Extraction of wearout model parameters using on-line test of an SRAM

Shu-Han Hsu*, Ying-Yuan Huang, Yi-Da Wu, Kexin Yang, Li-Hsiang Lin, Linda Milor

School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

A B S T R A C T

To accurately determine the reliability of SRAMs, we propose a method to estimate the wearout parameters of FEOL TDDB using on-line data collected during operations. Errors in estimating lifetime model parameters are determined as a function of time, which are based on the available failure sample size. Systematic errors are also computed due to uncertainty in estimation of temperature and supply voltage during operations, as well as uncertainty in process parameters and use conditions.

1. Introduction

Static Random Access Memories (SRAMs) are major components of systems-on-chips and are also used for memory in systems that require very low power consumption and easy access to data. To ensure that memory operation is stable, the reliability of SRAMs need to be considered. We focus on front-end gate oxide breakdown (FEOL TDDB), which is the failure due to the buildup of traps in the dielectric region, because it is one of the most important wearout mechanisms in semiconductors.

Prior work has focused on determining the impact of bias temperature instability (BTI) on circuit performance as a function of time and mitigating its impact [1–4]. In [1–3], the data retention voltage and read access time parameters are accessed to estimate the threshold voltage, which in turn is used to determine the extent of degradation due to BTI. In [4], the SRAM is put in a test state to measure the leakage current, which is also linked to degradation due to BTI and various SRAM performance metrics. BTI (and hot carrier injection (HCI)) causes degradation in circuit/device performance *en route* to failure. Hence, performance degradation provides an indication of wearout, and performance monitors are an appropriate way to detect wearout due to BTI and HCI.

In this study, we use data on failures in the SRAM to estimate the wearout model parameters of FEOL TDDB. FEOL TDDB models are not based on degradation, but instead involve only a time-to-failure. Hence, to detect failure rates, it is necessary to monitor actual failures and to link these failures to lifetime models. Because of the large number of identical cells, the SRAM can be used to detect the characteristics of wearout due to FEOL TDDB. Hence, the task of this paper is to appropriately estimate the model parameters, using the SRAM data as a vehicle.

The parameters to be extracted are those of the two parameter

Weibull distribution for FEOL TDDB: the characteristic lifetime, η, and the shape parameter, β. The model parameters are extracted from time-to-failure data from the cells in the SRAM.

This work determines the accuracy in extracting the model parameters by considering both random and systematic errors. The random errors are due to the availability of samples (failed SRAM cells). Clearly, more data provides more accurate model parameter extraction. Systematic errors are due to usage variations, such as supply voltage and operating temperature fluctuations, as well as variations due to process parameters and workload. The analysis of systematic errors can be used to determine when and if sensor data is needed to supplement analytical wearout models when estimating wearout model parameters.

This paper examines how to analyze FEOL TDDB failures in SRAMS during on-line operations. The rest of the paper is organized as follows. Section 2 describes the methodology for extracting the wearout parameters. Section 3 shows the error analysis, and the paper is concluded in Section 4.

2. Methodology

The lifetime distribution of a device due to wearout by front-end gate oxide breakdown (FEOL TDDB) can be modeled with a two-parameter Weibull distribution:
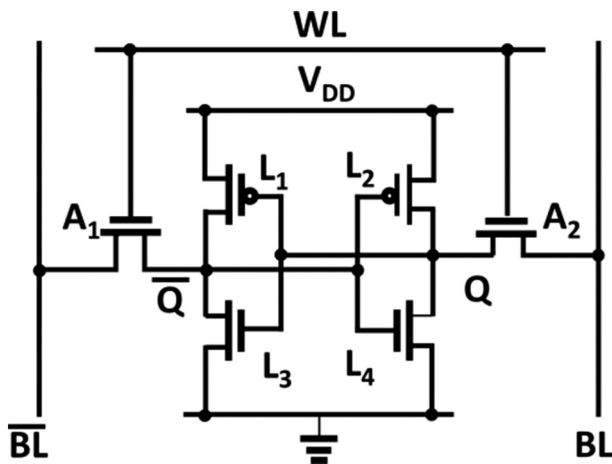
$$P(t) = 1 - exp\left(-\left(\frac{t}{\eta}\right)^{\beta}\right) \tag{1}$$

where η is the characteristic lifetime and β is the shape parameter. For FEOL TDDB:

$$\eta = A_{FEOL}(WL)^{\frac{-1}{\beta}}e^{\frac{-1}{\beta}}V^{a+bT}exp\left(\frac{cT+d}{T^2}\right)s^{-1} \tag{2}$$

Fig. 1. Circuit diagram of an SRAM, where A_1 and A_2 are the access transistors, and L_1, L_2, L_3 and L_4 are the latch transistors.

where a, b, c, d, and A_{FEOL} are process-dependent constants. V and T are voltage and temperature. W and L are the width and length of the MOSFET device; s is the probability of stress. The parameters used in this study were obtained from experimental data [5].

The characteristic lifetime of the SRAM, η_{SRAM}, is a combination of Weibull distributions for the components, and is the solution of [6]:

$$1 = \sum_{i=1}^{n} \left(\frac{\eta_{SRAM}}{\eta_i} \right)^{\beta_i} \tag{3}$$

where η_i, $i = 1, \ldots, n$ are the characteristic lifetime of all the circuit components, and β_i are the corresponding shape parameters. Similarly, it can be found that [6]:

$$\beta_{SRAM} = \sum_{i=1}^{n} \beta_i \left(\frac{\eta_{SRAM}}{\eta_i} \right)^{\beta_i} \tag{4}$$

The difference between stressing a device to determine its lifetime and an SRAM to determine its lifetime relates to the probability of stress, s, of each of the devices. When we build models for individual devices, we assume that s = 1. However, s depends on the use scenario of the SRAM. As shown in Fig. 1, there are access transistors and latch transistors present in the SRAM. Initially, if we assume that the SRAM stores logic "1" 50% of the time and logic "0" 50% of the time, then s = 0.5 in (2) for all cells' four transistors in the latch. s ≈ 0 for the access transistors, turning on only when the cell is accessed. 50% is set as the baseline for comparison. If the duty cycle is changed, then the SRAM will degrade at a different rate.

The SRAM failures due to FEOL TDDB during operation are calculated using Monte Carlo simulation. The random variable is the failure probability in (1). The resulting data are time stamps for the failures of SRAM cells. These SRAM failures can be permanent or soft errors. We must separate failures due to permanent wearout from soft failures. They can be separated with minor additions to the hardware [7]. We must also separate failures due to other wearout mechanisms from those caused by FEOL TDDB. The cause of failure can be identified via BIST [8]. Hence, it is possible to isolate a sequence of failures due to only FEOL TDDB for extraction of the FEOL TDDB wearout parameters.

Given a sequence of time stamps for SRAM failures due to only FEOL TDDB, we extract the Weibull parameters using generalized maximum likelihood estimation [9]. The original SRAM cell parameters for FEOL TDDB degradation in this study are with η = 20 years and β = 1.12. As shown in Fig. 2(a), the SRAM failure samples for a sample size of 94,000 (an SRAM with 94 k cells) due to FEOL TDDB are plotted on a Weibull plot. The FEOL TDDB extracted parameters are η = 19.966 and β = 1.119.

(a)

(b)
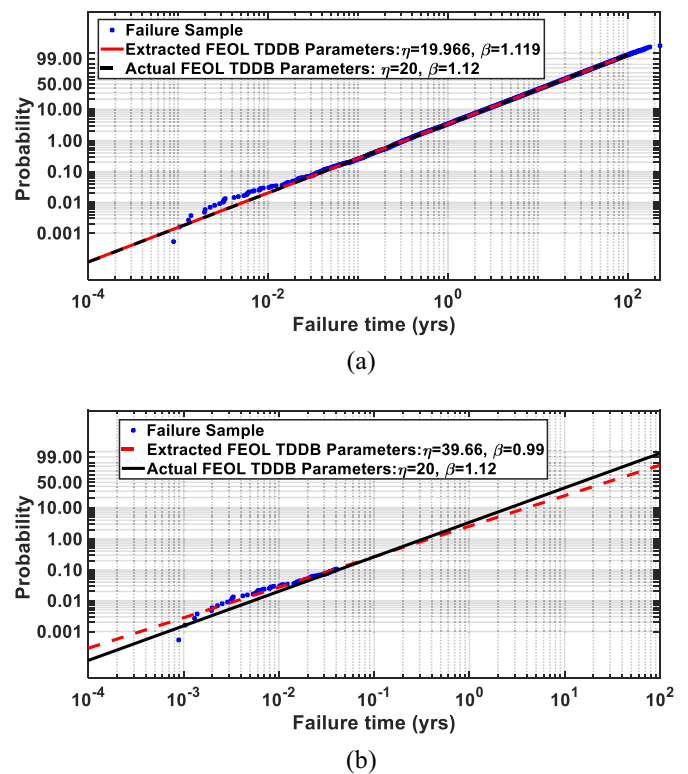
Fig. 2. SRAM failure samples due to FEOL TDDB for (a) a sample size of 94,000 with extracted FEOL TDDB failure parameters, η = 19.966 and β = 1.119, and (b) a sample size of 100 out of 94,000 SRAM cells with extracted FEOL TDDB failure parameters of η = 39.66 and β = 0.99.

If data is collected during operations, all samples are not available. Fig. 2(b) shows the case where only the first 100 failed samples are available. The FEOL TDDB extracted parameters are η = 39.66, β = 0.99 for the first 100 samples, which are far from the actual parameters, η = 20 and β = 1.12. As the sample size is increased to 94,000, the FEOL TDDB extracted parameters are η = 19.966 and β = 1.119, which are close to the actual parameters.

The observed data is from an SRAM cell. However, wearout models, such as in (1) are for single devices with s = 1. Therefore, if we extract the wearout model parameters from SRAM data (η and β), the observed parameters are not those of single devices. They are for collections of devices, as computed with (3) and (4). Therefore, we need to map the observed results to the device model. We use simulation to find the mapping between the process-level Weibull parameters and SRAM cell Weibull parameters, as shown in Fig. 3. The figures show that any set of device-level wearout model parameters maps to SRAM cell failure distribution parameters.

Since we observe only the SRAM parameters, the maps in Fig. 3 must be inverted to map the observed SRAM parameters into device model parameters, as shown in Fig. 4.

3. Error analysis

There are random and systematic errors in the computation of the extracted parameters. The random error comes from the limited sample size used to extract the parameters and properties of the map in Fig. 3. The systematic error comes from uncertainty in temperature, supply voltage, process parameters and use scenario.

3.1. Random error: the effect of sample size

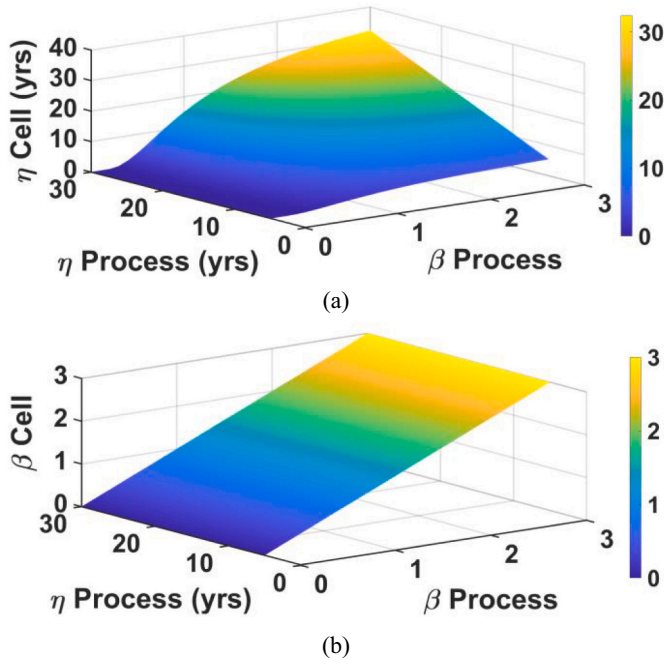The number of samples increases the accuracy in the estimation of

(a)

(b)

Fig. 3. Mapping between process-level Weibull parameters (η and β) and SRAM cell Weibull parameters for (a) η and (b) β for FEOL TDDB.
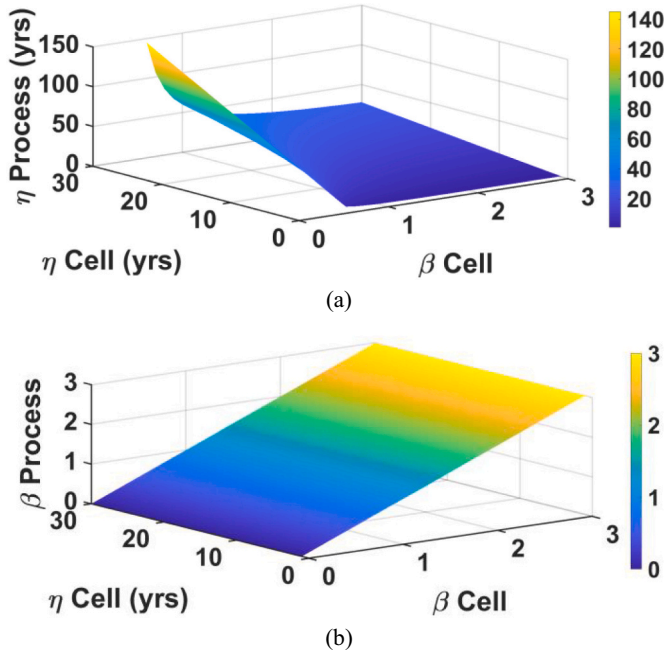

(a)

(b)

Fig. 4. Inverse mapping between process-level Weibull parameters (η and β) and SRAM cell Weibull parameters for (a) η and (b) β for FEOL TDDB.

the characteristic lifetime and shape parameter. Using model parameters and (1), we computed the expected number of samples as a function of time for an SRAM with 94 k cells. Fig. 5 shows the relative standard deviation (standard deviation/mean) for $\ln(\eta)$ as a function of time. The result for β is similar.

Fig. 5(a) is the relative error in estimating SRAM parameters. These are mapped to device model parameters with the functions in Fig. 4. The errors in estimating the parameters of the SRAM combine with the slope to determine the errors in estimating the model parameters, i.e.,
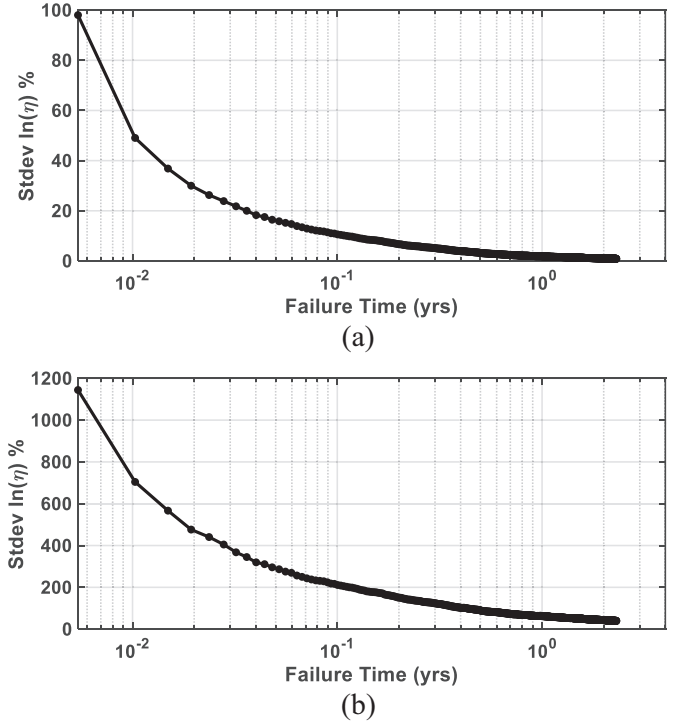

(a)

(b)

Fig. 5. Standard deviation error of the extraction of $\ln(\eta)$ as a function of time for the (a) SRAM with $\eta = 20$ yrs., $\beta = 1.12$ and (b) a single device.

$$\sigma^2(\ln(\eta_{device})) = \left(\frac{\partial(\ln(\eta_{device}))}{\partial(\ln(\eta_{SRAM}))}\right)^2 \sigma^2(\ln(\eta_{SRAM})) + \left(\frac{\partial(\ln(\eta_{device}))}{\partial(\beta_{SRAM})}\right)^2 \sigma^2(\beta_{SRAM}) \quad (5)$$

Fig. 5(b) shows an example standard deviation of extracting the characteristic lifetime for the device vs. time. The results for the shape parameter, β, are similar. The standard deviation errors for the device are larger than those of the SRAM, because the mapping from the cell to process-level parameters introduces large sensitivities. For the SRAM cell, a 30% error is observed at 0.019 years, a 20% error is seen at 0.036 years, and a 10% error is found at 0.111 yrs. For the device in Fig. 5(b), the standard deviation falls to 30% in 4.7 years.

3.2. Systematic errors

We consider four sources of systematic error that can cause the estimated device parameters to differ from their true values. These include variation in the environment, such as temperature and supply voltage. Systematic errors can also be due to process parameter variations and variation in the use scenario.

The percent changes in the SRAM characteristic lifetime errors due to the percent changes in operating temperature and voltage are shown in Fig. 6. When the operating temperature overshoots by 15% or voltage overshoots by 5%, the characteristic lifetime errors drop 82%. However, when the operating temperature undershoots by 15%, the SRAM characteristic lifetime errors can increase by 1373%. Similarly, when the operating voltage undershoots by 5%, the SRAM characteristic lifetime errors can increase by 1512%. This signals that undershooting the operating conditions has a larger effect on changing the SRAM lifetime compared to overshooting, which can be as large as 16.7 times larger for temperature with a 15% error in operating conditions, and 18.44 times larger for voltage with a 5% error in operating conditions, respectively. These errors translate into systematic errors in the estimation of device wearout parameters.

To see how the percent changes in characteristic lifetime due to variations in systematic errors translate into actual errors in device model parameters, we compute the sensitivity of device model

(a)

(b)
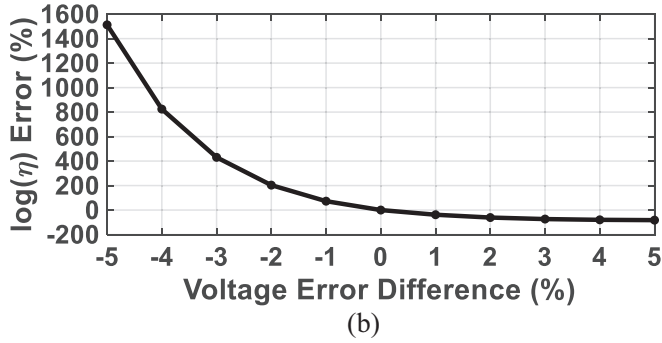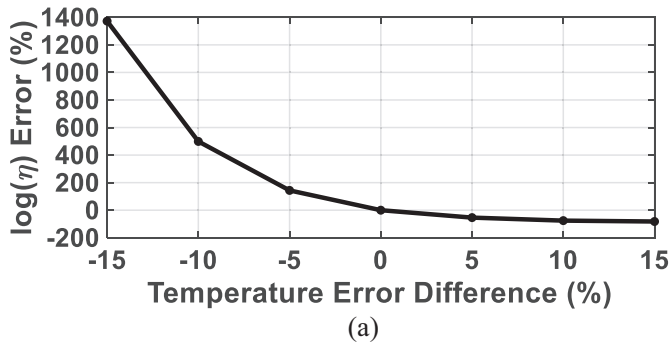
Fig. 6. Percent changes in errors in device characteristic lifetime estimation from variations in (a) temperature and (b) voltage. Voltage error differences above 5% causes the SRAM to fail upon startup and below −5% causes the SRAM to have essentially infinite characteristic lifetimes (e.g. above 300 years).
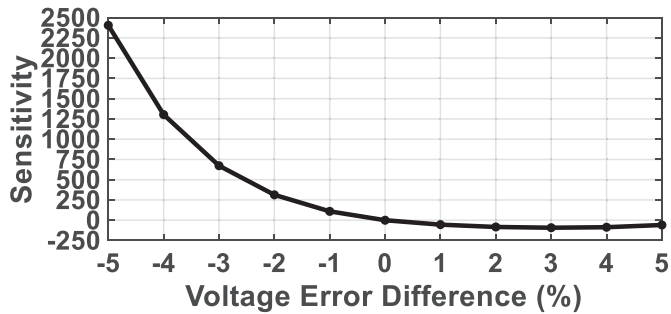

(a)

(b)
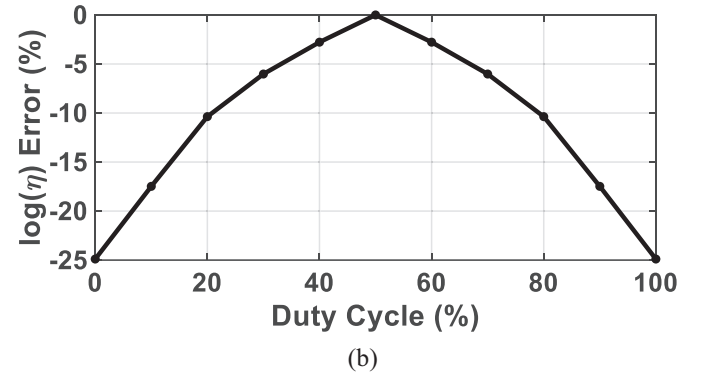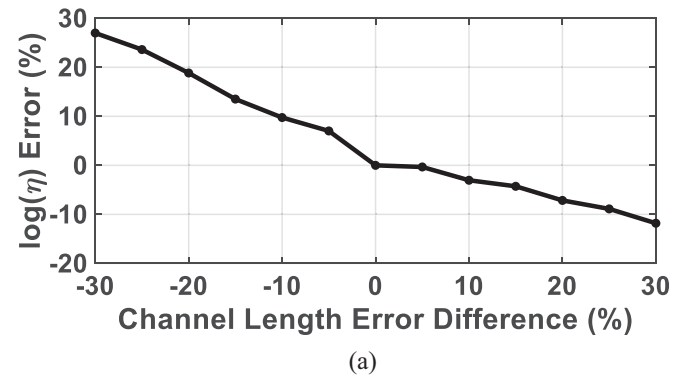
Fig. 9. Percent changes in errors for characteristic lifetime from variations in (a) channel length and (b) duty cycle (error calibrated to a duty cycle of 50%).
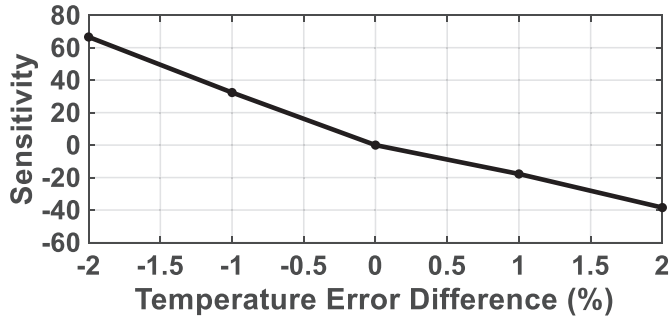
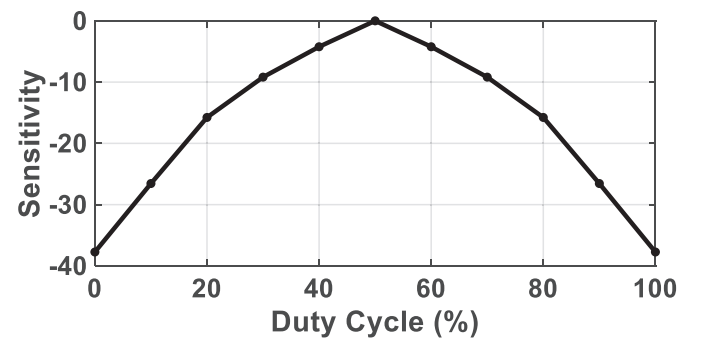**Fig. 7.** Sensitivity of the extraction of $\ln(\eta_{device})$ to changes in temperature, within the range of temperature sensor accuracy.
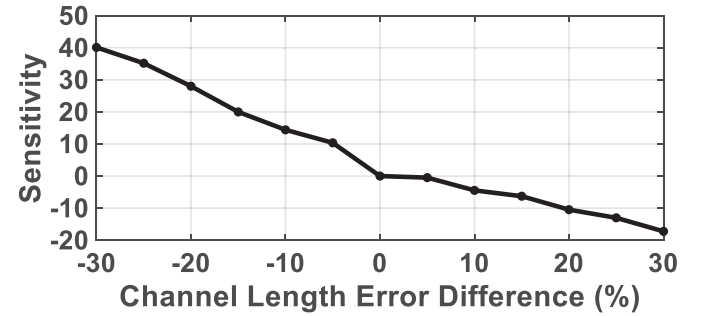
Fig. 10. Sensitivity resulting from changes in channel length.

Fig. 8. Sensitivity of the extraction of $\ln(\eta_{device})$ to changes in voltage, within the range of voltage sensor accuracy.

Fig. 11. Sensitivity from changes in duty cycle.

parameters to temperature and voltage, as follows:

$$\frac{\Delta \ln(\eta_{device})}{\ln(\eta_{device})} = \frac{\partial(\ln(\eta_{device}))/\ln(\eta_{device})}{\partial(\ln(\eta_{SRAM}))/\ln(\eta_{SRAM})} \cdot \frac{\partial\left(\ln\left(\frac{\eta_{SRAM}))}{\ln(\eta_{SRAM})}\right)\right)}{\frac{\partial(T)}{T}} + \frac{\partial\left(\ln\left(\frac{\eta_{device}))}{\ln(\beta_{device})}\right)\right)}{\frac{\partial(\beta_{SRAM})}{\beta_{SRAM}}} \cdot \frac{\frac{\partial(\beta_{SRAM})}{\beta_{SRAM}}}{\frac{\partial(T)}{T}}.$$

(6)

and

$$\frac{\Delta \ln(\eta_{device})}{\ln(\eta_{device})} = \frac{\partial(\ln(\eta_{device}))/\ln(\eta_{device})}{\partial(\ln(\eta_{SRAM}))/\ln(\eta_{SRAM})} \cdot \frac{\partial\left(\ln\left(\frac{\eta_{SRAM}))}{\ln(\eta_{SRAM})}\right)\right)}{\frac{\partial(V)}{V}} + \frac{\partial\left(\ln\left(\frac{\eta_{device}))}{\ln(\eta_{device})}\right)\right)}{\frac{\partial(\beta_{SRAM})}{\beta_{SRAM}}} \cdot \frac{\frac{\partial(\beta_{SRAM})}{\beta_{SRAM}}}{\frac{\partial(V)}{V}}.$$

(7)

Hence, systematic errors are calculated by combining the sensitivity to the shift in temperature and voltage and the sensitivity in Fig. 4, relating the SRAM and device wearout parameters.

It can be seen that systematic errors due to shifts in temperature and voltage are very large, especially for shifts towards lower temperatures and voltages. The systematic errors are not so large for positive shifts in voltage and temperature. This means that we have a much higher risk of overestimating lifetime than underestimating it. This also indicates that extracting model parameters must be used in conjunction with sensor data.

There are a wide variety of sensors that can be used to monitor temperature and voltage [10–14]. These sensors are widely used and embedded in system-on-chips (SoCs), because they are used to slow down operations when the temperature is too high, in order to prevent overheating, with a typical limit being 85 ℃, which is the limit for Raspberry Pi SOCs [15]. For a 45 nm process, the accuracy for a temperature and voltage sensor is 4.13 ℃ and 10.67 mV, respectively in the range from 0.91 V ~ 1.09 V and 0 ℃ ~ 120 ℃ [11]. Therefore, temperature variations can be detected within 2%. The sensitivity of extraction of wearout parameters to variation in temperature and voltage is illustrated in Figs. 7 and 8, respectively.

When the temperature or voltage variations are positive, the random errors dominate the temperature/voltage variations, and vise versa. The positive temperature/voltage changes make the samples fail in a very short time, so the random effects from sample size is hard to observe for these cases. Therefore, process monitors are needed to make sure the temperature and voltage do not have negative changes, which makes the errors increase greatly.

Environmental parameters, such as temperature and voltage, are not the only causes of systematic variation. Die-to-die process parameter variations and duty cycle (which is a function of the application run on the SRAM) can also cause systematic errors. We assume that die-to-die variation in channel length is the primary source of process parameter variations.

An application has a distribution of duty cycles among the SRAM cells. A duty cycle of 0, means that the cell stores only a 0, while a duty cycle of 0.5, corresponds to a cell that stores 0 50% of the time. The duty cycle impacts the stress distribution. If the mapping in Figs. 3 and 4 are constructed for a cell with a 50% duty cycle, and the actual cell experiences only a 30% duty cycle, the mapping back to device model parameters will produce errors. Most applications have duty cycle distributions that center around 30–50% [16].

As can be seen from Fig. 9, errors in process parameters and duty cycle cause smaller changes in errors in the lifetime, and as can be seen from Figs. 10 and 11, these errors cause smaller errors in the extracted parameters than voltage and temperature.

The errors in process parameters and duty cycle are also smaller than random variations. If cost and space are an issue, process and duty cycle monitors can be excluded, because they are not as dominant.

4. Conclusions

This paper has proposed to use the SRAM to count bit failures to determine the FEOL TDDB model parameters for a circuit operating in the field. Both random and systematic errors have been evaluated for a case study. Overall, systematic errors are larger than random errors when extracting device wearout parameters. Of the systematic errors, changes in supply voltage and temperature produce the largest errors. All four conditions should be monitored with sensors during operation to update the models accordingly. With appropriate sensors of operating conditions, the SRAM can be used to estimate wearout model parameters for individual chips using data from operation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M. Seok, P.R. Kinget, T. Yang, J. Li, D. Kim, Recent advances in in-situ and in-field aging monitoring and compensation for integrated circuits: Invited paper, 2018 IEEE International Reliability Physics Symposium (IRPS), 2018 11-15 March 2018. (pp. 5C.1-1-5C.1-6).

[2] T. Yang, P.R. Kinget, M. Seok, Register file circuits and post-deployment framework to monitor aging effects in field, ESSCIRC Conference 2016: 42nd European Solid-State Circuits Conference, 12–15 Sept. 2016, 2016, pp. 425–428.

[3] T. Yang, D. Kim, J. Li, P.R. Kinget, M. Seok, In-situ and in-field technique for monitoring and decelerating NBTI in 6T-SRAM register files, IEEE Transactions on Very Large Scale Integration (VLSI) Systems 26 (11) (2018) 2241–2253.

[4] F. Ahmed, L. Milor, Online measurement of degradation due to bias temperature instability in SRAMs, IEEE Transactions on Very Large Scale Integration (VLSI) Systems 24 (6) (2016) 2184–2194.

[5] E. Wu, J. Suñé, W. Lai, E. Nowak, J. McKenna, A. Vayshenker, D. Harmon, Interplay of voltage and temperature acceleration of oxide breakdown for ultra-thin gate oxides, Solid State Electron. 46 (11) (2002) 1787–1798.

[6] S. Hsu, K, Yang, and L. Milor, "reliability and accelerated testing of 14nm FinFET ring oscillators," in 2019 XXXIV Conference on Design of Circuits and Integrated Systems (DCIS), 20-22 Nov, 2019 (2019) 1–7.

[7] D. Kim, L. Milor, An ECC-assisted postpackage repair methodology in main memory systems, IEEE Transactions on Very Large Scale Integration (VLSI) Systems 25 (7) (2017) 2045–2058.

[8] W. Kim, C. Chen, D. Kim, L. Milor, Built-in self-test methodology with statistical analysis for electrical diagnosis of wearout in a static random access memory array, IEEE Transactions on Very Large Scale Integration (VLSI) Systems 24 (7) (2016) 2521–2534.

[9] D. Kim, S. Hsu, L. Milor, Optimization of experimental designs for system- level accelerated life test in a memory system degraded by time-dependent dielectric breakdown, IEEE Transactions on Very Large Scale Integration (VLSI) Systems 27 (7) (2019) 1640–1651.

[10] T. Kishimoto, T. Ishihara, H. Onodera, On-chip reconfigurable monitor circuit for process variation and temperature estimation, 2018 IEEE International Conference on Microelectronic Test Structures (ICMTS), 19–22 March 2018, 2018, pp. 111–116.

[11] Y. Miyake, Y. Sato, S. Kajihara, Y. Miura, Temperature and voltage measurement for field test using an aging-tolerant monitor, IEEE Transactions on Very Large Scale Integration (VLSI) Systems 24 (11) (2016) 3282–3295.

[12] A.K.M.M. Islam, J. Shiomi, T. Ishihara, H. Onodera, Wide-supply-range all-digital leakage variation sensor for on-chip process and temperature monitoring, IEEE J. Solid State Circuits 50 (11) (2015) 2475–2490.

[13] D.A. Kamakshi, H.N. Patel, A. Roy, B.H. Calhoun, A 28 nW CMOS supply voltage monitor for adaptive ultra-low power IoT chips, 2017 IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S), 16–19 Oct. 2017, 2017, pp. 1–2.

[14] C. Chung, M. Sun, An all-digital voltage sensor for static voltage drop measurements, 2016 IEEE Sensors Applications Symposium (SAS), 20–22 April 2016, 2016, pp. 1–4.

[15] R. Pi, Frequency management and thermal control documentation, https://www.raspberrypi.org/documentation/hardware/raspberrypi/frequency-management.md.

[16] R. Zhang, T. Liu, K. Yang, C. Chen, L. Milor, SRAM stability analysis and performance–reliability tradeoff for different cache configurations, IEEE Transactions on Very Large Scale Integration (VLSI) Systems 28 (3) (2020) 620–633.