# RTL Level Preparation of High-Quality / Low-Energy / Low-Power BIST

M. B. Santos, I.C. Teixeira and J. P. Teixeira
IST / INESC-id,
R. Alves Redol, 9, 1000-029 Lisboa, Portugal
jct@inesc.pt

S. Manich, R. Rodriguez and J. Figueras
Univ. Politecnica de Catalunya (UPC)
Barcelona, Spain
manich@eel.upc.es

## Abstract

*While high-quality BIST (Built-In Self Test) based on deterministic vectors often has a prohibitive cost, pseudo-random based BIST may lead to low DC (Defects Coverage) values, requiring however very long test sequences with the corresponding energy waste and possible overheating due to extra switching activity caused by test vectors. The purpose of this paper is to discuss how a recently proposed RTL (Register Transfer Level) test preparation methodology can be reused to drive innovative, high-quality / low-energy / low-power BIST solutions. RTL test generation is carried out through the definition of partially defined test vectors (masks) that, while targeting multiple detection of RTL faults lead to high DC values. An energy / power model is proposed to optimize the energy / power consumption of the test at RTL level. It is shown that the proposed method achieves better DC values with low-energy and low-power consumption, when compared to pseudo-random test excitation. The usefulness of the methodology is ascertained using the* VERIDOS *simulation environment in modules of the CMUDSP and TORCH ITC'99 benchmark circuits.*

## 1. Introduction

Product complexity, performance and quality requirements are ever increasing, while power, cost and time-to-market requirements are decreasing. This trend puts a heavy pressure on design productivity and quality, and leads the design process to higher levels of abstraction, and to HDL (Hardware Description Languages). *Low-power design* and *design reuse* techniques are currently being used, as well as IP (Intellectual Property) based methods. As a consequence, *embedded core reuse* also requires *core test reuse* and RTL (Register Transfer Level) test planning and preparation. Moreover, energy and power requirements are becoming very relevant in electronic design. In fact, low-energy operation is needed to extend battery lifetime in portable equipment. Low-power is needed to constrain the temperature of electronic devices under operation. Low-maximum-power is also needed to avoid power rail bouncing, hot spots and electromigration, which limit device reliability. *Low-energy / low-power requirements for the normal operation* mode should go together with *low-energy / low-power requirements in test mode* [1]. Test resource partitioning makes BIST (Built-In Self Test) an attractive solution, provided that high test-effectiveness can be obtained. Test-effectiveness is measured as the ability of the test pattern to uncover likely defects [2]. Accordingly, a test is said to be high-quality if its level of test-effectiveness is high.

The purpose of this paper is to present a methodology for *high-quality / low-energy / low-power BIST* preparation at RTL level. High-quality BIST is ascertained through likely physical *DC* (Defects Coverage) metrics. Low-energy / low-power BIST is accomplished by reducing the number of test vectors and the number of nodes being switched during test application. The methodology for *high-quality / low-energy / low-power BIST* preparation is cost-effective and useful for complex designs, as it is applied at RTL level.

RTL level test generation is carried out through the definition of a reduced set of partially defined test vectors (masks), forcing a limited subset of "care" bits. At this level, the energy is estimated using a model proposed in this paper that is specifically designed for this type of excitation. It uses two parameters, $\alpha$ and $\beta$ to model two energetic costs: first the energy due to the internal activity of the nodes caused by the pseudo-random excitation and second the energy spent to change the state of the internal nodes controlled by masks. This model is evaluated by a proprietary tool, VERIDOS [3].

The paper is organized as follows. In section 2, a review of the RTL level test preparation methodology and tools is conducted. Section 3 introduces the proposed mask-based BIST. In section 4, the model for estimating energy at RTL level is presented. In section 5, optimization of defects coverage, energy and power metrics at RTL level is discussed. Section 6 presents results using ITC'99 benchmarks. Finally, section 7 summarizes the conclusions.

## 2. RTL Test Preparation

In a previous paper [4], the authors showed that test generated at RTL can be rewardingly reused in a production environment to improve the coverage of physical defects. In fact, random pattern-resistant faults, which require prohibitively large numbers of equiprobable patterns or

multiple weighted sets [5], can be detected with significantly shorter test lengths, if test is derived using RTL information. Then, it dramatically reduces the required energy for the BIST session. In a previous paper [6], the authors provided evidence that multiple detection of hard to detect *RTL explicit and implicit faults* leads to the detection of *random pattern-resistant realistic faults* at logic level, that is, hard to detect bridging and open defects. Explicit (implicit) RTL faults are associated with variables explicitly (or not) included in the RTL code. RTL-TPG (Test Pattern Generation) is carried out by defining *partially specified test vectors* (**masks**), which drive the system under test into the functionality visited in a limited set of the input space. We refer to this functionality as **dark-corners** [4].

Test quality of digital systems is frequently evaluated using the LSA (Line Stuck-At) fault model. However, more accurate fault models are used in this paper. The simulation environment uses a commercial design system and `DOTLAB`, a proprietary set of defect-oriented tools, including `LOBS` (the proprietary defect extractor) and `VERIDOS`, which performs mixed-level (behavioral / structural) fault simulation using VHDL (Very high speed integrated circuit Hardware Description Language) or Verilog behavioral descriptions, and Verilog structural descriptions [3]. This simulation tool uses an extension of the biased-voting model for bridging faults, as described in [7]. Hence, gate-level Verilog fault models for bridging and line-open defects, both for interconnection and cell faults, are included in the `VERIDOS` tool for CMOS physical implementations [7]. `VERIDOS` generates RTL fault lists according to the RTL fault models defined in [4], performs mixed RTL / logic level fault simulation and the *WSA* (Weighted Switching Activity) computation (the metric for energy / power estimation) [8] [9]. Additionally, it computes the RTL *IFMB* (Implicit Functionality and Multiple Branch coverage) [10] and layout level *DC* coverage metrics.

## 3. Mask-Based BIST

Low cost BIST solutions require low area TPG, typically pseudo-random TPG. Random pattern resistant faults require that some degree of test determinism be considered. Different approaches have been proposed for random pattern resistant fault detection in digital circuits. These approaches basically perform logic level LSA fault simulation with pseudo-random vectors in order to identify hard to detect faults, which are subsequently detected using weighted random pattern generation [11] [12] [13] or deterministic approaches [14]. However, high LSA fault coverage does not guarantee high *DC* [15]. Moreover, hard accessibility to parts of the structural description is expected to result from the synthesis of functional parts seldom exercised. Nevertheless, this information can be obtained at RTL with low cost fault simulation.

At-speed BIST energy / power consumption can be reduced by means of: (I) vector selection and reduction of the number of vectors applied [14] [16] [17] [18], (II) TPG carried out for low-power BIST [19] [20], (III) circuit activity reduction during shift in the chain of a test-per-scan architecture [21] [22]. The proposed BIST strategy consists in the customization of the pseudo-random test vectors, generated on-chip with a LFSR (Linear Feedback Shift Register) for instance, with partially specified test vectors, referred to as masks. Usually, the number of masks, $R$, is limited, and the number of constrained positional bits, $w_i$ in mask $m_i$, is much smaller than the input word length, $w$. A merit factor $\psi_i = w_i / w$ is defined for each mask. The case studies used as test vehicles are modules of the CMUDSP [23] and TORCH [24] ITC'99 benchmark circuits. As an example, Table 1 shows the limited effort needed to customize pseudo-random patterns for the "pcu control" module (PCU_ctr) and for the "agu control" module (AGU_ctr) from CMUDSP, the "co-processor0" module (Cp0_ctr) and the "Booth multiplier or adder" module (MOAPpsum) from TORCH.

The TPG process is performed in such a way that, after mask generation, as described in [6], the test pattern $V = \{v_0, v_1, \cdots, v_N\}$ is built of $N = \{N_0(PR) + \Sigma N_i(\text{mask } m_i)\}$ vectors, in which $N_0$ are pseudo-random vectors and, for each mask $m_i$, $N_i$ vectors are generated. The unconstrained positional bits of the $N_i$ vectors are filled with the 0/1 values generated by the LFSR. The RTL-based methodology allows good estimations at RTL of the required length of the BIST session, which lead to high *DC* values.

| Module | *R* # masks | *w* # PIs | *w*x*R* | $\sum w_i$ tot. # fix bits |
|---|---|---|---|---|
| PCU_ctr. | 6 | 347 | 2082 | 241 |
| AGU_ctr. | 14 | 35 | 490 | 217 |
| Cp0_ctr. | 3 | 28 | 84 | 16 |
| MOA Ppsum | 3 | 272 | 816 | 270 |

Table 1 - Mask customization for ITC'99 benchmark modules.

In order to perform on-chip pseudo-random vector customization using RTL generated masks, additional test hardware is required, which implies extra silicon area and increased energy / power consumption. Two structural solutions for LFSR bit masking have been proposed and evaluated in [25], taking advantage of the reduced number of constrained bits in the masks. Their implementation is automated through a dedicated tool. However, the usage of masks is flexible and offer other interesting architectures not totally BIST like: (I) BIST only includes the pseudo-random generator and masks are obtained from an external source. This external source could be an ATE (Automatic Test Equipment) or a microprocessor if the BIST is embedded into a SoCs (System on a Chip). (II) BIST only includes the storage of masks and pseudo-random vectors are obtained externally. This configuration could be interesting to protect IP modules since the functional part of the test (masks) is

protected, while the pseudo-random part is obtained externally for example, using scan-path.

In next section the model to estimate the energy / power consumption at RTL level is presented. This model takes profit of the masked pseudo-random nature of the vectors.

## 4. Energy / Power Estimation in Mask-Based BIST

As mentioned above, energy and power consumption are evaluated during BIST mask preparation at RTL level. Different methods and models exist to estimate energy / power at this level of description, like [26] [27] [28]. However, they do not use the specific nature of the problem under consideration. In this paper, the model proposed for test energy / power estimation uses this special characteristic of test vectors to achieve a simplified model that allows very fast estimations. Since the metric used to evaluate the energy / power model is the *WSA* (Weighted Switching Activity) a brief summary on this metric is presented next.

### 4.1. Basic Concepts on the *WSA* Metric

The *WSA* is a metric that is extensively used to estimate energy and power consumption of CMOS circuits at logic level. It counts the number of transitions of internal nodes and makes a weighted addition of these values. The weights are related to the scale of parasitic capacitors associated to each node. This metric neglects any other source of power consumption different from the associated to the switching of the nodes [8] [9].

Assume a circuit excited by a set of test vectors $V = \{v_0, \cdots, v_{k-1}, v_k, \cdots, v_N\}$. If $t_{ck}$ is the time interval between vectors, then $t_k = k \times t_{ck}$ is the time instant when $v_k$ is applied, and thus is the time instant of $k^{th}$ cycle. Based on this fact, the weighted switching activity metric $WSA_k$ can be calculated for the transition between vectors $(v_{k-1}, v_k)$, and is named *cycle weighted switching activity* [29]. From this metric, energy and power consumption can be estimated if following facts are considered.

Let $E_k$ be the energy consumed by the circuit during the transition of input vectors $(v_{k-1}, v_k)$, that is named *cycle energy*. This $E_k$ is proportional to $WSA_k$ if it is assumed that the main part of the energy consumption comes from the switching of the internal parasitic capacitors. Let $P_k$ be the average power consumption measured during the same transition of input vectors, that is named *cycle power*. This $P_k$ is also proportional to $WSA_k$ since the clock period $t_{ck}$ is assumed constant. When $WSA_k$ is integrated through all input vectors, couples $(v_{k-1}, v_k)$ are chained up, the total weighted switching activity $WSA_N$, is obtained and it is proportional to the total energy consumption, $E_N$. Parameter $P_{max}$ is the maximum cycle power, and is proportional to $WSA_{max}$, which is the maximum value of $WSA_k$. The average power consumption of the full test is calculated from expression $P_N = E_N / (N \times t_{ck})$, that is named *total power*, it is also proportional to $WSA_N$ for a given length *N* of the test session. In brief, the following relations can be used.

$$\begin{aligned} E_k &\propto P_k \propto WSA_k \\ P_{max} &\propto WSA_{max} \\ E_N &\propto WSA_N \\ P_N &\propto WSA_N / N \end{aligned} \quad (1)$$

These relations demonstrate that $WSA_k$ is the key point on estimating energy and power consumption at logic level.

To clarify terminology, notice that power is a physical magnitude that indicates the flux of energy per unit time. Power consumption is usually related to device temperature and thus it is given as a quantity averaged at a certain time interval. To abbreviate nomenclature, in this paper the word "power" means always "average power". Notice that cycle power assumes a time interval of a *single* vector transition, while total power assumes a time interval equivalent to the total test application time.

### 4.2. Mask Influence on Energy / Power Consumption

The model presented is specifically designed to operate assuming a pseudo-random based excitation, customized with masks. Thanks to this, the expression of the model can be simplified taking profit of the stable statistical properties of the excitation.

To illustrate the essential idea of the model, first consider the example of Figure 1. It presents the energy consumption of the AGU_ctr module during a test session. This circuit is excited using two types of test vectors: pseudo-random vectors (*normal vectors*) and a given set of masked pseudo-random vectors (*masked vectors*). The x-axis of the plot corresponds to the index of the test vector. Test vectors are sequentially applied to the circuit, according to their index value.

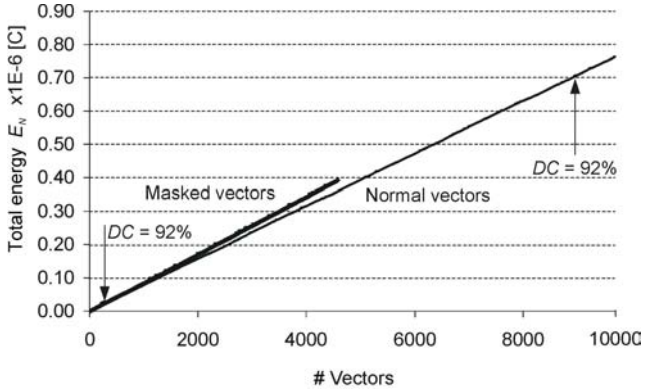


Figure 1 - Energy / power consumption of the AGU_ctr module during a test session. Two types of test vectors are used: pseudo-random (*normal vectors*) and masked pseudo-random (*masked vectors*).

The y-axis corresponds to the total energy $E_N$ estimated using the following expression: $Q = (1/2) \times WSA \times c_0 \times V_{DD}$, where $c_0$ is a minimal node load capacitance (a technological library parameter) and $V_{DD}$ the voltage swing of nodes. Notice that energy *E* is related to charge *Q* through the supply voltage, $E = Q \times V_{DD}$. The plot also shows the point where the *DC* level of 92% is achieved for each type of test.

Let us first focus on the *normal vector* case. An almost straight line beginning at zero and stopping near 0.8 $\mu C$ is observed. This linear shape can be explained as follows: (I) Since the TPG is of type pseudo-random, static probability and transition density of signals are both time invariant [30]. Consequently, the internal nodes of the circuit will present a similar situation as well. According to this, cycle energy $E_k$ will have a stable value if the test length is long enough. These facts explain the linear shape of the energy plot. (II) The total energy consumption $E_N$ is the addition of the individual cycle energies $E_k$ caused by each couple of consecutive test vectors. Since the number of applied vectors is large, these individual amounts of cycle energies become much smaller than the total energy consumption of the entire test. Accordingly, the staircase shape is not appreciable.

Now consider the *masked vector* experiment. In this second case, it is observed that the energy consumption slope slightly increases above the normal case. Moreover, the linear trend is again observed. This is explained by the fact that the mask introduces a change on the switching behavior of internal nodes of the circuit, which modifies its energy consumption profile accordingly. However, once the mask has changed the behavior of the circuit, it consumes again as a circuit excited by a pseudo-random TPG, which accounts for the linearity of the trend.

### 4.3. Mask-Based Energy / Power Model

The situation illustrated in this small example can be extended to a general case. Masks have influence on the slope of the energy (power). This slope may increase or decrease, depending on what parts of the circuit are enabled or disabled. If $\alpha$ parameter is associated to this slope, then each mask $m_i$ will have an associated $\alpha_i$ including the case without mask that has parameter $\alpha_0$. Notice that $\alpha$ is dimensionally equivalent to the cycle weighted switching activity $WSA_k$.

Another case not illustrated in the previous example should be considered in masked pseudo-random excitation. Assume a circuit is kept stable (not switching) because a constant test vector is placed at its input. No energy consumption should be observed in this situation. At a given moment, a mask is switched on and off or two masks are alternatively switched at the input while the test vector is still kept constant. During mask switching, a certain amount of energy consumption is detected that can be explained as the cost of having a new circuit behavior. A different parameter $\beta_{ij}$ is used to model this effect and it represents a given amount of energy necessary to switching from mask $m_i$ to mask $m_j$. Cases $\beta_{0j}$ and $\beta_{j0}$ represent switching on and off of mask $m_j$. This parameter is dimensionally equivalent to the cycle weighted switching activity $WSA_k$ as well.

The complete model combines the two previous parameters $\alpha$ and $\beta$. The expression of the model is as follows

$$WSA_N = \alpha_0 \times N_0 + \sum_{i=1}^{R} (\alpha_i \times N_i) + \\ + \sum_{i=1}^{R} \left( \beta_{i0} \times S_{i0} + \beta_{0i} \times S_{0i} + \sum_{j \geq 1, j \neq i}^{R} \beta_{ij} \times S_{ij} \right) \quad (2)$$

where *N* is the duration of the complete test that is the addition of: $N_0$ the length of the pseudo-random subsequence and $N_i$ the length of the subsequences of customized pseudo-random vectors using masks $m_i$. Parameter $S_{ij}$ is the number of times switching between masks $m_i$ and $m_j$ takes place. Similarly, $S_{0i}$ and $S_{i0}$ are the number of times switching on and switching off are performed by mask $m_i$.

The previous expression (2) can be simplified if some common situations are considered. (I) Mask energies $\beta_{0i}$ and $\beta_{i0}$, which in a general case may be different, can be assumed equal to an average value $\beta_i$ since $S_{0i} = S_{i0}$ is frequently found. (II) Consider the following inequality $\beta_{i0} + \beta_{0j} \geq \beta_{ij}$. It means that, during the switching from mask $m_i$ to mask $m_j$ some activity of the nodes may be overlapped, which would not occur if the switching of masks took place separately. If coefficients $\beta_{ij}$ are substituted by the left side of the inequality, it will be assumed that the calculation of an upper bound of the test energy consumption is made. In many cases, this upper bound will be acceptable. To summarize, the following simplified expression is proposed to estimate the energy / power at RTL level

$$WSA_N = \alpha_0 \times N_0 + \sum_{i=1}^{R} (\alpha_i \times N_i + \beta_i \times S_i) \quad (3)$$

where $S_i$ is the number of times mask $m_i$ is switched on and off.

### 4.4. Estimation of $\alpha$ and $\beta$ Parameters at RTL Level

Estimating $\alpha$ and $\beta$ parameters at RTL level is simple if *static probability* **P** and *transition density* **D** statistics are considered. Static probability is defined as the probability of a node to be equal to logic 1. Transition density is defined as the average number of transitions performed by a node per unit time. Usually these two statistical parameters are independent except in special cases, as it is shown later.

Since $\alpha$ and $\beta$ parameters are dimensionally equivalent to $WSA_k$, this metric can be used to compute former parameters. The computation of $WSA_k$ can be made according to the following expression

$$WSA_k = t_{ck} \times \sum_j F_j \times \mathbf{D}_j \quad (4)$$

where $F_j$ is the weight of node *j* and the summation is extended to all the internal nodes of the circuit. Probabilistic simulators exist which are able to propagate the transition density to all internal nodes of the circuit. Even when no detailed information about gates exists, estimation of the number of gates is appropriately made based on the complexity of the functionality [31] [32]. These types of simulators require of **P** and **D** statistics to be defined for each input node. According to this, $\alpha$ and $\beta$ parameters are obtained from expression (4) after modifying inputs **P** and **D** conveniently. Consider next two cases for each parameter, concerning unmasked and masked primary input nodes *l* of the circuit.

**Estimation of $\alpha_i$**

1. Unmasked nodes. Since these nodes are directly excited by pseudo-random vectors, the value of statistics are extracted from the TPG and thus
$$\mathbf{P}_l^{\alpha} = \mathbf{P}_{TPG}$$
$$\mathbf{D}_l^{\alpha} = \mathbf{D}_{TPG}$$
2. Masked nodes. Assuming that $x_l(m_i)$ is the masked value of input node *l* when mask $m_i$ is present then, the value of statistics are
$$\mathbf{P}_l^{\alpha} = x_l(m_i)$$
$$\mathbf{D}_l^{\alpha} = 0$$

After the definition of input statistics, the probabilistic simulation is executed to calculate the statistics $\mathbf{P}_j^{\alpha}$ and $\mathbf{D}_j^{\alpha}$ of internal nodes. After this, equation (4) is applied to calculate the coefficient

$$\alpha_i = t_{ck} \times \sum_j F_j \times \mathbf{D}_j^{\alpha} \quad (5)$$

with the summation extended to all internal nodes.

**Estimation of $\beta_i$**

1. Unmasked nodes. Since by definition of this parameter, the pseudo-random vector applied is kept constant during mask change, following values are assumed
$$\mathbf{P}_l^{\beta} = \mathbf{P}_{TPG}$$
$$\mathbf{D}_l^{\beta} = 0$$
2. Masked nodes. The value of these input nodes may change when mask is switched on or off. Then, the transition density will be related to the static probability as follows

$$\text{Switching on,} \begin{cases} \mathbf{P}_1^{\beta-\text{on}} = x_i(m_j) \\ \mathbf{D}_l^{\beta-\text{on}} = \dfrac{1}{t_{ck}} \times \left| x_i(m_j) - \mathbf{P}_{TPG} \right| \end{cases}$$

$$\text{Switching off,} \begin{cases} \mathbf{P}_l^{\beta-\text{off}} = \mathbf{P}_{TPG} \\ \mathbf{D}_l^{\beta-\text{off}} = \dfrac{1}{t_{ck}} \times \left| x_i(m_j) - \mathbf{P}_{TPG} \right| \end{cases}$$

Notice that a masked input node may switch if non-masked value is different. Thus, the probability to have a transition is $\mathbf{P}_{TPG}$ if masked value is 0, or $(1-\mathbf{P}_{TPG})$ if masked value is 1. Since the static probability of masked nodes is different with or without mask, $\beta$ is estimated averaging non-masked to masked and masked to non-masked transitions. Thus, after applying probabilistic simulator the statistics $\mathbf{P}_j^{\beta-\text{on}}$, $\mathbf{D}_j^{\beta-\text{on}}$ and $\mathbf{P}_j^{\beta-\text{off}}$, $\mathbf{D}_j^{\beta-\text{off}}$ are obtained for all internal nodes. Then the coefficient is obtained using (4) and thus

$$\beta_i = t_{ck} \times \frac{1}{2} \times \left( \sum_j F_j \times \mathbf{D}_j^{\beta-\text{on}} + \sum_j F_j \times \mathbf{D}_j^{\beta-\text{off}} \right) \quad (6)$$

with the summations extended to all internal nodes.

## 4.5. Model Validation

The model has been compared to values obtained from the VERIDOS simulator. It performs energy estimations at logic level of the circuit, using technological information from the layout extracted by the LOBS tool. Figure 2 shows the results of this comparison.

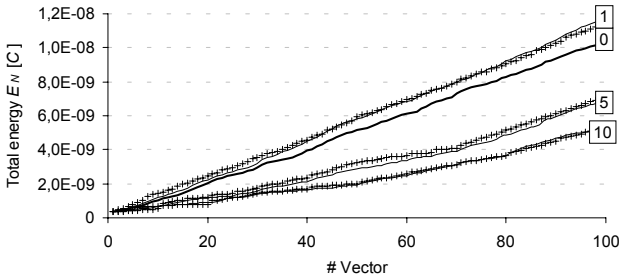


Figure 2 - Comparison between the proposed model (dots) and VERIDOS (lines) in AGU_ctr module. Case 0 is pure pseudo-random excitation. Cases 1,5,10 use same 14 masks with different arrangements.

Four test sequences consisting in a series of 100 vectors are used. Sequence 0 is pure pseudo-random. Sequences 1, 5 and 10 use the same 14 masks combined in a different way. In 1, masks are applied cyclically following the pattern {... pr, pr, pr&msk(i), pr, pr, pr&msk(i+1) ...}. In 5 and 10, masks are sequentially applied from 1 to 14 and, after mask 14, the test continues with pure pseudo-random vectors. In 5, each mask is combined with 5 pseudo-random vectors. In 10, each mask is combined with 10 pseudo-random vectors. As it can be observed, the model accurately predicts both the type of masks used and the pattern followed to apply them.

Table 2 lists the coefficients of model (3). Each line of the table corresponds to a mask, except the first line. Cases 0, 1, 5, and 10 of the table correspond to the same cases of Figure 2. In this table, $\alpha$ and $\beta$ parameters has been obtained by fitting with experimental data obtained from VERIDOS tool.

However, work is under way with probabilistic simulators to make the adjustment as proposed in previous section.

| | α | β | Case 0 | | Case 1 | | Case 5 | | Case 10 | |
|---|---|---|---|---|---|---|---|---|---|---|
| Mask | x10$^{-12}$ [C] | | *N* | *S* | *N* | *S* | *N* | *S* | *N* | *S* |
| 0 (no mask) | 98,15 | | 100 | | 68 | | 30 | | 0 | |
| 1 | 19,09 | 66,50 | 0 | 0 | 3 | 6 | 5 | 2 | 10 | 2 |
| 2 | 9,65 | 133,50 | 0 | 0 | 3 | 6 | 5 | 2 | 10 | 2 |
| 3 | 46,98 | 5,58 | 0 | 0 | 3 | 6 | 5 | 2 | 10 | 2 |
| 4 | 21,14 | 60,25 | 0 | 0 | 3 | 6 | 5 | 2 | 10 | 2 |
| 5 | 19,80 | 43,50 | 0 | 0 | 2 | 4 | 5 | 2 | 10 | 2 |
| 6 | 48,49 | 29,00 | 0 | 0 | 2 | 4 | 5 | 2 | 10 | 2 |
| 7 | 51,93 | 38,51 | 0 | 0 | 2 | 4 | 5 | 2 | 10 | 2 |
| 8 | 48,23 | 15,62 | 0 | 0 | 2 | 4 | 5 | 2 | 10 | 2 |
| 9 | 75,31 | 61,25 | 0 | 0 | 2 | 4 | 5 | 2 | 10 | 2 |
| 10 | 71,75 | 18,49 | 0 | 0 | 2 | 4 | 5 | 2 | 10 | 1 |
| 11 | 34,27 | 24,06 | 0 | 0 | 2 | 4 | 5 | 2 | 0 | 0 |
| 12 | 16,30 | 69,51 | 0 | 0 | 2 | 4 | 5 | 2 | 0 | 0 |
| 13 | 10,80 | 64,55 | 0 | 0 | 2 | 4 | 5 | 2 | 0 | 0 |
| 14 | 19,12 | 91,01 | 0 | 0 | 2 | 4 | 5 | 2 | 0 | 0 |

Table 2 - Coefficients of the mask-based energy / power model presented in Figure 2.

## 5. High-quality / Low-energy / Low-power BIST Optimization

The usefulness of model (3) is found during BIST preparation, in order to make mask selection achieve the low-energy / low-power goal while keeping a high *DC* level (the metric to evaluate the quality of test). Since energy estimation can be made at RTL level, which is the same as the BIST preparation level, greedy strategies can be used for optimization purposes. These greedy strategies become very powerful thanks to the fast evaluation of the mask-based energy / power model proposed and the *IFMB* metric, which is the RTL level indicator for high *DC* values of the final structure.

The optimization criterion is based on the trade-off existing between cycle power $P_k$ and the total number of vectors $N$ necessary to reach a given *DC* level, see Figure 3.

Total energy is proportional to the total number of transitions at test completion (target *DC* reached). Each curve of Figure 3 represents the trade-off between cycle power $P_k$ and total number of vectors $N$ for tests that require the same energy and achieve the same *DC*. Therefore, if cycle power increases then, more faults will be detected at each test vector and less vectors will be required to reach the *DC* level. Accordingly, the contrary situation is found when cycle power decreases. On the other hand, if with a different test the achievement of a the same *DC* level requires a larger quantity of total energy then, the trade-off will become worse. That is, for the same cycle power, more vectors will be required or, conversely higher cycle power will be necessary for the same total number of vectors. In this case, the trade-off curve will move up-right in the plot. Otherwise, if total energy decreases a better trade-off will be found, meaning that the test will require alternatively less cycle power or fewer vectors to reach the *DC* level. Same reasoning could be made for the case where $E_N$ is constant and *DC* level is variable.

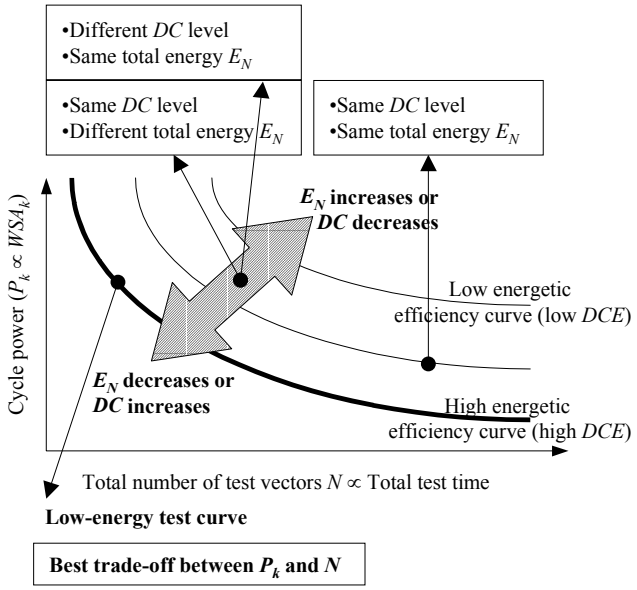


Figure 3 - Trade-off between cycle power $P_k$ and number of vectors $N$, for a given $E_N$ and *DC* level.

The position of the trade-off curve can be viewed as the *energetic efficiency* of the test. If the curve is high, then the energetic efficiency is low and thus most of the energy is not used to detect faults. However, if the position of the curve is low, then the energetic efficiency is high and thus the energy consumption is better used to increase the detection of faults. A ratio *DCE* (Defects Coverage to Energy) can be defined to quantify the energetic efficiency. Its definition is

$$DCE = \left.\frac{\Delta DC}{\Delta E_N}\right|_{\text{Last fault}} \quad (7)$$

and it calculates the slope quotient of the $E_N$ vs. *DC* curve when the last fault is detected. In Figure 5 the definition of the ratio is shown graphically for a typical example. Notice that high *DCE* means high energetic efficiency while low *DCE* means low energetic efficiency.

### 5.1. RTL Level Optimization Strategy

The optimization strategy has a triple objective: achievement of the target *DC*, limitation of the cycle power $P_k$ under a security level and reduction of the total energy $E_N$ (improvement of $P_k$ vs. $N$ trade-off curve and thus length of test). This triple objective is attained during the generation and arrangement of masks in the test sequence. These masks are used to focus the action of pseudo-random vectors in parts of the circuit, "dark corners" of the functionality. Two complementary strategies are applied to obtain the final test sequence: (I) *Generation strategy* that is applied in those parts of the circuit that are functionally dependent or nested. (II) *Arrangement strategy* that is applied in those parts of the circuit that are functionally independent. Figure 4 illustrates graphically these two types of parts.

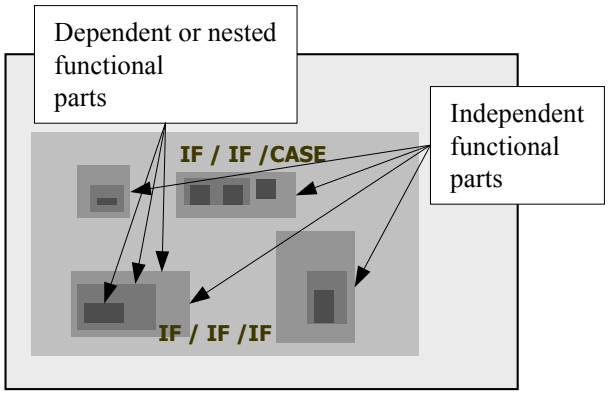


Figure 4 - Different parts of the functionality of a circuit.

BIST preparation applies these strategies in two steps: in the first step the generation strategy and in the second step the arrangement strategy.

**Generation strategy**

In this step, masks are generated for each independent functional part. Since masks must be customized with pseudo-random vectors, the number of pseudo-random vectors $N_i$ will be a function of the target *DC* corresponding to each part. Initially, pseudo-random excitation is applied. If the total energetic efficiency does not decrease excessively, no mask will be generated, contrarily a mask will be calculated. If energetic efficiency is still low, more masks will be forced until the level of energetic efficiency increases above a reasonable level. The criterion to determine if the energetic efficiency is low is based on the *DCE* ratio and it looks if the inequality $DCE < DCE_{PR}$ is fulfilled. The value $DCE_{PR}$ is a reference level which can be selected by the designer. Figure 5 illustrates the evolution of the $E_N$ vs. *DC* when the number of masks increases.

Usually, the value $DCE_{PR}$ is selected to permit a certain level of degradation of the energetic efficiency. This is translated in the usage of more test vectors to excite the circuit than the strictly necessaries, which has the added value of an extra detection of non-modeled faults and thus an increase of the quality of the test.

**Arrangement strategy**

In this second step, the trade-off $P_k$ vs. *N* that has been setup in the previous step is exploited in order to select a suitable level of the cycle power. The level of cycle power is controlled by increasing / decreasing the total number of vectors *N*.

This value *N* can be tuned by applying different arrangements to the masks of independent functional parts. The final test sequence is then constructed based on two possible configurations: *serial or parallel arrangements*. A serial distribution of masks will produce a large number of vectors *N* but a low level of cycle power. Contrarily, a parallel arrangement will produce a shorter number of vectors *N* but a level of cycle power higher than before.

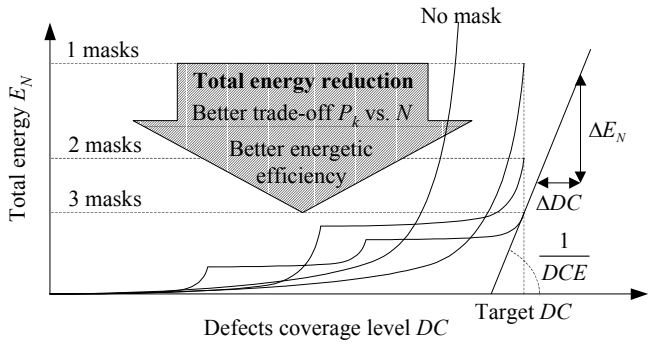


Figure 5 - Illustrative example of the trend of the total energy vs. defects coverage plot during a pseudo-random excitation using customization with 0, 1, 2 or 3 masks.

In order to illustrate the achievement of this triple objective of high-quality / low-energy / low-power BIST preparation, results from experiments performed in modules of the TORCH and CMUDSP ITC'99 benchmarks are presented in next section.

## 6. Experimental results

In this section, the results obtained in modules of the TORCH and CMUDSP are presented. Different test strategies have been used at RTL level. Results from the experiments include the defects coverage *DC*, total energy $E_N$ and cycle power $P_k$ metrics evaluated using `VERIDOS`, `DOTLAB` and `LOBS` tools. Since $P_k$, may largely fluctuate from one vector to the next, the average of this value is presented in the plots giving a softer curve closer to the evolution of the global temperature of the circuit. Jointly with these metrics, the total number of vectors *N* is given as well. This number is limited to 1000 for the AGU_ctr module and to 300 for the PCU_ctr module in order to have a clear view of details.

### 6.1. Results for the AGU_ctr module

A total of 14 masks has been generated for the AGU_ctr module. Six different arrangements and duration of masks have been used to illustrate the evolution of metrics. Cases 0, 1, 5 and 10 use the same patterns and masks as in Figure 2. Cases 15 and 20 are similar to 5 and 10, although here each mask is merged with 15 and 20 pseudo-random vectors. In all cases except pure pseudo-random (case 0), *DCE* metric is kept at a high level, and so the energetic efficiency of masks does not decrease excessivelly.

In Figure 6, *DC* and $E_N$ is presented vs. *N*. The influence of masks is clearly observed in the plots as it is discussed in the following points.

***DC* vs. *N* plot**. (I) Pseudo-random excitation does not achieve the defects-coverage level of 92%. This result would not improve significantly if ten times more vectors would have been applied. Using masks this level goes beyond 96% applying the same number of pseudo-random vectors.

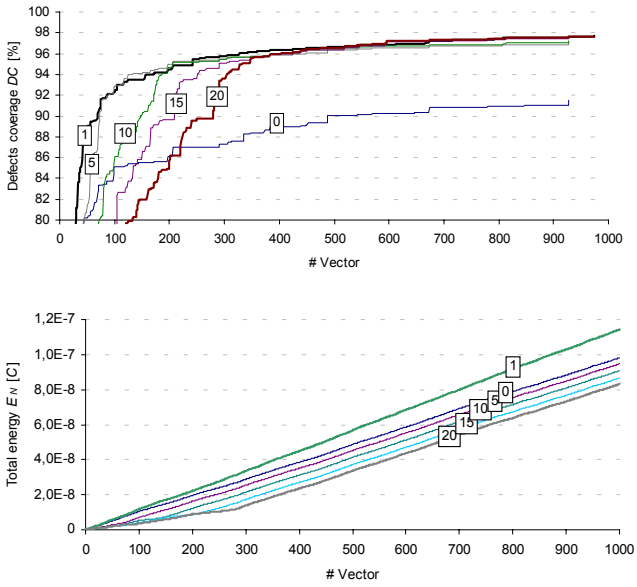


Figure 6 - *DC* level and $E_N$ plots vs. *N* in the AGU_ctr module.

(II) *DC* level increases at different speeds, depending on the arrangement of masks. Case 1 that combines masks cyclically and very quickly is the fastest to rise. It is explained because this arrangement acts as a "pseudo-parallel" configuration that despite not being "pure parallel" it allows the most balanced progression of all masks. (III) Cases from 5 to 20, compared to case 1 behaves like a more serial arrangement since each mask is kept stable during more pseudo-random vectors. This is the reason why from one case to the next more vectors are required to approach a similar *DC* level.

- ***$E_N$* vs. *N* plot**. (I) Case 0 energy consumption follows equation (3) with a single coefficient $\alpha_0$. (II) Case 1 energy increases more rapidly than case 0 because the fast switching of masks overweight $\beta_i$ coefficients of equation (3). This is the drawback of using a pseudo-parallel arrangement instead of a pure parallel one. (III) Remaining cases present an increase of the energy slower than case 0 because $\alpha_i$ coefficients of masks are smaller than $\alpha_0$. $\beta_i$ coefficients are almost unexisting because each mask is applied a single time. (IV) The total energy observed after 1000 vectors is different. These values of energy would change if masks were maintained during more time since $\alpha_i < \alpha_0$. However, care should be taken during the comparison of total energies of tests since *DC* levels are normally different at a given time instant.

In Figure 7 $E_N$ and average $P_k$ are presented as a function of *DC* level. Discussion on the most relevant points follows next.

- ***$E_N$* vs. *DC* plot**. (I) For a given value of *DC* level, say 91.55%, case 0 has spent more energy than other cases (higher $E_N$). Moreover, the *DCE* at this point is lower for case 0 (higher slope), which means that the energetic efficiency is decreasing faster and thus the detection capability is being exhausted before other cases.

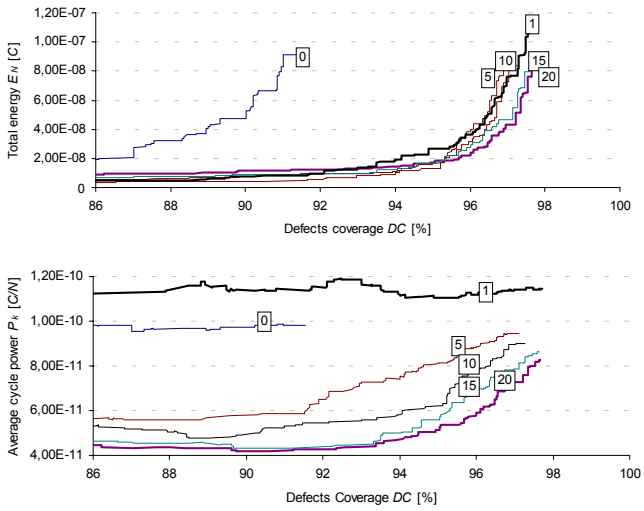


Figure 7 - $E_N$ and average $P_k$ vs. *DC* plots in the AGU_ctr module.

(II) Curves for cases 1 to 20 behave almost the same; the degradation of the energetic efficiency is similar. This means that masks are not totally exhausted and thus they could be used in combination with longer pseudo-random sequences to increase the *DC* level.

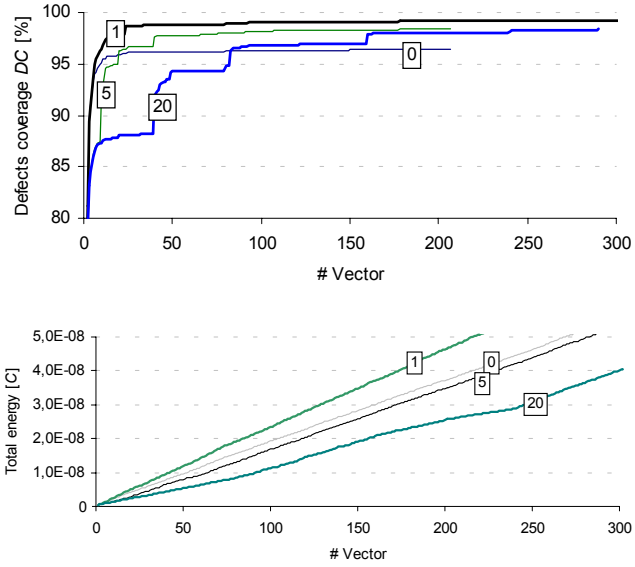


Figure 8 - *DC* level and $E_N$ plots vs. *N* in the PCU_ctr module.

- **Average *$P_k$* vs. *DC* plot**. (I) Case 0 is high power consuming, however this power usage is not translated to a lower *N* neither to a high *DC* due to the low energetic efficiency of the pseudo-random excitation (low *DCE*). (II) The average $P_k$ of cases from 1 to 20 decreases according to the changing of the arrangement configuration from parallel to serial.

### 6.2. Results for the PCU_ctr module

Figure 8 and Figure 9 present results for PCU_ctr module. In order to avoid repetitions, only the most important points will be discussed.

- ***DC* vs. *N* plot**. (I) From cases 1 to 20, parallel to serial arrangements are applied. This is translated to different rising speeds of the *DC* level. (II) In case 20, masks are maintained excessive time and thus their possibility to increase *DC* level is exhausted (they energetic efficiency decreases so much). Notice the staircase shape of the curve.

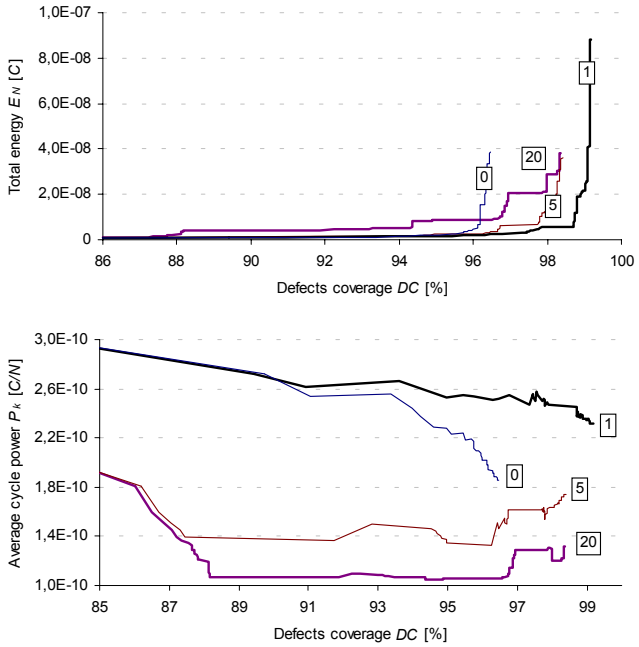


Figure 9 - $E_N$ and average $P_k$ vs. *DC* plots in the PCU_ctr module.

- ***$E_N$* vs. *N* plot**. Extract same conclusions as in previous experiment.

Comments on Figure 9 come below.

- ***$E_N$* vs. *DC* plot**. (I) Again, focussing on case 20, the excessive time each mask is applied can be observed in this plot. Since from a given point masks do not improve *DC* significantly, the curve turns up (*DCE* decreases) at the final stage of each mask. Notice that *DCE* level is restored (increased) with each new mask. This large oscillation of *DCE* makes the overall energetic efficiency of case 20 low. (II) Case 20 energetic efficiency would be improved reducing the duration of masks (compare to case 5).
- **Average *$P_k$* vs *DC* plot**. (I) Despite the low energetic efficiency of case 20, its average $P_k$ is low. Despite this apparent advantage, much more vectors than the strictly necessaries are required to achieve a similar *DC* level. Similar average $P_k$ and *DC* level could be achieved applying fewer vectors to each mask (improve of the energetic efficiency).

Table 3 presents a numerical summary of previous plots, from Figure 6 to Figure 9. Table is divided in two parts. In the top part, a selection of best values is shown (bold numbers). Best value means maximum *DC* level and minimum $E_N$ (Energy label in the table) and average $P_k$ (Power label). In the bottom part of the table, a comparison between test sequences is made. In order to make a correct comparison, *DC* levels are matched to case 0. Once these levels are balanced, $E_N$ and average $P_k$ values are compared. The best cases are indicated with bold numbers. Notice the important reduction of $E_N$ in both modules, -94,57% and -91,26% compared to case 0. Reductions in average $P_k$ with values of -56,47% and -42,89% are also significant. Finally, notice that highering the *DC* level of comparison of the table would lead to different best cases but selected between masked tests.

| | **Selection of Best Values** | | | | | |
|---|---|---|---|---|---|---|
| | AGU_ctr | | | PCU_ctr | | |
| Case | *DC* [%] | Energy [*C*] | Power [*C/N*] | *DC* [%] | Energy [*C*] | Power [*C/N*] |
| 0 | 91,55 | 9,10E-08 | 9,81E-11 | 96,47 | 3,84E-08 | 1,86E-10 |
| 1 | **97,73** | 1,11E-07 | 1,14E-10 | **99,18** | 8,83E-08 | 2,32E-10 |
| 5 | 97,12 | 8,76E-08 | 9,44E-11 | 98,41 | **3,60E-08** | 1,74E-10 |
| 10 | 97,26 | 8,36E-08 | 9,01E-11 | | | |
| 15 | 97,63 | 8,41E-08 | 8,63E-11 | | | |
| 20 | 97,68 | **8,06E-08** | **8,27E-11** | 98,36 | 3,83E-08 | **1,32E-10** |
| Best values | 97,73 | 8,06E-08 | 8,27E-11 | 99,18 | 3,60E-08 | 1,32E-10 |
| | Comparison to PR sequence for same *DC* | | | | | |
| | AGU_ctr | | | PCU_ctr | | |
| Case | *DC* [%] | Energy [*C*] | Power [*C/N*] | *DC* [%] | Energy [*C*] | Power [*C/N*] |
| 0 | 91,55 | 9,10E-08 | 9,81E-11 | 96,47 | 3,84E-08 | 1,86E-10 |
| 1 | 91,65 | 8,54E-09 | 1,14E-10 | 96,47 | **2,52E-09** | 2,52E-10 |
| 5 | 91,60 | **4,94E-09** | 6,04E-11 | 96,47 | 3,36E-09 | 1,46E-10 |
| 10 | 92,15 | 9,37E-09 | 5,45E-11 | | | |
| 15 | 91,69 | 9,21E-09 | 4,37E-11 | | | |
| 20 | 92,25 | 1,22E-08 | **4,27E-11** | 96,51 | 8,90E-09 | **1,06E-10** |
| Comparison | | -94,57% | -56,47% | | -91,26% | -42,89% |

Table 3 - Comparison between different test vector sessions in AGU_ctr and PCU_ctr modules. Case 0 is pure pseudo-random excitation. Cases 1-20 are masked pseudo-random.

## 7. Conclusions

An RTL level based TPG methodology has been used to derive high-quality / low-energy / low-power BIST solutions for digital systems. High correlation of *IFMB* (Implicit Functionality and Multiple Branch) and *DC* (Defects Coverage) test quality metrics allows RTL level TPG to reach a high *DC* value. Soft customization of pseudo-random tests (through masks) leads to high *DC*, low number of vectors, low energy and power comparable (or even lower) to those obtained with pseudo-random test. A model that allows a fast estimation of energy / power at RTL level has been proposed. Thanks to this, the preparation of BIST can be accelerated. Results show that the proposed method of BIST preparation achieves good levels of *DC*, low energy and low power if compared to pure pseudo-random tests. It has also been shown that, for the AGU_ctr module, the application of 14 masks increases the *DC* from 91,55% to 97,68%. If the same *DC* level is assumed, the application of masks allows reduction of total energy and average power by -94,57% and -56,47% respectively.

**Acknowledgments**

This work has been partially funded by CRUP (Portugal) and ME (Spain) under Portuguese/Spanish University Cooperation

Integrated Action: E 36/ 02 and HP01-05, by FCT Fundação para a Ciência e a Tecnologia projecto POCTI/41788/ESE/2001 - LPBIST and by CICYT Ministerio de Ciencia y Tecnología y fondos FEDER proyecto TIC2001-2246.

## References

[1] F. Corno, M. Rebaudengo, M. Sonza and M. Violante, "Optimal Vector Selection for Low Power BIST", *International Symposium on Defect and Fault Tolerance in VLSI Systems,* pp. 219-226, November 1999.

[2] L.C. Wang, R. Mercer, T.W. Williams, "On the Decline of Testing Efficiency as Fault Coverage Approaches 100%", *Proc. IEEE VLSI Test Symp. (VTS),* pp. 74-83, 1995.

[3] M. B. Santos, F.M. Gonçalves, I.C. Teixeira and J. P. Teixeira, "Defect-Oriented Verilog Fault Simulation of SoC Macros using a Stratified Fault Sampling Technique", *Proc. of the IEEE VLSI Test Symp. (VTS),* pp. 326-332, 1999.

[4] M.B. Santos, F.M. Gonçalves, I.C. Teixeira and J.P. Teixeira, "RTL-Based Functional Test Generation for High Defects Coverage in Digital Systems", *Journal of Electronic Testing, Theory and Application (JETTA),* vol. 17, Nº 3/4, pp. 311-319, Kluwer, June/August 2001.

[5] J. A. Waicukauski, E. Lindboom, E.B. Eichelberger and O.P. Forlenza, "A Method for Generating Weighted Random Test Patterns", *IBM J. Research & Development,* vol. 33, no. 2, pp 149-161, Mar. 1989.

[6] M. B. Santos, J. Braga, P. Coimbrão, J. P. Teixeira, S. Manich and L. Balado, "RTL Guided Random-Pattern-Resistant Fault Detection and Low Energy BIST", *Proc. IEEE Design and Diagnostic of Electronic Circuits & Systems (DDECS),* pp.37-43, 2001.

[7] M. B. Santos and J.P. Teixeira, "Defect-Oriented Mixed-Level Fault Simulation of Digital Systems-on-a-Chip Using HDL", *Proc. of the Design Automation and Test in Europe (DATE),* pp. 549-553, March 1999.

[8] G. Kissin, "Measuring Energy Consumption in VLSI: a Foundation", *14th ACM Symposium on the Theory of Computing,* pp. 99-104, 1982

[9] S. Devadas, K. Keutzer and J. White, "Estimation of Power Dissipation in CMOS Combinational Circuits.", *Proceedings of the IEEE Custom Integrated Circuits Conference,* pp. 19.7.1-19.7.6, 1990.

[10] M.B. Santos, F.M. Gonçalves, I.C. Teixeira and J.P. Teixeira, "Implicit Functionality and Multiple Branch Coverage (IFMB): a Testability Metric for RT-Level", *Proc. of the Int. Test Conf. (ITC)*, pp. 377-385, 2001.

[11] H. D. Schnurmann, E. Lindbloom and R.G. Carpenter, "The Weighted Random Test-Generator", *IEEE Trans. Computers,* vol. 24, no.7, pp.695-700, July 1975.

[12] H.-J. Wunderlich, "PROTEST: A Tool for Probabilistic Testability Analysis", *Design Automation Conference (DAC),* pp. 204-211, 1985.

[13] D. Neebel and C. Kime, "Cellular Automata for Weighted Random Pattern Generation", *IEEE Trans. On Computers,* vol. 46, no. 11, pp. 1219-1229, November 1997.

[14] S. Hillenbrand, B. Reeb, S. Tarnick, H.-J. Wunderlich, "Pattern Generation for a Deterministic BIST Scheme", *Proc. ACM/IEEE Int. Conf. on Computer-Aided Design (ICCAD95),* pp. 88-94, 1995.

[15] J.J.T. Sousa, F.M. Gonçalves, J.P.Teixeira, C. Marzocca, F. Corsi, T.W. Williams, "Defect Level Evaluation in an IC Design Environment", *IEEE Trans. on CAD,* vol. 15, nº. 10, pp. 1286-1293, 1996.

[16] F. Corno, M. Rebaudengo, M. Reorda, M. Violante, "A New BIST Architecture for Low Power Circuits", *Proc. IEEE European Test Workshop,* pp. 160-164, 1999.

[17] S. Manich, A. Gabarró, J. Figueras, P. Girard, L. Guiller, C. Landrault, S. Pravassoudovitch, P.Teixeira, M. Santos, "Energy and Average Power Consumption Reduction in LFSR Based BIST Structures", *Proc. Conf. On Design of Int. Circ. and Syst. (DCIS),* pp. 651-656, 1999.

[18] P. Girard, L. Guiller, C. Landrault, S. Pravossoudovitch, "A Test Vector Inhibiting Technique for Low Energy BIST Design", *Proc. IEEE VLSI Test Symposium,* pp. 407-413, 1999.

[19] X. Zhang and K. Roy, "Design and Synthesis of Low Power Weighted Random Pattern Generator Considering Peak Power Reduction", *Proc. Int. Symp. on Defect and Fault Tolerance in VLSI Systems,* pp. 148-156,1999.

[20] F. Corno, M. Rebaudengo, M. Reorda, G. Squillero, M. Violante, "Low Power BIST via Non-Linear Hybrid Cellular Automata", *Proc. Of VLSI Test Symp. (VTS),* pp. 29-34, 2000.

[21] S. Gerstendörfer, H.-J. Wunderlich, "Minimized Power Consumption for Scan-Based BIST", *Proc. Int. Test Conference (ITC),* pp. 77-84, 1999.

[22] S. Wang, S. K. Gupta, "LT-RTPG: A New Test-Per-Scan BIST TPG for Low Heat Dissipation", *Proc. Int. Test Conference (ITC),* pp. 85-94, 1999.

[23] CMUDSP benchmark (I-99-5, ITC 99), http://www.ece.cmu.edu/~lowpower/benchmarks.html.

[24] The Torch processor benchmark, http://www-flash.stanford.edu:80/torch/

[25] M.B. Santos, J. Braga, P. Coimbrão, J.P. Teixeira, S. Manich, L. Balado, J. Figueras, "Low Energy BIST Preparation at RT-Level", *Conf. On Design of Circuits and Integrated Syst. (DCIS),* pp. 451-456, November, 2001.

[26] Q. Qiu, Q. Wu, M. Pedram and C.-S. Ding, "Cycle-Accurate Macro-Models for RT-Level Power Analysis", *International Symposium on Low Power Electronics and Design,* pp. 125-130, 1997.

[27] J. Zhu, P. Agrawal and D. D. Gajski, "RT Level Power Analysis", *Asian and South Pacific Design Automation Conference,* 8A.3, 1997.

[28] S. Gupta and F. N. Najm , "Power Modeling for High-Level Power Estimation", *IEEE Transactions on Very Large Scale Integration (VLSI) Systems,* Vol. 8, No. 1, pp. 18-29, February, 2000.

[29] Anthony M. Hill and Sung-Mo (Steve) Kang, "Determining Accuracy Bounds for Simulation-Based Switching Activity Estimation", *Proceedings of the International Symposium on Low Power Design,* pp. 215-220, April, 1995.

[30] F. Najm, "Transition Density, A Stochastic Measure of Activity in Digital Circuits", *Proceedings of the IEEE Design Automation Conference,* pp. 644-649, June, 1992.

[31] E. M. Sentovich, K. J. Singh, L. Lavagno, C. Moon, R. Murgai, A. Saldanha, H. Savoj, P. R. Stephan, R. K. Brayton, A. S. Vincetelli, "SIS: A System for Sequential Circuit Synthesis", May, 1992, http://www-cad.eecs.berkeley.edu/Software/software.html

[32] F. Najm, "Towards a High-Level Power Estimation Capability", *Proceedings of the International Low Power Design*, pp. 87-92, April, 1995.