A Comprehensive Methodology for Stress Procedures Evaluation and Comparison for Burn-In of Automotive SoC

D. Appello[1], P. Bernardi[2], G. Giacopelli[1], A.Motta[1], A. Pagani[1], G.Pollaccia[1], C. Rabbi[1], M. Restifo[2], P. Ruberg[3], E. Sanchez[2], C.M. Villa[1], F. Venini[2]

[1] *STMicroelectronics, Italy* [2] *Politecnico di Torino, Italy* [3] *TALLINN University*

Abstract: **Environmental and electrical stress phases are commonly applied to automotive devices during manufacturing test. The combination of thermal and electrical stress is used to give rise to early life latent failures that can be naturally found in a population of devices by accelerating aging processes through Burn-In test phases. This paper provides a methodology to evaluate and compare the stress procedures to be run during Burn-In; the proposed method takes into account several factors such as circuit activity, chip surface temperature and current consumption required by the stress procedure, and also considers Burn-In flow and tester limitations. A specific metric called Stress Coverage is suggested summing up all the stress contributions. Experimental results are gathered on an automotive device, showing the comparison between scan-based and functional stress run by a massively parallelized test equipment; reported figures and tables quantify the differences between the two approaches in terms of stress.**

Keywords: **Burn-In, Scan-based stress, functional stress programs, stress coverage.**

1. Introduction

Electronic components involved in safety critical applications must comply with high reliability standards currently set to less than 1ppM failing. Burn-in (BI) is a test procedure, employed during manufacturing, which is specifically conceived to give rise to early latent faults that naturally affect a population of devices. BI testing is characterized by its long duration (several hours) making it a bottleneck of the entire manufacturing process.

Usually, two main types of stress are applied during the BI process: **external thermal** stress, where devices are tested into climatic chambers with temperatures reaching up to 125°C and **internal electrical** stress that targets the activation of internal functionalities of the device during the BI phase. Thermal stress can be defined as an environmental stress that comes from the outside of the device while the electrical stress is an internal stress because it exploits the self-heating phenomenon due to the finite electrical conductivity of metal alloys and the finite thermal conductivity of intra-layer dielectrics [3].

The combination of thermal and electrical stress is capable to accelerate the aging process unveiling affecting weak devices characterized by issues deriving from process variations. Assessing the quality and the effectiveness of the stress tests is not a trivial task. Many scientific works on this field tried to give a formal description of which are the metrics to be used, defining models that well describe the characteristics of the main contributors to aging effects for silicon devices. Besides the fact that aging depends on main factors, scientific studies [1, 3, 4, 6, 9] tend to focus on a restricted number of metrics at a time depending on the kind of stress they are addressing.

This paper focuses on electrical stress and provides a comprehensive methodology for evaluating its strength. The novelty of our study with respect to prior works consists in a strategy that takes into account several kind of measurements: circuit activity, internal temperature and current consumption generated by the execution of a stress procedure. In addition, this methodology allows taking into consideration a specific BI flow and the limitations imposed by the BI equipment. Furthermore, the calculation can be tuned in order to properly address specific failing mechanisms for a given technology, the stress quality measurement proposed allows comparing the different procedures executed during manufacturing BI phase and eventually reducing test times.

The paper is organized as follows. Section 2 is providing some notions about the Burn-In flow and aging phenomenon. Section 3 introduces the SoC stress coverage. Section 4 provides some experimental results based on simulation and execution on real silicon device and section 5 comments the study.

2. Background

2.1. Burn-in process

The BI process is composed of many phases; all of them are crucial for quality and cost effectiveness. The BI tester is an Automatic Test Equipment (ATE) able to drive a large number (up to thousands) of devices at a time; the devices, tested in parallel, are put into climatic chambers capable to control high temperatures for infant mortality screening.

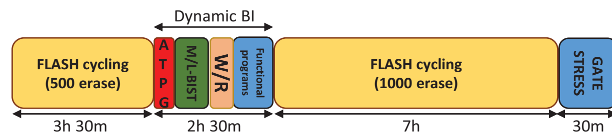

Figure 1. Typical Burn-In flow.

The BI process typically encompasses different stress and test steps that are shown in Figure 1. Test phases are scheduled during the entire session in order to spot any fault excited during stress phases. Most of the BI sections are constituted by short stress procedures that are executed in a loop. This important characteristic has to be taken into account later when stress quality will be assessed. Dynamic burn-in is intended as an internal stress process operated on the device. It is worth to mention that during dynamic burn-in higher supply voltages are applied with respect to the nominal ones. FLASH cycling consists in erasing the whole FLASH memory multiple times, for example 500, and finally checking that the correct functionality is still assured. This step represents the most time consuming phase in the whole process due to long erasing

times imposed by the technology and by reliability engineers' quality requirements.

2.2. Aging, Faults and physical perspective

Different types of defects can affect IC and some of them can be detected during early life stages. From a physical perspective [2] the most common defects are: resistive contacts/vias, resistive opens, resistive bridges, gate oxide shorts, improper implants and silicide breaks. Wafer and package level test can well address these defects, but they may miss failures that can arise only during mission time. In order to give rise to latent faults, aging models have to be carefully studied, understood and proper techniques able to excite them must be put in place during manufacturing tests. The most common aging effects we focused on are exhaustively discussed in previous works [1, 3 and 4]. Techniques for aging acceleration exploit these effects and are based on the variation of three main parameters during the manufacturing test of the Device under test (DUT): temperature, supply voltage and transistor activity [1]. These techniques aim at generating a controlled stress, both thermal and electrical, on the device in order to screen out infant mortalities of ICs. Stress, using general terms, can be defined as the condition under which an electronic device is experiencing electrical and physical degradations [1].

3. Proposed method for stress evaluation

The quality of an applied stress for System-on-chip (SoC) needs to be assessed before being validated from reliability engineers. This is the main motivation that led us to investigate on possible stress metrics enabling an effective evaluation of the stress capabilities of any procedure implemented along a BI flow.

Electrical stress is crucial for complementing the thermal stress externally induced by the climatic chamber of the BI tester. In the following paragraphs, we will discuss and examine three main contributors: switching activity (SW) of the SoC circuitries, temperature distribution on chip surface (T) and current consumption (I). The methodology also takes into consideration the BI flow and the limitations imposed by the BI tester.

Final result is a *Stress Coverage* (*S*) metric for SoC that is the combination of spatial distribution, temporal distribution and intensity of the application for a given stress procedure. In the proposed method, the Stress Coverage (*S*) metric combines two contributions:

- Stress Distribution ($S^{distribution}$): spatial indication about the performed stress procedure;
- Stress Strength ($S^{strength}$): intensity of the applied stress.

Independently of the stress type to measure, we propose to calculate *S* as a weighted sum of $S^{strength}$ and $S^{distribution}$, as in (1):

(1) $$S = \alpha \cdot S^{strength} + \beta \cdot S^{distribution}$$

Parameters α and β are weight values that can be used to tune the equation favouring one kind of stress component more than another. Capability and limitation of the BI equipment are strongly influencing the Stress Coverage calculation; in particular, frequency management and limited bandwidth available impose severe limitations to some kind of stress approaches. For the sake of comparison, stress coverage has to be normalized in time and effective frequency has to be taken into account in the global computation of stress components over a significant period. Current formula (1) should be limited to measurements of $S^{strength}$ and $S^{distribution}$ samples acquired during a certain evaluation time, thus leading to (2).

(2) $$S^{t_eval} = \alpha \cdot S^{strength} + \beta \cdot S^{distribution}$$

S^{t_eval} can then be described as the stress coverage achieved in time t_eval. A decision about the value of t_eval to be selected mainly depends on the type of stress to be evaluated and will be further discussed for specific cases. S^{t_eval} provides a very effective stress quality measurement and allows an intuitive comparison of several kind of stress based on Design-for-Testability strategies or functional approaches. Stress evaluation need to be carried out at the nominal frequency of final application by the BI tester. The next subsection 3.1 provides indications about the calculation of S^{t_eval} for the three kind of stress contributors, namely S_{sw} concerning switching activity, S_{temp} for the internal temperature and $S_{current}$ for current dissipation. Experimental results will compare single stress coverage measurements and a total value calculated as follows:

(3) $$S_{TOT} = \omega \cdot S_{sw} + \tau \cdot S_{temp} + \theta \cdot S_{current}$$

Tuning factors ω, τ and θ technology factors and the specific fault effect that the stress procedure is meant to excite. Section 3.2 illustrates how to take advantage of a stress composition technique during the BI flow.

3.1. Stress coverage calculation

The multiplication of the mean value of observed stress and the maximum value gives the information about the effective intensity of the stress (4), whereas the division of the mean value and the standard deviation gives the information on stress distribution over the die surface (5).

(4) $$S^{strength} = S^{mean} * S^{max}$$

(5) $$S^{distribution} = S^{mean} / S^{std-dev}$$

S^{mean}, S^{max} and $S^{std-dev}$ are computed according to the type of stress that has to be evaluated. Switching activity evaluations and temperature measurements give also information about the stress spatial distribution, thus requiring layout information, while current measurement gives only an absolute value on the chosen test point.

Switching activity analysis

Switching activity measurement is performed through logic simulation by counting the number of transitions observed at gate level for both state transitions and glitches (accounted for the 10), according to equation (6).

(6) $$SW_i^{t_eval} = \#transitions_i + \gamma \cdot \#glitches_i$$

A simple rule to determine *t_eval* is illustrated in (7), which is based on the research of the Lowest Common Multiple of the duration of the stress procedures under evaluation.

(7) $$t_eval \geq LCM\{stress_proc_length_{i=0 \to N}\}$$

In general, this time should not be very long as BI segments are usually composed of short stress procedures that are extensively repeated, even for long times.

In the case, topological information is also considered, the computation leads to a more accurate measurement; for this reason, for each single gate output, we suggest to introduce the contribution of its fan-out as described in (8). The SW_i is weighted by a multiplication with FO_i, which is the number of inputs of successive gates connected to the output of the *i-th* gate, finally obtaining the SWF metric.

(8) $$SWF_i = FO_i \cdot SW_i$$

S^{mean} is the average value of SWF_i calculated on the whole set of gates; S^{max} corresponds to the maximum SWF value observed and $S^{std-dev}$ is calculated as in (9)

(9) $S^{std-dev} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(SWF_i - \mu)^2}, where\ \mu = \frac{1}{N}\sum_{i=1}^{N} SWF_i$

All the calculations proposed above can be performed by restoring to the results obtained by a logic simulation in the gate level netlist. This is already a good stress measure, but it is not providing any spatial information so far. We suggest to use the layout information to precisely locate the switching activity over the chip surface. Dividing the die surface as a matrix of multiple macro-cells, each including several gates, lowers the computational complexity especially when big gate counts are involved. In (10) k is index of the current macro-cell and the $MSWF_k$ is the Macro Switching Activity weighted by Fan-out.

(10) $$\mathrm{MSWF}_k = \frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}} FO_i^k \cdot \mathrm{SWF}_i^k$$

Figure 2 shows the *MWSF* map (over the die area) of two stress procedure A and B. The points on the map correspond to macro-cells and the colour scale reflects the activity strength.

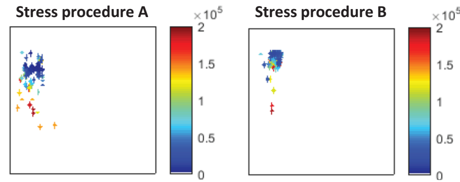

Figure 2. A comparison of the switching activity developed by two stress procedure on the device area. For sake of readability, the map reports only the first 100 most switching macro-cells, thus pinpointing the most activated areas.

Temperature distribution

Temperature stress evaluation is much more complex and requires either very effective estimation models or sophisticated measurement equipment such as thermo-camera. Layout information are crucial in this case, as well as the subdivision of the die area in macro-cells because it is unfeasible to model the temperature for each single gate. Similarly, thermo-cameras have limited resolution which leads to a finite number of measurement points per frame.

Concerning BI of mature technologies, the main objective is to warm up the device as much as possible (within the nominal physical limits). Conversely, temperature gradients result more beneficial for exacerbating physical defects in newest technologies.

Same formulas are still valid for the computation of S^{mean}, S^{max} and $S^{std-dev}$ related to thermal stress. When extensive high temperature application is the main stress criteria, S^{max} is the highest value of temperature observed on the chip surface, S^{mean} is the average of the temperature of the measurement matrix and $S^{std-dev}$ is obtained again according to (9). The proposed calculation may thus be flexibly used for other criteria of evaluation, such as the gradient with respect to a specific area or point to point intra-chip gradients.

When dealing with thermal measurement, the application time needs to be carefully kept into consideration. Time accounts differently for thermal stress measurement, because the calculations are performed on a set of samples (temperature maps) acquired during a defined amount of time. Therefore, *t_eval* value takes only into account the stress measured after *t_warmup*, the time threshold after which a stable temperature of all modules of the chip surface has been reached. Given these timing points, if the stress procedure is applied for a shorter time, the quality of the stress is eventually compromised.

Current consumption

The third and last evaluation is the single point measure of the current during the execution of the stress pattern at package level on the most suitable pin-out grouping. All those data are collected in order to create a dataset that characterises all the stress procedures from different perspectives and helps to find correlation between these measurements. Current consumption measurements are carried out using a multi-meter measuring a single value that may vary depending on which patterns are applied, i.e., functional programs or scan-based ones.

3.2. Burn-In flow and stress composition technique

According to the collected results, quality and reliability engineers can take decisions about the stress to be applied targeting BI time minimization. For each test phase *j* in a set composed by M test phases, we will refer to t_apply_j as the application time budget for test phase *j*.

A practical comment concerning this BI conception phase is that any single evaluated procedure is not sufficient for a complete satisfaction of stress requirements. Either DfT based procedures or functional programs will show heavily stressed areas and regions where the stress is less effective, for this reason, a stress pattern composition is strongly advised. Stress composition technique allows higher stress coverage at the expense of time; therefore, a crucial challenge is to adopt parallel stress solutions that can result into an acceptable BI time. In terms of switching activity, stress coverage of single procedures may be characterized by good $S^{strength}$ but less effective $S^{distribution}$. Figure 3 illustrates the concept by showing the composition of the stress procedures reported in figure 2. The map reports the best MSWF value selected among procedure A and B for each macro-cell, the layout is now covered in a better way and the intensity of the stress is a combination of the two.

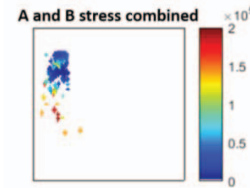

Figure 3. Switching activity developed by a composition of the two stress procedures shown in figure 2.

We propose to calculate the S^{mean} for a composite procedure as in (11):

(11) $$S^{mean} = \sum_{i=0}^{N} \frac{\mathrm{MAX}\{proc_{j=0 \to M}\}}{N}$$

4. Experimental Results

A real case study has been considered in order to demonstrate the effectiveness of our metric; an automotive device powered by a 32-bit processor, on-chip Flash and RAM memories, FPU and many peripheral cores, mainly used for ABS and other critical parts of the vehicle. In order to provide a meaningful set of measurement, we have compared two types of stress: Scan-Based and Functional stress procedures.

All the measures were collected on a debug station that is built to reproducing the same environment of the burn-in tester in terms of driver vector rate performances and connection between the driver and the host PC. The measurements have been taken at ambient temperature (around 25°C) that is different from the temperature in the real test environment (125°C). This difference will not affect the experiment because the temperature developed inside the device does not depend on the external temperature. Concerning the scan-based stress,

the vector rate at which stress patterns are fed to the target depends on the capabilities of the tester; in our case, the maximum frequency for external stimuli is 10MHz. Regarding the functional stress, tester intervention is limited to the loading phase and to the final communication in which devices notify the end of the program. Frequency gap between scan (10MHz) and functional (at speed 128MHz) is crucial for the evaluation and comparison of the stress capacities.

Experimental results demonstrate that DfT approaches like scan based may allow a better distribution of the activity as they simultaneously exercise many cells in the circuit. On the other side, the functional approach may be limited by the locality of a program execution. In our case study, functional stress encompasses 8 programs, each one tailored to stress a specific area of the SoC. Three Fault Class scenarios have been taken into account by carefully selecting values for tuning parameters of equations (1), (2), and (3). Such parameters have been chosen depending on the stress mechanism that the test engineer is focusing on. Class A is temperature oriented, class B focuses on the switching activity and class C is a combination of the previous two.

Table I reports data about switching activity related stress. The SoC is composed by 33345 macro-cells (TOT) and 17803 are considered to be functionally or scan excitable (Active). The value of *t_eval* in our case study is 10ms, thus we have a very high absolute values for max and mean MSWF. It can be noticed that the values for Functional procedures are remarkably higher than Scan. This is due to low frequency of Scan application. Composition achieves best numbers, meaning that an effective stress can be obtained in a longer time.

TABLE I
Switching activity stress measurement with t_eval = 10 ms.

STRESS procedure	MSWF			S_{sw} class A
	MAX	*MEAN*	*STD DEV*	
SCAN	38013	1366	425	0.039
FUNCTIONAL - composition	1874855	81177	3617	3.505

Area Coverage can be computed as percentage of macro cells switched at least one time. Table II provides measurement from this point of view, showing that Scan loses the comparison in the activation of gates with respect to Functional composition.

TABLE II
Area coverage comparison.

STRESS procedure	Switched macro-cells	% on TOT	% on Active
SCAN	7201	22%	40%
FUNCTIONAL - composition	11520	35%	65%

Table III reports data about temperature related stress. It is again considered a grid composed of 33345 macro-cells. The value of *t_eval* in our case study is about 10 min, as at that time the chip surface temperature recorded by a thermo-camera is stable. The *t_apply* time for this kind of stress varies according to the stress and for Functional may range among 5 and 10 min. It can be noticed that again the values obtained for the Functional approach are better than Scan.

TABLE III
Temperature oriented stress measurement with t_eval = 10 min.

STRESS procedure	TEMPERATURE [°C]			S_{temp} class B
	MAX	*MEAN*	*STD DEV*	
SCAN	33.80	30.85	1.14	1.525
FUNCTIONAL - composition	51.38	47.89	1.17	2.412

3D thermal maps in figure 4 visualize the difference between SCAN and Functional - composition from area coverage and temperature intensity point of view.

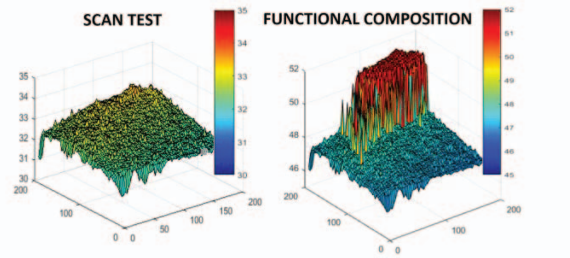

Figure 4. Thermal map comparison after t_apply.

A similar trend is maintained by current consumption measurement, reported in table IV.

TABLE IV
Current consumption measurements.

STRESS procedure	*Current [mA]*	$S_{current}$ *class C*
SCAN	28.82	0.039
FUNCTIONAL - composition	306.08	0.088

Final Stress Coverage values are reported in table V, where they are normalized to 1 for better readability.

TABLE V
Normalised Stress Coverage for the three fault classes.

STRESS procedure	*Fault class A*	*Fault class B*	*Fault class C*
SCAN	0.303	0.045	0.134
FUNCTIONAL - composition	1.000	1.000	1.000

5. Conclusion and Future Works

A metric for stress evaluation, has been introduced. Several types of stress measurement have been taken into account and deeply investigated. Stress coverage makes the comparison among stress procedures easier and helps to devise a scheduling strategy according to the collected final values. The evaluation of the real effectiveness of stress needs large volumes analysis. Monitoring and understanding of production data may represent a feedback to refine stress selection.

6. References

[1] H. Zhang, M. A. Kochte, E. Schneider, L. Bauer, H. J. Wunderlich and J. Henkel, "STRAP: Stress-aware placement for aging mitigation in runtime reconfigurable architectures," Computer-Aided Design (ICCAD), 2015 IEEE/ACM International Conference on, Austin, TX, 2015, pp. 38-45.

[2] P. Nigh and A. Gattiker, "Test method evaluation experiments and data," Test Conference, 2000. Proceedings. International, 2000, pp. 454-463.

[3] M. R. Casu, M. Graziano, G. Masera, G. Piccinini and M. Zamboni, "An electromigration and thermal model of power wires for a priori high-level reliability prediction," in IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 12, no. 4, pp. 349-358, April 2004.

[4] X. Guo, W. Burleson and M. Stan, "Modeling and experimental demonstration of accelerated self-healing techniques," 2014 ACM/EDAC/IEEE Design Automation Conference (DAC), 2014, pp. 1-6.

[5] A. Benso, A. Bosio, S. D. Carlo, G. D. Natale and P. Prinetto, "ATPG for Dynamic Burn-In Test in Full-Scan Circuits," 2006 15th Asian Test Symposium, Fukuoka, 2006, pp. 75-82.

[6] M. d. Carvalho, P. Bernardi, E. Sanchez and M. S. Reorda, "An Enhanced Strategy for Functional Stress Pattern Generation for System-on-Chip Reliability Characterization," 2010 11th International Workshop on Microprocessor Test and Verification, Austin, TX, 2010, pp. 29-34.

[7] D. Appello et al., "Automatic Functional Stress Pattern Generation for SoC Reliability Characterization," 2009 14th IEEE European Test Symposium, Seville, 2009, pp. 93-98.

[8] D. Appello, P. Bernardi, R. Cagliesi, M. Giancarlini and M. Grosso, "An Innovative and Low-Cost Industrial Flow for Reliability Characterization of SoCs," 2008 13th European Test Symposium, Verbania, 2008, pp. 140-145.

[9] A. Amouri, J. Hepp and M. Tahoori, "Built-In Self-Heating Thermal Testing of FPGAs," in IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems, vol.35, no.9, pp.1546-1556, Sept.2016.

[10] J. R. Black, "Electromigration—A brief survey and some recent results," IEEE Trans. on Electron Devices, vol. 16, no. 4, pp. 338-347, Apr 1969.