Information-Theoretic Bounds for Switching Activity Analysis in Finite-State Machines under Temporally Correlated Inputs

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

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

Abstract - *This paper presents information theoretic bounds for Finite-State Machine (FSM) switching activity under high-order temporal correlations on the primary inputs. As the main theoretical contributions: 1) we extend the previous work done on switching activity bounds calculation to account for complex spatiotemporal effects which are present when the target machine models real hardware and receives data from real applications; 2) we show that, for the special case of unifilar FSMs, the order and probabilities characterizing the Markov chain on the state lines of the FSM can be easily deduced and the case of general FSMs can be reduced to this simple scenario. As the experimental results show, the lower and upper bounds are very sensitive to the input environment and if appropriately derived, provide a reliable measure of the switching activity on FSM state lines.*

1. Introduction

Since power consumption has become a critical issue in the design of digital systems, tools that can control the power budget during the various phases of the design process are in high demand [1]. Synthesis systems typically take a hardware description language model of a design as the initial input and then the synthesis path follows several steps: *high-level synthesis*, *state assignment*, *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 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. Much less work has been done to find good theoretical bounds for the average switching activity in FSMs or on buses. Preliminary results in this direction have been recently reported in [7-9].

The subject of this paper is the derivation of *theoretical bounds* for FSM switching activity under complex temporal correlations on primary inputs. 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 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.

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 target in this paper is to provide lower and upper bounds for the average Hamming distance on the state lines which are valid *regardless* of the state encoding used in the final implementation. 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 in [7]. 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 [9], 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 limited by the fact that they are achievable merely through *compression* techniques rather than *encoding* techniques (which is the case for FSMs). In [8], tight theoretical lower and upper bounds have been derived for FSM switching and their use in a high-level power model has also been demonstrated.

In this paper we improve over the work in [8] by showing that the effect of complex temporal correlations on the primary inputs cannot be neglected. Based on the new concept of *unifilar* FSM, we provide an efficient technique for determining the exact order and probabilities characterizing the Markov chain on the state lines of the FSM.

The paper is organized as follows. In Section 2, we present an overview of the theoretical lower and upper bounds, the complexity of their calculation and discuss their relationship with other bounds. In Sections 3 and 4, we discuss the effect of temporal correlations on the accuracy of the bounds and provide an efficient method to capture them using *unifilar* FSMs. The experimental results are given in Section 5. Section 6 concludes the paper and summarizes our main contribution.

2. Informational and combinatorial bounds on FSM switching

In this section, we summarize some of the previous work on theoretical bounds for FSM switching activity [8]. 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 of 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 in 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 [10] (applicable only to uncorrelated inputs) or [11] (applicable to any high-order Markov source at the primary inputs). Alternatively, having the STG of the FSM, a 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. In Sections 3-4, we will also present an alternative method to derive the order and probabilities characterizing the Markov chain of the state lines of the FSM.

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 d_{ij} is the Hamming distance between the codes assigned to states i and j.

Definition 2. [12] (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 3. (Informational energy of a state) Given a Markov chain characterized by the stochastic matrix $Q = (q_{ij})_{1 \le i,j \le n}$, the *informational energy of state* i is defined as $E_i = \sum_{j=1}^{n} q_{ij}^2$.

Based on this information-theoretic measure, we can compute *encoding-independent* lower and upper bounds for the average Hamming distance of a Markov chain:

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

$$\sum_i p_i \cdot \tilde{d}_{i,\,min} \le \tilde{d} \le \sum_i p_i \cdot \tilde{d}_{i,\,max} \tag{1}$$

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

$\tilde{d}_{i,\,max} = \sqrt{(E_i - q_{ii}^2) \cdot \sum_{l=1}^{n_i} (k-l+1)^2 \cdot \binom{k}{l-1}}$, and m_i, n_i are chosen such that:

$$\sum_{l=1}^{m_i - 1} \binom{k}{l} < t_i \le \sum_{l=1}^{m_i} \binom{k}{l} \text{ and } \sum_{l=1}^{n_i - 1} \binom{k}{l-1} < t_i \le \sum_{l=1}^{n_i} \binom{k}{l-1}.$$

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. ■

The bounds presented above 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 of the average Hamming distance for a fixed length encoding. An alternative way to derive tighter bounds is based on the *topological* properties of the FSM. In addition, the following bounds are better than the *simple bound* $\sum_{i \ne j} p_i \cdot q_{ij}$ used in [2].

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

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

$$\sum_{i=1}^{n} p_i \sum_{l=1}^{m_i} l \sum_{j=a_l}^{a_{l+1}-1} q_{ij} \le \tilde{d} \le \sum_{i=1}^{n} p_i \sum_{l=1}^{n_i} (k-l+1) \sum_{j=b_l}^{b_{l+1}-1} q_{ij} \quad (2)$$

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 Th.1. 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. ■

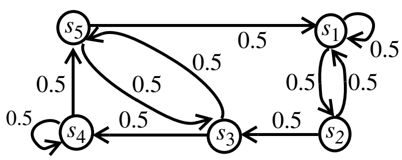


Fig.1: A simple Markov chain

Note 1: The bounds in Th.2 may not be achievable for the whole graph due to the constraints which result from the specific structure of the Markov chain. In Fig.1, if *code* (s_3) = '011', to achieve the lower bound for state s_3, we must 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 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 O(t) time), but they are not as tight as the bounds given in Th.2 (computed in O($t$log$n$) time).

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

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

$$\sum_{i \ne j} p_i \cdot q_{ij} \le \sum_{i=1}^{n} p_i \cdot \sum_{l=1}^{m_i} l \cdot \sum_{j=a_l}^{a_{l+1}-1} q_{ij} \le \tilde{d} \quad (3)$$

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

Differently stated, although our lower bound may not be achievable, it is still tighter than the simple bound. We also note that the bounds presented here can be used in a high-level power estimation framework. Preliminary results in this direction are reported in [8].

3. The effect of input temporal correlations on the theoretical bounds

The bounds presented in the previous section are valid as long as the conditional and state probabilities are extracted correctly. As it has been shown previously [11], the effect of temporal correlations on the state probabilities is significant and by not considering them, errors of up to 500% can be obtained. Thus, as expected, ignoring the existing correlations or underestimating the actual order of the Markov chain which models the inputs will produce looser or incorrect bounds. To capture the exact correlations on the primary inputs, one can use the approach presented in [11] to get the correct values of state and conditional probabilities of the FSM, and thus the correct values for the theoretical bounds. However, that approach requires an additional step of *constrained reachability analysis* to reduce the state space and may be computationally intensive. In what follows, we propose a more efficient technique which holds for any *unifilar* FSM. In addition, we show that with appropriate transformations, any *non-unifilar* FSM can be transformed into a unifilar one and thus, can be efficiently handled.

4. Unifilar FSMs

Definition 4. (Unifilar FSM) Let $\mathcal{M} = (I, O, S, \delta, \lambda)$ be a Moore-type FSM[1] where I is the input space, O is the output space, S is the state space, $\delta : S \times I \to S$ is the next state function and $\lambda : S \to O$ is the output function. M is *unifilar* if for any fixed state $s \in S$ and for any $x_1 \in I$, $x_2 \in I$, $x_1 \ne x_2$, then $\delta(s, x_1) \ne \delta(s, x_2)$. In other words, there are no two transitions going to the same state for different input combinations.

For unifilar FSMs, the following result holds:

Theorem 3. Let $\mathcal{M}$ be a *unifilar* FSM. If the inputs $\{x_n\}_{n \ge 1}$ are modeled by a lag-k Markov chain, then the state space is modeled by a lag-k+1 Markov chain and the following holds:

$$p(s_n | s_{n-1} \ldots s_{n-k-1}) = p(x_n | x_{n-1} \ldots x_{n-k}) \quad (4)$$

where $\delta(s_{n-i-1}, x_{n-i}) = s_{n-i}$ for any $i = 0,...,k$.

Hint for proof: It uses the definition of k-step conditional probabilities and Definition 4. ■

The above result provides the basis for an efficient

1.The case of a Mealy-type machine can be treated similarly.

Table 1: Different theoretical bounds for the average Hamming distance and the effect of ignoring temporal correlations (*mcnc'91*)

		With temporal dependencies							Without temporal dependencies	
Circuit	PIs/FFs	LBI	UBI	LBC	UBC	Looseness (%)	LBR	UBR	Looseness (%)	%Violations (>UBC or <LBC)
cse	7/4	0.2768	0.9169	0.2834	0.9133	5.77	0.2834	0.8188	11.11	5.00
dk17	2/3	0.7777	2.2234	0.8428	2.1853	17.98	1.0371	1.9477	29.05	3.00
ex5	2/3	0.7017	1.8275	0.7547	1.7981	16.51	0.9222	1.6222	45.45	18.00
keyb	7/5	0.2957	2.2621	0.4488	2.1417	1.41	0.4567	2.0315	21.62	5.00
planet	7/6	0.7971	6.0511	0.9999	5.8889	99.90	2.4249	3.7159	114.50	0.00

state probability computation. Instead of performing constrained reachability analysis, followed by effectively solving the Markov chain on the primary inputs and state lines jointly (as in [11]), we can directly get the stochastic matrix characterizing the state lines, thus saving computation time and memory requirements. In practice, few FSMs have the unifilar property; however, any non-unifilar FSM can be transformed in an equivalent unifilar FSM based on the following result.

Theorem 4. For any non-unifilar FSM, there exists an equivalent unifilar FSM with at most p times more states, where p is the largest number of input combinations that map a fixed current state to the same next state.

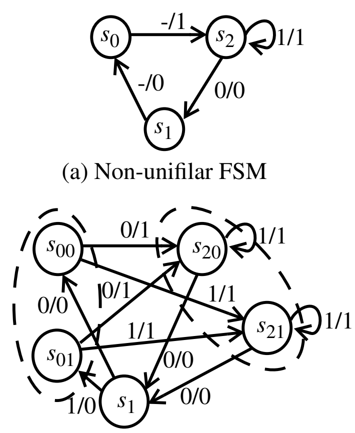
Hint for proof: If there are p input combinations that map a fixed current state to the same next state, the latter can be cloned p times to satisfy the conditions in Definition 4. Note that the state cloning can be done only once for each state, such that state duplicates are reused. ■

Note 3: We observe that the maximum number of states created by the above procedure is upper-bounded by $|S| \times p$, whereas in [8] it is upper-bounded by $|S| \times |I|$. Since usually p is much less than $|I|$, the space/time complexity is drastically reduced.

For example, for the non-unifilar FSM in Fig. 2(a), by duplicating states s_2 and s_0, we get an equivalent unifilar FSM (Fig.2(b)).

The above result suggests a more efficient way of finding the steady state probabilities of the target FSM, without resorting to reachability analysis or advanced numerical techniques for Markov chain solving [11]. For every FSM, an equivalent unifilar FSM is built and based on Th.3, the stochastic matrix for the state lines is obtained. The Markov chain is solved for the state space only[1] and thus, based on the input statistics, exact probabilities are obtained for the states of the FSM.

1.Previous approaches solve for the joint (*input, state*) Markov chain.



(a) Non-unifilar FSM

(b) Equivalent unifilar FSM

Fig.2: Two equivalent FSMs

5. Experimental results

In this section, we provide experimental results for a subset of *mcnc'91/ISCAS89* circuits. The data are first-order temporally correlated vectors at the primary inputs of the *mcnc'91* benchmarks and data with correlations spanning more than one time step for the *ISCAS'89* benchmarks. The bit probabilities range from 0.3 to 0.7 in all cases. We are interested in assessing the effectiveness of the proposed bounds in FSM analysis and the effect of ignoring the input temporal correlations on the accuracy and correctness of the proposed bounds. To this end, we target two sets of experiments:

a) The first one shows comparatively the two theoretical bounds for average Hamming distance and the effect of not considering the temporal correlations on the primary inputs (Table 1). Having the STG of each circuit, we first extract the values of the conditional and steady state probabilities for the underlying equivalent unifilar FSM Markov chain; then, we compute the minimum and maximum values of the switching activity considering a *minimal length encoding* (i.e., log$n$ bits where n is the total number of reachable states). For columns 3-9, the actual temporal correlations on the primary inputs are considered. Each FSM is transformed into an equivalent unifilar FSM and then, based on the order and probabilities of the inputs, the order and probabilities of the states are derived exactly.

We give in Table 1[1] the values of the lower/upper bounds based on informational energy (col. 3, 4) and the tighter bounds computed as in Section 2 (col. 5, 6) for some *mcnc'91* benchmarks. In column 7, we give the *average looseness* of the proposed bounds, while the minimum and maximum values of the switching activity obtained over 1000 random encodings are given in columns 8, 9.

As it can be seen, our bounds are sufficiently close to the ones obtained through (almost) exhaustive search (reported in col. 8, 9). For columns 10-11, the Markov chain for the state lines is built by looking only at the occurrence probability of the input patterns and by ignoring the input temporal correlations. As it can be seen, for this case, the bounds obtained are either more loose (col. 10) or they are actually incorrect and produce a significant percentage of violations (col. 11). Thus, considering the exact temporal correlations are important for the accurate bound computation.

b) The second set of experiments shows the effectiveness of our approach on some large *ISCAS'89* benchmarks (Table 2). For each FSM, the exact probabilities are extracted by first building an equivalent unifilar FSM and then, using Theorems 1 and 2, we compute the bounds. (In Table 2, LBI, UBI, LBC, UBC denote the same measures as in Table 1; LBS denotes the simple lower bound.) .

Table 2: Different theoretical bounds for the average Hamming distance (*ISCAS'89*)

Circuit	PIs/FFs	LBI	LBC	UBC	UBI	LBS
s298	3/14	0.0227	0.9167	12.6670	13.5165	0.9167
s400	3/21	0.6667	0.6667	13.8333	13.8354	0.6667
s526	3/21	-0.4592	0.9167	19.0831	20.4130	0.9167
s641	35/19	-1.8305	0.8910	16.2268	18.8702	0.8865
s820	18/5	0.4339	0.6831	3.2592	3.4772	0.6831
s1196	14/18	0.0238	1.0761	17.0544	18.0621	0.9914
s1238	14/18	0.0193	1.0762	17.0561	18.0683	0.9915

The results are consistent with the above observations. We note that due to the large size of the circuits, we could not provide the bounds obtained by (almost) exhaustive search.

1. PIs/FFs in column 2 denote the number of primary inputs and flip-flops, respectively.

6. Conclusion

In this paper, we have presented an efficient technique to account for input temporal correlations in the computation of accurate theoretical bounds for the average switching activity on the state lines in FSMs. As the main theoretical contribution, we improve over the previous work by accounting for complex spatiotemporal effects which are present when the target machine models real hardware and receives data from real applications. We also show that, for the special case of unifilar FSMs, the order and probabilities characterizing the Markov chain on the state lines of the FSM can be easily deduced and the case of general FSMs can be reduced to this simple scenario. The net result is a more efficient technique for estimating accurate theoretical bounds for FSM switching activity.

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

[8] D. Marculescu, R. Marculescu, and M. Pedram, 'Theoretical Bounds for Switching Activity Analysis in Finite-State Machines,' in *Proc. Intl. Symposium on Low-Power Electronics and Design*, Aug. 1998.

[9] S. Ramprasad, N.R. Shanbhag, and I.N. Hajj, 'Achievable Bounds on Signal Transition Activity,' in *Proc. Intl. Conf. on Computer-Aided Design,* Nov. 1997.

[10] 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.

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

[12] 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.