EXACT POWER ESTIMATION USING WORD LEVEL TRANSITION PROBABILITIES

Ana T. Freitas* Arlindo L. Oliveira† José C. Monteiro Horácio C. Neto
IST/INESC
Av. Alves Redol, 9
1000 Lisboa, Portugal
{atf,aml,jcm,hcn}@inesc.pt

Abstract

We propose a model and an algorithm to perform exact power estimation taking into account all temporal and spatial correlations of the input signals. The proposed methodology is able to accurately model temporal and spatial correlations at the logic level, with the input signal correlations being specified at the word level using a simple but effective formulation.

This paper includes three independent contributions: a proposal of a simple specification language that can represent complex input correlations in a compact form; a methodology for module characterization that is independent of a specific input trace; and an extension of the model to make it applicable to RT-level power estimation by abstracting the details of the internal structure of the combinational modules.

Although the complexity associated with the exact characterization of the modules restricts the applicability of the approach to medium sized circuits, it provides an important tool to validate approximate methods. We are currently researching heuristic approximations that can be used to extend the range of the applicability of the methodology while minimizing the loss in accuracy.

1 Introduction

In the last few years, research on techniques for low power at various levels of design has intensified. Initial work was done at the logic and circuit levels. However, much research is now being conducted at the register-transfer and behavioral levels as it has been recognized that decisions taken during high-level synthesis can have a much larger impact on the power dissipation of the final circuit. During the design exploration phase efficient power estimation tools are required. In this work we are concerned with this specific task, that of estimating the power dissipation of combinational modules, both at the logic level and at the register-transfer levels.

Existing tools are able to estimate, in an accurate way, the power dissipated in a circuit described at the logic level. However, for efficiency reasons, the model used for the statistics of the input signals includes simplifying assumptions that may not be realistic in many cases.

*This work was supported in part by FCT and the PRAXIS-XXI program.

†This author is also affiliated with the Cadence European Laboratories.

One common simplification is the assumption that the primary inputs are uncorrelated in time and space.

The basic model used by most power estimation techniques for a functional module is of the form:

$$P = C_{eff} V_{DD}^2 f \tag{1}$$

where V_{DD} is the voltage of the power supply of the module and f is the clock frequency at which it operates. The factor C_{eff} is the average switched capacitance of the functional unit.

In this work, a power estimation approach that is able to model exactly temporal and spatial correlations at the logic level is presented. Three separate contributions can be identified. In the first place, we propose a simple but powerful formalism that is adequate to express spatial and temporal correlations commonly observed in practice. In the second place, we propose an algorithm to perform pattern independent characterization of combinational modules. Finally, we present an extension of the method to make it applicable to perform accurate estimation at the RT level. These contributions are described in sections 3, 4 and 5, respectively. Experimental results are reported in section 6 and section 7 concludes the work.

2 Related work

Existing approaches for power estimation can be divided into two categories: simulation based methods and probabilistic methods.

Simulation based methods can model very accurately the spatial and temporal correlations present at the input and internal signals. However, they suffer from the significant drawback that they require the existence of a trace that accurately matches the conditions of operation of the system. When the objective is to characterize a module that can be used in a variety of conditions, simulation based methods may fail to yield the required precision.

Probabilistic methods do not require the existence of a trace and can be used to characterize circuits in a manner that is independent of the inputs presented. However, the most efficient approaches presented to date allow only for very simplified models of temporal and spatial correlations, and, therefore, limit the accuracy of the estimates they present. Other, more precise approaches, are computationally inefficient but perform an exact modeling of the circuit characteristics. To extend their range of applicability, approximations to the exact models can be used.

Our approach fits in this second class, in the sense that it performs an exact modeling of the circuit characteristics and is independent of the existence of a trace.

2.1 Simulation based methods

When a trace that reflects accurately the spatial and temporal input correlations is available, the effective switching capacitance can be computed in a trivial way by simulating the circuit and computing directly the summation

$$C_{eff} = \sum_k C_k T_k \tag{2}$$

where T_k represents the total number of transitions in node k. Using simulation, this computation is straightforward, since the waveforms at each node can be obtained from the input waveforms. However, it may be very inefficient if the trace is very long.

Methods have been proposed to speed up this computation, without incurring in a significant loss of precision. One such possibility is random sampling where only a subset of the transitions present in the trace is used. More sophisticated methods include sequence compaction [1] and sequence synthesis. These approaches aim at reproducing accurately the input correlations with traces that are much smaller than the initial trace.

Another possibility is to model the temporal and spatial correlations at the primary inputs by generalizing the concept of a trace and creating a finite state machine that creates a sequence of inputs with the statistics of that trace [2]. This generalized trace is not necessarily obtained from an actual run of the circuit since it may include don't cares, allowing the user to model accurately a large number of input dependences. However, this approach requires a complex procedure to synthesize the finite state machine and to perform the power estimation.

Entropy based methods [3, 4] can also be included in this class since they use an input sequence to extract input and output entropies of a circuit. These methods use the concept of computational work (which is obtained from the entropy of the input and output signals), as a measure of the average activity in a circuit. Najm observes that average power is proportional to the product of the circuit area and the average node transition activity [4]. However, these methods may incur in a significant loss of accuracy since they do not take into account the details of the circuit structure.

All these methods require the existence of a trace to be applicable. In many applications, the specification of a trace is simply not practical, either because the design is in the early stages of design exploration or because a trace independent characterization of the modules is needed. In these cases, we must resort to probabilistic based methods.

2.2 Probabilistic methods

Probabilistic methods have the potential to be used independently of the existence of an actual trace. The simplest approach is based on the assumption that the primary inputs are uncorrelated in time and space [5]. The user may specify a probability P_1 that a given bit is at 1, which, under the independence assumptions used, gives the probability of a transition in that bit as $2P_1(1 - P_1)$. This modeling is crude, and even an exact computation of P_1 will give results that may be very far from actual dissipated power observed in the presence of significant spatial and temporal correlation. Regrettably, ignoring these correlations leads, in many cases, to estimates that are not precise enough.

Several methods that aim at modeling spatial and temporal correlations more accurately have been presented. One such method [6] models the pairwise spatial correlations of the primary inputs and propagates them through the circuit. The advantage of this procedure is that, by using these coefficients, the process only involves local calculations, and therefore can be applied to larger circuits. On the other hand, there is no control over the total error introduced in successive computations.

Another approximation that has the potential to model accurately spatial and temporal correlations is based on the use of *algebraic decision diagrams* (ADDs) to characterize the dissipation incurred by each module under each particular input transition [7]. This method has the significant advantage that it permits the exact computation of the power dissipated in a combinational module given the ADD that represents the switched capacitance as a function of the primary inputs at time t and time $t + 1$ and the probability of all possible input transitions. The main problem of this method is that ADD representations can easily become unmanageably large.

Our approach to module characterization can be viewed as an alternative approach to ADD based characterization methods but offers different possibilities and limitations since the detailed computations are different. Since the method is based on the computation of a characterization function using BDDs, it incurs in the advantages and drawbacks inherent to this representation. A significant advantage is that the computation of the required functions can be easily performed using one of the many available BDD packages. The most serious drawback is that the size of the circuits that can be handled directly, (i.e., without using any approximation method) is restricted to small and medium sized blocks.

The techniques proposed in this work perform the exact computation of the power dissipated in a combinational module independently of the existence of a trace and can be combined with an input modeling that can represent, in a very compact way, complex temporal and spatial correlations.

3 Input correlation modeling

Consider a combinational circuit with n input nodes and m internal nodes. The variables that correspond to the input nodes are $X = \{x_1, \ldots x_n\}$, while the circuit internal nodes are represented by variables $Y = \{x_{n+1}, \ldots x_{n+m}\}$. A subset of the nodes in Y, $Z \subseteq Y$, are output nodes of the circuit. For each node i, we will use two variables, one representing the value $b$efore the transition (x_i^b) and the other one representing the value $a$fter the transition (x_i^a).

The input correlation modeling we propose is based on the definition of a set of functions F_i and associated probabilities P_i. F_i, defined over the variables $\{x_1^b, \ldots x_n^b, x_1^a, \ldots x_n^a\}$ represents all transitions that take place with probability P_i (a real number between 0 and 1, and satisfying $\sum_i P_i = 1$).

By specifying a set of pairs (F_i, P_i), it is possible to model accurately, and in a compact form, relatively complex temporal correlations between input words. Each F_i is specified using a simple formalism: for each possible transition, a description of the input word before and after a transition is given; individual bits in the word can be represented by:

- `0` : The bit takes the value 0
- `1` : The bit takes the value 1
- `-` : The bit takes any value.
- `.` : The bit keeps its value after the transition.
- `#` : The bit changes value after the transition.

As an example, the following segment of a specification:

```
.10-  .10-  0.2
#000  #000  0.1
```

means that:

- With probability 0.2, the first bit stays the same, the second and third bits are fixed at 10 and the last bit can take any value.
- With probability 0.1, the first bit changes value while the last three bits stay at the value 0.

Gray			Binary			Uniform		
...#	...#	0.500	...0	...1	0.500	----	----	1.0
..#.	..#.	0.250	..01	..10	0.250			
.#..	.#..	0.125	.011	.100	0.125			
#...	#...	0.125	#111	#000	0.125			

Table 1: Specification of input word statistics for Gray, binary and uniform distribution codes.

Vectors representing transitions with the same probability are modeled by an unique F_i function. Clearly, this representation can be extended, but, as presented, it can be used to model, in a compact way, complex temporal and spatial correlations. As an example, a 4 bit Gray code and a 4 bit binary code are modeled by the specifications of the F_i's shown in table 1.

Note that, under the uniform assumption, each of these codes have $P_1 = 0.5$ for all the bits. In general, the three distributions described above will generate very different estimates for power dissipation when word level correlations are taken into account.

To illustrate the potential of this representation let us consider another example. Consider a circuit with a datapath module and a controller. The behavior of the input signals of the two modules can differ significantly. Usually controller input signals are more deterministic. Consider that the datapath inputs change with a random sequence and that controller inputs change with a binary sequence. If the circuit has a datapath with six inputs and a controller with four, we can specify, with our representation, all possible transitions as follows:

...0------	...1------	0.5
..01------	..10------	0.25
.011------	.100------	0.125
#111------	#000------	0.125

With this description only 4 lines are needed to consider all transitions. An equivalent description using a complete trace requires 65537 lines to describe the same transitions. Moreover, the trace grows exponentially as the number of bits increase.

The generation of the F_i and P_i can be accomplished in several ways. In many cases the user may be able to specify a good approximation, by considering the operating environment of the circuit under consideration. In other cases, they may be extracted automatically from typical traces. Finally, and this represents the most interesting case from the point of view of RT level power estimation, the pairs (F_i, P_i) can be obtained automatically by considering the structure of the circuit and the correlations imposed by other modules in the signal path. This procedure is outlined in section 5.

At present, the functions F_i, as well as all other functions required by the approach, are manipulated using a standard *binary decision diagram* (BDD) package [8].

4 Pattern independent characterization of modules

This section presents a new approach for the characterization of combinational modules that can be combined with the input modeling described in the previous section to perform accurate power estimation.

This approach can be viewed as a generalization of the work of Ghosh [5]. His approach is based on the computation, for each node, of a function (represented by a BDD) that describes the conditions under which that node switches. With this function it is possible to compute the dissipated power by computing expression 2. The value of T_i, in this expression, is obtained by summing the probability of all input transitions that cause a transition in node i.

Our approach avoids the need to explicitly sum over all these transitions by using the functions F_i and the associated probabilities P_i, defined in the previous section. To use this information in an effective way, we compute a single function that characterizes the circuit, the Transition Consistency Function or TCF. This function, which depends on the input and internal signals at time t and time $t+1$, represents all possible transitions of the circuit under analysis.

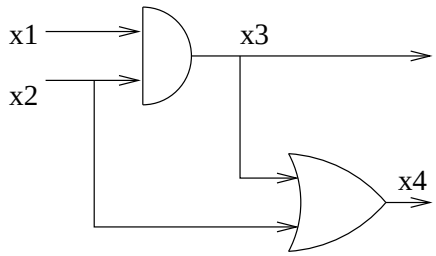
Using this function and the description of the probability of the input transitions, it is possible to compute exactly the dissipated power. This approach models precisely all temporal and spatial correlations of the input signals which are specified by the values of F_i and P_i. The spatial correlations of the internal signals induced by the structure of the circuit and by the correlation of the inputs is taken into account since the TCF describes exactly the transitions that take place in the network.

4.1 The transition consistency function

We are interested in the computation of the dissipated power when the input variables X change from a value X^b to a value X^a. For simplicity, we will consider only the zero delay model. Since there exists a total of $2(n+m)$ variables, one needs to compute a probability distribution defined over a space of dimension $2^{2(n+m)}$.

Specifically, we will consider the function $\mathcal{T}(x_1^b, \ldots, x_{n+m}^b, x_1^a, \ldots, x_{n+m}^a)$ that describes all possible transitions. We defined $\mathcal{T}(x_1^b, \ldots, x_{m+n}^a)$ to be 1 if the values of the variables in the circuit $\{x_1, \ldots, x_{n+m}\}$ may experiment a transition from $x_1^b, x_2^b, \ldots, x_{n+m}^b$ to $x_1^a, x_2^a, \ldots, x_{n+m}^a$. Function $\mathcal{T}$ is called the Transition Consistency Function.

Consider the circuit with two inputs and two outputs, as shown in figure 1, and the correspondent list of all possible transitions.



00000000
00000101
01010000
01010101
00001000
00001111
01011000
01011111
10000000
10000101
11110000
11110101
10001000
10001111
11111000
11111111

Figure 1: Example of a two input/two output circuit and the list of minterms in the Transition Consistency Function.

In this circuit, it is possible to observe a transition in the variables x_1, x_2, x_3, x_4 from the combination $1, 0, 0, 0$ to the combination $1, 1, 1, 1$. Therefore, $\mathcal{T}(1, 0, 0, 0, 1, 1, 1, 1)$ is

1. On the other hand, a transition from $0,0,0,0$ to $0,1,1,1$ is not possible, and therefore, $\mathcal{T}(0,0,0,0,0,1,1,1)$ is 0.

4.2 Computation of the TCF

The actual computation of $\mathcal{T}(x_1^b,\ldots,x_{n+m}^b,x_1^a,\ldots,x_{n+m}^a)$ is straightforward. Let the function at internal node i be $f_i(x_1,\ldots,x_n)$. The function $\mathcal{T}$ can now be computed as

$$\mathcal{T} = \prod_{i=n+1}^{n+m} x_i^b \equiv f_i(x_1^b,\ldots,x_n^b) \cdot \prod_{i=n+1}^{n+m} x_i^a \equiv f_i(x_1^a,\ldots,x_n^a) \tag{3}$$

It is easy to show that this function is non-zero only at 2^{2n} minterms, and therefore the BDD size of f is bounded by $2(n+m)2^{2n}$ [9].

4.3 Computation of dissipated power

The dissipated power can now be easily computed using expression (2). The total number of transitions in node k, T_k, is given by

$$T_k = \sum_i |(x_k^b \oplus x_k^a)\mathcal{T}(x_1^b,\ldots,x_{n+m}^a)F_i| \times \frac{P_i}{|F_i|} \tag{4}$$

where we used $|g|$ to denote the number of minterms in g and F_i and P_i were defined in section 3.

The computation of expressions (2) and (4) requires an iteration over all the circuit nodes and the computation of an AND operation for each node in the circuit. This makes it computationally expensive, and hard to apply to higher level power estimation.

However, by exchanging the summations, one can re-write the total power as:

$$C_{eff} = \sum_i \frac{P_i}{|F_i|} \sum_k C_k |(x_k^b \oplus x_k^a)S_i| \tag{5}$$

where S_i is given by:

$$S_i = \mathcal{T}F_i \tag{6}$$

The computation of the inner summation of expression (5) can be performed with a single pass over the BDD defining S_i. Therefore, $\mathcal{T}$ can be used to fully specify the power dissipation of a module, without the need to keep explicitly its internal structure.

4.4 Generalization to unit and nominal delay

As described, the method takes into account only functional transitions, i.e., transitions between steady states of the signals. This is equivalent to the use of a zero-delay model for all the gates in the circuit. However, power dissipation due to spurious transitions is also important and may account for a significant fraction of the total dissipated power.

It is possible to generalize the above method to handle non-zero delay models. Although a detailed description of the proposed generalization is outside the scope of this work, we briefly outline a possible way to handle different delay models.

The method can be generalized by replacing variables x_i^b and x_i^a by a set of variables $x_i^0, x_i^1 \ldots, x_i^{f-1}, x_i^f$, that represent the value of variable x_i at time instants $t_0, t_1, \ldots, t_{f-1}, t_f$.

Expression 3 will now become the product of $f + 1$ sub-expressions, and all the other computations will carry through with minor modifications.

This mechanism can be used to deal with unit delay models, as well as with generic delay models by applying discretization to the time axis.

5 Application to RT level power estimation

At the RT level the modeling of spatial and temporal correlations at the module inputs is of critical importance. Ideally, at this abstraction level one would compute word-level statistics (from functional simulation) and use them on library-stored models of power consumption. For an accurate estimate, this model must take into account both the correlations between the bits in a word and the correlation in time between words.

Even if the details of the specific implementation of a module are known, for the sake of efficiency a model that abstracts away part of the information for the module should be used. The most commonly used approach to determine the coefficients used in the RTL power models is to simulate the module using random inputs and to adjust the model using linear regression [10, 11, 12]. This procedure has several disadvantages:

- the model may be biased due to the input-pattern dependency problem associated with simulation-based techniques
- the model is independent of the topology of the circuit and may become inaccurate if the module input probabilities are not close to uniform
- spatial and temporal correlations of the module inputs are not taken into account

The methodology we propose exhibits a very significant feature, namely the fact that the power dissipation of a combinational circuit can be easily computed abstracting the internal details of the combinational modules. This ability to abstract the details of the modules is obtained, in our case, at the expense of storing the TCF. Clearly, this approach will only be effective at the RT level if the TCF can be somehow reduced, without compromising the accuracy of the estimation. Consider the simple situation where two combinational modules are connected, as shown in figure 2:

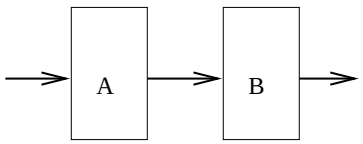


Figure 2: Two connected combinational blocks.

Given the TCF $\mathcal{T}$ for block A, it is possible to compute functions F_i^B for block B that can be used to compute the power dissipated in block B using expression 5. These can be computed using:

$$F_i^B = \forall_{j \in Y \setminus Z} \; \exists_{x_j^b x_j^a} \; \mathcal{T}(x_1^b, \ldots, x_{n+m}^a) F_i^A(x_1^b, \ldots, x_n^a) \tag{7}$$

Although F_i^B can be used directly in expression 5, this function depends on the input variables of A, since smoothing away these variables would lead to an inaccurate computation of the value of $|\mathcal{T} F_i^B|$. This, however, does not represent a strong limitation of the method since only the dependence on the primary input variables needs to be kept.

6 Experimental results

The following table presents preliminary results for power dissipation at the logic level obtained using the proposed approach. As test cases, we used circuits from the MCNC-mlex benchmarks together with two small adders. These circuits were mapped to the MSU library before power estimation. The results obtained assuming both spatial and temporal correlations of the signals are compared with those obtained if independence between input bits, both in time and space, is considered [5]. The fourth and fi fth column of the table show the power estimated using word level transition statistics, where the input words follow binary and Gray codes, respectively. These estimations use the specifi cation of the pairs (F_i, P_i) as described in table 1. The last column shows the power estimated with the power estimate algorithm in SIS that assumes independence between bits (the same results are obtained with our method if the Fi is specifi ed according to the uniform distribution).

	# inputs	# outputs	# literals	Power (μW)		
				Binary	Gray	Uniform
add_cla_4	8	4	47	127.5	81.3	137.9
add_cla_8	16	8	117	218.7	125.5	352.4
cm162a	14	5	72	45.6	23.7	205.2
cm85a	11	3	69	66.5	43.5	194.1
cu	14	11	115	111.0	57.1	406.2
z4ml	7	4	131	248.2	142.3	395.1
lal	26	19	264	167.4	91.8	905.6
pcler8	27	17	142	68.7	36.0	306.0
pm1	16	13	91	28.7	15.2	322.8
sct	19	15	222	77.5	39.5	779.2
f51m	8	8	245	494.6	329.5	809.3
9symml	9	1	359	599.1	364.7	1437.7
cc	21	20	112	42.0	21.2	394.3
count	35	16	219	60.2	28.8	609.6
unreg	36	16	193	56.3	21.9	692.5
alu2	10	6	680	865.0	544.1	1738.2
alu4	14	8	1322	1535.5	1023.6	3068.0

Table 2: Power dissipation results obtained using word level statistics (Binary and Gray) and uniform bit distribution (Uniform).

As the experimental results show, the assumption of temporal and spatial independence of input signals can introduce a signifi cant error on power estimation, and therefore an accurate power dissipation model must take these into account.

7 Conclusions and future work

In this paper we proposed a probabilistic approach that uses a simple but powerful formalism for power estimation, which considers both spatial and temporal correlations of the primary input signals. If independence between input signals is assumed, a signifi cant error can be introduced in the power estimates. In many cases, it is of critical importance to use an

estimation method that accounts for both types correlations in order to obtain a realistic value for the power dissipation. This is particularly true for RTL power estimates.

The novel formalism introduced in this paper allows for an easy description of the input statistics. Given this description, the method is able to compute the exact switching activity in every node in the circuit.

The amount of memory required by the BDDs used to represent the transition consistency function, limits the size of the modules that can be analyzed using these methods. Alternative data structures and approximation methods that aim at circumventing this limitation are the topic of active research.

References

[1] D. Marculescu, R. Marculescu, and M. Pedram. Hierarchical sequence compaction for power estimation. In *33rd Design Automation Conference (DAC'97)*, pages 570–575, New York, June 1997. Association for Computing Machinery.

[2] J. Monteiro and S. Devadas. Techniques for the Power Estimation of Sequential Logic Circuits Under User-Specified Input Sequences and Programs. In *Proceedings of the International Symposium on Low Power Electronics and Design*, pages 33–38, April 1995.

[3] D. Marculescu, R. Marculescu, and M. Pedram. Information Theoretic Measures of Energy Consumption at Register Transfer Level. In *Proceedings of the International Symposium on Low Power Electronics and Design*, pages 81–86, April 1995.

[4] M. Nemani and F. Najm. Toward a High-Level Power Estimation Capability. *IEEE Transactions on Computer-Aided Design*, 15(6):588–598, June 1996.

[5] A. Ghosh, S. Devadas, K. Keutzer, and J. White. Estimation of Average Switching Activity in Combinational and Sequential Circuits. In *Proceedings of the* 29^{th} *Design Automation Conference*, pages 253–259, June 1992.

[6] R. Marculescu, D. Marculescu, and M. Pedram. Efficient Power Estimation for Highly Correlated Input Streams. In *Proceedings of the* 32^{nd} *Design Automation Conference*, pages 628–634, June 1995.

[7] A. Bogliolo, L. Benini, and G. De Micheli. Characterization-free behavioral power modeling. In *Proceedings of the Design Automation and Test in Europe*, pages 767–773, 1998.

[8] K. Brace, R. Rudell, and R. Bryant. Efficient implementation of a BDD Package. In *Proc. of the ACM/IEEE Design Automation Conference*, pages 40–45. ACM Press, June 1990.

[9] A. L. Oliveira, L. Carloni, T. Villa, and A. Sanviovanni Vincentelli. Exact minimization of binary decision diagrams using implicit techniques. *IEEE Transactions on Computers*, 47(11):1282–1296, November 1998.

[10] S. Powell and P. Chau. Estimating Power Dissipation of VLSI Signal Processing Chips: the PFA Technique. In *VLSI Signal Processing IV*, pages 250–259, 1990.

[11] Q. Qiu, Q. Wu, and M. Pedram. Cycle-Accurate Macro-Models for RT-Level Power Analysis. In *Proceedings of the International Symposium on Low Power Electronics and Design*, pages 125–130, August 1997.

[12] A. Raghunathan, S. Dey, and N. Jha. Register-Transfer Level Estimation Techniques for Switching Activity and Power Consumption. In *Proceedings of the International Conference on Computer-Aided Design*, pages 158–165, November 1996.