

[22] J. Neter, M.K. Kutner, C.J. Nachtsheim, and W. Wasserman, 'Applied Linear Regression Models,' Irwin, Burr Ridge, Illinois, 1996.

[23] A. Tyagi, Personal communication, Feb. 1998.