2007

Statistical Models for Hot Electron Degradation in Nano-Scaled MOSFET Devices

Suk Joo Bae

Seong-Joon Kim

Way Kuo

Paul H. Kvam
University of Richmond, pkvam@richmond.edu

Statistical Models for Hot Electron Degradation in Nano-Scaled MOSFET Devices

Suk Joo Bae **Seong-Joon Kim**

Department of Industrial Engineering, Hanyang University, Seoul, Korea

Way Kuo

Department of Electrical & Computer Engineering, The University of Tennessee, Knoxville

Paul H. Kvam

School of Industrial & Systems Engineering, Georgia Institute of Technology

Abstract

In a MOS structure, the generation of hot carrier interface states is a critical feature of the device's reliability. On the nano-scale, there are problems with degradation in transconductance, shift in threshold voltage, and decrease in drain current capability. Quantum mechanics has been used to relate this decrease to degradation and device failure. Although the lifetime and degradation of a device are typically used to characterize its reliability, in this paper we model the distribution of hot-electron activation energies, which has appeal because it exhibits two-point discrete mixture of logistic distributions. The logistic mixture presents computational problems that are addressed in simulation.

Index Terms– EM Algorithm, Logistic Distribution, Maximum Likelihood, Mixture Distribution, Nanotechnology, Reliability.

ACRONYMS

cdf	cumulative distribution function
HCI	hot carrier injection
IC	integrated circuit
MLE	maximum likelihood estimator
$\mathcal{MN}$	multivariate normal
MOS	metal-oxide-semiconductor
MOSFET	metal-oxide-semiconductor field effect transistor
pdf	probability density function

NOTATION

C_i	constant of hot carrier induced degradation model for $i = 1, 2, 3$
$D_{IT}(t_0)$	original interface trap density
$\Delta D_{IT}(t)$	hot carrier activated trap density at time t
$\mathcal{D}(t)$	degradation of a MOSFET device at time t
D_{IT}	interface traps density
E_m	electrical field
$F(\cdot)$	distribution of the hot-electron activation energies
$I(\cdot)$	Fisher information matrix
I_D	drain current
I_{Sub}	substrate current
k	reaction constant
$l(\cdot)$	log-likelihood function
n_b	concentration of Si-H bonds at the interface
n_0	initial concentration of Si-H bonds at the interface
N_T	total concentration of Si bonds
$N_{IT}(t_0)$	initial concentration of interface traps for $t_0 = 0$

$N_{IT}(t)$	concentration of interface traps at time t
$\Delta N_{IT}(t)$	generated interface traps
p	probability of higher activation energies; $p \in (0, 1)$
τ, τ_1, τ_2	lifetime constants
Si^*	Si dangling bond
W	channel width of a device
V_{DD}	power supply voltage
β	coefficient for the I_{Sub}-V_{DD} relationship
μ, σ	parameters of a logistic distribution
μ_i, σ_i	parameters of a mixture logistic distribution for $i = 1, 2$
$\Theta(\Theta_n)$	parameter space (with samples size n)
φ_{IT}	critical energy in electronvolts (eV) for generating an interface trap
φ_0	minimum energy (eV) that an electron must possess to create impact ionization
$\bar{\varphi}_{IT}$, $\bar{\varphi}_{IT,1}$, $\bar{\varphi}_{IT,2}$	mean defect energies
q	elementary charge with the value $1.60218 \times 10^{-19} C$
λ	hot-electron mean-free-path

1 Introduction

The study of reliability has played a vital role in the engineering of products, both large scale and micro scale. In the next decade, it will play an even bigger role for industries in *nanofabrication*, which amounts to designing and manufacturing devices on the nanometer scale; a nanometer (1 nm = one billionth of a meter) is approximately the length of a row of ten hydrogen atoms.

Actually, standard reliability analysis is already essential for the efficient manufacture of nano-devices, but the field of nanotechnology is virtually devoid of results that address reliability issues that are unique to this scale of product. In fact, just as basic physics principles must be rethought at the quantum level, current reliability theories and methods are only partially applicable to systems operating on a nanometer scale. On the molecular level, familiar material properties like conductivity no longer obey laws based on macro scale materials (e.g., Ohm's law). In the same sense, the essential metrics of reliability analysis - material degradation, fatigue, and basic failure mechanisms assume new meaning on the nanometer scale. Sennhauser [25] noted that traditional reliability models may be insufficient due to quantum effects, thermal processes and defect diffusion processes. Experimenters need to consider additional sources of variation such as thermal fluctuations, quantum statistics and Heisenburg uncertainty [3].

There is great potential for reliability improvement if only because current nano-devices are riddled with defects that cause frequent failure problems; the devices are easily damaged by defects that are otherwise harmless to larger micro-devices. A full understanding of the physics and statistics of the defect generation is required in order to investigate the ultimate reliability limitations for nano-devices.

In a MOS structure, for example, the generation of hot carrier interface state is a critical feature of the device's lifetime measurement. Gate current of MOSFETs is made up of electrons injected into the gate oxide by quasi-elastic scattering [13]. However, electrons with high kinetic energies (called "hot carriers") can generate electron-hole pairs near the drain due to impact ionization from atomic-levelled collisions. Those carriers may be injected into the gate oxide and trapped

on defect sites in the oxide. It results in creation of interface states at Si-SiO_2 interface which leads to degradation in transconductance, shift in threshold voltage, and decrease in drain current capability [15]. Understanding the physical mechanisms of HCI will provide meaningful clues for backtracking from observed macro-defects to inferred nano-defects scattered inside the MOSFET devices. This is analogous to reliability problems in which system failure data are used to infer properties about the system's components. In this paper, we investigate physical models of the defects (hot carriers) generation leading to failure based on statistical properties for MOSFETs.

The main results are contained in Section 2. A model for hot electron degradation is achieved via the mixture distribution of hot-electron activation energies. Procedures for statistical inference are outlined in Section 3, and Section 4 contains a discussion of its computation.

2 Physical Models for Hot Carrier Interface State Generation

In a MOS structure, a thin layer of silicon dioxide (SiO_2) forms the insulating layer between the control gate and the conducting channel of transistors used in modern ICs (see Figure 1). As circuits have been made denser to meet the increasing demand for faster logic and memory devices, the dimensions of the transistors have been reduced ("scaled") correspondingly. For example, SiO_2 layer thickness has decreased to 2.0 nm or less, but technology cannot shrink these dimensions indefinitely because thinning down the oxide thickness raises severe technological problems: dielectric thickness variation, penetration of impurities from the highly doped polysilicon gate, reliability and lifetime problems for devices made with the ultrathin oxides, etc [23].

In particular, hot carrier induced degradation in SiO_2 films is perceived as a main potential obstacle for the continued down-scaling of MOSFET devices. During device operation, the film is subjected to electrical stress, and electronic defects like hot carriers that limit device lifetimes are more likely to be created for short-channel devices. Generally, silicon-based transistors are

annealed[1] in a hydrogen-rich environment in order to passivate[2] defects at the Si-SiO_2 interface. However, hydrogen (H) is known to play a key role in the HCI degradation of the transistors with smaller geometries. Lyding et al. [15] proposed to replace hydrogen with deuterium during the final wafer sintering process[3] in order to reduce susceptibility to hot electron degradation effects. The details as to how hydrogen degrades a MOSFET device will be illustrated at the following.

2.1 Mechanisms of Hydrogen Release from the Si-SiO_2 Interface

A principal mechanism of MOSFET degradation is the creation of an interface state (or traps) at the Si-SiO_2 interface. The creation is mainly caused by desorption of hydrogens from the passivated dangling bonds at the mismatched Si-SiO_2 interface. This depassivation is activated directly by the hot electrons that exist during transistor switching. The hot electrons near the drain (see Figure 1) in short-channel devices can generate electron-hole pairs via impact ionization[4] [9]. Figure 2 describes the depassivation procedure at Si-SiO_2 interface. First, electrons (e) or holes (h) with high kinetic energies are attracted to the Si-SiO_2 interface, which weaken the Si-H bond until it breaks. As a result, the hydrogen diffuses into the oxide or Si substrate, subsequently creating interface traps with density D_{IT}. The hydrogen (H: atom and H^+: ion) release reaction breaking Si-H bonds is described in the following equation:

$$Si - H + H^+ + e^- \rightarrow Si^* + H_2 \qquad \text{or} \qquad Si - H + H \rightarrow Si^* + H_2, \tag{1}$$

where Si^* represents the Si dangling bond that is an "interface trap".

The hot-carrier-induced trap density D_{IT} is directly proportional to the concentration of Si dangling bonds at the interface. Because the amount of degradation of a MOSFET switching current that leads to device failure is a function of the variation in the interface trap density activated by the hot carriers, the amount of degradation of a MOSFET device can be represented

[1]The annealing is a process that the transistors are heated at sufficiently high temperatures and slowly cooled down

[2]To treat a subject in order to reduce the chemical reactivity

[3]A process of forming a coherent mass by heating without melting

[4]The formation of or separation into ions by heat, electrical discharge, radiation, or chemical reaction

as the concentration of Si dangling bonds at the interface, which can be measured by D_{IT} via charge pumping technique [8].

Electrons need sufficient activation energies to surmount a surface energy barrier to generate interface traps, and the activation energies are directly linked to interface trap defects. Defects do not necessarily have the same activation energy; in fact, there exists a distribution of activation energies for the HCI generation failure mechanism [17]. From a chemistry viewpoint, there are two reasons to expect a distribution of activation energies for MOSFET devices: the variation in the bond energies due to Si-SiO_2 interface disorder, and the possibility of multiple pathways to activation. If the bonding energies are homogeneously distributed at the interface, the activation energy distribution will be of unimodal form, but the distribution will be bimodal if there exist competing mechanisms of interface state formation following multiple pathways to activation [9].

2.2 Activation Energy Distribution of Hot Carrier Induced Defects

Because the activation of hydrogen at the passivated Si-SiO_2 interface is caused by collisions with electrons (or holes) flowing in the channel, it is crucial to identify the energy distribution of these electrons as a function of the number of interface traps over time to evaluate reliability of MOSFET devices.

N_{IT} is proportional to the concentration of the Si dangling bonds at the interface $N_T - n_b$, where N_T is a total concentration of Si bonds which are able to appear as dangling ones if hydrogen leaves the bond. The time dependent trap generation can be described by a simple version of power law [19]:

$$\Delta N_{IT}(t) = N_{IT}(t) - N_{IT}(t_0) = \frac{n_0}{1 + (kt)^{-\alpha}}. \tag{2}$$

The reaction constant k and the power α are values which can be estimated from the experimental data. In terms of the concentration of interface traps, the degradation of a MOSFET device can be approximated by

$$\mathcal{D}(t) = \frac{\Delta D_{IT}(t)}{D_{IT}(t_0)} \simeq \frac{\Delta N_{IT}(t)}{N_{IT}(t_0)} = \frac{1}{1 + (t/\tau)^{-\alpha}}, \tag{3}$$

where $\tau = 1/k$ is a lifetime constant that has units of time. Note that the model (3) is identical to the degradation model for grating decays in optical interconnects derived from Bragg grating theory in Erdogan et al [6]. For the hot carrier degradation mechanism, lifetime determination is based on the observed accelerated degradation of drain voltage. This is because hot carrier degradation is not accelerated by an increase in temperature [10]. The degradation of a device in terms of $\Delta N_{IT}(t)$ can be related to at I_D [13] as

$$\Delta N_{IT}(t) = C_1 \left[t \frac{I_D}{W} \exp\left(-\frac{\varphi_{IT}}{q\lambda E_m} \right) \right]^n . \tag{4}$$

Introducing an easily measurable I_{Sub} to monitor the device degradation, E_m can be represented with the multiplication factor

$$\frac{I_{Sub}}{I_D} = C_2 \exp\left(-\frac{\varphi_0}{q\lambda E_m} \right). \tag{5}$$

A lifetime is defined as the time to reach a fixed number of interface traps. By combining (4) and (5),

$$\frac{tI_D}{W} \propto \left[\frac{I_{Sub}}{I_D} \right]^{-\varphi_{IT}/\varphi_0} \qquad \text{or} \qquad t \propto I_{Sub}^{-\varphi_{IT}/\varphi_0}. \tag{6}$$

The substrate current is a function of the power supply voltage as

$$I_{Sub} \propto \exp\left(-\frac{\beta}{V_{DD}} \right). \tag{7}$$

Combining (6) and (7), the lifetime of the device can be written as

$$t = C_3 \exp\left(\frac{\varphi_{IT}\beta}{\varphi_0 V_{DD}} \right), \tag{8}$$

and replacing $C_3 \exp(\bar{\varphi}_{IT}\beta/(\varphi_0 V_{DD}))$ by experimentally observed τ in (3), finally the degradation of a MOSFET device can be approximated by using the following distribution on the hot-electron activation energies:

$$F(\varphi_{IT}) = \left[1 + \exp\left(-\frac{\varphi_{IT} - \bar{\varphi}_{IT}}{\sigma} \right) \right]^{-1}, \tag{9}$$

where $\sigma = (\varphi_0 V_{DD})/(\alpha\beta)$. Note that the degradation model $\mathcal{D}(t)$ in (3) that is represented as the proportion of activated defects before time t is equivalent to the probability that activated defects have activation energy less than or equal to φ_{IT}.

Although the lifetime and degradation of a device are typically used to characterize its reliability, in this case the distribution of hot-electron activation energies has a unique appeal because it has a common logistic structure. The logistic distribution, derived from disorder-induced variations in the *Si-H* activation energies, is identical to a Fermi-derivative distribution of the energies of electronic states [5]. Figure 4-(a) shows how the degradation model caused by interface traps (or defects) varies as a function of power supply voltage(V_{DD}) of a MOSFET device for fixed values of β and φ_0, along with the distributions of defect activation energies in 4-(b). The parameter values in the figure are simulated from experimental observations of short-time-tests for 180 nm MOSFET devices in Haggag et al. [9].

2.3 Bimodal Distribution of Activation Energies

Existence of multiple paths and competing mechanisms for the release of hydrogen yields inhomogeneous activation energy distributions. Figure 3 shows the energy level of hydrogen release to different activation pathways (the hydrogen may be attracted to Si or SiO_2 or the Si-SiO_2 interface). Through atomic simulations based on density functional theory, Tuttle et al. [28] showed that the activation energy of hydrogen is distributed around 3.5 eV if the hydrogen desorbs into the SiO_2, but below 3 eV if the final hydrogen state is closer to the silicon bulk. As a result, the time-dependent HCI degradation model is a *mixture* of the model (3) [11]:

$$\mathcal{D}(t) = \frac{\Delta D_{IT}(t)}{D_{IT}(t_0)} = \frac{p}{1 + (t/\tau_1)^{-\alpha_1}} + \frac{1-p}{1 + (t/\tau_2)^{-\alpha_2}}. \tag{10}$$

By letting $\tau_1 = C_{3,1} \exp(\bar{\varphi}_{IT,1}\beta_1/(\varphi_0 V_{DD}))$ and $\tau_2 = C_{3,2} \exp(\bar{\varphi}_{IT,2}\beta_2/(\varphi_0 V_{DD}))$ in (10), the degradation model of a MOSFET device can be represented through φ_{IT} as a mixture of logistic distributions:

$$F(\varphi_{IT}) = p \cdot \left[1 + \exp\left(-\frac{\varphi_{IT} - \bar{\varphi}_{IT,1}}{\sigma_1}\right)\right]^{-1} + (1-p) \cdot \left[1 + \exp\left(-\frac{\varphi_{IT} - \bar{\varphi}_{IT,2}}{\sigma_2}\right)\right]^{-1}, \tag{11}$$

where $\sigma_1 = (\varphi_0 V_{DD})/(\alpha_1\beta_1)$ and $\sigma_2 = (\varphi_0 V_{DD})/(\alpha_2\beta_2)$.

Tuttle et al. [28] experimentally observed a higher mean energy $\bar{\varphi}_{IT,1} \approx 3.5$ eV as well as a lower mean energy $\bar{\varphi}_{IT,2} \approx 2.9$ eV. The higher energy band comes from "single collisions" with

higher energetic electrons and a consequent release of the hydrogen through a higher energy path in the MOSFET. On the other hand, the lower energy band comes from "multiple collisions" with lower energetic electrons and a consequent release of the hydrogen through a lower energy path in the MOSFET [9]. Figures 5-(a) and 5-(b) display the mixture of time-dependent HCI degradation model (10) and the mixture distribution of defect activation energies, respectively, at varying p values with $\bar{\varphi}_{IT,1} = 3.5$ eV, $\bar{\varphi}_{IT,2} = 2.9$ eV, and $V_{DD} = 3.0V$.

3 Parameter Estimation

In this section, we outline the procedures for statistical inference for different characteristics of the MOSFET lifetime. Using the measurement of hot-electron activation energies, we rely on the method of maximum likelihood to estimate logistic model parameters, or more precisely, parameters for the logistic mixture distribution. While the inference for the logistic distribution is straightforward, there are important issues in dealing with estimation for the mixture distribution.

3.1 Logistic Mixture Distribution

The cdf of the random variable X having the logistic distribution is given by

$$F(x;\boldsymbol{\theta}) = \frac{1}{\left\{1+\exp\left(-\frac{x-\mu}{\sigma}\right)\right\}}, \qquad -\infty < x < \infty, \tag{12}$$

for $\boldsymbol{\theta} = (\mu, \sigma)^T$, where μ and σ are location and scale parameters. The corresponding pdf is

$$f(x;\boldsymbol{\theta}) = \frac{1}{\sigma}\frac{\exp\left(-\frac{x-\mu}{\sigma}\right)}{\left\{1+\exp\left(-\frac{x-\mu}{\sigma}\right)\right\}^2} = \frac{1}{\sigma}\left\{F(x)[1-F(x)]\right\}. \tag{13}$$

The pdf of the logistic distribution is symmetric and bell-shaped like that of the normal distribution. Since the logistic distribution has slightly longer tails, it would require an extremely large number of observations to accurately assess whether data come from a normal or logistic distribution. The logistic random variable X has mean $E[X] = \mu$, variance $\mathrm{Var}(X) = (\pi^2\sigma^2)/3$ and coefficient of variation $\pi\sigma/(\mu\sqrt{3})$. The log-likelihood function for a sample of size n from the logistic distribution

is given by

$$l(x_1,\ldots,x_n;\boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(x_i;\boldsymbol{\theta}) = -n\log\sigma - \sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right) - 2\sum_{i=1}^{n}\log\left\{1+\exp\left(-\frac{x_i-\mu}{\sigma}\right)\right\},$$

and the MLEs, $(\hat{\mu},\hat{\sigma})$ of the parameters (μ,σ) satisfy the following likelihood equations:

$$\begin{aligned} \sum_{i=1}^{n}\left\{1+\exp\left(-\frac{x_i-\hat{\mu}}{\hat{\sigma}}\right)\right\}^{-1} &= \frac{n}{2}, \\ -\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\hat{\mu}}{\hat{\sigma}}\right) + \sum_{i=1}^{n}\left(\frac{x_i-\hat{\mu}}{\hat{\sigma}}\right)\left\{1+\exp\left(-\frac{x_i-\hat{\mu}}{\hat{\sigma}}\right)\right\}^{-1} &= \frac{n}{2}. \end{aligned} \tag{14}$$

Taking advantage of the similarity in shape between the logistic and normal distributions, initial values of $\hat{\mu}$ and $\hat{\sigma}$ might be taken as $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$, and $\sqrt{n^{-1}\sum_{i=1}^{n}(X_i-\bar{X})^2}$, respectively. Then solutions could be improved by applying the Newton-Raphson method. When both μ and σ are unknown, the Newton-Raphson method converges quickly to the solutions $\hat{\boldsymbol{\theta}} = (\hat{\mu},\hat{\sigma})^T$. Since the logistic-likelihood function is quasi-concave, the solutions are unique for distinct values of x_i. [1].

The MLEs $\hat{\boldsymbol{\theta}} = (\hat{\mu},\hat{\sigma})^T$, as consistent roots of the likelihood equations (14), satisfy

$$\left(\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\right) \xrightarrow{L} \mathcal{MN}(\mathbf{0}, I^{-1}(\boldsymbol{\theta}_0)), \tag{15}$$

where $\boldsymbol{\theta}_0$ is the true value of $\boldsymbol{\theta}$ and the Fisher information $I(\boldsymbol{\theta}_0)$ is given by

$$I(\boldsymbol{\theta}_0) = -E\begin{bmatrix} \frac{\partial^2 l}{\partial^2\mu^2} & \frac{\partial^2 l}{\partial\mu\partial\sigma} \\ \frac{\partial^2 l}{\partial\mu\partial\sigma} & \frac{\partial^2 l}{\partial^2\sigma^2} \end{bmatrix} = \begin{bmatrix} \frac{1}{3\sigma^2} & 0 \\ 0 & \frac{3+\pi^2}{9\sigma^2} \end{bmatrix}. \tag{16}$$

It is common in practice to estimate the inverse of the covariance matrix of the MLE by the observed information matrix $I(\hat{\boldsymbol{\theta}})$ rather than the expected information matrix $I(\boldsymbol{\theta}_0)$ evaluated at $\boldsymbol{\theta}_0 = \hat{\boldsymbol{\theta}}$. In general, the observed information matrix is more convenient to use than the expected information matrix as it does not require an expectation to be taken. However as shown in (16), the expectations are trivial in the logistic case and we can easily derive the covariance matrix of the MLE from the expected information matrix. When only μ is unknown, a MLE $\hat{\mu}$ can be uniquely

determined by replacing $\hat{\sigma}$ with known σ value in the likelihood equation. Alternatively, we can use an estimator

$$\zeta_n = \bar{X}_n - \frac{l'(\bar{X}_n)}{l''(\bar{X}_n)}, \tag{17}$$

instead of $\hat{\mu}$ because they have the same asymptotic distribution. Here $\bar{X}_n$ is the average of n samples as a $\sqrt{n}$-consistent estimator (denoted as $\tilde{\mu}$) and

$$l'(\bar{X}_n) = \frac{\partial}{\partial \mu} l(\mathbf{x}; \boldsymbol{\theta})\big|_{\mu=\bar{X}_n} = n - 2\sum_{i=1}^{n} \frac{\exp\left(\frac{x_i - \bar{x}_n}{\sigma}\right)}{1 + \exp\left(\frac{x_i - \bar{x}_n}{\sigma}\right)},$$

$$l''(\bar{X}_n) = \frac{\partial^2}{\partial \mu^2} l(\mathbf{x}; \boldsymbol{\theta})\big|_{\mu=\bar{X}_n} = -2\sum_{i=1}^{n} \frac{\exp\left(\frac{x_i - \bar{x}_n}{\sigma}\right)}{1 + \exp\left(\frac{x_i - \bar{x}_n}{\sigma}\right)^2}.$$

Theorem 3.1 *Let $\hat{\mu}$ be the MLE of μ, and let ζ_n be given by (17), then $\sqrt{n}(\zeta_n - \hat{\mu}) \to 0$ as $n \to \infty$.*

The proof is listed in the appendix. When only μ is known, the MLE $\hat{\sigma}$ can be uniquely determined by replacing $\hat{\mu}$ with known μ value in the likelihood equations. With moment estimator $\tilde{\sigma} = \{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2/n\}^{1/2}$, which is a $\sqrt{n}$-consistent estimator of σ,

$$\eta_n = \tilde{\sigma} - \frac{l'(\tilde{\sigma})}{l''(\tilde{\sigma})} \tag{18}$$

has the same asymptotic distribution as the MLE $\hat{\sigma}$ (from the theorem) with $\tilde{\sigma}$ instead of $\bar{X}_n$. Here,

$$l'(\tilde{\sigma}) = \frac{\partial}{\partial \sigma} l(\mathbf{x}; \boldsymbol{\theta})\big|_{\sigma=\tilde{\sigma}} \qquad \text{and} \qquad l''(\tilde{\sigma}) = \frac{\partial^2}{\partial \sigma^2} l(\mathbf{x}; \boldsymbol{\theta})\big|_{\sigma=\tilde{\sigma}}.$$

The degradation model of a MOSFET device can represented in (11) is a two-point discrete mixture of logistic distributions. For a random variable X generated from this mixture of logistic distributions, then in terms of $\boldsymbol{\theta}_1 = (\mu_1, \sigma_1)^T$, $\boldsymbol{\theta}_2 = (\mu_2, \sigma_2)^T$ and $\boldsymbol{\Psi} = (p, (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T))^T$, X has pdf

$$f(x; \boldsymbol{\Psi}) = p \cdot f_1(x; \boldsymbol{\theta}_1) + (1 - p) \cdot f_2(x; \boldsymbol{\theta}_2), \tag{19}$$

and corresponding CDF

$$F(x; \boldsymbol{\Psi}) = p \cdot F_1(x; \boldsymbol{\theta}_1) + (1 - p) \cdot F_2(x; \boldsymbol{\theta}_2), \tag{20}$$

where $F_j(x; \boldsymbol{\theta}_j)$ and $f_j(x; \boldsymbol{\theta}_j)$ are from (12) and (13), respectively, with parameters μ_j and σ_j for $j = 1, 2$.

4 Solving the MLE

Several methods have been proposed to estimate the parameter $\mathbf{\Psi}$. The MLE, based on maximizing the log-likelihood function

$$l(x_1, \ldots, x_n; \mathbf{\Psi}) = \sum_{i=1}^{n} \log \left\{ p \cdot f_1(x_i; \boldsymbol{\theta}_1) + (1-p) \cdot f_2(x_i; \boldsymbol{\theta}_2) \right\},$$

possesses a number of desirable statistical properties. The MLE can be solved via the likelihood equation $\partial l(x_1, \ldots, x_n; \mathbf{\Psi})/\partial \mathbf{\Psi} = \mathbf{0}$. The resulting solution $\hat{\mathbf{\Psi}}$ satisfies

$$\hat{p} = \sum_{i=1}^{n} \kappa(x_i; \hat{\mathbf{\Psi}})/n \qquad \text{and} \qquad \sum_{j=1}^{2} \sum_{i=1}^{n} \kappa(x_i; \hat{\mathbf{\Psi}}) \partial \log f_j(x_i; \hat{\boldsymbol{\theta}}_j)/\partial \boldsymbol{\theta} = \mathbf{0}, \tag{21}$$

where $\kappa(x_i; \hat{\mathbf{\Psi}}) = \hat{p} f_1(x_i; \hat{\boldsymbol{\theta}}_1)/[\hat{p} f_1(x_i; \hat{\boldsymbol{\theta}}_1) + (1-\hat{p}) f_2(x_i; \hat{\boldsymbol{\theta}}_2)]$ is the posterior probability that x_i belongs to the first component of the mixture.

The EM algorithm (see McLachlan and Krishnan [16], for example) can be used to find the MLE for mixtures by solving the likelihood equations (21) iteratively. To apply the EM method, we imagine each observation from the mixture distribution comes with an indicator variable that tells us which of the two logistic distributions the observation was generated. In this case, such an indicator is treated as a missing value.

Starting from an arbitrary initial guess, the algorithm operates in two repeated steps. The E-step estimates missing values as they appear in the log-likelihood, then the M-step finds a local optimum to the likelihood using the estimated data in place of what was missing. With the estimated indicator functions, the MLE is solved more simply using (14) in the M-step (as discussed in the last section). Convergence properties for finite mixture models are discussed in Tanaka and Takemura [22]. In this case, because of the heavy tails in the logistic density and the large variances in (16), the convergence can be slow, depending on how good the initial guess is.

In Table 1, we present summary results of the simulation comparing the performance of different MLE/EM algorithms for mixtures of logistic components. Comparisons between varying degrees of separation in the mixture distributions (complete separation, moderate overlap, and large overlap) are illustrated in Figures 6 (a) - (c), respectively. In these simulations, the mixing proportion p

takes on the values .20, .50, and .80. For a given mixture, the component distributions differ from each other only by location and scale differences. For each set of parameter configurations, samples of size $n = 1,000$ were generated from the corresponding mixture of logistic distributions.

Data generation and parameter estimation for logistic mixture were executed using the *mle* program [12], and for maximizing the likelihood function in the EM algorithm, four different methods were used; simplex, direct, conjugate gradient, and simulated annealing (see [20] for details of the four methods).

Each of the methods has strengths and weakness for different types of functions. In the case of the logistic likelihood, they have peculiar differences; Table 1 shows a summary of the simulation comparing the performance of MLE/EM algorithms for mixtures of logistic components. The computation stopped when the relative errors ϵ, of all 5 parameters, $\mathbf{\Psi} \equiv (p, \mu_1, \sigma_1, \mu_2, \sigma_2)'$ reached $10^{-4} \cdot \epsilon \equiv [\mathbf{\Psi}^{(h+1)} - \mathbf{\Psi}^{(h)}]/\mathbf{\Psi}^{(h)}$. The symbol $(*)$ represents result tends to a pathological solution and the relative bias, $|y - \hat{y}|/y$ is calculated over every parameter in parentheses. The direct and simulated annealing methods provide better results than the other two methods for this parameter estimation. However, simulated annealing requires longer process time for the annealing function. In general, the estimation precision increases when the mixing proportion is neutral ($p = 0.5$).

One way of obtaining standard errors of the estimates of the parameters in a mixture model is to approximate the covariance matrix of $\hat{\mathbf{\Psi}}$ by the inverse of the observed information matrix. For mixture models, the sample size has to be very large to guarantee the asymptotic theory of maximum likelihood, hence a resampling approach such as bootstrapping method can be considered to construct standard errors of the estimates of the parameters. Standard error estimation of $\hat{\mathbf{\Psi}}$ can implemented via a bootstrap procedure; Chapter 13 of [18] for an analogous resampling approach.

5 Discussion

Statistical models based on known physical principles (e.g., the power-rule model, the Arrhenius rule, Eyring Model, etc.) have provided strong methods for parametric inference for various testing

problems in manufacturing. Nano-manufacturing will provide more twists to these traditional models due to the nature of nano-defects and Heisenberg uncertainty. This paper provides basic physical modeling for MOSFET devices based on the nano-level degradation that takes place at defect sites in the MOSFET gate oxide. The distribution of hot-electron activation energies proves to be more accessible than analogous measures of degradation or lifetime, and is derived as a logistic mixture distribution using physical principles on the nanoscale. Although the inference problem is ridden with computational challenges, the derivation of MLEs is straightforward using the EM Algorithm.

6 Appendix

Proof of Theorem : For the true value of μ (denoted as μ_0), Taylor expansion of $l'(\hat{\mu})$ about $l'(\mu_0)$ is

$$l'(\hat{\mu}) = l'(\mu_0) + (\hat{\mu} - \mu_0)l''(\mu_0) + \frac{1}{2}(\hat{\mu} - \mu_0)^2 l'''(\mu^*),$$

where μ^* lies between μ_0 and $\hat{\mu}$. $l'''(\cdot)$ exists for the three-times differentiable logistic likelihood function with respect to μ, and it is bounded. Since the left side is zero for the MLE $\hat{\mu}$, we have

$$\sqrt{n}(\hat{\mu} - \mu_0) = -\frac{\sqrt{n}l'(\mu_0)}{l''(\mu_0)} + R_n,$$

where

$$R_n = -\frac{\sqrt{n}l'(\mu_0)}{l''(\mu_0)} \left[\frac{1}{1 + \frac{n}{l''(\mu_0)} \frac{1}{2n}(\hat{\mu} - \mu_0) l'''(\mu^*)} - 1 \right].$$

Analogously, Taylor expansion of $l'(\tilde{\mu})$ about $l'(\mu_0)$ can be represented as

$$l'(\tilde{\mu}) = l'(\mu_0) + (\tilde{\mu} - \mu_0)l''(\mu_0) + \frac{1}{2}(\tilde{\mu} - \mu_0)^2 l'''(\mu^*).$$

It follows from (15) that

$$\zeta_n = \tilde{\mu} - \frac{1}{l''(\tilde{\mu})} \left[l'(\mu_0) + (\tilde{\mu} - \mu_0)l''(\mu_0) + \frac{1}{2}(\tilde{\mu} - \mu_0)^2 l'''(\mu^*) \right],$$

and by using the expansion $l''(\tilde{\mu}) = l''(\mu_0) + (\tilde{\mu} - \mu_0)l'''(\mu^*)$,

$$\sqrt{n}(\zeta_n - \mu_0) = -\frac{\sqrt{n}l'(\mu_0)}{l''(\mu_0)} \left(1 + \frac{(\tilde{\mu} - \mu_0)l'''(\mu^*)}{l''(\mu_0)} \right)^{-1} + R'_n,$$

where

$$R'_n = \sqrt{n}(\hat{\mu} - \mu_0)\left[1 - \frac{l''(\mu_0)}{l''(\tilde{\mu})} - \frac{1}{2}(\tilde{\mu} - \mu_0)\frac{l'''(\mu^*)}{l''(\tilde{\mu})}\right].$$

As $n \to \infty$, R_n and R'_n tend to 0 in probability since $\hat{\mu} \to \mu_0$, $\frac{l''(\mu_0)}{l''(\tilde{\mu})} \to 1$, and $\tilde{\mu} \to \mu_0$, and additionally $\left(1 + \frac{(\tilde{\mu}-\mu_0)l'''(\mu^*)}{l''(\mu_0)}\right) \to 1$.

References

[1] Antle, C. E., Klimko, L., and Harkness, W. (1970), "Confidence Intervals for the Parameters of the Logistic Distribution", *Biometrika*, Vol. 57, No. 2, 397–402.

[2] Balakrishnan, N. (1992), *Handbook of the Logistic Distribution*, Dekker, New York.

[3] Birge, R. R., Lawrence, A. F., and Tallent, J. R. (1991), "Quantum effects, thermal statistics and reliability of nanoscale molecular and semiconductor devices", *Nanotechnology*, Vol. 2, 73–87.

[4] DeCani, J. S., and Stine, R. A. (1986), "A Note on Deriving the Information Matrix for a Logistic Distribution", *The American Statistician*, Vol. 40, No. 3, 220–222.

[5] Devine, R. A. B., Autran, J. L., Warren, W. L., VAnheusdan, K. L., and Rostaing, J. C. (1997), "Interfacial Hardness Enhancement in Deuterium Annealed 0.25 μm Channel Metal Oxide Semiconductor Transistors", *Applied Physics Letters*, Vol. 70, No. 22, 2999–3001.

[6] Erdogan, T., Mizrahi, V., Lemaire, P. J., and Monroe, D. (1994), "Decay of Ultraviolet-Induced Fiber Bragg Gratings", *Journal of Applied Physics*, Vol. 76, No. 1, 73–80.

[7] Feng, Z. D., and McCulloch, C. E. (1996), "Using Bootstrap Likelihood Ratios in Finite Mixture Models", *Journal of the Royal Statistical Society B*, Vol. 58, No. 3, 609–617.

[8] Groeseneken, G., Maes, H. E., Beltran, N., and DE Keersmaecker, R. F. (1984), "A Reliable Approach to Charge- Pumping Measurements in MOS Transistors", *IEEE Transactions on Electron Devices*, Vol. 31, No. 1, 42–53.

[9] Haggag, A., McMahon, W., Hess, K., Cheng, K., Lee, J., and Lyding, J. W., (2001), "High-Performance Chip Reliability from Short-Time-Tests: Statistical Models for Optical Interconnect and Deep-Submicron Transistor Failures",*International Reliability Physics Symposium*, Orlando, FL.

[10] Heremans, P., Van den Bosch, G., Bellens, R., Groeseneken, G., Maes, H. E. (1990), "Temperature Dependence of the Channel Hot-Carrier Degradtion of n-Channel MOSFETs", *IEEE Transactions on Electron Devices.*, Vol. 37, No. 4., 980–993.

[11] Hess, K., Haggag, A., McMahon, W., Cheng, K., Cheng, K., Lee, J., and Lyding, J. W. (2001), "The Physics of Determining Chip Reliability", *IEEE Circuits Devices Magazine*, Vol. 17, No. 3, 33–38.

[12] Holman, D. J. (2003), *A Programming Language for Buliding Likelihood Models*, User's manual, Vol. 1, Ver. 2.1.

[13] Hu, C., Tam, S. C., Hsu, F., Ko, P., Chan, T., and Terrill, K. W. (1985), "Hot-Electron-Induced MOSFET Degradation-Model, Monitor, and Improvement', *IEEE Journal of Solid-State Circuits*, Vol. 20, No. 1, 295–305.

[14] Lehmann, E. L., and Casella, G. (1998), *Theory of Point Estimation*, Springer, New York.

[15] Lyding, J. W., Hess, K., and Kizilyalli, I. C. (1996), "Reduction of Hot Electron Degradation in Metal Oxide Semiconductor Transistors by Deuterium Processing", *Applied Physics Letters*, Vol 68. No. 18, 2526–2528.

[16] McLachlan, G. J., and Krishnan, T. (1997), *The EM Algorithm and Extensions*, Wiley, New York.

[17] McMahon, W., Haggag, A., and Hess, K. (2003), "Reliability Scaling Issues for Nanoscale Devices", *IEEE Transactions on Nanotechnology*, Vol. 2, No. 1, 33–38.

[18] Meeker, W. Q. and Escobar, L. A. (1998), *Statistical Methods for Reliability Data*, Wiley, New York.

[19] Penzin, O., Haggag, A., McMahon, W., Lyumkis, E., and Hess, K. (2003), "MOSFET Degradation Kinetics and Its Simulation", *IEEE Transactions on Electron Devices*, Vol. 50, No. 6, 1445–1450.

[20] Press, W. H., Flannery B. P., Teukolsky S. A., and Vetterling W. T. (1989), *Numerical Recipes in Pascal: The Art of Scientific Programming*, Cambridge University Press, Cambridge.

[21] Quandt, R. E., and Ramsey, J. B. (1978), "Estimating Mixtures of Normal Distributions and Switching Regressions", *Journal of American Statistical Association*, Vol. 73, 730-738.

[22] Tanaka, K., and Takemura, A. (2003), "Strong Consistency of MLE for Finite Mixtures of Location-Scale Distributions When the Scale Parameters are Exponentially Small", *Mathematical Engineering Technical Report*, University of Tokyo.

[23] Waser, R., (2003), *Nanoelectronics and Information Technology: Advanced Electronic Materials and Novel Devices.*, Wiley, Weinheim.

[24] Sennhauser, U., Reiner, J., and Nellen, P. M. (2004), "Nanoreliability", *13th NID Workshop*, Athens, Greece.

[25] Sennhauser, U. (2004), "Reliability of Nanostructed Materials and Devices", Project No. 5202.1, Swiss Federal Laboratories for Materials Testing and Research.

[26] Shukla, S. K., Norman, G., Parker, D., and Kwiatkowska, M. (2005), "Evaluating the Reliability of Defect-Tolerant Architectures for Nanotechnology with Probabilistic Model Checking", Submitted for Publication in *IEEE Transactions on Computer Aided Design* (under 2nd revision).

[27] Stathis, J. H. and DiMaria, D. J. (1998), "Reliability Projection for Ultra-Thin Oxides at Low Voltage", *IEDS Technical Digest*, Vol. 6, 167–170.

[28] Tuttle, B., and Van d Walle, C. G. (1999), "Structure, Energetics, and Vibrational Properties of Si-H Bond Dissociation in Silicon", *Physical Review B-Condensed Matter and Materials Physics*, Vol. 39, 12884–12889.

[29] Vallett, D. P. (2002), "Failure Analysis Requirements for Nanoelectronics", *IEEE Transactions on Nanotechnology*, Vol. 1, No. 3, 117–121.

Authors

Suk Joo Bae is an Assistant Professor in the Department of Industrial Engineering at Hanyang University, Seoul, Korea. He received his Ph.D. from the School of Industrial and Systems Engineering at the Georgia Institute of Technology in 2003. He worked as a reliability engineer at the SDI, Korea, from 1996 to 1999. His research interests are centered on reliability evaluation of light displays and nano-devices via accelerated life and degradation testing, statistical robust parameter design, and process control for large-volume on-line processing data. He is a member of INFORMS, ASA, and IMS.

Seong-Joon Kim is currently a Graduate Student in the Department of Industrial Engineering at Hanyang University, Seoul, Korea. His research interests are centered on reliability evaluation of nano-devices, spatial modeling of point process.

Way Kuo is the University Distinguished Professor and Dean of Engineering at the University of Tennessee. He was the Wisenbaker Chair of Engineering and head of the industrial engineering department at Texas A&M. He is an elected member of the National Academy of Engineering. He received his B.S. in nuclear engineering from National Tsing Hua University in Taiwan and his Ph.D in industrial engineering from Kansas State University. He is a fellow of IIE, IEEE, and ASQ, and an Academician of International Academy for Quality. He has chaired the Council of Industrial Engineering Academic Department Heads and the IIE Council of Fellows.

Paul H. Kvam is a Professor in the School of Industrial and Systems Engineering at the Georgia Institute of Technology. He received his Ph.D. from the University of California, Davis in 1991. His research interests include reliability evaluation in engineering applications, including accelerated life testing, degradation testing, analysis of dependent systems, and nonparametric inference. He is a fellow of the ASA.

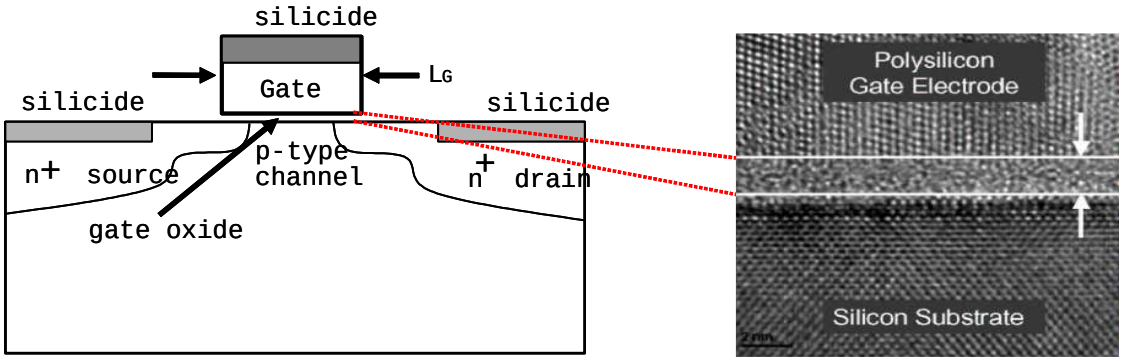


Figure 1: Basic structure of a planar MOSFET: For 90 nm generation gate oxide, the thickness of silicon oxide (SiO_2) is less than 2 nm.

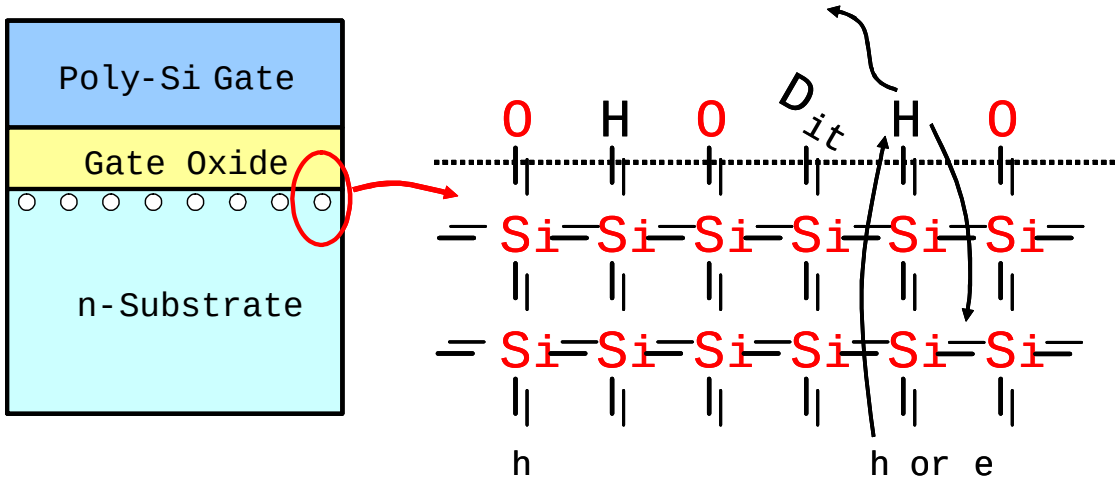


Figure 2: Desorption procedure of hydrogens at Si-SiO_2 interface.

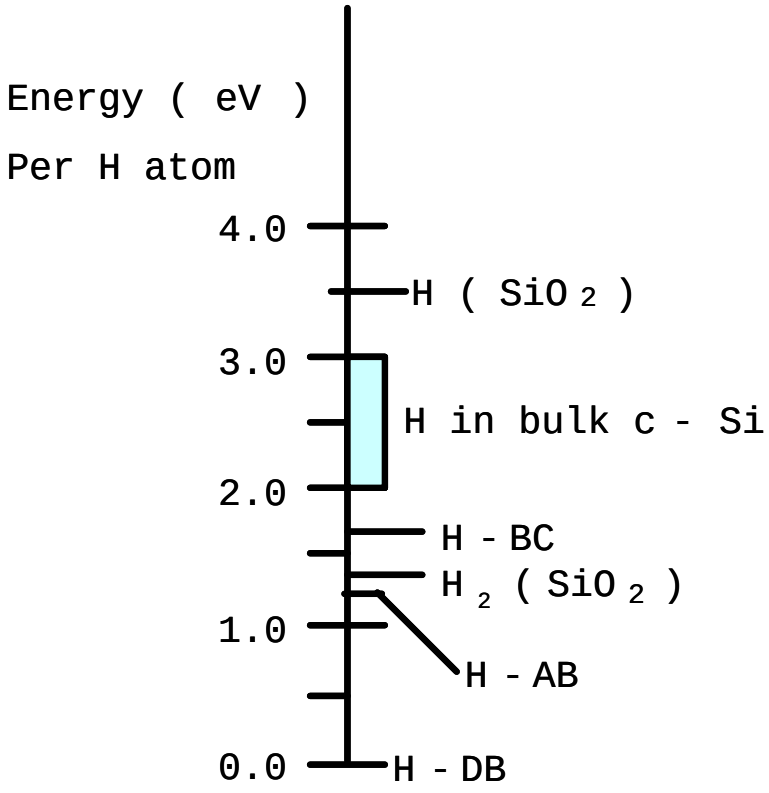
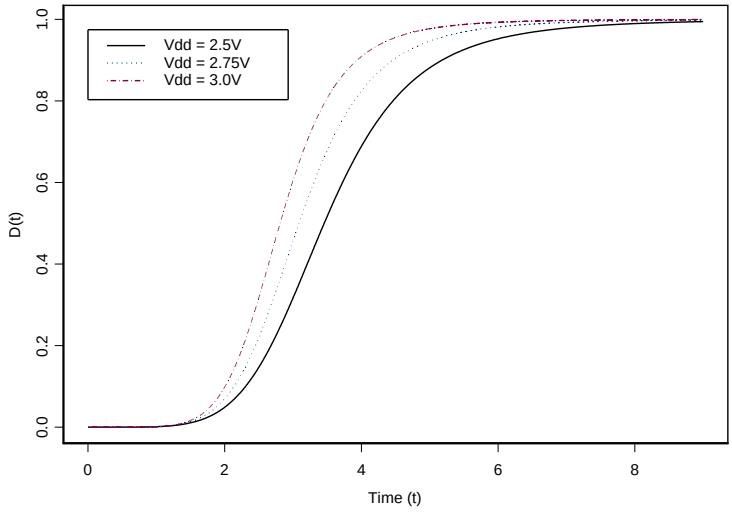
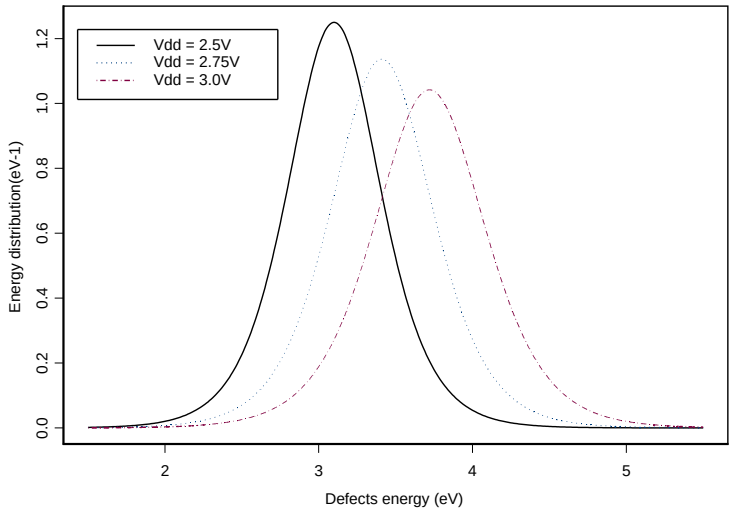


Figure 3: Energy to release hydrogen to different locations (courtesy of Tuttle et al [28]).

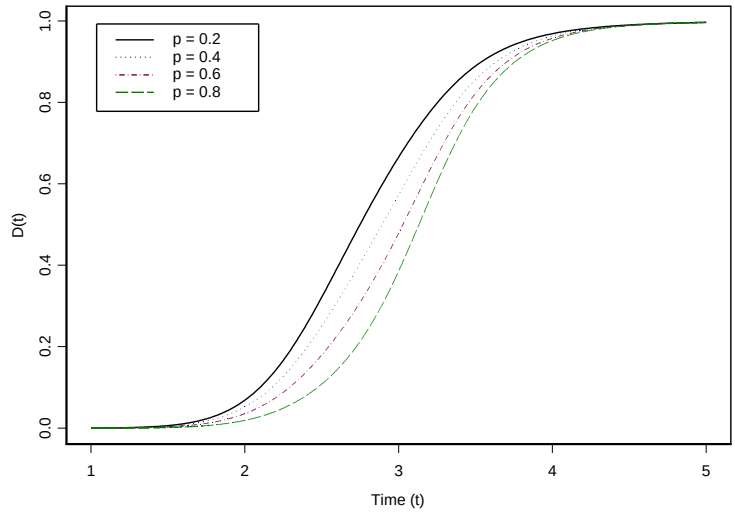


(a)

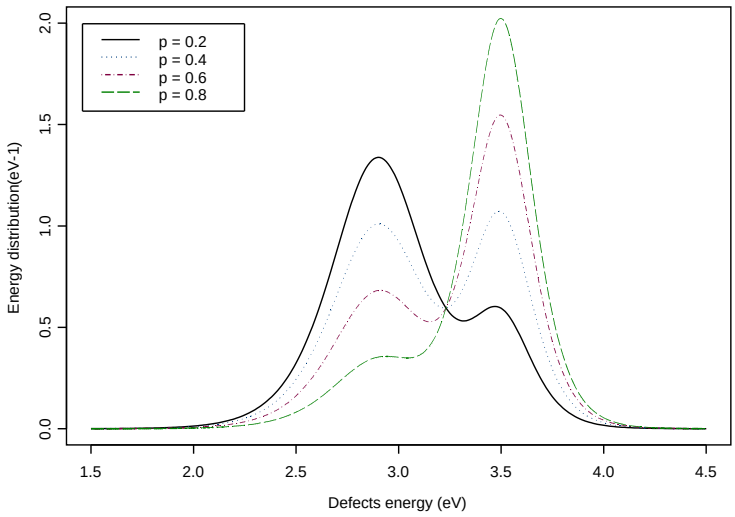


(b)

Figure 4: (a)-Defects degradation model (3); (b)-Distribution of activation energies for interface traps.

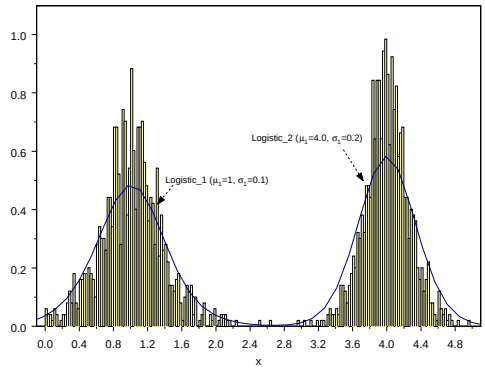


(a)

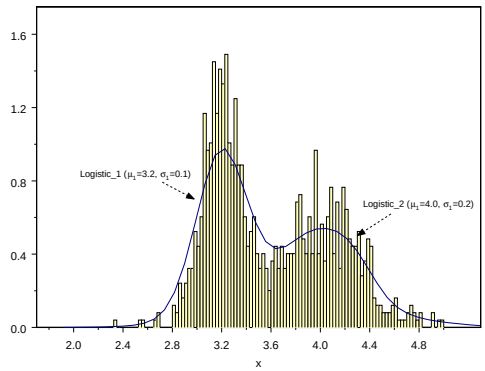


(b)

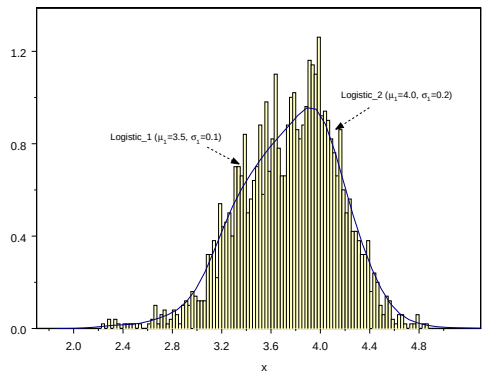
Figure 5: (a)-Defects degradation model (10) and (b)-Mixture distribution of activation energies for defects.



(a) Complete separation



(b) Moderate overlap



(c) Large overlap

Figure 6: Simulation of mixture of two logistic components with varying degrees of separation ($p = 0.5$).

Table 1: Simulation results for mixtures of logistic components: (∗) result tends to a pathological solution.

			Simplex	Direct	Conjugate Gradient	Simulated Annealing
		$\hat{p}$	0.2135 (0.0676)	0.2000 (0.0000)	0.0623 (0.6887)	0.2024 (0.0118)
		$\hat{\mu}_1$	0.9798 (0.0202)	0.9845 (0.0155)	3.4516 (2.4516)	0.9846 (0.0154)
	$p = 0.2$	$\hat{\sigma}_1$	0.0649 (0.3508)	0.0701 (0.2991)	(*)	0.0701 (0.2986)
		$\hat{\mu}_2$	4.0297 (0.0074)	4.0326 (0.0081)	3.2261 (0.1935)	4.0324 (0.0081)
		$\hat{\sigma}_2$	0.2558 (0.2790)	0.2459 (0.2294)	(*)	0.2459 (0.2294)
		$\hat{p}$	0.5380 (0.0759)	0.5000 (0.0000)	0.3000 (0.4000)	0.4988 (0.0025)
		$\hat{\mu}_1$	0.9899 (0.0101)	1.0077 (0.0077)	(*)	1.0077 (0.0077)
Complete Separation:	$p = 0.5$	$\hat{\sigma}_1$	0.1098 (0.0976)	0.1065 (0.0.0652)	(*)	0.1065 (0.0654)
$(\mu_1, \sigma_1, \mu_2, \sigma_2) =$		$\hat{\mu}_2$	3.9789 (0.0053)	3.9793 (0.0052)	2.9987 (0.2503)	3.9793 (0.0052)
(1.0, 0.1, 4.0, 0.2)		$\hat{\sigma}_2$	0.1686 (0.1570)	0.1878 (0.0609)	(*)	0.1878 (0.0611)
		$\hat{p}$	0.2365 (0.7043)	0.8000 (0.0000)	0.3000 (0.6250)	0.8013 (0.0016)
		$\hat{\mu}_1$	0.9227 (0.0773)	1.0036 (0.0036)	(*)	1.0036 (0.0036)
	$p = 0.8$	$\hat{\sigma}_1$	0.0336 (0.6638)	0.1037 (0.0367)	(*)	0.1036 (0.0364)
		$\hat{\mu}_2$	1.4642 (0.6340)	4.1985 (0.0496)	2.9690 (0.2578)	4.1988 (0.0497)
		$\hat{\sigma}_2$	(*)	0.2399 (0.1997)	(*)	0.2399 (0.1993)
		$\hat{p}$	0.1159 (0.4204)	0.1407 (0.2965)	(*)	0.1401 (0.2994)
		$\hat{\mu}_1$	3.1986 (0.0004)	3.2030 (0.0009)	3.7810 (0.1816)	3.2030 (0.0009)
	$p = 0.2$	$\hat{\sigma}_1$	0.0269 (0.7313)	0.0317 (0.6829)	(*)	0.0318 (0.6820)
		$\hat{\mu}_2$	3.8942 (0.0265)	3.9043 (0.0239)	3.8523 (0.0369)	3.9044 (0.0239)
		$\hat{\sigma}_2$	0.2274 (0.1369)	0.2448 (0.2238)	0.2361 (0.1807)	0.2447(0.2233)
		$\hat{p}$	(*)	0.5354 (0.0707)	0.5994 (0.1989)	0.5358 (0.0717)
		$\hat{\mu}_1$	1.5640 (0.5112)	3.2309 (0.0097)	3.1778 (0.0069)	3.2310 (0.0097)
Moderate overlap:	$p = 0.5$	$\hat{\sigma}_1$	0.0059 (0.9407)	0.1058 (0.0577)	0.0828 (0.1723)	0.1058 (0.0580)
$(\mu_1, \sigma_1, \mu_2, \sigma_2)=$		$\hat{\mu}_2$	3.5660 (0.1086)	4.0194 (0.0049)	4.0274 (0.0068)	4.0194 (0.0049)
(3.2, 0.1, 4.0, 0.2)		$\hat{\sigma}_2$	0.2820 (0.4102)	0.1707 (0.1463)	0.1327 (0.3365)	0.1707 (0.1466)
		$\hat{p}$	0.7890 (0.0138)	0.9900 (0.2375)	0.9132 (0.1415)	0.7380 (0.0776)
		$\hat{\mu}_1$	3.3924 (0.0601)	3.3168 (0.0365)	3.2198 (0.0062)	3.1909 (0.0028)
	$p = 0.8$	$\hat{\sigma}_1$	(*)	0.0993(0.0070)	0.1679 (0.6794)	0.1038 (0.0383)
		$\hat{\mu}_2$	3.1957 (0.2011)	3.8810 (0.0297)	4.4663 (0.1166)	3.9269 (0.0183)
		$\hat{\sigma}_2$	0.0469 (0.7654)	0.2503 (0.2515)	0.3196 (0.5981)	0.2224 (0.1120)
		$\hat{p}$	0.0667 (0.6663)	0.0100 (0.9501)	0.0968 (0.5162)	0.1158 (0.4211)
		$\hat{\mu}_1$	3.3662 (0.0382)	3.7802 (0.0800)	3.7811 (0.0803)	4.6594 (0.3313)
	$p = 0.2$	$\hat{\sigma}_1$	0.0784 (0.2161)	0.2193 (1.1930)	(*)	0.1880 (0.8796)
		$\hat{\mu}_2$	3.9515 (0.0121)	3.9235 (0.0191)	3.8212 (0.0447)	3.8520 (0.0370)
		$\hat{\sigma}_2$	0.2224 (0.1120)	0.2317 (0.1583)	0.3885 (0.9426)	0.1843(0.0784)
		$\hat{p}$	0.6691 (0.3382)	0.5753 (0.1506)	0.6052 (0.2103)	0.5734 (0.1469)
		$\hat{\mu}_1$	3.5146 (0.0042)	3.6536 (0.0439)	3.4518 (0.0138)	3.4909 (0.0026)
Large overlap:	$p = 0.5$	$\hat{\sigma}_1$	0.1309 (0.3085)	0.1124 (0.1240)	0.1227 (0.2269)	0.0892 (0.1078)
$(\mu_1, \sigma_1, \mu_2, \sigma_2)=$		$\hat{\mu}_2$	4.1835 (0.0459)	4.7028 (0.1757)	4.2314 (0.0578)	4.0164 (0.0041)
(3.5, 0.1, 4.0, 0.2)		$\hat{\sigma}_2$	0.1309 (0.3457)	0.1957 (0.0215)	0.3010 (0.5048)	0.2492 (0.2462)
		$\hat{p}$	0.7272 (0.0911)	0.8403 (0.0504)	0.8389 (0.0487)	0.8396 (0.0495)
		$\hat{\mu}_1$	3.6389 (0.0397)	3.5123 (0.0035)	3.5120 (0.0034)	3.5120 (0.0034)
	$p = 0.8$	$\hat{\sigma}_1$	0.2027 (1.0274)	0.1011 (0.0107)	0.1009 (0.0090)	0.1009 (0.0087)
		$\hat{\mu}_2$	3.5063 (0.1234)	4.0418 (0.0105)	4.0389 (0.0097)	4.0378 (0.0095)
		$\hat{\sigma}_2$	0.0405 (0.7973)	0.2912 (0.4558)	0.2907 (0.4534)	0.2904 (0.4521)