A Board-Level Prognostic Monitor for MOSFET TDDB

D. Goodman, B. Vermeire, J. Ralston-Good, R. Graves
Ridgetop Group, Inc.
3580 West Ina Road
Tucson, AZ 85741
520-742-3300
doug@ridgetop-group.com

Abstract— A prognostic cell to monitor time-dependent dielectric breakdown and electronic aging of integrated circuits has been designed and fabricated.

TABLE OF CONTENTS

1. INTRODUCTION

The ability to identify performance degradation and failure modes in integrated circuit (IC) components significantly enhances the safety and reliability of a component, system, or sub-system. Specifically, the ability to monitor electronic aging at the board level, and to detect an impending IC failure, facilitates the advent of corrective actions necessary to avert a catastrophic event. Based on the application of a board-level prognostic monitor chip, accurate predictions of electronic aging and end-of-life failure modes can be extracted.

A prognostic chip has been designed to monitor the time dependent dielectric breakdown (TDDB) of MOS transistors. The self-stressing integrated MOSFETs that are monitored by the prognostic chip act as the TDDB aging sensors for the host application. The monitored MOS transistors are identical to those used in the host IC, to insure that the extracted data maps to the condition of the key components of the host IC. The prognostic circuitry biases the MOS transistors to accelerate aging under certain environmental conditions. For this reason the monitor, or sensor devices in the prognostic chip are designed to be sacrificial, and are not used in the host application, but instead are packaged separately, mounted, and biased at the board level.[1,2]

2. PROGNOSTIC HEALTH MAINTENANCE

All integrated circuits have non-zero failure probability, that is, there is at every instance of use, there exists a finite probability that the component will fail. A particular product line's failure rate is the number of such products that is expected to fail per unit time. Since this number is an average for a group of ICs, the failure rate is equivalent with the failure probability for a particular, individual IC. The failure probability during the life of an IC typically follows a classic "bathtub" failure rate curve as shown in Figure 1, consisting of three distinct regions: (i) an initial region where the failure probability is high, called the burn-in or infant mortality region, (ii) a useful life region where the failure rate is minimum and typically constant, and (iii) a wear out or end-of-life region where the product's failure rate increases. The bathtub curve describes the aging process, and can be measured for a particular product line when that product is used under well-defined conditions. This allows a manufacturer to predict the product lifetime.

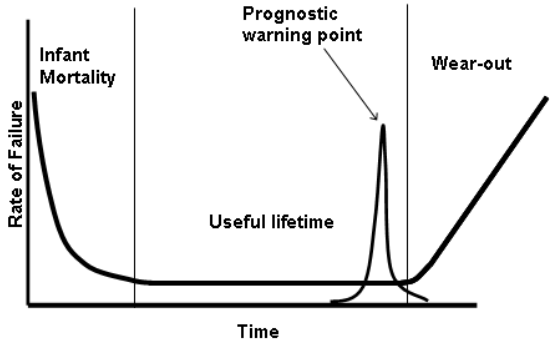


Figure 1 – Failure rate vs. time for a typical electronic component. The prognostic warning point is generated by the prognostic cell described in this work.

Diagnostics are implemented to determine the state of a component and whether it can function as designed. A prognostic is a predictive diagnostic that includes the ability to determine the remaining life or time span of proper operation and performance of a component. Health management is the added capability to make appropriate decisions about maintenance actions based on diagnostics/prognostics information, available resources, and operational demand.

Aging is a function of time, intensity of use, environment, and the quality of the manufacturing process. For microelectronic devices, operating environments include

1

[1] 0-7803-9546-8/06/$20.00© 2006 IEEE

[2] Ridgetop Group Paper #RG101905, Version 2, Updated Oct. 24, 2005

variables that can reduce IC lifetime such as temperature, operating voltage, incident ionizing radiation, humidity, presence of corrosives, etc. The actual lifetime of an IC in the field may be quite different than the lifetime measured under controlled and specified conditions. This means that the bathtub curve failure rate characteristic for a particular instance of a product (which is the probability that that particular IC will fail as a function of time) may be shifted substantially relative to the baseline. For electronics applications where system reliability is important, system designers often assume worst-case conditions for reliability calculations. Integrated circuits can then be selected that meet the worst-case requirements. However, worst-case conditions are typically poorly defined and not well known, and also may not be continuously present. This concern may be addressed by adding a larger design margin. It is also possible, however, that an IC is used outside its specified operating range (due to a system shock of some kind, for example). Such an excursion may lead to unknown effects on IC aging, resulting in questions about the correct maintenance requirements.

Since a worst-case design approach often results in over-specification of reliability requirements, and since excursions outside specified operating ranges may occur in service life, it would be particularly useful if an integrated circuit or integrated circuit board could be equipped with an aging process monitor that can determine if the component is entering, or is about to enter, the wear out region of its life. This implies that the probability of failure is no longer constant (as it is during its useful life), but is getting increasingly larger. Such a monitor is called a prognostic cell, and this cell is capable of predicting impending failure.

Increasing the intensity of use or environmental stress may shift the failure distribution of an integrated circuit along the time axis. In this case the device will have a shorter useful life, and may also exhibit degradation of its operational performance parameters. This acceleration of device performance degradation, or associated failure rate, can be used to design and build a prognostic cell. A prognostic cell for a given integrated circuit and a given failure mechanism can be designed to apply different bias intensities to a group of components. Excess stress causes the detrimental effects of aging to appear at shorter times in the particular devices. A measurable parameter that is indicative of the amount of cumulative wear of the stressed sub-circuit, such as a current or voltage, is compared to a specific reference value within the prognostic circuitry.

3. THE PROGNOSTIC CELL

The prognostic cell consists of a family of prognostic devices, each with different trigger points. An impending failure is indicated when the monitored value is greater than a preset reference value. Ideally, the prognostic detection method used will trigger at a calibrated time, or prognostic distance, before the onset of the end-of-life region, as shown in Figure 1.

Time dependent dielectric breakdown (TDDB) of silicon dioxide has been found to be exponentially dependent on the electric field across the gate oxide [1]. Although the exact mechanism for every failure mode is not known, it is generally accepted that breakdown is caused by oxide charge trapping [2]. Acceleration of the breakdown of an oxide can therefore be achieved by applying a voltage higher than the supply voltage, to increase the electric field across the oxide. When the test monitor device fails, a certain fraction of the circuit lifetime has been used up. The fraction of useful circuit life that has been used up is dependent on the amount of overvoltage applied and can be estimated from the known distribution of failure times.

We note also that because of the improvement in the yield of modern commercial semiconductor processes, oxide breakdown is typically of the intrinsic type. Contamination or particle related failures are not common today, so we do not consider in detail the extended models for defect-related oxide failure.

4. CIRCUIT DESCRIPTION

The design challenge for the TDDB cell is to apply sufficient, but not overly large, stress voltage to the test transistor to enhance its gate oxide failure without stressing the oxides of other transistors in the prognostic cell. The constraint of having only a single power supply, e.g.,
V_{DD} = 2.5 volts, forces the need to generate a higher on chip potential in such a way that it does not affect any other component on the chip. This is accomplished by designing a well bias circuit that will pump the well containing the test circuitry to create a larger total potential across the oxide of the test transistor than the power supply voltage, as illustrated in Figure 2. Two diodes and capacitors form the charge pump. A feedback loop accurately controls the stress voltage.

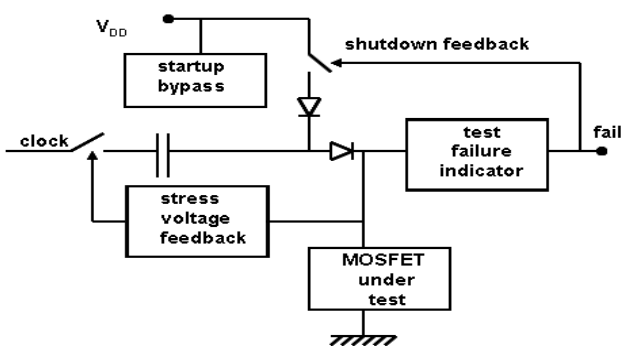


Figure 2 - Schematic circuit for the TDDB prognostic circuit.

If failure of the test capacitor is indicated (by a drop in the stress voltage), the output bit is set low and a second feedback loop shuts down the charge pump (to avoid excess power consumption in this stage).

For this experiment, a suitable 0.25 μm process was chosen. The prognostic circuit average power dissipation is less than 50 μW.

5. CALCULATION OF STRESS VOLTAGE

There are two field-dependent TDDB models in widespread use: The E and 1/E models. Both models fit the available data equally well for large oxide fields (E_{ox} > 8 MV/cm), however, the model that best describes oxide breakdown at both high and low oxide fields typically seen in circuit operation is the E model [3]. The E model can be expressed as:

$$t_{BD63} = C\exp(-\gamma E_{ox})\exp\left(\frac{E_a}{kT}\right), \quad (1)$$

Where: t_{BD63} = time-to-breakdown with a failure fraction of 63% (cumulative failure F =1-1/e), C is a constant, γ = field acceleration parameter, E_{ox} = oxide field, E_a = temperature activation energy, k = Boltzmann's constant, and T = absolute temperature.

The field acceleration parameter is dependent on temperature:

$$\gamma = b + \frac{c}{T}, \quad (2)$$

where b and c are constants from [3]. Additionally, the temperature activation energy is dependent on the electric field:

$$E_a = k - lE_{ox}, \quad (3)$$

where k and l are empirical constants.

Time-to-breakdown values are distributed around t_{BD63} in a manner following Weibull statistics [4, 5]. The time-to-breakdown value, at which a portion F of the capacitors has failed, t_{BDF}, can be determined from

$$\log[-\ln[1-F]] = \log\left(A\frac{t_{BDF}{}^{\beta}}{\alpha^{\beta}}\right), \quad (4)$$

Where F = cumulative failure percentage, A = capacitor area, α = scale parameter, β = shape parameter.
The scale parameter α is related to the mean of the distribution while the shape parameter β is related to its standard deviation. While the shape parameter is largely independent of the oxide electric field and temperature, the scale parameter shifts with both. From equations (1) and (4) it follows that there is a functional relationship between the scale parameter α and the constant C. The complete cumulative distribution can be written as:

$$t_{BDF} = \frac{C'Exp(-\gamma E_{ox})Exp\left(\frac{E_a}{kT}\right)\left[Ln\left(\frac{1}{1-F}\right)\right]^{1/beta}}{A^{1/beta}} \quad (5)$$

where C' is a (new) constant.

Intrinsic breakdown is the dominant failure mode during the end of lifetime region. For intrinsic breakdown, the parameters necessary to describe the oxide failure distribution (γ, E_a, C' and β) depend only on the oxide thickness. It is therefore possible to estimate the parameters for a particular process by reviewing literature data for similar oxide thickness values.

The γ value and the C' value from data reported in [3] have been used in this work. It is not possible to determine β values accurately using capacitors that are all of the same size unless extremely large sample sizes are used. Because of this, reliable estimates of β could not be obtained using literature data, so capacitor samples were obtained from the foundry. Time to breakdown was measured on these samples for five sets of capacitors with different areas at 8V stress voltage. The areas ranged between 1.07 x 10^{-6} cm^2 and 1.0 10^{-2} cm^2. Equation (4) can be rewritten,

$$\log[-\frac{\ln[1-F]}{A}] = \beta\log\left(\frac{t_{BDF}}{\alpha}\right) \quad (6)$$

This means that a plot of log[-ln[1-F]/A] versus log[t_{BDF}] will give a straight line with slope β. From Equation (6) it follows that using a range of areas (different values of A) will result in a much wider range of ordinate values in such a plot, meaning that the slope can be determined more accurately. A plot of the TDDB data obtained from the foundry samples is shown in Figure 3. Because the wide range in available capacitor areas is exploited, this plot contains the same information (for purposes of determining the TDDB statistics) as a set of 50,000 capacitors of a single size. The β value that is extracted from the fit (the solid line in Figure 3 is 3.2.

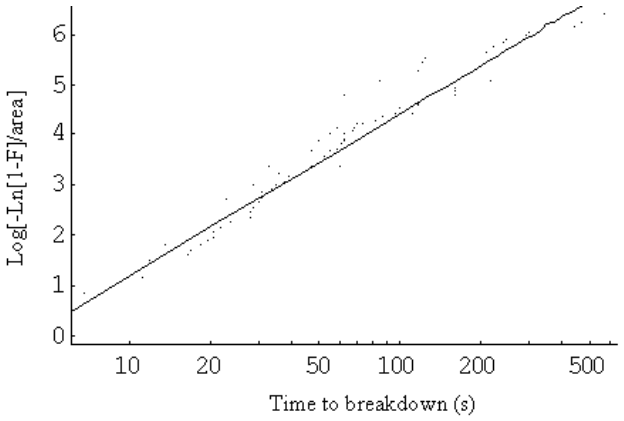


Figure 3 - TDDB results for sample capacitors. The stress voltage was 8 V. The β value obtained of the fit line is 3.2.

Since the oxide breakdown model relates oxide field to the breakdown time, a relationship between gate voltage and oxide field is required. A first order approximation for the oxide field for an nMOS transistor that is turned on is

$$E = \frac{V_G + \left(E_g / 2 - \phi_p \right)}{t_{ox}}, \quad (7)$$

where V_G = gate voltage, E_g = bandgap energy, ϕ_p = substrate potential, t_{ox} = oxide thickness.

Since the stressed device is biased in strong inversion, any gate voltage over and above the supply voltage ΔV_g will appear across the oxide (not across the depletion layer) and result in an increase in the electric field ΔE:

$$\Delta E = \frac{\Delta V_G}{t_{ox}} \quad (8)$$

A prognostic cell must trigger before the accompanying circuit fails. This means that the cumulative failure probability of the prognostic cell must be close to unity before the cumulative failure probability of the circuit becomes appreciable, as shown in Figure 4.

This can be realized by operating the prognostic cell at a higher oxide field than the circuit. The probability of correctly predicting circuit failure using a prognostic cell is equal to the probability that the prognostic cell triggers (fails) before the circuit. If the following variables are defined: p(t) = trigger (failure) probability density function of the prognostic cell, c(t) = failure probability density function of the circuit, then the probability of correct prediction is given by,

$$p(correct) = \int_0^\infty c(t) \left(\int_0^t p(u)du \right) dt \quad (9)$$

The probability that the prognostic cell does not correctly predict failure of the circuit is equal to the probability that the circuit fails before the prognostic cell fails. This is given by,

$$p(incorrect) = \int_0^\infty p(t) \left(\int_0^t c(u)du \right) dt \quad (10)$$

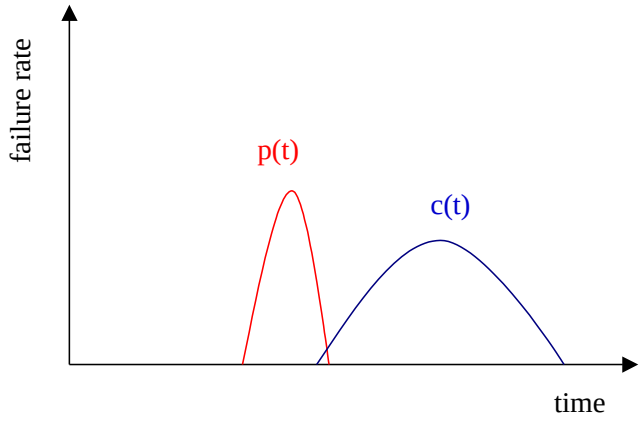


Figure 4 – The failure probability density of the circuit c(t) and the trigger rate of the prognostic cell p(t) as a function of time.

The prognostic distance is the time between the 1% cumulative failure probability point of the circuit and the 99% cumulative failure probability point of the prognostic cell, as shown in Figure 5.

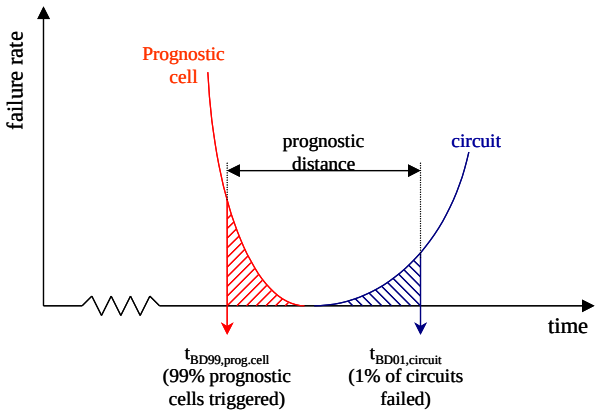


Figure 5 – Prognostic distance in terms of cumulative failure probabilities $t_{BD01,circuit}$ and $t_{BD99,prog.cell}$.

The current design attempts to achieve a prognostic distance of 10%. This means that before 90% of the time has elapsed at which 1% of the circuits have failed, 99% of the prognostic cells must have triggered. The reliability requirement can be stated in terms of time-to-breakdown of the circuit and prognostic cell:

$$\frac{9}{10} t_{BD01,circuit} = t_{BD99,prog.cell} \quad (11)$$

Applying Equation (5) for $t_{BD01,circuit}$ and $t_{BD99,prog.cell}$ results in the basis for a design equation for a prognostic cell.,

$$\frac{9}{10} \frac{C' Exp(-\gamma E_{ox,circuit}) Exp\left(\frac{E_a}{kT}\right)\left[Ln\left(\frac{1}{1-0.01}\right)\right]^{1/beta}}{A_{circuit}^{1/beta}}$$

$$= \frac{C' Exp(-\gamma E_{ox,prog.cell}) Exp\left(\frac{E_a}{kT}\right)\left[Ln\left(\frac{1}{1-0.99}\right)\right]^{1/beta}}{A_{prog.cell}^{1/beta}}$$

(12)

From this, the amount of overvoltage required is obtained:

$$\Delta V_G = \Delta E \;\; t_{ox} = \left(E_{ox,prog.cell} - E_{ox,circuit}\right) t_{ox}$$

$$= \frac{t_{ox}}{\gamma} Ln\left[\frac{10}{9}\left(\frac{A_{circuit}}{A_{prog.cell}}\right)^{1/\beta}\left(\frac{Ln(1-0.99)}{Ln(1-0.01)}\right)^{1/\beta}\right] \quad (13)$$

Equation 13 is a design equation for the TDDB prognostic cell, since it gives the excess stress voltage (ΔV_G) that is required for a given prognostic distance. For this experiment, a particular 0.25 µm process was chosen and the excess stress voltage for the design is calculated using Equation 13 to be necessarily equal to 0.9 V, for a total stress voltage of 3.4 V. This stress voltage is achievable using the designed TDDB prognostic circuit.

6. CONCLUSIONS

The goal of the improving electronic circuit and system reliability using embedded prognostics is to develop self-stressing cells that operate autonomously and give advance warning of impending failure of integrated circuits. The prognostic circuit design monitors sensor devices to determine the cumulative aging effects in the form of performance parameter degradation that is dependent on the environment that the board is subjected to.

The prognostic circuit described in this work is configured for simple buffered logic high or low output, facilitating integration into a governing control system or prognostic health management (PHM) system.

A feedback module is designed into the prognostic circuit to remove power from the stressing circuit at the point of the prognostic cell triggers, thereby preventing a current drain and increased power dissipation by the prognostic cell in the event of an oxide failure.

The prototype prognostic cell has been fabricated in a silicon process, and is scheduled for testing.

REFERENCES

[1] A.M. Yassine, "Time dependent dielectric breakdown of ultrathin gate oxide," *IEEE Trans. Electron Devices*, vol. 47, pp. 1416-1420, 2000.

[2] J.C. Lee, I.-C. Chen, "Modeling and Characterization of Gate Oxide Reliability," *IEEE Trans. Electron Devices*, vol. 35, pp. 2268-2279, 1988.

[3} M. Kimura, "Field and Temperature acceleration model for time-dependent dielectric breakdown," *IEEE Trans. Electron Devices*, vol. 46, pp. 220-229, 1999.

[4] D.R. Wolters and J.F. Verwey, "Breakdown and wear-out phenomena in SiO2 films," *Instabilities in Silicon Devices*, vol. 1, G. Barbottin and A. Vapaille, Eds. Amsterdam: Elsevier Science Publishers, 1986, pp. 315-362.

[5]E.J. Gumbel, Statistics of Extremes: Columbia University Press, 1958.

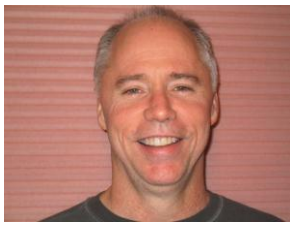
BIOGRAPHIES

Doug Goodman holds a B.S. in Electronic Engineering from California Polytechnic State University, San Luis Obispo, and has a Masters degree from the University of Portland. He has over 25 years of experience in low-noise instrumentation design, DFT and fault simulation techniques, and integrated circuit tool development. His previous affiliations include Tektronix, Delco Systems, Bunker Ramo Corporation, Analogy, Opmaxx, Inc and Fluence Technology Inc. He was on the patent committee at Tektronix, Inc. while in the firms' Laboratory Instrument Division. He has also served as an

adjunct faculty member at the University of Arizona.

Dr. Bert Vermeire is a Research Professor at Arizona State University and a consultant to Ridgetop Group. Bert has prior experience as Principal Investigator on several projects dealing with advanced semiconductor processes, BIST Development and radiation hardening. Prior to the University of Arizona, Bert was with IMEC in Belgium. He received his Ph.D. from the University of Arizona, Tucson in 1999 and has written 27 technical papers.

Jeremy Ralston-Good holds a B.S. and M.S. in Electrical Engineering from Vanderbilt University. Jeremy has prior research experience in the field of physical mechanisms of failure of semiconductor devices that result from electrostatic discharge events and radiation environmental effects with Dynamics Research Corporation, San Diego, CA



Russell Graves has over 10 years of experience in radiation electronics, modeling, and simulation. He holds a BSEE from Northern Arizona University and an MSEE from Vanderbilt University.