EasyChair Preprint

№ 1509

IPP: Applying Vstress and defect activation coverage to produce zero-defect mixed-signal automotive ICs

Wim Dobbelaere, Frederik Colle, Anthony Coyette,
Ronny Vanhooren, Nektar Xama, Jhon Gomez and Georges Gielen

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

September 13, 2019

Applying Vstress and defect activation coverage to produce zero-defect mixed-signal automotive ICs

Wim Dobbelaere, Frederik Colle, Anthony Coyette, Ronny Vanhooren
ON Semiconductor, Oudenaarde, Belgium
wim.dobbelaere@onsemi.com

Nektar Xama, Jhon Gomez and Georges Gielen
Department of Electrical Engineering, KU Leuven
Leuven, Belgium

***ASTRACT* - The test and design methodology currently used to develop automotive mixed-signal integrated circuits, is not sufficient to achieve the < 10 PPB quality target. In particular it does not allow to activate and detect all latent defects, which are a significant cause of vehicle failures in the field. This paper discusses whether full burn-in will be needed in order to meet the quality goal, or whether Vstress in combination with a defect activation coverage methodology will do the job.**

I. PROBLEM STATEMENT

The emergence of assisted and autonomous driving not only increases the numbers of integrated circuits (ICs) in a car to well above 400 ICs, but the reliability and safety requirements also result in demanding quality requirements with 10 PPB or fewer test escapes, both in digital and mixed-signal automotive ICs. The introduction of defect-oriented mixed-signal test and design-for-test methods [1][2][3] has in recent years resulted in a significant reduction of the number of application board failures. The current level of IC returns in the automotive IC industry is around the PPM level. The remaining customer returns occur while the IC is already inside an automotive vehicle. Some failures occur immediately after the assembly of the car (0-km failures, representing about 33%), but the majority of the failures (67%) occur in the field after several (thousands of) kilometers of driving.

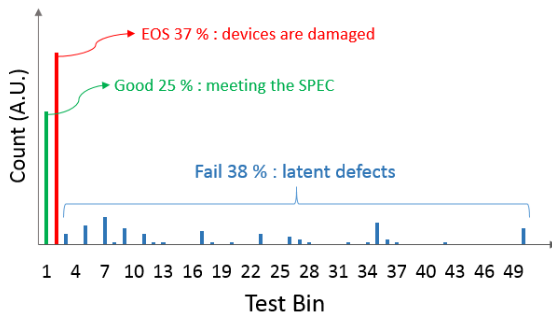

Fig. 1. Test bin analysis of IC vehicle failures (both at 0 km and in the field).

Fig. 1 shows the retest results of a number of returned ICs after vehicle failure (both at 0 km and in the field). On 25% of the ICs no issue was found and the IC was shown to meet the specifications. 38% of the ICs passed the time-zero outgoing tests right after production, but failed on retest after their return. This means that a (partial) defect must have been present in the part, which was not detected at time zero, either because of insufficient test fault coverage or more likely because the defect did not yet express itself in full (i.e. a latent defect). A detailed study on 71 such failures indicates that 42% are due to MOS defects, 22% are due to defects in capacitors, 19% are related to routing defects and 15% are due to via/contact issues, while 19% are due to node-to-node shorts [4]. The remaining 37% of the returned ICs (see Fig. 1) are completely damaged devices (called electrical overstress (EOS) failures), where further analysis is impossible. Several of these are expected to be due to latent defects as well, but the evidence is gone due to the destructive nature of the failure. The number of returned ICs is still too high to guarantee the targeted 10-PPB level, and therefore further improvements to the mixed-signal test and design-for-test (DfT) methodologies are needed, especially to detect latent defects.

This paper investigates what would be the best industrial approach to produce ICs without undetected latent defects. Historically two approaches have been used to signal out ICs with latent defects: burn-in and Vstress. Both approaches will be compared both experimentally and using theoretical models, and the defect activation coverage will be introduced as a driving force to make future mixed-signal automotive circuits meet the <10 PPB industry goal.

II. METHODS FOR LATENT DEFECT DETECTION

A. Latent defect failure mechanisms

A defect is defined as an unexpected difference between a fabricated circuit and its layout. A defect can generate a fault, which in turn can generate a circuit failure. A latent defect is characterized by the fact that it does not generate a fault at time zero, but it generates one after some time of operation. Some common defects are metal mousebites, resistive vias, resistive contacts, metal cracks, pinholes, crystalline defects, and particles. If such defects are not caught at time zero, as for instance an open circuit, they may become a reliability risk as the connection may be degrading over time by electromigration [5]. Electromigration is the transport of material caused by the continuous movement of ions in a conductor. As a result, an open (loss of material) or a short circuit (accumulation of material) can be generated. On the other hand, pinholes, crystalline defects and surface roughness generate a degradation of the insulator properties of the gate oxide or the interlayer dielectric. This accelerates the time-dependent dielectric breakdown (TDDB), which is the complete and sudden loss of isolation of the dielectric material at a given spot. Several equations predicting the time to breakdown (t_{BD})

This research is funded by the Flemish Agency for Innovation and Entrepreneurship (VLAIO) under the NORMA project

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

have been developed. The model presented in [6] is preferred for our analysis, as it allows a direct link to the defect by making use of an effective oxide thickness X_{eff} that changes according to the severity of the defect (e.g. the depth of the pinhole through the oxide). The time to breakdown t_{BD} can then be expressed as :

$$t_{BD} = \tau_0 \exp(G \cdot X_{eff}/V_{ox}) \quad (1)$$

where G and τ_0 are constants that indicate the slope and intercept of the $\ln(t_{BD})$ versus $1/E_{ox}$ plot, and V_{ox} is the voltage across the oxide. The parameters G and τ_0 can be measured at room temperature, but are temperature dependent. Therefore, to predict t_{BD} accurately, it is necessary to include:

$$t_{BD} \propto \exp(E_a/kT) \quad (2)$$

where E_a is the activation energy, k is the Boltzmann constant, and T is the junction temperature in kelvin.

B. Industry approach to detect latent defects

Testing for latent defects in industry today is primarily performed using two techniques: burn-in and Vstress. Both methods at their core aim to activate latent defects, which are then screened out by subsequent testing. The difference between burn-in and Vstress lies in the applied conditions and the necessary equipment. Burn-in requires a furnace that can contain the dies, which are stressed at nominal voltage and at temperatures above 100°C for several hours, causing all potential latent defects to be activated at the same time. The required burn-in conditions, on top of the testing time and the effort needed to put the dies under these conditions, lead to significant costs, making this technique very expensive. Furthermore, there is the risk that the dies experience thermal and mechanical stress that can cause permanent damage (see examples in Fig. 2). The second method, Vstress, is performed on a general-purpose Automated Test Equipment (ATE) and the testing time is in the range of milliseconds to seconds. This results in a low test cost. A key difference is that a much smaller part of the circuit is stressed effectively, requiring a structured design-for-test approach to get the stress voltage applied at all critical locations in the circuit. Also, Vstress cannot be applied for a too long time, in order not to jeopardize the overall lifetime of the IC.

C. Theoretical comparison of the burn-in and Vstress methods

As described by equations (1) and (2), the temperature and the voltage have an influence on the time to breakdown. In the case of burn-in, the voltage plays a limited role as the IC is subjected to its mission conditions, resulting in long activation times (hours). In contrast, in the case of Vstress, the IC is operated at a higher supply voltage, in one or several operational conditions, possibly enabled by DfT circuitry. As illustrated in Fig. 3, the voltage stress dramatically speeds up the activation and therefore is more effective to reach the final aim of forcing the latent defects in the oxides to break down before testing. The figure illustrates the equivalent Vstress time needed to obtain the same activation as burn-in.

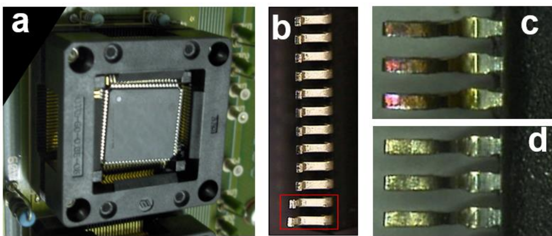

Fig. 2. The burn-in process is semi-manual, leading to tilted devices (a) which can cause bent leads (b). Also, the long exposure at high temperature may lead to solderability issues as a result of oxidation ((c) versus (d)).

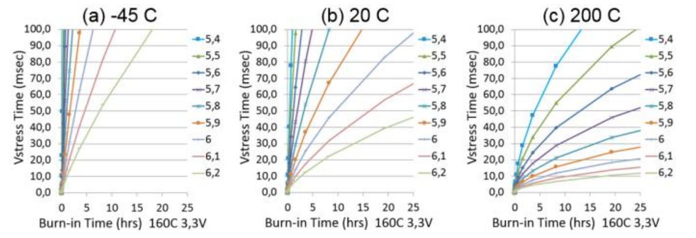

Fig. 3. Equivalent time to get the same activation as burn-in, indicating the strong acceleration by voltage and temperature during Vstress.

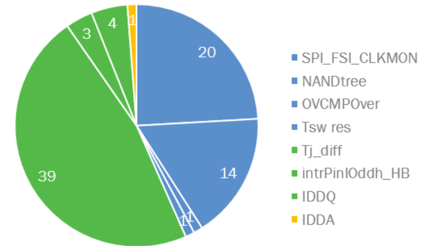

Fig. 4. Burn-in test (24 hours at 120°C) on 103000 ICs during the safe launch of a new automotive product has resulted in 83 failures.

III. Experiments and analysis of burn-in and Vstress

A. Case study : burn-in during the launch of a new product

During the safe launch phase of a new automotive product, 103000 devices in total have been subjected to burn-in at 120°C for a 24-hour time period. All devices were fully tested prior and after burn-in with the same test program.

Fig. 4 shows an overview of 83 encountered burn-in rejects. After electrical verification 36 rejects were found to be caused by test instabilities (blue color). The remaining confirmed electrical rejects have been submitted to failure analysis and extensive data review. A total of 39 devices failed due to a too large difference between the calibrated internal temperature measurement and the temperature as derived from an ESD protection diode. Fig. 5(a) depicts 4 devices at post-burn-in which seem to have failed without any commonality: the handler temperature was found to be stable and the test sites were behaving similarly. When reviewing the pre-burn-in data, all failing devices were found to be originating from a single test site. Only for a specific combination of handler and load board occasionally outliers were observed for the test bus leakage measurement. This additional leakage is not compensated for during calibration of the internal temperature calibration, thus causing incorrect temperature coefficients to be zapped during the pre-burn-in. At post-burn-in these devices with incorrect temperature coefficients appear as tailing devices on the difference between the internal and the ESD-based measurement (TJ_DIFF), with some devices exceeding

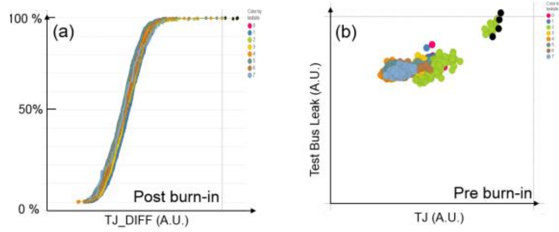

Fig. 5. Incorrect zapping of Test Site 2 during pre-burn-in testing (b) is causing unjustified post-burn-in failures (a).

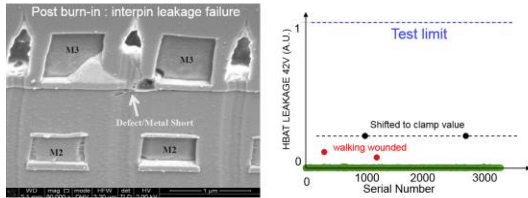

Fig. 6. Post-burn-in interpin leakage failures could have been caught pre-burn-in after Vstress, but the measurement was clamped.

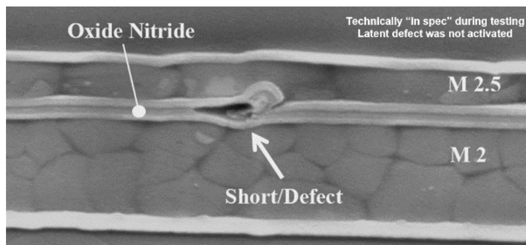

Fig. 7. Latent defect in a MIMC capacitor activated during burn-in.

the allowed difference between both. After accounting for test bus leakage and performing outlier screening, no more of these rejects were encountered. Three devices were rejected post-burn-in due to excessive leakage on a particular pin. Fig. 6(a) shows a metal-to-metal defect in a capacitor structure which is part of the affected pin. This is a type of defect that is typically targeted by Vstress. Fig. 6(b) reveals that these devices (black dots) already broke pre-burn-in while applying Vstress on the affected pin. Unfortunately, the measurement clamps these devices within the test limits and therefore these broken devices were submitted to burn-in. The presence of two additional outliers in this plot (red dots in Fig. 6(b)) illustrates the necessity of statistical post-processing to prevent "walking wounded" devices to be delivered to customers. Finally, only one true burn-in failure was encountered during this evaluation. This failure was caused by a latent defect on a plate capacitor (see Fig. 7). While this device failed the test program used for post-burn-in screening, it still performed within the customer specifications.

B. Case study : Vstress

Vstress consists of applying a voltage to a node in the circuit under test (CUT) that is higher than the nominal level. Usually this is done at the supply voltage node. In this way, the voltage at the other nodes is also increased and thus many devices are stressed. This increase, however, depends heavily on which devices are used in which circuit topology. To increase the effectiveness of Vstress in stressing more transistors, topology modification has been proposed in [8], where a number of "stress" transistors used only during

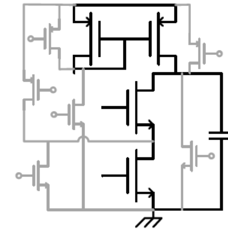

Fig. 8. Latent defects can be activated by adding additional stress transistors to a circuit, used only during testing [8].

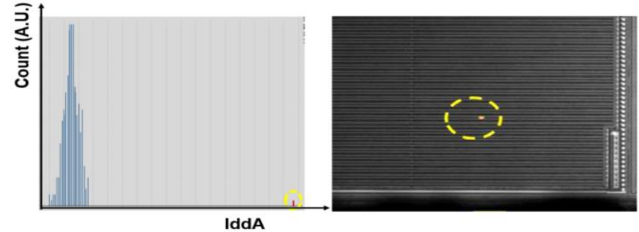

Fig. 9. The power-down IddA current is increasing significantly after a 100-mseconds Vstress of 5.4 V for one device that has a gate-source defect.

Vstress, have been added between well-chosen internal circuit nodes and the supply or ground rail. An example of this technique is shown for a simple, illustrative circuit in Fig. 8: the extra stress transistors are colored in gray and are connected between an internal node and one of the two supply rails. Depending on the required activation coverage, some of these extra transistors can be removed. Depending on the circuit and on which node is connected, these stress transistors should also be made bigger, resulting in a larger overhead area. These considerations combined result in a three-dimensional trade-off between activation coverage, test time and silicon area cost [8]. In Fig. 9 an example is shown of a gate-source defect that is detected by measuring the power-down IddA current before and after Vstress. In this particular case a large high-voltage LPDMOS is subjected to two stress conditions: (a) gate-source stress by setting 5.4V as V_{GS}; (b) gate-drain

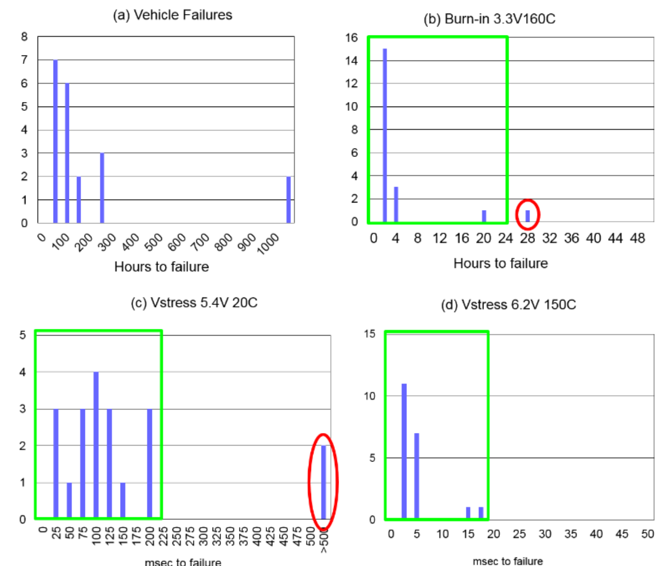

Fig. 10. Equivalent burn-in time and Vstress time to obtain the same activation as actual field failures (a-c) By using a sufficiently high Vstress voltage (6.2V) in combination with a high temperature (150°C) it is possible to achieve a zero-PPB failure rate (d) in this experiment.

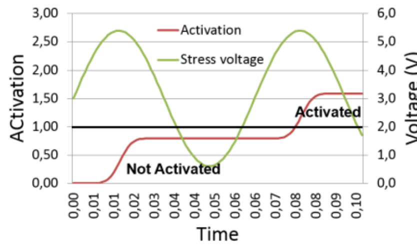

Fig. 11. Illustration of how a sinusoidal stress voltage is cumulatively activating latent defects.

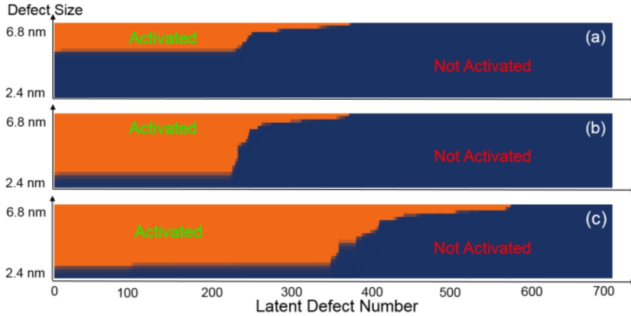

Fig. 12. Defect activation coverage diagram showing the activation status versus defect size for all ~700 potential latent oxide defects in a communication block of an automotive circuit: (a) Vstress at 3.3V, (b) Vstress at 6.2V, (c) Vstress at 6.2V with improved toggling (DfT).

stress by setting the LPDMOS off (V_G=V_S=maximum rating voltage) and the drain to 0V. The tests before and after stress aim at detecting an increase in current consumption for: (a) the drain-source leakage of the LPDMOS, (b) the current consumption of the LPDMOS pre-driver in its ON state, which channels a possibly activated gate leakage.

C. Comparison between burn-in and Vstress

In order to complete the comparison between burn-in and Vstress, the data have been collected from 18 customer returns that were analyzed to be caused by a latent oxide defect. Assuming that the cars were driving at an average speed of 50 km/h, the time needed to activate the latent defects and the time to failure has been calculated and is plotted as a histogram in Fig. 10(a). Using equations (1) and (2), X_{eff} has been calculated, which allows to calculate the corresponding activation time by burn-in (Fig. 10(b)), and Vstress (Fig. 10(c) and Fig. 10(d)). We conclude that it would have been possible to catch all returns if a sufficiently high Vstress voltage (6.2 V) would have been applied at a slightly elevated test temperature (150°C). In comparison, even with a full burn-in of 24 hours, not all defects would have been activated.

IV. DEFECT ACTIVATION COVERAGE

The main gap in the current industry practice is that there is no structured method to aim for certain latent defects. During burn-in a limited device toggling is executed, but the activation impact on the latent defects is merely limited to seeing the nominal voltage for a long time at an elevated temperature. The Vstress method is much more oriented towards catching a particular defect, but in many cases the designer needs to manually check if a certain node is actually feeling the Vstress voltage or not. A new test methodology is therefore needed that allows to simulate all latent defects and find out if all particular defects are sufficiently activated within the applied Vstress test time. Based on the model of the physical underlying breakdown mechanism, and using an actual defect simulator, the time to breakdown can be simulated at the component level. For each latent defect, an activation parameter can be calculated as:

$$\textit{Activation} = \int_0^{\textit{Test Time}} \left[\frac{1}{MTF}\right] dt \tag{3}$$

where MTF is the function modeling the breakdown over time of the latent defect, as presented in section II.A. The integration calculates the total activation, or the effectiveness, of the particular test applied to the circuit. In case the calculated activation reaches 1, it is established that the latent defect is effectively stressed by the test and will be activated during the Vstress production test. Fig. 11 illustrates the integration of this activation on a sinusoidal Vstress waveform. Based on this calculated activation for a sweep of the defect size (the complementary of X_{eff}), an activation diagram can be drawn, as illustrated in Fig. 12 for the communication block of an automotive IC. Such diagram allows to estimate and compare the effectiveness of a test strategy in terms of detecting latent defects. It can be seen that Vstress applied with a sufficiently high voltage, in combination with DfT techniques to allow defect toggling, results in a high activation coverage.

V. CONCLUSION

The automotive industry requires test and design-for-test methodologies that allow to produce mixed-signal ICs that are free of latent defects. This paper has described a method that offers such zero-defect perspective. The industry-used traditional methods of burn-in and Vstress have been compared, and it has been demonstrated that burn-in by itself will never allow to produce defect-free circuits. In contrast, the Vstress method, applied at a sufficiently high voltage and temperature, and combined with structured DfT methods and defect activation coverage simulations, promises to achieve a near-perfect latent defect coverage of oxide defects.

REFERENCES

[1] B. Kruseman et al., “Defect Oriented Testing for analog/mixed-signal devices,” in proceedings IEEE ITC, pp. 1–10, 2011.

[2] B. Esen, A. Coyette, N. Xama, W. Dobbelaere, R. Vanhooren, and G. Gielen, “A very low cost and highly parallel DFT method for analog and mixed-signal circuits,” in proceedings IEEE ETS, pp. 1–3, 2017.

[3] A. Coyette, B. Esen, R. Vanhooren, W. Dobbelaere, and G. Gielen, “Automatic generation of autonomous built-in observability structures for analog circuits,” in proceedings IEEE ETS , pp. 1–6, 2015.

[4] W. Dobbelaere, R. Vanhooren, N. Xama, A. Coyette, B. Esen, G. Gielen, “Testing for latent defects in the analog : does the spec matter ?” invited presentation at IEEE ITC, 2017.

[5] J. Segura et al.,“Parametric failures in CMOS ICs - A defect-based analysis,” in IEEE ITC, pp. 90–99, 2002.

[6] R. Moazzami, J. C. Lee, and C. Hu, “Temperature acceleration of time-dependent dielectric breakdown,” IEEE Transactions on Electron Devices, vol. 36, no. 11, pp. 2462–2465, 1989.

[7] N. Xama, A. Coyette, B. Esen, W. Dobbelaere, R. Vanhooren, and G. Gielen, “Automatic testing of analog ICs for latent defects using topology modification,” in proceedings ETS, pp. 1-6, 2017.