

Maximizing the Weighted Switching Activity in Combinational CMOS Circuits under the Variable Delay Model.

S. Manich and J. Figueras

DEE-UPC

Diagonal 647, P9, 08028 Barcelona, SPAIN

email: manich@eel.upc.es

Abstract

A methodology to find the couple of vectors maximizing the weighted switching activity in combinational CMOS circuits under variable delay model is presented. The weighted switching activity maximization problem is shown to be equivalent to a fault testing problem on a transformed circuit. A maximum weighted switching activity is achieved by test vectors covering a selected set of faults of the transformed circuit. Automatic Test and Pattern Generation tools are used to find the maximizing pair of vectors. The validity of the proposal is demonstrated on the ISCAS-85 benchmark circuits and the results show that the simulation time is reduced by an order of magnitude and the estimation of the maximum weighted switching activity is improved in comparison with pseudo-random sample simulation.

1 Introduction

Large peak power consumption has been associated to reliability problems in the CMOS VLSI circuits [1]. Electromigration, hot-carrier degradation and oxide breakdown problems may be caused by high maximum power consumption and are well-known sources of failure in high performance circuits [2]. To avoid some of these failures, the sizing of the power and ground buses needs the worst case maximum power consumption in order to properly size them [2].

The exact solution to this power maximization problem is NP-complete [3] [4] and, in consequence, heuristic methods based on different strategies have been proposed to compute maximum consumption bounds, such as Monte-Carlo simulation, Boolean Function Manipulation, Propagation of Uncertainty Waveforms and Automatic Test and Pattern Generation algorithms.

Monte-Carlo simulations [5] [6] [7] are used at times like tools to bound the power consumption and fre-

quently provide a reference to compare with which alternative methods [8]. The extensive use of CPU time is the main limitation of these methods.

Other methods are based on Boolean Function Manipulation as proposed by Devadas in [3]. The Boolean difference of each node in the circuit is derived and a Weighted Max-satisfiability algorithm is used to cover the maximum on-set of a certain logic function. For small circuits, this method is able to find the exact value of the maximum and the input excitation; however, it becomes impractical for medium and large circuits.

The propagation of uncertainty waveforms through the circuit is used in [9] [2] to estimate the maximum switching activity by counting transitions at every node. The algorithm proposed is linear with the number of gates thanks to the assumption of space-uncorrelated signals. This algorithm gives an upper bound which in some cases is very inaccurate. For small circuits, the estimation may be improved to an almost exact result by using partial enumeration techniques [10], but this requires significant CPU time for large circuits.

Techniques based on Automatic Test and Pattern Generation algorithms have been recently proposed. In [8] the implication and justification strategies of an ATPG algorithm are modified in order to maximize the number of switching in the internal nodes of a circuit. In [11], we translate the maximization problem to a fault detection problem and an ATPG technique is used to find the vectors covering a selected set of faults. The methods are able to treat large circuits in a reasonable computation time. The limitation of these works is the assumption of the zero delay model for the circuit which makes the evaluation of multiple node transitions impossible.

The maximum power consumption changes significantly if the delay and filtering model for the logic circuit is changed. For instance, experiments per-

formed with the ISCAS benchmark circuits show increases ranging from 98,1% to 309,8% in power consumption when the delay is changed from zero to unit delay. This extreme sensitivity has also been noted by [12], [13] and variations of the power consumption due to multiple transitions have been reported in [14] [15] [16], as well.

Different models of gate delays are considered in this paper. The maximum and its excitation vectors can be computed for **zero**, **unit** and **variable delay** models.

The rest of the paper is organized as follows. In Section 2 the statement of the problem and the basic definitions are presented as well as the description of the method, which is based on a circuit transformation and a heuristic maximization algorithm using a commercial ATPG tool. In Section 3, the circuit transformation is described and in Section 4 the heuristic using the ATPG tool is detailed. The proposed method is applied to benchmark circuits and the results are reported in Section 5 for the different delay models. Finally, the conclusions are reported in Section 6.

2 Logical Estimation of the Maximum Weighted Switching Activity in one period

The three main sources of power consumption in digital CMOS circuits are: dynamic capacitive switching power, P_D , the short circuit power, P_S , and the leakage power, P_L . The addition of these three sources of consumption gives the total power consumption, $P = P_D + P_S + P_L$. The dynamic power, P_D , is the main part in today's CMOS technologies [17]. The energy consumed at node i per switching is: $\frac{1}{2}C_i V_{DD}^2$ where C_i is the equivalent output capacitance and V_{DD} the voltage of the power source and the voltage swing of the node [18]. Therefore, a good approximation of the energy consumed in a period is $\frac{1}{2}s_i C_i V_{DD}^2$, where s_i is the number of switchings during the period.

Nodes connected to more than one gate are nodes with higher parasitic capacitance. Based on this fact, and in a first approximation, capacitance C_i is assumed to be proportional to the fanout of the node F_i [8]. Therefore, an estimation of the consumed energy at node i is $\frac{1}{2}s_i F_i c_0 V_{DD}^2$ where c_0 is the minimum output load capacitance.

2.1 Logic level Modeling of Weighted Switching Activity

At logic level, the estimation of the energy consumption requires the calculation of the fanout F_i and the switchings of the node, s_i . The fanout of the nodes is defined by the interconnection of the gates, and

the switchings can be estimated by a logic simulator. Hence, since the product $s_i F_i$ is proportional to the power consumption, the product is used at the logic level as an estimation of the energy and it is named Weighted Switching Activity (WSA) of node i .

Given a circuit G described at the gate level, the activity will depend on the input vectors and the delay and filtering model assumed for each gate of the circuit [12]. Assume that the internal nodes are preset to an initial state by an *anterior vector*, V_a . After this, at time zero, the input is changed by a *present vector*, V_p . During a period of time the internal signals may repeatedly change before achieving the new stable state due to the variable delay models of each gate. Let $N_i t$ be the stabilization time where t is the minimum time interval between two transitions and N_i the number of switchings of node i . Because of these transient states, the internal gates g_i present N_i different logic values, i.e. $\{g_i(0t), \dots, g_i(jt), \dots, g_i(N_i t)\}$. Accordingly, the WSA of the circuit is calculated as

$$A(V_a, V_p) = \sum_i^{all\ gates} \left(F_i \cdot \sum_{j=1}^{N_i} (g_i((j-1)t) \oplus g_i(jt)) \right)$$

where the summation of "XOR equal to 1" is the number of switchings, s_i , when the input vector is changed from V_a to V_p .

In the next subsection, the statement of the problem is presented.

2.2 Statement of the problem

Suppose a circuit, G , described at the gate level. A delay model is assumed for the gates. The maximum energy transition consumption is formulated as the finding of the couple of vectors $\{V_a^*, V_p^*\}$, producing the maximum WSA, A^* .

Since the maximum activity is extremely sensitive to the delay model [12], three cases have been considered in this paper for comparison: **zero delay**, **unit delay** and **variable delay**.

In the **zero delay** model the time domain of the circuit is eliminated and hence an instantaneous propagation of the signals is assumed. As a result, internal nodes either change or remain constant. In the **unit delay** model, the signal propagation is equal in all the gates. Hazards may occur. In the **variable delay** model different delays are assumed for the gates, and they are estimated to be proportional to the fanout F_i of the outputs.

In addition, the load capacitance also behaves like a filter, eliminating the transitions which arrive near in time. In order to model this effect, an *inertial filtering* model has been considered with a filtering window equal to the delay of the gate [6].

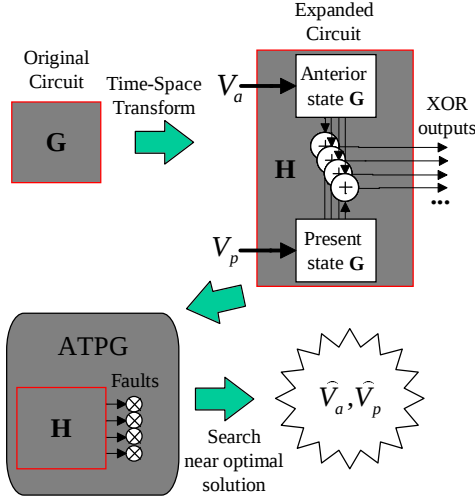


Figure 1: Basic steps in the estimation of the maximum Weighted Switching Activity.

2.3 Method to Maximize the Weighted Switching Activity. General description.

The maximization of the WSA is an NP-complete problem. Consequently, a nearly optimal solution is searched, and the corresponding vectors, $\{\hat{V}_a, \hat{V}_p\}$, are obtained. Two steps are the basis of the method. See Figure 1 for a representation of the main steps.

First, a **time-space** transformation translates the time sequence of the two vectors, $\{\hat{V}_a, \hat{V}_p\}$, of the original circuit, G , into an expanded combinational circuit, H , where inputs $\hat{V}_a$ and $\hat{V}_p$ are simultaneous. Then, two auxiliary input XOR gates are added in the expanded circuit to each internal node and its successive representation in the time-space transformation. The outputs of the XOR gates are the outputs of the expanded circuit.

Second, the **searching** for the pair $\hat{V}_a$ and $\hat{V}_p$ maximizing the WSA is undertaken, using a standard ATPG tool. The maximization of the WSA is equivalent to covering the maximum number of stuck-zero faults of a certain subset of the XOR outputs.

At the following sections, more detailed information about the **time-space** transformation and the **searching** steps is provided.

3 Time-Space transformation of the circuit

The main goal of the time-space transformation is to translate the structure and the time domain of the

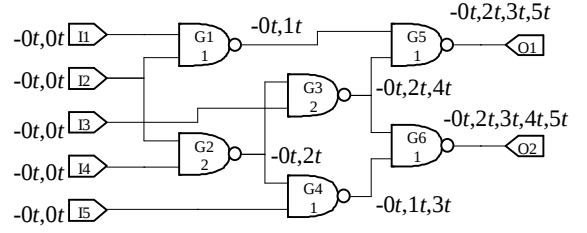


Figure 2: C17 benchmark circuit with transition times for the **variable delay** model. Delays are indicated in the gates.

original circuit into a new expanded circuit with simultaneous signals. In the expanded circuit there is a one-to-one correspondence between the transitions in the original circuit, G , and the XOR outputs of the expanded circuit, H . An illustrative example is presented. The example uses the C17 ISCAS benchmark circuit and the **variable delay** model is considered in the transformation.

In Figure 2, the C17 circuit with **variable delay** model is shown. The value of the delay is written inside the gates and is equal to the fanout of the gates. It is assumed that the anterior vector, V_a , is connected in the inputs prior to $0t$ time. The input is changed to the present vector V_p at time $0t$. The signal propagation times are shown in the signal lines. They indicate the length of the paths arriving at the node. The input transition propagation may produce in gates G1 and G2 one transition, in gates G3, G4 two, in gate G5 three and in gate G6 four transitions.

In Figure 3 the expanded circuit is presented. It has as many expansions as possible propagation times in the original circuit. In this case the expansion covers instants $(-0t, 1t, 2t, 3t, 4t, 5t)$. Each gate has a replica at the corresponding instant of the expansion. For example, look at the gate G5 which has propagation times $(-0t, 2t, 3t, 5t)$. Accordingly, it will have in the expanded circuit copies in the sections $(-0t, 2t, 3t, 5t)$. The inputs of each copy of G5 are connected at the replicas of the fanin gates G1 and G3, i.e. the inputs of the replica G5(3t) will be linked to the copies G1(1t) and G3(2t).

The detection of transitions is made by the auxiliary XOR gates which compare the signals of each gate and its successive replicas. The outputs of the XOR gates are the output of the expanded circuit.

In Table 1 the number of XOR outputs in the expansion of the benchmark circuits is listed for each delay model. Note that this number represents the max-

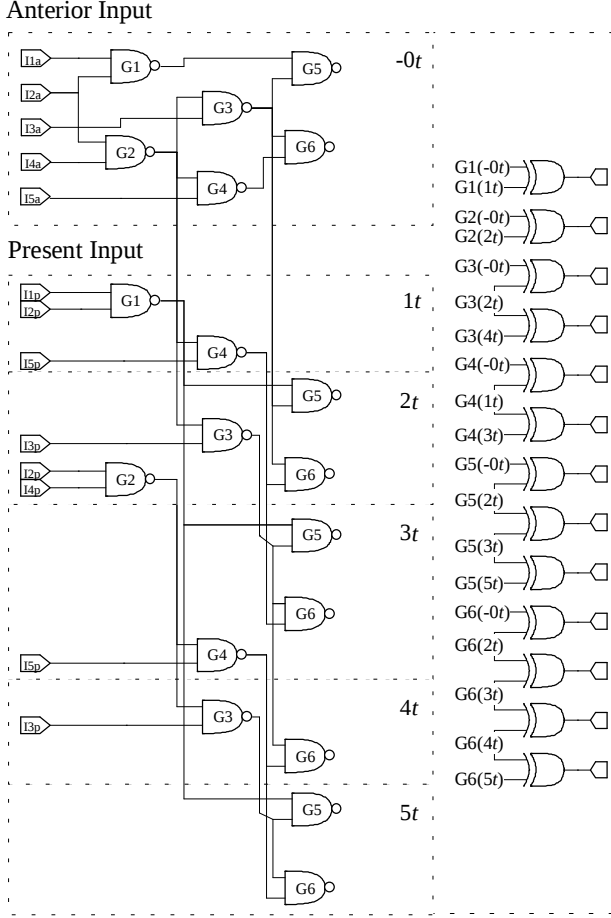


Figure 3: Expanded C17 circuit for the **variable delay** model.

imum number of transitions the circuit may achieve.

In the next Section, the methodology to search for the input vectors to maximize the WSA is described.

4 Heuristic method using the ATPG strategy

After the transformation explained in the previous section, stuck at zero faults are placed in the XOR outputs of the expanded circuit, H , in order to find vectors forcing 1's at these nodes. Then an ATPG tool is used to find an input excitation maximizing the weighted number of 1's. Since the nodes in the original circuit, G , are weighted by the fanout, F_i , the output signals of the expanded circuit will also be affected by the same weights.

The ATPG is driven to the maximum fault detection by forcing simultaneous detection to a selected set

of outputs. The simultaneous detection is achieved by connecting the outputs to an auxiliary AND gate and searching for a "1" at its output. This amounts to detecting a stuck at zero at the output of the AND. In our heuristic method, this subset is modified dynamically in order to move the search for the vector towards the maximum WSA.

The key point is the selection of the subset of outputs for simultaneous detection. A few pseudo-random simulations are made in the original circuit, G , to perform an initial estimation of the switchings, s_i , in the internal nodes. Then the XOR outputs in the expanded circuit, corresponding to the most active internal nodes of the original circuit, are taken as candidates for the subset.

The search algorithm starts with an initial subset of XOR outputs grouped in the auxiliary AND gate. The ATPG is run and the vector detecting this simultaneous fault is stored. The subset is increased and the ATPG is executed again and the new detection vector is stored. This routine ends when the detection fails a preestablished number of times or when the activities achieved by the stored vectors, do not improve the best result.

In the next section the results of the method are presented.

5 Experimental results

Two types of results are presented. First, the maximum WSA of some benchmark circuits is shown and compared to the activities obtained from pseudo-random sample simulation. And, second, the dependence of the maximum WSA input vectors on the delay model is analyzed.

In the first experiment, the maximum WSAs are obtained using the method proposed in this paper. The results are compared with the maximum WSAs

Circuit	Zero	Unit	Variable
C17	6	10	13
C432	160	867	636
C499	202	744	386
C880	383	2297	2112
C1355	546	4944	1938
C1908	880	8777	3331
C2670	1269	6250	7121
C3540	1669	14353	9850
C5315	2307	16127	11016
C6288	2416	95610	64756
C7552	3513	25472	19526

Table 1: Number of auxiliary XOR gates in the expanded ISCAS circuits when using **zero**, **unit** and **variable delay** models.

Circuit	Zero Delay			Unit Delay			Variable Delay		
	Mx	R	%	Mx	R	%	Mx	R	%
C432	184	150	22,7	681	498	36,7	405	350	15,7
C499	211	199	6,0	418	348	20,1	253	233	8,6
C880	374	223	67,7	751	462	62,6	463	455	1,8
C1355	441	362	21,8	1390	1395	-0,4	666	650	2,5
C1908	907	817	11,0	2986	2772	7,7	1867	1864	0,2
C2670	1244	418	197,6	3715	866	329,0	3020	554	445,1
C3540	1189	695	71,1	4873	1996	144,1	3489	1982	76,0

Table 2: Maximum WSA obtained with ATPG method (Mx) is compared to the maximum obtained from random simulations (R) for the **zero**, **unit** and **variable** delay models. Inertial Filtering is assumed for the **variable** delay model.

obtained from the simulation of 10000 pseudo-random vectors; this corresponds in execution time to an order of magnitude slower than an ATPG execution time for the most complex circuit. In Table 2 the results are presented.

In the **zero** delay model, the minimum improvement in front of pseudo-random simulations is 6,0% for the C499 and the best result is for the C2670 with 197,6%. In the **unit** delay model the minimum improvement is for the C1355 where the random simulation gives a better result. The best result is for the C2670 with a maximum increase of 329,0%. Finally, in the **variable** delay model the worst result is for the C1908 circuit with 0,2% and the best is for the C2670 with an improvement of 445,1%.

In Figure 4, the maximum WSAs for the three delay models are compared. In all the circuits, the **unit** delay model gives higher values than the **variable** delay model. The average increase in the maximum WSA when the model is changed from **zero** to **unit** delay is about 225,6% with a maximum increase in the C3540 of 309,8%. The average reduction in the maximum WSA when the delay model is changed from **unit** to **variable** is 31,4% with a maximum reduction in the C1355 of 52,1%.

In the second experiment the variations of the maximum WSA as a consequence of using a simpler delay model for the computation of the input vectors is evaluated. Two cases are analyzed. In the first case, the maximum WSA for the **unit** delay model is calculated using the input patterns of the **zero** delay's maximum WSA. The result is compared to the maximum WSA of the **unit** delay model. In the second case, a similar procedure is followed, but the maximum WSA is calculated for the **variable** delay model using the input patterns of the **unit** delay's maximum WSA. Then, the result is compared to the maximum

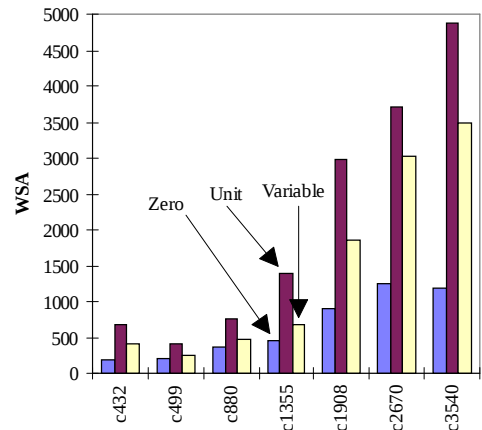


Figure 4: Comparison between the maximum WSA's for the three delay models.

WSA using the **variable** delay model. Table 3 summarizes the results.

Circuit	Unit delay's WSA			Variable delay's WSA		
	Mx	Mx Zero pattern	%	Mx	Mx Unit pattern	%
C432	681	515	24,4	405	157	61,2
C499	418	238	43,1	253	112	55,7
C880	751	520	30,8	463	259	44,1
C1355	1390	832	40,1	666	578	13,2
C1908	2986	1940	35,0	1867	1382	26,0
C2670	3715	3372	9,2	3020	2727	9,7
C3540	4873	2968	39,1	3489	3177	8,9

Table 3: Maximum WSA estimation differences due to the searching for the input patterns with a simpler delay model. Left: **zero** delay instead of **unit** delay. Right: **unit** delay instead of **variable** delay.

These data show that using a simpler delay model produces estimation errors averaging more than 30% for the experimented circuits. Due to the sensitivity to the delay model, a **variable** delay model can not be substituted by a simpler **unit** delay or a **zero** delay to estimate the maximum activity in CMOS circuits.

6 Conclusions

A method to obtain the Maximum Weighted Switching Activity (WSA) of a combinational circuit has been presented. The technique transforms the original circuit under analysis to a new expanded circuit containing the structural and temporal information of the original circuit. Afterwards, stuck at zero faults are placed in the outputs of the auxiliary XOR

gates of the expanded circuit and a standard Automatic Test and Pattern Generation tool (ATPG) is used to obtain the input sequence detecting a selected set of faults. This pattern maximizes the Weighted Switching Activity in the original circuit.

The method is able to maximize the WSA for **zero**, **unit** and **variable delay** models because according to our experiments the maximum WSA and the excitation patterns are extremely sensitive to the delay and they must be obtained using accurate delay models.

Experiments have been performed on ISCAS benchmark circuits. According to the results, the method improves the computational time in one order of magnitude approximately in comparison with pseudo-random sample excitations for circuits with more than 1000 gates. Moreover, the estimation of the maximum WSA made on a set of benchmark circuits improves the estimation an average of 67,8% with a maximum improvement of 445,1% for the C2670 with variable delay model and inertial filtering.

Acknowledgments

This work has been partially supported by the project CICYT-Project No. TIC 94-0561 and by CIRIT Project Ref. 1995SGR 00192.

References

- [1] J. R. Black. Electromigration - A brief survey and some recent results. *IEEE Transactions on Electron Devices*, 16:338+, 1969.
- [2] Anthony Martin Hill. *Switching Density Analysis for Power and Reliability in VLSI Circuits*. PhD thesis, PhD, University of Illinois at Urbana-Champaign, 1996.
- [3] Srinivas Devadas, Kurt Keutzer, and Jacob White. Estimation of power dissipation in CMOS combinational circuits using boolean function manipulation. *IEEE Transactions on Computer-Aided Design*, II(3):373-383, March 1992.
- [4] H. Kriplani, F. Najm, and I. Hajj. Maximum Current Estimation in CMOS Circuits. In *29th ACM IEEE Design Automation Conference*, pages 2-7, 1992.
- [5] M.Xakellis and F.Najm. Statistical Estimation of the Switching Activity in Digital Circuits. In *Proceedings of the 31th ACM/ IEEE Design Automation Conference*, pages 728-733, 1994.
- [6] C. Y. Wang, T. L. Chou, and K. Roy. Maximum Power Estimation Under Arbitrary Delay Model. In *Proceedings of the IEEE International Symposium on Circuits and Systems*, 1996.
- [7] Anthony M. Hill and Sung-Mo (Steve) Kang. Determining Accuracy Bounds for Simulation-Based Switching Activity Estimation. In *Proceedings of the International Symposium on Low Power Design*, pages 215-220, April 1995.
- [8] Chuan-Yu Wang and Kaushik Roy. Maximum Power Estimation for CMOS Circuits Using Deterministic and Statistical Approaches. In *Proceedings of the VLSI Conference*, 1996.
- [9] Chin-Chi Teng, Anthony M. Hill, and sung Mo Kang. Estimation of Maximum Transition Counts at Internal Nodes in CMOS VLSI Circuits. In *Proceedings of the ICCAD*, pages 366-370, November 1995.
- [10] Harish Kriplani, Farid N. Najm, and Ibrahim N. Hajj. Pattern Independent Maximum Current Estimation in Power and Ground Buses of CMOS VLSI Circuits: Algorithms, Signal Correlations, and Their Resolution. *IEEE Transaction on Computer-Aided Design of Integrated Circuits and Systems*, 14(8):998-1012, August 1995.
- [11] S. Manich and J. Figueras. Maximizing the Weighted Switching Activity in Combinational CMOS Circuits. In *Proceedings of the Sixth International Workshop Power and Timing Modeling, Optimization and Simulation*, September 1996.
- [12] Farid N. Najm and Michael Y. Zhang. Extreme Delay Sensitivity and the Worst-Case Switching Activity in VLSI Circuits. In *Proceedings of the 32th Design Automation Conference*, pages 623-627, June 1995.
- [13] A. Ghosh, S. Devadas, K. Keutzer, and J.White. Estimation of Averages Switching Activity in Combinational and Sequential Circuits. In *Proceedings of the 29th Design Automation Conference*, pages 253-259, June 1992.
- [14] L. Benini, M. Favalli, and B. Ricc. Analysis of hazard contributions to power dissipation in CMOs ICs. In *Proceedings of the International Workshop on Low Power Design, Napa Valey, California*, pages 27-32, April 1994.
- [15] Jan M. Rabaey and Massoud Pedram. *Low Power Design Methodologies*. Kluwer Academic Publishers, 1996.
- [16] M.A.Ortega and J.Figueras. Extra power consumed in static CMOS circuits due to unnecessary logic transitions. In *IV International Design Russian Workshop*, pages 40-42, June 27-29 1994.
- [17] A.P.Chandrakasan, S.Sheng, and Robert W. Brodersen. Low-power CMOS digital design. *IEEE Journal of Solid-State Circuits*, 27(4):473-483, April 1992.
- [18] Mehmet A. Cirit. Estimating Dynamic Power Consumption of CMOS Circuits. In *Proceedings of International Conference on Computer Aided Design*, pages 534-537, 1987.