Unintrusive Aging Analysis based on Offline Learning

Frank Sill Torres*†
*Dept. of Electronic Engineering
Universidade Federal de Minas Gerias
Belo Horizonte, Brazil

Pedro Fausto Rodrigues Leite Jr.*
†Institute of Computer Science
University of Bremen
Bremen, Germany

Rolf Drechsler†‡
‡Cyber-Physical Systems
DFKI GmbH
Bremen, Germany

***Abstract*— Runtime aging analysis of integrated circuits enables adaptive approaches in order to enhance the system's life time and permits the user to be aware of critical states. Common approaches utilize sensors that are integrated invasively into critical paths or report experienced aging. This work presents a lightweight supportive technique that correlates environmental and internal conditions with learned data in order to predict the actual wear-out of the system. Simulation results indicate the feasibility of the approach with prediction errors below 10%.**

Keywords*— *Aging, Reliability, Remaining Useful Lifetime, NBTI, TDDB

I. Introduction

In the current era of nanometer scale technologies, consideration of *circuit aging* is of rising importance in order to assure availability and reliability of integrated systems. Related mechanisms not only lead to degradation of circuit performance over time, but also might result into failing components. For example, wear-out effects like *Hot Carrier Injection* (HCI) and *Negative / Positive Bias Temperature Instability* (NBTI/PBTI) mostly affect performance and timing behavior [1]. Effects like *Electromigration* (EM) and *Time Dependent Dielectric Breakdown* (TDDB) might even lead to sudden delay increase or failure [2].

NBTI, dominant *aging* effect in latest technology nodes, increases the threshold voltage of PMOS devices that are stressed with negative-biased gate voltage [2]. HCI, which also shifts the transistor threshold voltage, results from source-drain voltage stress [1]. TDDB, based on continuous voltage stress over the thin gate oxide, increases the threshold voltage and might result in permanent device failures [1]. EM is provoked by current density stress and might result in higher wire resistance or even permanent faults [3]. It should be noted, that dominating factors for most of the introduced mechanisms are the supply voltage, temperature, signal probabilities and switching activity [3].

Principal countermeasure against the impact of *aging* are guard-banding, circuit hardening, adaptive voltage and frequency scaling, as well as load balancing [4, 5]. However, most of these strategies require adequate *aging* prediction and/or monitoring. Therefore, following main approaches can be identified:

- *In-situ slack sensors*: Circuitry that is added invasively to selected critical paths in order to detect or preview failing timing constraints [2].
- *Online self-testing*: Systems enter a test-mode in which Built-In Self-Test (BIST) routines or adaptive tests are executed in order to determine the current degradation [6].
- *Aging sensors*: Reference circuits that are spread over the system and report experienced *aging* in terms of delay degradation or parameter shift [1].
- *Stress monitors*: Sensors for monitoring of stress factors like temperature, voltage, or workload are distributed over the system and its data are accumulated during runtime. The extracted data are then correlated with *aging* models for determination of system degradation [7].

In-situ slack sensors and *online self-testing* offer accurate prediction of actual circuit timing behavior. The costs, however, are area offset and reduced performance due to insertion of logic into critical paths and/or required downtimes of the system. Further, both techniques fail to predict failures that occurred outside the monitored paths. In contrast, *aging* and *stress monitors* offer extraction of degradation effects of the whole system without alterations of the critical paths. This comes, though, at the costs of reduced prediction quality as both only capture partly the stress on the monitored circuit.

This work presents a strategy for stress monitoring and *aging* prediction based on learned data that requires no additional logic in the critical paths. In contrast to related works, the proposed methodology considers *aging* profiles with higher granularity and considers also permanent faults. Please note that this approach is understood as supportive technique that shall not replace on-chip sensors, but offer an additional measure to monitor *aging* effects in integrated systems.

The rest of the paper is organized as follows. Section II describes the idea of this work. Section III presents simulation results and Section IV concludes this work.

II. Proposed Method

The reliability of a system can be quantified via its *Mean Time To Failure* (MTTF), which is the average time the system runs until it fails. At run-time, the *Remaining Useful Life* (RUL) defines the expected time the system still works until it fails [8]. Usually, the RUL follows from the difference between the current lifetime of the system and its MTTF, which is the same for all samples of a system. However, this estimation ignores that each system is subjected to individual stress, like high temperatures, switching activities, or elevated supply voltages.

978-1-5386-0362-8/17/$31.00

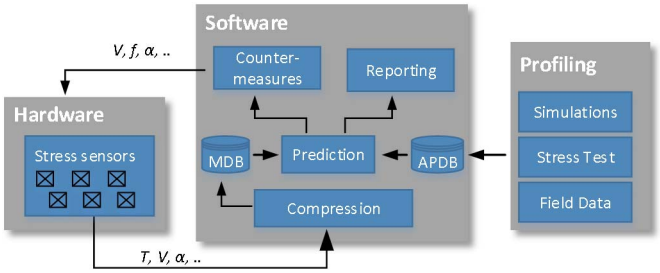

Fig. 1. Proposed approch for *aging* analysis

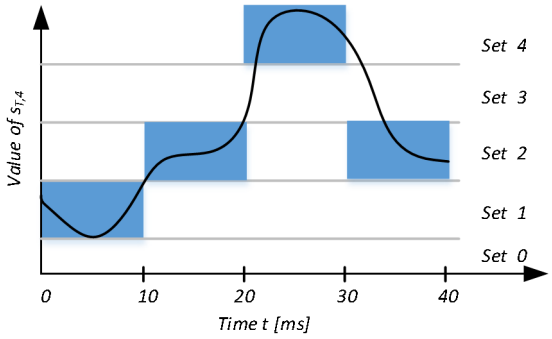

Fig. 2. Compression of measured data of a sensor $s_{T,4}$ into value *sets*

The principal idea of the proposed approach is depicted in Fig. 1. It consists of following sections:

- *Hardware*: Contains *stress sensors* that monitor stress parameters at run-time.
- *Software*: Contains a *Compression* instance for the measured data, which stores its result in a database for measured data (*MDB*), the actual *Prediction* of the RUL based on correlation between measured data and *aging* profiles stored in an *Aging Profile DataBase* (APDB), and instances that apply the results for *Countermeasures* and *Reporting*.
- *Profiling*: Estimation of *aging* profiles due to simulations, data extraction from post-silicon accelerated stress testing, and/or data extraction from real systems in the field.

In the following, the principal elements of the approach shall be detailed.

A. Stress sensors

Stress sensors can be temperature sensors, sensors that measure the supply voltage, or sensors that monitor the activity based on data of the circuit's primary inputs (PI) or pseudo-primary inputs (PPI) [7]. All sensors $s_{k,i}$ are grouped following its type k, with k being T for temperature, V for voltage, or α for activity. Further, it is assumed that all sensors average the acquired data over an acquisition period Δt.

It could be shown that these kinds of sensors can be realized with area overheads in the range of several percent [7, 9].

B. Compression

A principal challenge of the proposed approach is the compression of the acquired data. Opposing constraints are size of the required memory, which holds measured and profile data, and the accuracy of the estimated profiles. Thus, we propose the definition of *sets*, i.e., value ranges, and the accumulation of the time each parameter remains within a *set* (see Fig. 2). That means, for each sensor $s_{k,i}$ exists an N-level vector $S_{k,i}[]$ with N being the amount of *sets* for the sensor type k. Each element n of $S_{k,i}[]$ refers to the time the measured value of $s_{k,i}$ remained in the value range of set n.

Example 1. *Consider the exemplary curve of the temperature sensor $s_{T,4}$ depicted in Fig. 2, an acquisition period of Δt = 10 ms, and an initial empty vector $S_{T,4}[]$. From 0 to 10 ms the average temperature value of $s_{T,4}$ remained in set 1, and thus $S_{T,4}[1]$ is set to 10. Following this scheme, the resulting vector after 40 ms is $S_{T,4}$ = [0, 10, 20, 0, 10].*

It should be noted that this approach leads to a reduction of accuracy due to averaging and disregard of temporal relations amongst the values. In contrast, though, memory sizes can be considerably reduced.

C. Profiling and Database Creation

Profiling refers to the determination of circuit degradation due to *aging* under different stress conditions. Here, we quantify stress via temperature, voltage, and switching activity a circuit is experiencing. Principal idea is the extraction of the relation between parameter profiles and the expected lifetime of the circuit or circuits that are monitored by the sensors.

Each profile is included into an *Aging Profile DataBase* (APDB), as it is shown in Table I. Here, each profile q consists of the percental amounts of time $p_T(q,T_{set})$, $p_V(q,V_{set})$, and $p_\alpha(q,\alpha_{set})$ the monitored circuit(s) experienced a temperature within *set* T_{set}, a supply voltage within *set* V_{set}, and an activity within *set* α_{set}. Additionally, each profile possesses a related expected time to failure MTTF.

Example 2. *Consider the first row of Table I. In this profile, the monitored circuit experiences for $p_{T(1,1)}$ percent of its lifetime a temperature within T_{set1}, for $p_{V(1,1)}$ percent of its lifetime a voltage within V_{set1}, for $p_{\alpha(1,1)}$ percent of its lifetime an activity within set α_{set1}, and so on. It is expected that a circuit with that stress profile has a lifetime of $MTTF_1$.*

This example highlights the advantage of granular profiles, i.e., consideration of varying parameters during circuit lifetime, which allows more accurate prediction compared to static models. It follows further, *aging* must not be defined solely as delay degradation, but can also be quantified via occurrence of permanent faults. However, it also evident that the result of this analysis is an estimation and not an exact prediction of the RUL of the present circuit.

As mentioned above, the extraction of a profile can be executed via three different methods, i.e., *aging* simulation, accelerated stress testing, and/or observation of real systems in the field.

1) Aging Simulation

An *aging* simulator is a tool capable to simulate *aging* induced degradation of integrated circuits [3]. Prominent

Table I - Aging Profile DataBase (APDB) for a circuit monitored by a Temperature (T), Voltage (V), and activity (α) sensor.

T_{set1}	...	T_{setX}	V_{set1}	...	$V_{setY\_}$	α_{set1}	...	α_{setZ}	*MTTF*
$p_{T(1,1)}$	···	$p_{T(1,X)}$	$p_{V(1,1)}$	···	$p_{V(1,Y)}$	$p_{\alpha(1,1)}$	···	$p_{\alpha(1,Z)}$	$MTTF_1$
$p_{T(2,1)}$	···	$p_{T(2,X)}$	$p_{V(2,1)}$	···	$p_{V(2,Y)}$	$p_{\alpha(2,1)}$	···	$p_{\alpha(2,Z)}$	$MTTF_2$
...	...	...	...	...	...	...	...	...	...
$p_{T(M,1)}$	···	$p_{T(M,X)}$	$p_{V(M,1)}$	···	$p_{V(M,Y)}$	$p_{\alpha(M,1)}$	···	$p_{\alpha(M,Z)}$	$MTTF_M$

examples are BERT (UC Berkeley), RelXpert (Cadence), Eldo (Mentor Graphics), and MOSRA (Synopsys).

Each simulation lasts until a predefined period is passed or a failure is detected. This failure can be based on failing timing constraints or permanent faults. During each simulation, the values of temperature, voltage and/or activity are varied such that it corresponds to a predefined profile. Further, it is possible to vary the technology parameters via the Monte Carlo method [3]. Consequently, for each profile several *aging* simulations are executed, whereas the resulting MTTF is the average time until the circuit failed.

2) Accelerated stress testing and field test

Accelerated Stress Testing (AST) is a test method conducted after fabrication at which high levels of stress are applied for a short period of time in order to predict long-term behavior [10]. The proposed approach permits the integration of profiles extracted during AST as well as from devices, which had been already shipped to customers, into a APDB.

Please note that the practical realization of both latter methods is outside the scope of this work.

D. Prediction methods

In order to predict the circuits MTTF, a proper fitting method is required to correlate measured data with the entries of an APDB. In the following, we propose three related approaches.

1) Generalized Linear Model

The principal idea of the *Generalized Linear Model* (GLM) with *Feature Transformation* and *Partial Least Square* is the reduction of a given APDB to independent predictors and the definition of a linear model [11, 12]. That means, the input data of the APDB, which are the times the circuit remained in the *sets* (see also Table I), are transformed to N estimator variables t_i and the $MTTF_{GLM}$ is estimated via:

$$\mathrm{MTTF_{GLM}} = \sum_{i=1}^{N} w_i t_i + R \qquad (1)$$

with w_i are coefficients and R is a residue value, all determined by *Multiple Linear Regression* [11].

2) Euclidean Distance

This model applies *Euclidean Distance* (ED), i.e., the representation of a distance between two points in a Euclidean space, for determination of the entry of the APDB that represents best the applied stress profile.

The distance $d(a,b)$ between the two vectors $a = (a_1, a_2, \dots, a_N)$ and $b = (b_1, b_2, \dots, b_N)$ with each N values can be represented as follows:

$$d(a,b) = \sqrt{\sum_{i=1}^{N} (a_i - b_i)^2} \qquad (2)$$

This distance is determined for each entry of the APDB and the measured profile. Then, the MTTF of the APDB entry with the smallest distance to the measured profile is chosen.

3) Correlation

The *Correlation* method (CR) calculates the dependence and correlation (but not necessarily causality) between two vectors [13]. This correlation is qualified by a coefficient $\rho_{a,b}$ that represents the covariance *cov* of the two vectors a and b divided by the product of their standard deviations σ_a and σ_b and follows from:

$$\rho_{a,b} = \frac{\mathrm{cov}(a,b)}{\sigma_a, \sigma_b} \qquad (3)$$

After calculation of the correlation coefficients between every APDB entry and the measured profile, the MTTF of the APDB entry with the highest correlation is chosen.

E. Countermeasures and Reporting

The results of the proposed approach can be applied for countermeasures like load-balancing, adaptive voltage and/or frequency scaling, or graceful degradation [2]. Further, a potential user can be informed about critical system conditions as well as the expected RUL.

III. Simulation Results

In order to verify the feasibility of the proposed approach an inverter chain (comprised of hundred inverters) and selected ISCAS'85 circuits were implemented in a predictive 90 nm technology [14, 15]. Next, the circuits were aged for one year in steps of 0.01 years for random voltage and temperature profiles using the tool RelXpert and simulating the *aging* effects HCI, TDDB, and NBTI/PBTI. The chosen ranges for supply voltage and temperature were 0.9 V to 1.3 V and 10 °C to 90 °C, respectively.

The resulting delay shift Δt_d of the critical path was applied for definition of the expected time to failure MTTF via:

$$\mathrm{MTTF} = \frac{\Delta t_{d,\max}}{\Delta t_d} \cdot 1 year \qquad (4)$$

With $\Delta t_{d,\max}$ is the maximum acceptable delay shift, which was considered with 20 %. Based on these data, an APDB with 250 entries was created for each circuit. Fig. 3 illustrates the maximum and minimum determined expected lifetimes ranging from 1.5 years (c1355) to 6.1 years (c880).

The accuracy of the results was determined via the *leave-one-out cross-validation* method [16]. This means, we removed one entry of the APDB, used the remaining ones for generation of the models presented in section II.D, estimated the MTTF for the removed entry, and calculated the error between predicted and measured MTTF. This was repeated for all entries of the APDB. Further, for each circuit we executed static *aging*

Table II – Normalized Root Mean Square Error (NRMSE) for all proposed methods and a static analysis

Circuit	#Cells	NRMSE			
		GLM	ED	CR	Static
Inv	100	9.1 %	13.0 %	13.1 %	50.4 %
c499	474	8.3 %	8.1 %	10.1 %	61.4 %
c880	393	7.9 %	7.3 %	9.6 %	64.7 %
c1355	738	14.7 %	13.5 %	18.1 %	63.5 %
c5315	1781	8.1 %	7.9 %	10.6 %	63.2 %

simulations with constant supply voltage (V_{DD} = 1.1V) and temperature (T = 70 °C).

Subsequently, we estimated the *Normalized Root Mean Square Error* (NRMSE) for all acquired results via:

$$NRMSE = \frac{rms\left(\mathrm{MTTF}_{meas} - \mathrm{MTTF}_{pred}\right)}{\overline{\mathrm{MTTF}_{meas}}} \qquad (5)$$

with *rms* is the *Root Mean Square* function, MTTF_{meas} and MTTF_{pred} are the measured and predicted expected time to failure, and $\overline{\mathrm{MTTF}_{meas}}$ refers to the mean of the measured data.

The results listed in Table II indicate an average NRMSE of 9.6 % for the GLM model, 10 % for the ED method, and 12.3 % for the CR method. That means, for a given circuit the GLM model permits with a certainty of ca. 90 % the prediction of the expected lifetime if temperature and voltage continue to behave as in the observed period. This enables proactive strategies in order to extend the lifetime of a system. However, these results demonstrate, as expected, that the proposed method does not allow an exact determination of failing delays.

Results indicate further, that the proposed method outperforms the commonly applied static methods, which showed an average NRMSE of 60.6 %.

IV. Conclusions

Aging of integrated circuits under environmental stress is an important challenge for current and future technologies. This work proposes the prediction of the *Remaining Useful Lifetime* (RUL) of systems by monitoring environmental parameters and relating these to stored profiles. The acquired data can be utilized for proactive strategies in order to extend the expected lifetime of a system and/or inform the user about critical system states. The proposed approach is characterized by low area costs, the avoidance of interference with critical paths, and a flexible software part, that can be updated with improved models.

Exemplary *aging* simulations indicate that a *Generalized Linear Model* is a reasonable prediction method with a certainty of roughly 90 %.

Acknowledgements

We gratefully thank CAPES, CNPq, and FAPEMIG and for the financial support.

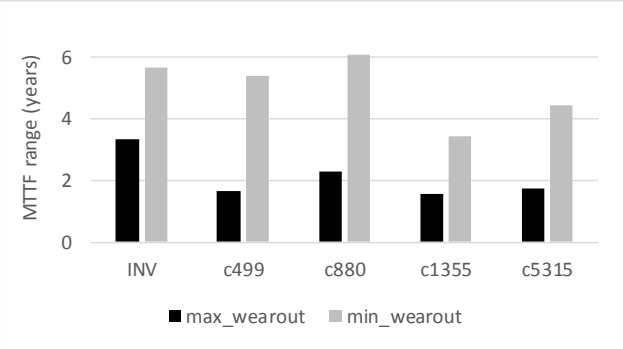

Fig. 3. Expected lifetime in years for chosen profiles with maximum and minimum wearout

References

[1] J. Keane, X. Wang, D. Persaud, and C. H. Kim, "An All-In-One Silicon Odometer for Separately Monitoring HCI, BTI, and TDDB," *IEEE Journal of Solid-State Circuits,* vol. 45, pp. 817-829, 2010.

[2] M. Agarwal, B. C. Paul, M. Zhang, and S. Mitra, "Circuit Failure Prediction and Its Application to Transistor Aging," in *25th IEEE VLSI Test Symposium (VTS'07)*, 2007, pp. 277-286.

[3] E. Maricau and G. Gielen, *Analog IC reliability in nanometer cmos*. New York: Springer, 2012.

[4] S. Mitra, K. Brelsford, Y. M. Kim, H. H. K. Lee, and Y. Li, "Robust System Design to Overcome CMOS Reliability Challenges," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems,* vol. 1, pp. 30-41, 2011.

[5] E. Mintarno, J. Skaf, R. Zheng, J. Velamala, Y. Cao, S. Boyd*, et al.*, "Optimized self-tuning for circuit aging," in *2010 Design, Automation & Test in Europe Conference & Exhibition (DATE 2010)*, 2010, pp. 586-591.

[6] F. Firouzi, F. Ye, A. Vijayan, A. Koneru, K. Chakrabarty, and M. B. Tahoori, "Re-using BIST for circuit aging monitoring," in *2015 20th IEEE European Test Symposium (ETS)*, 2015, pp. 1-2.

[7] R. Baranowski, F. Firouzi, S. Kiamehr, C. Liu, M. Tahoori, and H. J. Wunderlich, "On-line prediction of NBTI-induced aging rates," in *2015 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2015, pp. 589-592.

[8] A. Urmanov, "Electronic Prognostics for Computer Servers," in *2007 Annual Reliability and Maintainability Symposium*, 2007, pp. 65-70.

[9] Z. Zhang, Y. Ren, L. Chen, N. J. Gaspard, A. F. Witulski, T. W. Holman*, et al.*, "A Bulk Built-In Voltage Sensor to Detect Physical Location of Single-Event Transients," *Journal of Electronic Testing,* vol. 29, pp. 249-253, 2013// 2013.

[10] M. L. Bushnell and V. D. Agrawal, *Essentials of electronic testing for digital, memory, and mixed-signal VLSI circuits*. Boston: Kluwer Academic, 2000.

[11] H. L. Bu, G. Z. Li, X. Q. Zeng, J. Y. Yang, and M. Q. Yang, "Feature Selection and Partial Least Squares Based Dimension Reduction for Tumor Classification," in *2007 IEEE 7th International Symposium on BioInformatics and BioEngineering*, 2007, pp. 967-973.

[12] V. Esposito Vinzi and G. Russolillo, "Partial least squares algorithms and methods," *Wiley Interdisciplinary Reviews: Computational Statistics,* vol. 5, pp. 1-19, 2013.

[13] J. Lee Rodgers and W. A. Nicewander, "Thirteen Ways to Look at the Correlation Coefficient," *The American Statistician,* vol. 42, pp. 59-66, 1988/02/01 1988.

[14] J. P. Hayes, M. C. Hansen, and H. Yalcin, "Unveiling the ISCAS-85 benchmarks: A case study in reverse engineering," *IEEE Design & Test of Computers,* vol. 16, pp. 72-80, Jul-Sep 1999.

[15] "Reference Manual for Generic 90nm Salicide 1.2V/2.5V 1P 9M Process Design Kit (PDK) Revision 4.3.," ed: Cadence Design Systems, Inc., 2008.

[16] R. R. Picard and R. D. Cook, "Cross-Validation of Regression Models," *Journal of the American Statistical Association,* vol. 79, pp. 575-583, 1984/09/01 1984.