Machine Learning for Detection of Competing Wearout Mechanisms

Shu-han Hsu, Kexin Yang, and Linda Milor
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA 30332, USA
shsu33@gatech.edu

Abstract— **Because data from a variety of wearout mechanisms is confounded in circuits, we apply machine learning techniques to detect the parameters of competing failure mechanisms in ring oscillators, which more closely mimic circuit behavior than test structures. This is the first known application using data analysis to distinguish competing wearout mechanisms in circuit-level data. To quickly and efficiently analyze failure data, we propose to use maximum likelihood estimation to separately determine the parameters of each underlying distribution by only observing the time-to-failure of samples. The quasi-Newton method is used to update and optimize the parameter extraction.**

Index Terms—**Failure analysis; Quasi-Newton optimization; Time-dependent dielectric breakdown (TDDB); Wearout; Weibull distribution**

I. Introduction

Because emerging technologies, such as autonomous vehicles and wearable sensors for health monitoring, are becoming increasingly interrelated with public safety, the need for the assessment of highly reliable complex systems is important. To meet this demand, it is important to not just check if test structure data meet lifetime requirements, but to also test circuit-like test structures, such as ring oscillators. Because of the presence of confounded wearout mechanisms in circuit-like test structures, new data analysis methods are required. In this work, machine learning techniques are used to analyze failure data.

Ring oscillators and other small circuit blocks, which have behaviors similar to circuits, can be used to test for failure modes. However, unlike test structures that isolate failure modes, in circuits, various wearout mechanisms are convoluted with each other. Generally, invasive diagnostic methods for failure analysis, such as transmission electron microscopy (TEM), e-beam, or scanning electron microscopy (SEM), are used to study the failure modes, which require samples to be cut open, for example, using focused ion beam (FIB) techniques [1], [2]. When devices are built at the FinFET node or smaller, the metrology for TEM and other failure analysis techniques can become intricate and complex, requiring significant time to prepare and analyze samples. This causes a wait time to receive the failure results and high costs, which can impact product costs if done too often. Therefore, it is necessary to find a quick and non-invasive method to separate the causes of failure, so that efforts for process improvement can be prioritized.

This work uses 14nm FinFET ring oscillators as the circuit vehicle to extract wearout data, caused by a wide variety of wearout mechanisms, including time-dependent dielectric breakdown (TDDB), electromigration (EM), and stress-induced voiding (SIV). The ring oscillators are based on the 14nm FinFET pdk technology node design kit jointly developed by IBM, GlobalFoundries (GF), and Samsung. We use lifetime simulation of circuits [3]-[5] to generate the data sets to test our methodology. The lifetime simulation models are calibrated to test structure data. They take into account the stress of layout geometries and transistors, based on circuit operating conditions.

Using the data on time-to-failure of samples, maximum likelihood estimation analysis (MLE) [6] and the Weibull distribution are applied to detect the Weibull parameters of the primary and secondary wearout mechanisms, which are competing and independent of each other. We investigate the application of the quasi-Newton method and modifications to the Davidon-Fletcher-Powell (DFP) algorithm for implementing MLE.

In the next section, we provide background information on machine learning and the quasi-Newton method algorithm that we will use to extract parameters. Section 3 summarizes the wearout models and circuit example used in this paper. Section 4 discusses the modeling of competing and independent wearout mechanisms, followed by its application to our circuits in Section 5. Section 6 summarizes the parameter extraction method, followed by a discussion in Section 7. The paper concludes in Section 8 with a summary.

II. Machine Learning

Machine learning is an automated data analysis method that uses algorithms to iteratively learn from data and improve performance measure, such as prediction accuracy, patterns and/or key trends [7]. The advantages for using machine learning include quicker, faster, scalable, and more cost-effective analysis of complex data. For reliability applications, in this work, machine learning is also used to identify model parameters based on collected time-to-failure data.

978-1-5386-9504-3/19/$31.00 ©2019 IEEE

The quasi-Newton and gradient descent methods are the two most common updating techniques in machine learning [8], [9]. We employ the quasi-Newton method instead of a gradient descent method in MLE to find the parameters of the competing Weibull distributions. The quasi-Newton method, which is a second-order method, uses fewer steps to find the optimal value, taking more time to execute each step, while the gradient descent method, which is a first-order method, has the opposite properties [10]. The gradient descent is more suitable when there is a large data set or numerous parameters (ex. millions of samples or thousands of parameters), whereas the quasi-Newton method is used for problems with the opposite characteristics. Generally, industrial data sets vary from on the order of 10 to 1000 samples, and there are usually two wearout mechanisms that are most dominant. Therefore, only a few parameters need to be determined. Consequently, we have chosen to use the quasi-Newton method in our examples.

A. Quasi-Newton Method

The quasi-Newton method is a class of optimization algorithms that can find the minima or maxima of functions. It is based on Newton's method, which uses the first and second derivative (Hessian matrix) values to find the roots of a function. Newton's method can be thought of as similar to a gradient descent method with the addition of using second-order information to change the step size and direction. This addition of the Hessian information can help avoid descent directions that plateau too quickly. Newton's method is generally computationally expensive and slow, because it is more difficult to calculate the second derivative. The quasi-Newton method overcomes this problem by approximating the Hessian matrix instead of computing it directly. A variety of algorithms are available that involve different approximation methods for the Hessian matrix.

B. Algorithms for the Quasi-Newton Method

Two of the most common updating algorithms are the Broyden–Fletcher–Goldfarb–Shanno (BFGS) and Davidon-Fletcher-Powell (DFP) algorithms. Both algorithms are almost the same, differing only in the Hessian matrix update, and are complementary [11], [12]. Because our data set is generally on the order of 10 to 1000 samples, we use the DFP algorithm, which is less complicated than BFGS in computation. DFP uses the secant method to approximate the Hessian matrix. The complete procedure and modifications will be discussed in Section VI.

III. Wearout Models and Circuit Example

Wearout mechanisms for devices, which include front-end/middle-end/back-end of line time-dependent dielectric breakdown (FEOL/MOL/BEOL TDDB), electromigration (EM), and stress-induced voiding (SIV) can be described using the Weibull distribution with two parameters. The parameters are the characteristic lifetime, η, which is the time-to-failure of a sample at 63% failure probability, and the shape parameter, β, which describes the spread of the distribution of failure samples.

A. TDDB

Front-end of the line time-dependent dielectric breakdown (FEOL TDDB, also referred to as GTDDB) occurs when traps build up in the gate oxide region, and can be modeled as [13], [14]:

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

where a, b, c, d, and A_{FEOL} are process-dependent constants. V, T, and s are the voltage, temperature, and probability of stress, respectively. W and L are the width and length of the device. Experimental data were used to determine the constants [15].

Back-end of line time-dependent dielectric breakdown (BEOL TDDB, also referred to as BTDDB) occurs in the dielectric region between adjacent metal interconnect lines, described as [3]:

$$\eta = A_{BEOL} L_i^{-1/\beta} exp(-\gamma E^m + E_a/kT) s^{-1} \quad (2)$$

where A_{BEOL} is a constant that depends on the dielectric material properties, L_i is the vulnerable length (distance where metals are parallel to each other), T is temperature, γ (8.723113 cm/MV) is the field acceleration factor, k is the Boltzmann constant, s is the probability of stress, E_a is activation energy (~0.5eV), and m is ½ for the $\sqrt{E}$ model [16]. The electric field is dependent on voltage and distance between the conductors, S_i, i.e., $E = V/S_i$. The parameters are from experimental data [17].

Middle-of-the-line time-dependent dielectric breakdown (MOL TDDB, also referred to as MTDDB) is now of a concern due to dimension scaling and structure change in FinFET devices. It occurs in the spacing between the gate and source/drain contacts, and is modeled similarly to BEOL TDDB [18]:

$$\eta = A_{MOL} L_i^{-1/\beta} exp(-\gamma E^m + E_a/kT) s^{-1} \quad (3)$$

where A_{MOL} is a constant, E_a is activation energy (~0.5eV), γ (8.723113 cm/MV) is the field acceleration factor, and m is 1 for the E model [19]. The other parameters are similar to the BEOL TDDB model, and the parameters are from experimental data [19], [20].

B. EM

Electromigration (EM) is the dislocation of atoms in interconnect metals from the momentum transfer of electrons, modeled as [21]:

$$\eta = A_{EM} J^{-n} exp(E_a/kT) \quad (4)$$

where A_{EM} is a constant, T is temperature, J is current density, E_a is activation energy (0.85eV for copper [22]), n=1 (void growth), and k is Boltzmann constant. The experimental data for the parameters is from [21].

C. SIV

Stress-induced voiding (SIV) is device failure from the formation of voids by directionally biased motion of atoms induced from thermal mechanical stress between the metal and dielectric as follows [23]:

$$\eta = A_{SIV} W^{-M} (T_o - T)^{-N} exp(E_a/kT) \quad (5)$$

where W is linewidth, M is geometry stress component, T_o is stress-free temperature, T is temperature, N is thermal stress

component, E_a is activation energy, and A_{SIV} is a constant. The experimental data for the parameters is from [24].

D. Ring Oscillator Circuit

Because signals propagate through a circuit in various ways depending on the circuit design, the stress probabilities for each circuit will be different. The stress probability is the percentage of time the circuit is under stress. As shown in Fig. 1, the ring oscillator is composed of identical inverters and physical connections. For all three TDDB mechanisms, the duty cycle was calculated through Cadence Virtuoso and SPICE simulations to all have the same value of 0.5 for both NMOS and PMOS devices in each stage. Because the duty cycle determines the percentage of time the NMOS or PMOS is on, this is equivalent to calculating the stress probabilities for the TDDB mechanisms for each device in the circuit. The characteristic lifetime of the circuit can be calculated by [5]:

$$1 = \sum_{i=1}^{n} \left(\frac{\eta_{circuit}}{\eta_i}\right)^{\beta_i} \quad (6)$$

where η_i, i=1, …, n are the characteristic lifetimes of all the devices; and β_i, i=1, … n are the corresponding shape parameters. Since all devices are identical, with identical structures,

$$\eta_{circuit} = \left(x * \eta_i^{-\beta}\right)^{-1/\beta} \quad (7)$$

where x is the number of stages in the ring oscillator, and η_i is η of an invertor, $\eta_{inv} = \left(\eta_{pmos}^{-\beta} + \eta_{nmos}^{-\beta}\right)^{-1/\beta}$.

However, for EM, the degradation is dependent on the current flowing through the backend layers, and the current is the stress. For the case of a ring oscillator circuit, the current is only present instantaneously when the device is turned on to pull internal nodes to power or the ground rail. The average current density, J, in (4), is computed based on the instantaneous current. For SIV, the temperature is the stress.

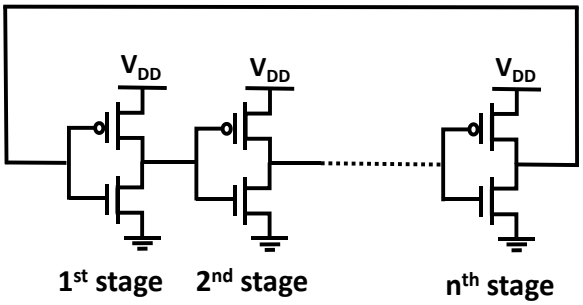


Fig.1. The circuit diagram of an n-stage ring oscillator

IV. Competing Wearout Mechanisms

Competing wearout mechanisms occur when failures are due to more than one degradation mode. These mechanisms are independent of each other. Suppose mechanism 1 (primary breakdown mode) has a probability density function, $f_1(t)$ and cumulative distribution function, $F_1(t)$. The survival function is $R_1(t) = 1 - F_1(t)$. Similarly, mechanism 2 (secondary breakdown mode) has the probability density function, $f_2(t)$; cumulative distribution function, $F_2(t)$; survival function, $R_2(t) = 1 - F_2(t)$. The competing failure probability density function, f(t), can be described as below [6].

$$\begin{aligned} f(t) &= P\{T_1 = t, T_2 \geq t\} \cup P\{T_1 \geq t, T_2 = t\} \\ &= P(T_1 = t, T_2 \geq t) + P(T_1 \geq t, T_2 = t) \\ &= P(T_1 = t)\, P(T_2 \geq t) + P(T_1 \geq t)\, P(T_2 = t) \\ &= f_1(t) * R_2(t) + f_2(t) * R_1(t) \end{aligned} \quad (8)$$

The competing failure probability density function should not be confused with the mixed Weibull probability density function, which has a similar equation shown below [25]:

$$f(t) = a * f_1(t) + b * f_2(t) \quad (9)$$

where a and b are the mixing weights. The mixed Weibull probability density function is where the breakdown occurs due to both mechanisms being present at the same time. However, the competing Weibull probability density function is where only one mechanism contributes to the breakdown at a specific failure time, but the cause could be either from mechanism 1 or 2, not both.

V. Parameter Extraction for Competing Wearout Mechanisms

In order to develop our methodology, we consider 3 ring oscillator structures, with the Weibull parameters for each set of competing mechanisms shown in Table I. Set I is based on a 11-stage ring oscillator, where the dominant and secondary wearout mechanisms are GTDDB and BTDDB, respectively. The dominant mechanism is the one that has the highest number of failures. Set 2 is based on a 501-stage ring oscillator, with EM being the dominant mechanism, and SIV as the secondary wearout mechanism. Set 3 is based on a 331-stage ring oscillator, with GTDDB as the dominant mechanism, and MTDDB as the secondary mechanism.

To model the failure distribution of these competing mechanisms, a point is picked randomly from each distribution. The smaller value is set as the lifetime, because it is the dominant wearout mechanism that fails first at that time point. Then, the points are plotted as ordered pairs: $(\ln(t_1), \ln(-\ln(1-(\frac{1}{2N}))))$, $(\ln(t_2), \ln(-\ln(1-(\frac{3}{2N}))))$, etc. Sample sizes N of 10, 100, and 1000 were generated for each set. The resulting datasets are shown in Fig. 2.

For a small sample size of 10, generally the failure times will be scattered closer together in the middle. When the sample sizes are increased, there is a higher probability of getting failure times at the extreme ends, with a lower or higher failure time. As the sample sizes are increased further to 1000, this results in the failure points landing closer together to the true Weibull parameters due to the sampling time range being enlarged, which is also why the accuracy in prediction increases for larger data sets.

VI. Methodology To Extract Competing Weibull Parameters

A. Maximum Likelihood Estimation

Maximum likelihood estimation (MLE) is a method that determines the parameters of a model from given observations, which would be failure times in our examples, by finding the parameter values that maximizes the likelihood or highest

TABLE I
COMPETING WEIBULL PARAMETERS FOR RING OSCILLATORS

Set #	Ring Oscillator Type	Mechanism 1 (Primary)	β_1	η_1 (yrs)	Mechanism 2 (Secondary)	β_2	η_2 (yrs)
1	11-STAGE	GTDDB	1.64	101.1959	BTDDB	1.9	149.8532
2	501-STAGE	EM	1.2	25.1296	SIV	1.2	30.2113
3	331-STAGE	GTDDB	1.64	12.6963	MTDDB	0.98	23.1412

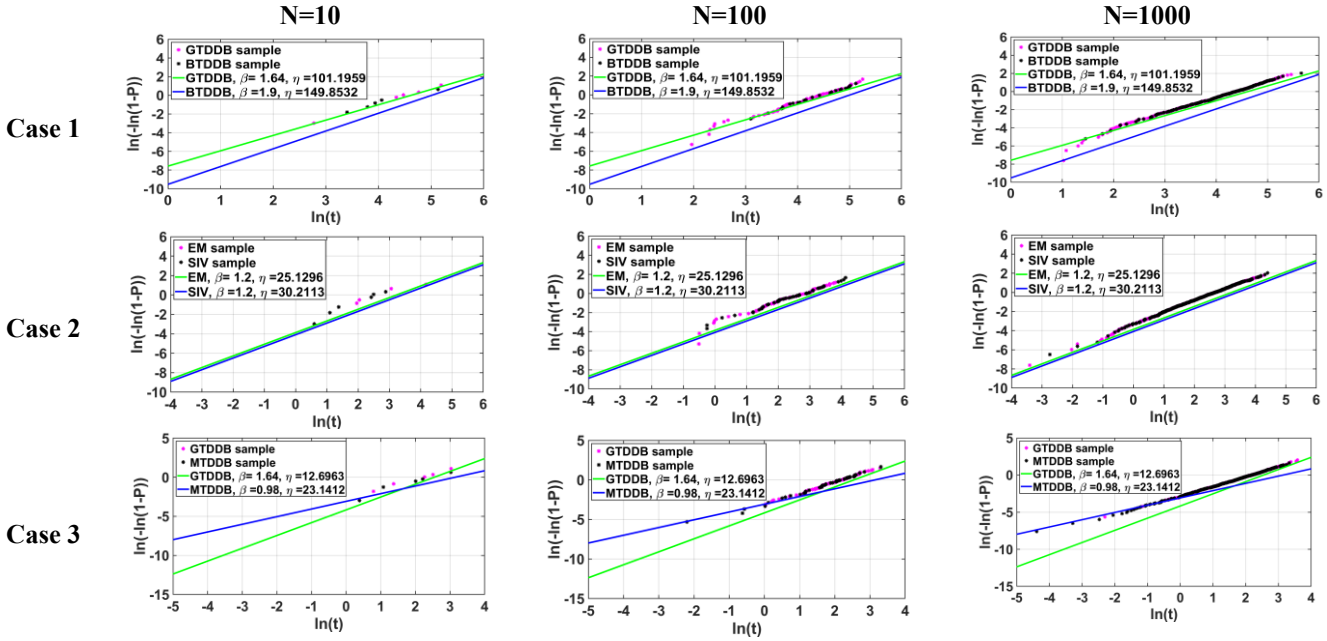


Fig. 2. Distributions of data for the three case studies (x axis: unit years), varying sample size. P is probability. The pink markers correspond to the primary wearout mechanism, while the black markers correspond to the secondary wearout mechanism.

probability of getting the observations given the parameters. The intuition is that the estimate which explains the data best will be the best estimator. MLE is employed to estimate the competing Weibull parameters for each set of distributions from Fig. 2.

The likelihood function for uncensored data can be simplified from [6] as:

$$\mathcal{L}(\theta) = C \prod_{i=1}^{N} f(t_i) \tag{10}$$

where θ is the set of competing Weibull parameters, β_1, β_2, η_1, η_2, and the log likelihood function can be written as

$$\ln \mathcal{L}(\theta) = \sum_{i=1}^{N} \ln f(t_i) + \ln C \tag{11}$$

where C is a constant.

The derivatives can also be simplified from [6] as:

$$\frac{\partial \ln \mathcal{L}(\theta)}{\partial \beta_1} = \sum_{i=1}^{N} \left\{ \frac{dR_1(t_i)}{d\beta_1} f_2(t_i) + \frac{df_1(t_i)}{d\beta_1} R_2(t_i) \right\} / f(t_i) \tag{12}$$

$$\frac{\partial \ln \mathcal{L}(\theta)}{\partial \beta_2} = \sum_{i=1}^{N} \left\{ R_1(t_i) \frac{df_2(t_i)}{d\beta_2} + f_1(t_i) \frac{dR_2(t_i)}{d\beta_2} \right\} / f(t_i) \tag{13}$$

$$\frac{\partial \ln \mathcal{L}(\theta)}{\partial \eta_1} = \sum_{i=1}^{N} \left\{ \frac{dR_1(t_i)}{d\eta_1} f_2(t_i) + \frac{df_1(t_i)}{d\eta_1} R_2(t_i) \right\} / f(t_i) \tag{14}$$

$$\frac{\partial \ln \mathcal{L}(\theta)}{\partial \eta_2} = \sum_{i=1}^{N} \left\{ R_1(t_i) \frac{df_2(t_i)}{d\eta_2} + f_1(t_i) \frac{dR_2(t_i)}{d\eta_2} \right\} / f(t_i) \tag{15}$$

$$\frac{dR_k(t)}{d\beta_k} = -R_k(t) \left(\frac{t}{\eta_k}\right)^{\beta_k} \ln\left(\frac{t}{\eta_k}\right) \tag{16}$$

$$\frac{df_k(t)}{d\beta_k} = \left(\frac{t}{\eta_k}\right)^{\beta_k - 1} R_k(t) \left\{ \left(\frac{1}{\eta_k}\right) + \left(\frac{\beta_k}{\eta_k}\right) \ln\left(\frac{t}{\eta_k}\right) \right\} - f_k(t) \left(\frac{t}{\eta_k}\right)^{\beta_k} \ln\left(\frac{t}{\eta_k}\right) \tag{17}$$

$$\frac{dR_k(t)}{d\eta_k} = R_k(t) \left(\frac{\beta_k}{\eta_k}\right) \left(\frac{t}{\eta_k}\right)^{\beta_k} \tag{18}$$

$$\frac{df_k(t)}{d\eta_k} = \left(\frac{\beta_k}{\eta_k}\right)^2 \left(\frac{t}{\eta_k}\right)^{\beta_k - 1} R_k(t) \left\{ -1 + \left(\frac{t}{\eta_k}\right)^{\beta_k} \right\} \tag{19}$$

where k=1 or 2.

B. *Quasi-Newton Algorithm*

The quasi-Newton method using the Davidon-Fletcher-Powell algorithm is used to optimize eq. (11).

Our implementation is shown in Fig. 2. We made modifications to step 3 by choosing a step size different from conventional choices that find the step size through quadratic interpolation [6]. The quadratic interpolation method may be better suited for large data sets on the order of millions of samples, because it should converge faster than a fixed step size. However, sample sizes for reliability tests in industry usually range from 10 to 1000 samples, so the relatively small sample size here may cause the quadratic interpolation method to be stuck or lost in local minima/maxima points during the changing step size calculations. In this case, switching to a

fixed constant step size can be considered, since the speed is still fast due to the relatively small sample size. For our cases, using the quadratic interpolation method often resulted in changing step sizes that could not converge. However, when we choose our step size to be a fixed constant, such as α=0.0001, the algorithm would converge easily.

Algorithm Procedure [6]

1. Initial condition: $\theta^0 = [\beta_1, \beta_2, \eta_1, \eta_2]^T$, $L^0 = lnL(\theta^0)$

2. Set optimization direction:

$d^i = -S_i g^i$, where $S_i = I(the\ unit\ matrix)$,

$g^i = -\nabla \ln L\,(\theta^i)^T$, i=0

3. Line search: $\theta^{i+1} = \theta^i + \alpha_i d^i$, where α_i is the optimal step length

4. Calculate parameters for Hessian matrix and new direction:

$p^i = \alpha_i d^i$, $g^{i+1} = -\nabla \ln L\,(\theta^{i+1})^T$, $q_i = g^{i+1} - g^i$

5. Estimate inverse Hessian matrix:

$S_{i+1} = S_i + \frac{p^i p^{iT}}{p^{iT} q^i} - \frac{S_i q^i q^{iT} S_i}{q^{iT} S_i q^i}$

6. Find set of Weibull parameters: Set i=i+1. If i=4 (number of Weibull parameters) then go to Step 7; otherwise go back to step 2.

7. Iteration procedure for optimization and stop condition:

If $\left|\ln L(\theta^i) - L^0\right| < \varepsilon$, then stop. Otherwise, set

$L^0 = \ln L\,(\theta^i)$, i=0 and go back to step 2.

Fig. 3. Implementation of the maximum likelihood based parameter extraction algorithm.

VII. Analysis of Extracted Competing Weibull Parameters

A. Case Studies

The initial conditions for β and η were both varied at the same increments for both competing mechanisms at the same time for sample sizes of 10, 100, and 1000. For example, in set 1, the β and η of both mechanisms were both set at 5% deviation from the initial condition, and MLE was employed to obtain the estimation results. Then, this procedure was repeated by setting both β and η of both mechanisms at 10% deviation from the initial condition, and the process was repeated again increasing the deviation by 5% each time up to the 15% deviation from the initial condition. The entire procedure was also repeated for deviation from -5% to -15% from the initial condition. For a sample size of 10, usually the MLE was unable to separate the competing Weibull parameters after deviating more than $\pm$5%. For the sample sizes of 100 and 1000, the MLE usually was unable to separate the competing Weibull parameters beyond $\pm$15%. The results are shown in Table II.

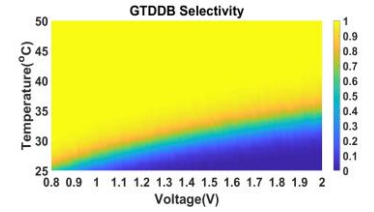
For all three competing mechanism sets, regardless of sample size, the error in estimation of β values is not correlated to the amount of deviation in the initial guess. However, the percentage error in the estimated η values corresponds proportionally to the deviation in the initial guess. For both cases, there may be an occasional point where there is a large error, which may be due to the MLE being stuck in a local minimum.

In all sets, the MLE was able to distinguish between the two competing mechanisms. However, the initial guess should be within $\pm$5% of the final result if the sample size is 10 for more accurate results. When the sample size is increased to 100, the initial condition can be farther from the final solution, i.e., $\pm$ 15%, while still being able to separate the competing mechanisms. Generally, when the sample size is increased further from 100 to 1000, the accuracy in determining the β values does not have much of an effect, but the accuracy in determining η is increased slightly. Sometimes, the range of competing Weibull parameters may be known beforehand due to previous experience. In this case, taking the average of all calculated estimated parameters over the range in deviations in initial conditions may lower the estimated errors, as shown in Table III.

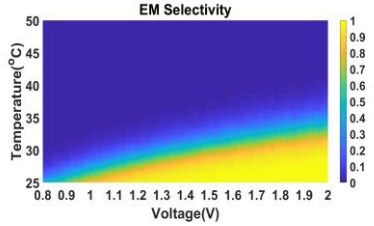
B. Application to Accelerated Testing

To study the competing mechanisms in accelerated testing, the 9-stage ring oscillator for GTDDB vs EM was chosen as an in-depth case study, because these were the only two detectable mechanisms over the test domain. The selectivity of each mechanism, which is defined as the probability of that mechanism failing first at the corresponding voltage and temperature, can found by using the Monte Carlo method, by using the steps described in Section V over all temperature and voltage conditions, as shown in Fig. 4.

To pick a GTDDB dominant region, the test condition of 2V, 37°C was selected, with GTDDB β=1.64, η= 26.92 min and EM β=1.2, η= 94.42 min, respectively, which has a GTDDB selectivity of 0.826. To pick an EM dominant region, the test condition of 2V, 32°C was selected, with GTDDB β=1.64, η= 160.49 min and EM β=1.2, η= 94.64 min, respectively, which has an EM selectivity of 0.7. For a region with a selectivity of 0.5 for both GTDDB and SIV, the test condition of 2V, 34°C was selected, with GTDDB β=1.64, η= 87.79 min and EM β=1.2, η= 94.55 min, respectively. These test conditions were picked to ensure that the ring oscillators failed in a reasonable and short time range.



(a)



(b)

Fig. 4. Selectivity for 9-stage ring oscillator (a) GTDDB and (b) EM

TABLE II
Maximum and Average Errors in Parameter Estimation vs Respective Deviations in Initial Conditions for Case Studies

Case	Sample Size	Mechanism	Parameter	Max Error	% Error in Initial Condition at Max Error	Average of Absolute Error for All Deviations in Initial Conditions	Corresponding Ranges used for Deviations in Initial Conditions
#1 (11-stage RO)	10	GTDDB	β	-16.605%	-5%	13.409%	-5% to +5%
			η	5.044%	5%	4.984%	
		BTDDB	β	9.747%	-5%	8.127%	
			η	5.008%	5%	4.993%	
	100	GTDDB	β	-25.868%	15%	12.356%	-15% to +15%
			η	15.008%	15%	9.914%	
		BTDDB	β	-61.884%	15%	37.969%	
			η	14.997%	15%	9.973%	
	1000	GTDDB	β	30.550%	15%	9.717%	-15% to +15%
			η	14.728%	15%	9.865%	
		BTDDB	β	52.369%	10%	30.178%	
			η	14.935%	15%	9.960%	
#2 (501-stage RO)	10	EM	β	-65.037%	-5%	40.957%	-5% to +5%
			η	-5.298%	-5%	5.211%	
		SIV	β	-50.348%	-5%	32.752%	
			η	-5.211%	-5%	5.132%	
	100	EM	β	-16.275%	-15%	10.360%	-15% to +15%
			η	15.361%	15%	9.342%	
		SIV	β	-18.458%	15%	6.076%	
			η	14.902%	15%	9.485%	
	1000	EM	β	-8.036%	-5%	4.472%	-15% to +15%
			η	-14.509%	-15%	8.310%	
		SIV	β	-14.855%	15%	8.227%	
			η	-14.701%	-15%	9.174%	
#3 (331-stage RO)	10	GTDDB	β	-22.072%	-5%	20.770%	-5% to +5%
			η	6.140%	5%	4.692%	
		MTDDB	β	58.133%	-5%	43.308%	
			η	4.986%	5%	4.876%	
	100	GTDDB	β	-35.044%	-10%	11.019%	-15% to +15%
			η	-14.191%	-10%	8.458%	
		MTDDB	β	79.540%	-10%	23.848%	
			η	-15.135%	-15%	9.958%	
	1000	GTDDB	β	-25.465%	-15%	9.110%	-15% to +15%
			η	-11.228%	-15%	8.293%	
		MTDDB	β	72.419%	-15%	19.656%	
			η	-14.658%	-15%	9.635%	

TABLE III
Averaged Estimated Values Over All Deviations in Initial Conditions vs Respective Errors for Case Studies

Case	Sample Size	Mechanism	Parameter	Average of All Estimated Values	Error of Averaged Estimated Value
#1 (11-stage RO)	10	GTDDB	β	1.420	-13.409%
			η	101.256	0.060%
		BTDDB	β	2.054	8.127%
			η	149.876	0.015%
	100	GTDDB	β	1.517	-7.526%
			η	101.283	0.086%
		BTDDB	β	2.229	17.341%
			η	149.891	0.025%
	1000	GTDDB	β	1.667	1.653%
			η	101.272	0.075%
		BTDDB	β	2.205	16.047%
			η	149.894	0.027%
#2 (501-stage RO)	10	EM	β	0.709	-40.957%
			η	25.108	-0.087%
		SIV	β	0.807	-32.752%
			η	30.187	-0.079%
	100	EM	β	1.115	-7.080%
			η	25.327	0.785%
		SIV	β	1.166	-2.818%
			η	30.354	0.471%
	1000	EM	β	1.205	0.416%
			η	24.894	-0.938%
		SIV	β	1.164	-3.036%
			η	30.093	-0.393%
#3 (331-stage RO)	10	GTDDB	β	1.299	-20.770%
			η	12.880	1.448%
		MTDDB	β	1.404	43.308%
			η	23.167	0.110%
	100	GTDDB	β	1.586	-3.308%
			η	12.794	0.768%
		MTDDB	β	1.184	20.782%
			η	23.121	-0.089%
	1000	GTDDB	β	1.505	-8.202%
			η	12.262	-3.418%
		MTDDB	β	1.148	17.100%
			η	23.082	-0.254%

TABLE IV
Maximum and Average Errors in Parameter Estimation vs Respective Deviations in Initial Conditions for Accelerated Testing Application

Case	Sample Size	Mechanism	Parameter	Max Error	% Error in Initial Condition at Max Error	Average of Absolute Error for All Deviations in Initial Conditions	Corresponding Ranges used for Deviations in Initial Conditions
GTDDB dominant	**10**	**GTDDB**	β	-20.79%	-5%	20.71%	-5% to +5%
			η	-4.33%	-5%	3.84%	
		EM	β	9.06%	-5%	7.67%	
			η	-5.19%	5%	3.25%	
	100	**GTDDB**	β	-24.17%	5%	19.03%	-15% to +15%
			η	13.82%	15%	7.58%	
		EM	β	7.94%	5%	5.43%	
			η	17.03%	-10%	9.38%	
	1000	**GTDDB**	β	-36.63%	10.00%	19.21%	-15% to +15%
			η	14.96%	5%	9.04%	
		EM	β	31.99%	-10%	12.67%	
			η	10.80%	-15%	7.90%	
EM dominant	**10**	**EM**	β	26.71%	5%	26.57%	-5% to +5%
			η	29.72%	5%	29.12%	
		GTDDB	β	-17.16%	5%	17.13%	
			η	-5.34%	5%	4.92%	
	100	**EM**	β	-6.73%	-5%	6.59%	-15% to +15%
			η	-15.62%	15%	14.59%	
		GTDDB	β	27.75%	-10%	25.61%	
			η	-14.98%	-15%	9.88%	
	1000	**EM**	β	8.30%	-10%	7.57%	-15% to +15%
			η	2.77%	-10%	1.23%	
		GTDDB	β	-11.36%	15%	7.66%	
			η	44.13%	15%	18.90%	
GTDDB & EM Selectivity of 0.5	**10**	**GTDDB**	β	-14.31%	5% & -5%	14.31%	-5% to +5%
			η	18.16%	-5%	17.66%	
		EM	β	17.11%	5% & -5%	17.11%	
			η	14.64%	5%	14.13%	
	100	**GTDDB**	β	18.90%	-15%	16.46%	-15% to +15%
			η	-10.46%	5%	7.91%	
		EM	β	1.25%	-5%	1.10%	
			η	9.86%	15%	5.26%	
	1000	**GTDDB**	β	-22.07%	-10%	9.39%	-15% to +15%
			η	-5.69%	10%	2.61%	
		EM	β	34.37%	-15%	14.42%	
			η	6.57%	-5%	3.28%	

The analysis for the 9-stage ring oscillator at accelerated conditions is summarized in Table IV. Similar to the case studies, the MLE was able to discern between the two competing Weibull parameters. Since the MLE can distinguish at 50% selectivity with low absolute error, then it is not necessary to do testing at the GTDDB dominant and EM dominant conditions to find these failure modes, which can lower the sample sizes and costs. If the difference in η for the two mechanisms is larger than an order of magnitude, such as when the dominant mechanism has a selectivity of 90%, it will become difficult for MLE to distinguish the parameters for the two mechanisms, because there will be few samples coming from the non-dominant mechanism. The maximum error for the case studies and accelerated conditions does not have any correlation with the deviations in initial conditions, which may be due to the MLE being stuck in a local minimum. However, if necessary, the MLE algorithm can be further optimized by changing the step size and error limit for stopping the calculation, which may further decrease the errors.

VIII. Conclusions

Maximum likelihood estimation and Weibull distribution were used to distinguish primary and secondary wearout mechanisms for 14nm FinFET ring oscillators through quasi-Newton algorithms. The error in estimation of β values for both mechanisms are uncorrelated to the error in initial condition values, and the error in estimation of η values for both mechanisms were found to be directly correlated to the error in initial condition values. For a sample size of 10, the wearout mechanisms were able to be distinguished up to a deviation error of $\pm$5% from the actual value for the initial condition, while the sample sizes of 100 and 1000 could be separated for up to a deviation error of $\pm$15% from the actual value for the initial condition. This is the first known procedure for separating wearout mechanisms in circuit level data using machine learning applications, which provides a quick and non-invasive method to perform failure analysis at low costs.

Acknowledgments

The authors would like to acknowledge Dr. Yi-da Wu and Li-Hsiang Lin for their discussions. The authors would like to thank the NSF for support under Award Number 1700914.

References

[1] H. Choi, S. Heo, H. Hong, S. Yang, Y. Han, *et al.*, "High resolution short defect localization in advanced FinFET device using EBAC and EBIRCh," in *IEEE International Symposium on the Physical and Failure Analysis of Integrated Circuits (IPFA)*, 2017, pp. 1-4.

[2] Y. Greenzweig, Y. Drezner, A. Raveh, O. Sidorov, and R. H. Livengood, "E-beam invasiveness on 65 nm complementary metal-oxide semiconductor circuitry," *Journal of Vacuum Science & Technology B,* vol. 29, p. 021202, 2011.

[3] C. Chen and L. Milor, "Microprocessor Aging Analysis and Reliability Modeling Due to Back-End Wearout Mechanisms," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems,* vol. 23, pp. 2065-2076, Oct. 2015.

[4] T. Liu, C. Chen, and L. Milor, "Comprehensive Reliability-Aware Statistical Timing Analysis Using a Unified Gate-Delay Model for Microprocessors," *IEEE Transactions on Emerging Topics in Computing,* vol. 6, no. 2, pp. 219-232, 2018.

[5] K. Yang, T. Liu, R. Zhang, and L. Milor, "A Comprehensive Time-Dependent Dielectric Breakdown Lifetime Simulator for Both Traditional CMOS and FinFET Technology," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems,* vol. 26, no.11, pp. 2470-2482, 2018.

[6] T. Ishioka and Y. Nonaka, "Maximum likelihood estimation of Weibull parameters for two independent competing risk," *IEEE Transactions on Reliability,* vol. 40, pp. 71-74, Apr 1991.

[7] S. Angra and S. Ahuja, "Machine learning and its applications: A review," in *International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)*, 2017, pp. 57-60.

[8] L. Bottou, F. E. Curtis, and J. Nocedal, "Optimization Methods for Large-Scale Machine Learning," *SIAM Review,* vol. 60, pp. 223-311, 2018.

[9] P. Domingos, "A few useful things to know about machine learning," *Communications of the ACM,* vol. 55, pp. 78-87, 2012.

[10] Q. V. Le, J. Ngiam, A. Coates, A. Lahiri, B. Prochnow, *et. al*., "On optimization methods for deep learning," *Proceedings of the International Conference on Machine Learning*, 2011, pp. 265-272.

[11] P. Hennig and M. Kiefel, "Quasi-Newton Methods: A New Direction," *Journal of Machine Learning Research,* vol. 14, pp.843-865, 2013.

[12] R. Schoenberg, "Constrained Maximum Likelihood," *Computational Economics,* vol. 10, pp. 251-266, 1997.

[13] X. Li, J. Qin, and J. B. Bernstein, "Compact Modeling of MOSFET Wearout Mechanisms for Circuit-Reliability Simulation," *IEEE Transactions on Device and Materials Reliability,* vol. 8, pp. 98-121, 2008.

[14] K. Yang, T. Liu, R. Zhang, and L. Milor, "Circuit-level reliability simulator for front-end-of-line and middle-of-line time-dependent dielectric breakdown in FinFET technology," in *IEEE VLSI Test Symposium (VTS)*, 2018, pp 2470-2482.

[15] E. Wu, J. Suñé, W. Lai, E. Nowak, J. McKenna, A. Vayshenker*, et al.*, "Interplay of voltage and temperature acceleration of oxide breakdown for ultra-thin gate oxides," *Solid-State Electronics,* vol. 46, pp. 1787-1798, Nov. 1, 2002.

[16] F. Chen, O. Bravo, K. Chanda, P. McLaughlin, T. Sullivan*, et al.*, "A Comprehensive Study of Low-k SiCOH TDDB Phenomena and Its Reliability Lifetime Model Development," in *IEEE International Reliability Physics Symposium (IRPS)*, 2006, pp. 46-53.

[17] C. C. Chen, M. Bashir, L. Milor, D. H. Kim, and S. K. Lim, "Backend dielectric chip reliability simulator for complex interconnect geometries," in *IEEE International Reliability Physics Symposium (IRPS)*, 2012, pp. BD.4.1-BD.4.8.

[18] G. S. Haase and J. W. McPherson, "Modeling of Interconnect Dielectric Lifetime Under Stress Conditions and New Extrapolation Methodologies for Time-Dependent Dielectric Breakdown," in *IEEE International Reliability Physics Symposium (IRPS)*, 2007, pp. 390-398.

[19] K. Yiang, H. W. Yao, and A. Marathe, "TDDB Kinetics and their Relationship with the E- and √E-models," in *International Interconnect Technology Conference (IITC)*, 2008, pp. 168-170.

[20] T. Kauerauf, A. Branka, G. Sorrentino, P. Roussel, S. Demuynck*, et al.*, "Reliability of MOL local interconnects," in *IEEE International Reliability Physics Symposium (IRPS)*, 2013, pp. 2F.5.1-2F.5.5.

[21] W. Ahn, H. Zhang, T. Shen, C. Christiansen, P. Justison, *et al.*, "A Predictive Model for IC Self-Heating Based on Effective Medium and Image Charge Theories and Its Implications for Interconnect and Transistor Reliability," *IEEE Transactions on Electron Devices,* vol. 64, pp. 3555-3562, 2017.

[22] C. Hu, L. Gignac, B. Baker, E. Liniger, R. Yu*, et al.*, "Impact of Cu microstructure on electromigration reliability," in *IEEE International Interconnect Technology Conference (IITC)*, 2007, pp. 93-95.

[23] C.-C. Chen and L. Milor, "System-level modeling and reliability analysis of microprocessor systems," in *IEEE International Workshop on Advances in Sensors and Interfaces (IWASI)*, 2013, pp. 178-183.

[24] H. W. Yao, K.-Y. Yiang, P. Justison, M. Rayasam, O. Aubel*, et al.*, "Stress migration model for Cu interconnect reliability analysis," *Journal of Applied Physics,* vol. 110, p. 073504, 2011.

[25] D. B. Kececioglu and W. Wang, "Parameter estimation for mixed-Weibull distribution," in *Annual Reliability and Maintainability Symposium (RAMS),* 1998, pp. 247-252.