Analytical Macro-Modeling for High-Level Power Estimation

Giuseppe Bernacchia
Dipartimento di Elettronica
Elettrotecnica e Informatica
University of Trieste
via A.Valerio 10, I-34127 Trieste, Italy
e-mail: bernac@ipl.univ.trieste.it

Marios C. Papaefthymiou
Department of Electrical Engineering
and Computer Science
University of Michigan
Ann Arbor, Michigan 48109
e-mail: marios@eecs.umich.edu

1   Abstract

In this paper, we present a new analytical macro-modeling technique for high-level power estimation. Our technique is based on a parameterizable analytical model that relies exclusively on statistical information of the circuit's primary inputs. During estimation, the statistics of the required metrics are extracted from the input stream and a power estimate is obtained by evaluating a model function that has been characterized in advance. Thus, our model yields power estimates within seconds, since it does not rely on the statistics of the circuit's primary outputs and, consequently, does not perform any simulation during estimation. Moreover, our model achieves significantly better accuracy than previous macro-modeling approaches, because it takes into account both spatial and temporal correlations in the input stream.

In experiments with ISCAS-85 combinational circuits, the average absolute relative error of our power macro-modeling technique was at most 3%. For all but one circuit in our test suite, the worst-case error was at most 19.5%. In comparison with power estimates for a ripple-carry adder family that were obtained using Spice, the average absolute and worst-case errors of our model's estimates were 5.1% and 19.8%, respectively.

In addition to power dissipation, our macro-modeling technique can be used to estimate the statistics of a circuit's primary outputs. Our experiments with ISCAS-85 circuits yield very low average errors. Thus, our technique is suitable for fast and accurate power estimation in core-based system with pre-characterized blocks. Once the metrics of the primary inputs are known, the power dissipation of the entire system can be estimated by simply propagating this information through the blocks using their corresponding model functions.

2   Introduction

Low power consumption has become a primary objective in VLSI design, primarily due to reliability concerns and portability considerations. The core component of almost every low-power design methodology is power estimation. Several power estimation techniques have been proposed over the past few years [7, 9]. In general, high-accuracy estimation schemes are computationally demanding, whereas fast estimation schemes tend to sacrifice accuracy. As circuit size and complexity continue increasing with exponential pace, however, there is a growing need for fast and accurate power estimators that can be used at high abstraction levels in the early design stages.

High-level power estimation techniques fall into two categories. In the *top-down* approaches, the circuit is described as a set of Boolean functions. Information about the circuit's activity is used to optimally decompose these functions and minimize power dissipation [10]. In the *bottom-up* approaches, the circuit is described as a set of *blocks* with known internal structure. The power dissipation of each block is estimated using a *macro-model*.

In this paper, we present a new analytical macro-modeling technique for bottom-up power estimation. Our technique does not use any simulation during the estimation phase and can thus provide power dissipation estimates extremely fast. Moreover, our technique takes into account spatial and temporal input correlations, thus achieving significantly better accuracy than previous macro-modeling approaches. In experiments with the ISCAS-85 combinational circuits, the average absolute relative error of our macro-modeling technique was at most 3%. For all but one circuit in our test suite, the worst-case error was at most 19.5%. For a family of ripple-carry adders whose dissipation was estimated using Spice simulations, the average and worst-case errors of our model were 5.1% and 19.8%, respectively. Our model is parameterizable and can provide accurate estimates of a block's output statistics. It is thus suitable for fast and accurate power estimation in core-based design.

Macro-modeling for high-level power estimation has attracted considerable attention over the past few years. A look-up table (LUT) model was introduced in [3]. The parameters of that model were the *average input signal probability* P_{in}, the *average input transition density* D_{in}, and the *average output transition density* D_{out}. The metric D_{out} was evaluated using zero-delay simulation. The LUT reported estimates for equi-spaced discretized values

of the parameters. For input characteristics that did not correspond to any LUT values, estimates were obtained using an interpolation scheme. Several modifications to the LUT method were introduced in [1] to improve its accuracy.

Both LUT-based approaches are general and can be easily used for different kinds of circuits without any modification of the model itself. They suffer from two potentially serious disadvantages, however. First, a large LUT may be required to ensure good accuracy. If n is the number of equi-spaced values chosen for P_{in} and D_{in}, and m is the number of simulations required to obtain D_{out}, the size of the LUT is approximately $m \times n^2$ [1]. Second, the LUT schemes do not consider input correlations which have been found to severely influence the overall power consumption [5, 6]. Furthermore, they both require the use of simulation during the estimation phase to obtain D_{out}. Even if these simulations are performed with zero-delay models, they are time-consuming, particularly when circuit size is large.

An analytical model for power estimation has been introduced in [4]. This model avoids the requirements of a large LUT. Moreover it uses a fourth input parameter, the *average input spatial correlation coefficient* SC_{in}, to account for spatial correlation of the inputs. A zero-delay estimation is still required during the estimation phase, however, to obtain D_{out}. Moreover, temporal correlations are not taken into account.

The remainder of this paper has four sections. In Section 3 we give the input parameters of our analytical macro-model. In Section 4 we describe a procedure for building the macro-model of any given circuit. Experimental results from the application of our macro-modeling technique on ISCAS-85 circuits and a family of ripple-carry adders are given in Section 5. Our results show that our analytical macro-model can provide excellent power dissipation estimates within a few seconds. Section 6 concludes our paper with on-going research.

3 Macro-Model Metrics

One of the most challenging aspects in the construction of a power macro-model is the choice of the model's input parameters, or metrics. These metrics should be capturing the features that are primarily responsible for a system's dissipation and can thus help in obtaining good estimates of its power consumption and output statistics. In this section, we describe the metrics of our macro-model.

Similar to the approaches in [1, 3, 4], our model uses the *average input probability* P_{in} and the *average input transition density* D_{in}. Given an *input stream* $x = \langle (x_{11}, x_{12}, \ldots, x_{1M}), (x_{21}, x_{22}, \ldots, x_{2M}), \ldots, (x_{N1}, x_{N2}, \ldots, x_{NM}) \rangle$, these metrics are defined as:

$$P_{in} = \frac{\sum_{j=1}^{M} \sum_{i=1}^{N} x_{ij}}{MN},$$

$$D_{in} = \frac{\sum_{j=1}^{M} d_j}{M},$$

where M is the number of primary inputs to the circuit, N is the length of the stream of input vectors, and d_j is the transition density of the jth input bit. The transition density d_j is defined as the number of $0 \to 1$ and $1 \to 0$ transitions per unit time for bit j. P_{in} corresponds to an average probability only if the 0's and 1's at each input are uniformly distributed. Nevertheless, this metric captures the input features effectively even if the input distribution is not uniform.

Input correlation plays a significant role in power dissipation, especially when the circuit is part of a wide datapath. We have therefore decided to incorporate input correlation to the metrics of our model. Specifically, we use two correlation metrics S_{in} and T_{in} to capture spatial and temporal correlations, respectively, of the inputs.

The spatial correlation metric S_{in} is defined as the average of the bit-wise XNOR between all possible *channel streams* $x_i = \langle x_{1i}, x_{2i}, \ldots, x_{Ni} \rangle$ and $x_j = \langle x_{1j}, x_{2j}, \ldots, x_{Nj} \rangle$ in the stream x:

$$S_{in} = \frac{\sum_{j=1}^{M} \sum_{k=1}^{M} \sum_{i=1}^{N} x_{ij} \oplus x_{ik}}{N \times M \times (M-1)}.$$

This choice was motivated by the following observation. The metric SC_{in} in [4] was defined as the average of the bit-wise AND between all pairs of input channel streams x_j and x_k. According to this definition, two channel streams x_j and x_k will be highly correlated only if they comprise matching 1's in the two streams. Therefore, the AND operator will miss correlated channel streams that comprise matching 0's. The choice of the XNOR captures correlations of both 1's and 0's.

The temporal correlation metric T_{in} captures input features that are missed by S_{in}. Given a channel stream of length N, we take a *window* of length L from the stream and convolve them together. Each term in the convolution represents the number of times that two signals are both simultaneously high. Each term indicates how well the chosen subset is reproduced in the sequence, and therefore provides an estimate of the signal's temporal correlation. The main issue in the definition of T_{in} is the choice of L, since a value that is either too big or too small may result in missing correlation information. In [2] the authors use a similar window technique to estimate the temporal correlation of input streams with arbitrary distributions. Their study shows that a suitable value for L is 10. The final value of T_{in} is the average over all the inputs of the convolution means:

$$T_{in} = \frac{\sum_{j=1}^{M} \sum_{L-1}^{N-L+1} (w_j \otimes x_j)}{N \times M},$$

where w_j is an arbitrary window of length L in the channel stream x_j.

Our model does not use D_{out}, thus avoiding any time-consuming simulations during the estimation phase. D_{out}
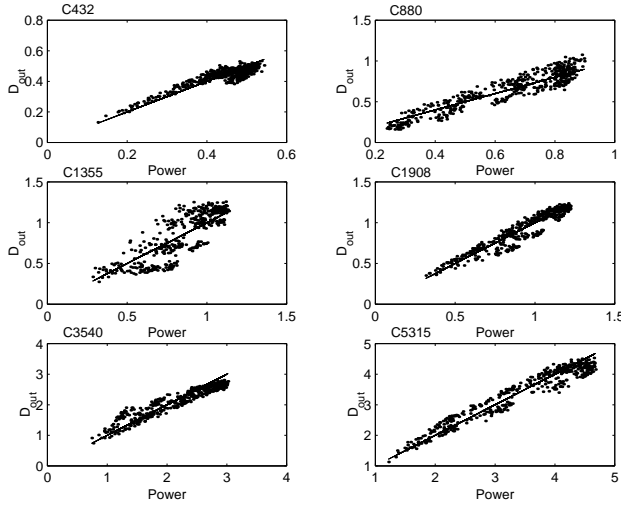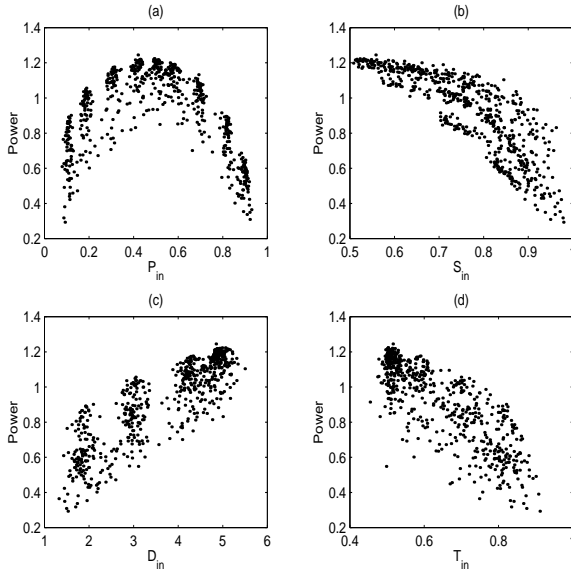
Figure 1: D_{out} versus average power.

Figure 2: Power dissipation of C1908 with respect to the four input metrics of our macro-model.

provides valuable information about a circuit's dissipation and has been included in all previous macro-models. Figure 1 shows D_{out} versus power dissipation for the ISCAS-85 becnhmark circuits we used in our tests. In most cases, the two quantities are related extremely well. Even though our macro-model does not use D_{out}, it still achieves remarkable accuracy, superior to that of previous macro-models.

4 Macro-Model Characterization

To avoid the large memory requirements of a large look-up table, our macro-model uses a nonlinear function

$$PD_{avg} = f(P_{in}, D_{in}, S_{in}, T_{in}) \qquad (1)$$

to estimate the average dissipated power. For the sake of efficiency in the estimation phase, we opted for the use of simple functions. Figure 2 gives the dissipation of the circuit C1908 from the ISCAS-85 benchmark and provides evidence supporting a low-order polynomial dependency of PD_{avg} with respect to the four input metrics of our macro-model. For example, it is evident that the dependency with respect to P_{in} is quadratic-like, while that with respect to D_{in} seems almost linear. We therefore used a 3rd degree complete polynomial as a template function for our model. Such a function requires the calculation of 35 coefficients. A similar function can be used to estimate the output statistics $P_{out}, D_{out}, S_{out}$, and T_{out} of each block.

The characterization phase of our model is quite straightforward. Given a circuit, we first specify a value for the metrics P_{in} and S_{in} in the range [0.1, 0.9]. A sequence of inputs streams is subsequently generated using a random number generator. Unless the circuit has very few inputs, it is very difficult to control all four metrics simultaneously. We thus generated a sufficiently high number of streams to ensure that results are collected for a reasonable subset of values for D_{in} and T_{in}. To that effect, we used 650 streams, each consisting of 200 input vectors.

The reference power values are obtained using a zero-delay model simulation. Since the simulation of large circuits with accurate, low-level simulators is computationally demanding, we built our proposed macro-model following this approach, although it does not take into account phenomena like glitches and short-circuit power dissipation. As shown by the Spice simulations in Section 5, our macro-model is nevertheless very accurate. In [4], characterization was performed using Monte Carlo simulation with general delay models, thus resulting in very accurate reference values. During the estimation phase, however, zero-delay simulation was used for the sake of efficiency. Thus, the estimation potential of the model was compromised, since the input data were not accurate enough.

Once the reference values have been obtained, a standard nonlinear fitting algorithm can be used to compute the coefficients of the model. The total time required for the characterization of each model from the ISCAS-85 circuits is less than 2 minutes, including simulation and coefficient computation.

The main advantage of our model is that it does not need any simulations during estimation. For the design of highly-complex systems with several different pre-defined blocks, our technique can provide fast and accurate estimates, thus enabling designers to explore different block arrangements in real time. Alternative macro-modeling techniques that require even zero-delay simulation during estimation are prohibitively time-demanding, because they must perform a complete simulation each time a block is moved. In our method, all we need to know is the metrics of the primary inputs. This information can be propagated extremely fast through the system by evaluating simple functions.

3

Circuit	G	PI	PO	Max ϵ	Avg ϵ	$\sigma(\epsilon)$
C432	160	36	7	28.7%	3.0%	8.6
C880	383	60	26	10.0%	1.5%	1.8
C1355	546	41	32	19.5%	1.3%	2.5
C1908	880	33	25	12.0%	1.5%	2.3
C3540	1669	50	22	9.2%	1.5%	2.1
C5315	2307	178	123	5.5%	0.9%	0.6

Table 1: Accuracy of power estimates.

5 Results

This section presents results from the application of our macro-model to several ISCAS-85 benchmark circuits and to a family of ripple-carry adders.

Table 1 gives the accuracy of the power estimates obtained with our model for several ISCAS-85 circuits. The first two columns give the names of the circuits and their corresponding gate counts. Columns three and four give the number of primary inputs and primary outputs, respectively. The last three columns give the maximum, average, and standard deviation of the absolute relative error ϵ for the estimates obtained with our macro-model. Apart from C432, the maximum error was no more than 19.5%. Moreover, for all circuits, the average error was at most 3%. Averages were computed over 500 data points, each of which was collected by applying a randomly generated stream of length 200.

Table 2 gives the accuracy with which our model estimated the output characteristics of the circuits in our test suite. These results mirror those obtained for power dissipation, showing that the technique could be used effectively to achieve fast and accurate results in the early stages of system design.

In addition to the benchmark circuits, we applied our model to simple ripple-carry adder circuits whose dissipation was estimated separately using Spice. Our simulations were performed using the same criteria for the input vectors as with the benchmark circuits. Our model was applied on three circuits: a 1-bit adder, a 2-bit adder, and a 4-bit adder. Our goal was to demonstrate that the proposed method is highly reliable even with more accurate models for the gates, and that it can be easily used to describe very regular circuits, that is, circuits comprising several replicas of a basic template block.

To encompass parameterization, a fifth metric n must be included to the function of our model, which identifies the number of replicas (or number of input bits) in the circuit:

$$PD_{avg} = f(n, P_{in}, D_{in}, S_{in}, T_{in}) \qquad (2)$$

Table 3 gives the accuracy of our macro-model with respect to Spice estimates. The estimates are not as accurate as in the case of zero-delay estimation. They are nevertheless quite close to Spice and were obtained within a few seconds. Worst-case absolute relative error never exceeded 20%. Average absolute relative error was at most 5.1%. The circuits used in this experiment were particularly glitchy, because XOR gates had been replaced with their NAND equivalent. Further improvements could be achieved by modifying the template function.

Circuit	Metric	Max ϵ	Avg ϵ	$\sigma(\epsilon)$
C432	P_{out}	29.4%	3.0%	11.0
	S_{out}	6.6%	1.3%	1.2
	D_{out}	38.7%	3.9%	13.8
	T_{out}	27.1%	3.5%	9.4
C880	P_{out}	4.6%	7.8%	0.4
	S_{out}	2.3%	0.4%	0.1
	D_{out}	14.0%	2.2%	3.7
	T_{out}	12.7%	1.7%	2.5
C1355	P_{out}	6.8%	0.5%	0.5
	S_{out}	1.1%	0.2%	0.0
	D_{out}	5.0%	0.7%	0.5
	T_{out}	5.3%	0.5%	0.3
C1908	P_{out}	4.7%	0.9%	0.7
	S_{out}	1.9%	0.5%	0.1
	D_{out}	10.1%	1.7%	2.3
	T_{out}	9.1%	1.5%	1.6
C3540	P_{out}	5.4%	1.3%	1.1
	S_{out}	2.2%	0.4%	0.1
	D_{out}	19.7%	2.8%	5.9
	T_{out}	9.0%	1.9%	2.5
C5315	P_{out}	4.2%	0.6%	0.2
	S_{out}	1.0%	0.2%	0.0
	D_{out}	9.7%	1.7%	2.3
	T_{out}	5.3%	1.2%	1.0

Table 2: Accuracy of output statistics.

Circuit	PI	PO	Max ϵ	Avg ϵ	$\sigma(\epsilon)$
1 bit	3	2	15.5%	3.1%	6.6
2 bit	5	3	14.7%	2.9%	7.0
4 bit	9	5	19.8%	5.1%	10.3

Table 3: Accuracy of adder power estimates in comparison with Spice.

6 Conclusion

We presented a new analytical macro-modeling technique for high-level power estimation. Due to its speed and accuracy, our macro-modeling approach is suitable for real-time exploration of architectural alternatives using predefined blocks. The metrics of the proposed model are obtained exclusively from input statistics. Therefore, there is no need of simulations during estimation. Our technique was evaluated on combinational benchmark circuits demonstrating very good accuracy in comparison with zero-delay and Spice simulation estimates. Preliminary results with sequential circuits are equally promising.

References

[1] M. Barocci, L. Benini, A. Bogliolo, B. Riccó, G. De Micheli, "*Lookup Table Power Macro-Models for Behavioral Library Components*", *Proc. IEEE Alessandro Volta Workshop on Low Power Design*, Mar. 1999.

[2] N. Dragone, R. Zafalon, C. Guardiani, C. Silvano, "*Power Invariant Vector Compaction Based on Bit Clustering and Temporal Partitioning*", *Proc. Int. Symp. on Low Power Electronics and Design*, pp. 118–120, 1998.

[3] S. Gupta, F.N. Najm, "*Power Macromodeling for High Level Power Estimation* ", *Proc. 34th ACM/IEEE Design Automation Conf.*, pp. 365–370, June 1997.

[4] S. Gupta, F.N. Najm, "*Analytical Model for High Level Power Modeling of Combinational and Sequential Circuits*", *Proc. IEEE Alessandro Volta Workshop on Low Power Design*, Mar. 1999.

[5] E.D. Kyriakis-Bitzaros, S. Nikolaidis, A. Tatsaki, "*Accurate Calculation of Bit-Level Transition Activity Using Word-Level Statistics and Entropy Function*", *Digest Tech. Papers IEEE/ACM International Conf. on Computer-Aided Design*, Nov. 1998.

[6] P.E. Landman, J.M. Rabaey, "*Black-Box Capacitance Models for Architectural Power Analysis*", *Proc. Int. Workshop Low Power Design*, pp. 165–170, Apr. 1994.

[7] P. Landman, "*High-level Power Estimation*", *Int. Symp. Low Power Electronics*, pp. 20–35, Aug. 1996.

[8] F.N. Najm, "*Transition Density: A Stochastic Measure of Activity in Digital Circuits*", *Proc. 28th ACM/IEEE Design Automation Conf.*, pp. 664–649, June 1991.

[9] F.N. Najm, "*A survey of power estimation techniques in VLSI circuits*", *IEEE Trans. VLSI Systems*, pp. 446–455, Dec. 1994.

[10] S.B.K. Vrudhula, H.-Y. Xie, "*Techniques for Cmos Power Estimation and Logic Synthesis for Low Power*", *Proc. Int. Workshop Low Power Design*, pp. 21–26, Apr. 1994.