Statistics of retention failure in the low resistance state for hafnium oxide RRAM using a Kinetic Monte Carlo approach

Nagarajan Raghavan a,*, Daniel D. Frey b, Michel Bosman c, Kin Leong Pey a

a Engineering Product Development, Singapore University of Technology and Design, 487 372, Singapore
b Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
c A*STAR Institute of Materials Research and Engineering (IMRE), 117 602, Singapore

ARTICLE INFO

Article history:
Received 27 May 2015
Received in revised form 17 June 2015
Accepted 18 June 2015
Available online xxxx

Keywords:
Conductive filament
Diffusion barrier
Markov model
Retention failure
RRAM

ABSTRACT

Retention is one of the key reliability metrics for non-volatile memory devices. In oxygen ion transport based resistive switching memory (OxRAM), the retention phenomenon has been well studied from an electrical perspective and physical mechanisms explaining the origin of retention loss have also been speculated to support the observed data. However, the stochastic aspects of retention loss and its variability deserve to be investigated so that the time-dependent shift in the resistance distribution and the retention failure time statistics can be better quantified and estimated for a given set of operating conditions. We propose here a phenomenological Markovian multi-state model combined with the percolation framework and ion diffusion theory to analyze the distributions of retention failure in the low resistance state for OxRAM devices.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Reliability study of resistive switching memory has gained a lot of attention over the past few years [1–4], as successful process optimization has helped in the realization of ultra-small devices of 10 × 10 nm2 dimensions [2], that make RRAM a potential candidate for replacement of 3D NAND Flash technology, which suffers from severe variability issues due to the large spread in the defect count of different discrete devices in the memory array. There are three key degradation mechanisms that are of primary concern in RRAM — endurance degradation [5,6], retention loss [7,8] and read disturb immunity [9,10]. While the different possible physical mechanisms explaining these degradation phenomena have been well studied, there is limited know-how on the right statistical models that are needed to fit the accelerated test data. Though some initial effort has been put in towards this direction [7], the models still tend to be non-physical, which limits their general application to reliability and variability assessment of the technology on the whole.

The motivation of this study is to develop a statistical framework for analyzing the distributions of retention loss failure in the low resistance (ON) state (LRS) for HfOx RRAM, which is a bipolar switching device governed by the reversible drift/diffusion transport of oxygen ions. We make use of the Markov Chain theory in conjunction with the percolation cell model [11] to describe the stochastic dynamics of the retention loss phenomenon. The statistical spread is assessed by using a Kinetic Monte Carlo (KMC) routine and the physics of ion transport is described by the Mott theory [12].

The structure of this work is organized as follows. In Section 2, the statistical framework developed and the equations governing the retention loss process are described in detail. This is followed by an analysis of the time-based shift in the filament resistance distribution and the final retention failure time distribution for different switching compliance (\(I_{\text{comp}}\)) settings in Section 3. The impact of filament shape on the distribution is assessed and a likelihood analysis is also carried out in this section to determine which of the standard statistical distributions best fits the simulated data as well as electrical test data on a Ti-doped HfOx device. Finally, we conclude the study in Section 4 with the different inferences that can be made and possible suggestions for further work in extending and applying the model developed here to other scenarios. Our results tend to show that the onset of retention loss failure is determined by the defects residing close to the top electrode interface irrespective of the shape of the filament. Furthermore, the Generalized Extreme Value (GEV) distribution seems to be the most suitable one to describe the failure data. The analysis in this study is carried out on a structure with physical dielectric thickness of \(t_{\text{ox}}\) ~ 3.6 nm and defect size, \(a_0\) ~ 6 Å.

2. Statistical framework for retention loss modeling

The detailed methodology developed to model the retention phenomenon is schematically shown in Fig. 1. The filament in the LRS is modeled as a collection of oxygen vacancy defects arranged in a 2D
lattice of possible defect locations (based on the percolation framework). The defects are placed randomly across the rows and columns (without any spatial preference) of the cell network to “create” the filament. Once the filament is created, the criterion for retention failure is defined as the first instance at which a complete row of defects is annihilated by the diffusion of oxygen ions (with electric field absent). Oxygen ions are assumed to jump across multiple potential energy wells from the top electrode before they finally reach a defect in the filament. The space between adjacent rows of defects and the metal–dielectric interface are considered to serve as diffusion energy barriers. Every oxygen ion residing in the top electrode is considered to move (hop) based on a random walk across the multiple potential barriers.

Each row of defects is considered as a discrete state of a Markov Chain and the top (TE) and bottom electrodes (BE) are referred to as the source state and sink state respectively. We only consider the $O_2^{2-}$ ions, determined using a random number generator and the normalized state probabilities for each row, as shown schematically in the color coded horizontal bar.

\[
\frac{dp_{TE}(t)}{dt} = \lambda_{DIFF} \cdot (p_{K-1}(t) - p_k(t)) 
\]

\[
\frac{dp_{BE}(t)}{dt} = \lambda_{DIFF} \cdot p_N(t) 
\]

Given the initial conditions, these differential equations can be solved analytically to obtain the state probability equations as given by Eqs. (4)–(6), which is graphically represented by Fig. 2. It is interesting to note that the state probability expressions (Eq. (5)) are Poisson-like. At every time instant, the row in which the defect is passivated is determined by the range of the cumulative state probabilities where the random number falls into. As an example, if the value \( \text{rand} \) at a time instant, \( t = k \times \Delta t \), falls in the range between \( p_1(k\Delta t) + p_2(k\Delta t) \) and \( p_1(k\Delta t) + p_2(k\Delta t) + p_3(k\Delta t) \), then we consider the defect to be passivated in the third row of the filament with reference to TE. The value of \( \Delta t \) is chosen to be infinitely small when compared to \((1/\lambda_{DIFF})\), which can be considered as the effective time constant for the diffusion phenomena. In our analysis, we arbitrarily choose \( \Delta t = \chi \times (1/\lambda_{DIFF}) \), where \( \chi = 1/5000 \). The smaller the value of the scaling factor \( \chi \), the more accurate the simulation would be, though at the expense of increased computational load. The expressions for \( \lambda_{DIFF} \) and \( \lambda_{REC} \) are given by Eqs. (7) and (8) where \( \nu \) is the lattice vibration frequency (-1014 Hz), \( a \) is the lattice periodicity, \( k_B \) is the Boltzmann constant and \( T \) is the absolute temperature.

\[
p_{TE}(t) = \exp(-\lambda_{DIFF} \cdot t) 
\]

\[
p_k(t) = \lambda_{DIFF} \cdot \nu \cdot a \cdot \exp\left(-\frac{E_{a,REC}}{k_B T}\right) \left(\frac{t}{k_B T}\right) 
\]

\[
p_{BE}(t) = 1 - p_{TE}(t) - \sum_{k=1}^{N} p_k(t) 
\]

\[
\lambda_{DIFF} = \nu \cdot \left(\frac{a}{h}\right) \cdot \exp\left(-\frac{E_{a,DIFF}}{k_B T}\right) 
\]

Please cite this article as: N. Raghavan, et al., Statistics of retention failure in the low resistance state for hafnium oxide RRAM using a Kinetic Monte Carlo approach, Microelectronics Reliability (2015), http://dx.doi.org/10.1016/j.microrel.2015.06.090


\[ \lambda_{REC} = \nu \cdot \exp\left(-\frac{E_{a-REC}}{k_BT}\right) \]  

(8)

3. Retention failure and filament resistance statistics

Having developed the model to describe retention loss, let us examine the results of the KMC simulation for both the retention failure time as well as the time-dependent shift in the filament resistance distribution. The main operational parameter considered in our simulation is the compliance setting \( I_{comp} \) for the switching process which determines the size of the filament. Values for \( I_{comp} \) ranging from 1 μA all the way to 100 μA have been explored. We assume here that the defect count \( N_{defects} \) exhibits a Gaussian trend with a mean and standard deviation of \( \mu_{N-LRS} \) and \( \sigma_{N-LRS} \) respectively. The assumed values of these two parameters for different \( I_{comp} \) is listed in Table 1, and these values can be extracted from analytical fitting of the Quantum Point Contact (QPC) model described by Degraeve et al. [13] to the measured I–V data for multiple switching cycles. Fig. 3 plots the resistance-time trend from several simulation trials for two \( I_{comp} \) values of 10 μA and 100 μA respectively.

The distribution trend of the resistance of the filament at the LRS state to begin with. Also notice the increasing deviation from the conventional Lognormal trend as \( I_{comp} \) is reduced. For ultra-low power applications (very low \( I_{comp} \)), the variability issues seems to become a key constraint as this plot shows.

The gradual evolution of \( R_{fil} \) during the baking process is probed from the results of our model as shown in Fig. 5 for the case of \( I_{comp} = 100 \mu A \), taking \( T = 398 \text{ K (125 °C)} \) and \( E_{a-DIFF} = 1.4 \text{ eV} \). Note that the initial resistance trend follows the Lognormal trend very well up to a baking time of around 800 s, following which the filaments start to “shrink” and “rupture” at some necking point causing the trend to appear more convexial in the high percentile region from 800–1400 s. At the far right end of the plot, the data correspond to resistance of filaments that already ruptured. The shape of the distribution becomes clearly bimodal as the rupture begins.

Having looked at the evolution of resistance trends, we now plot the retention time statistics on a standard Weibull plot in Fig. 6 for widely different values of \( I_{comp} \). Note that although the data seems to fit in

<table>
<thead>
<tr>
<th>Compliance current ( (I_{comp}) )</th>
<th>( \mu_{N-LRS} ) (μ)</th>
<th>( \sigma_{N-LRS} ) (μ)</th>
<th>( \sigma_{N-LRS}/\mu_{N-LRS} ) (no units)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 μA</td>
<td>80</td>
<td>5</td>
<td>6.25%</td>
</tr>
<tr>
<td>50 μA</td>
<td>50</td>
<td>8</td>
<td>16.0%</td>
</tr>
<tr>
<td>25 μA</td>
<td>25</td>
<td>6</td>
<td>24.0%</td>
</tr>
<tr>
<td>10 μA</td>
<td>15</td>
<td>7</td>
<td>46.7%</td>
</tr>
<tr>
<td>1 μA</td>
<td>11</td>
<td>6</td>
<td>54.5%</td>
</tr>
</tbody>
</table>

Please cite this article as: N. Raghavan, et al., Statistics of retention failure in the low resistance state for hafnium oxide RRAM using a Kinetic Monte Carlo approach, Microelectronics Reliability (2015), http://dx.doi.org/10.1016/j.microrel.2015.06.090

![Fig. 2. Markov chain state probability trends for different rows (layers) of vacancies in the percolation based filament matrix. Stages I, II ... refer to the first and second rows of \( V_t \) closest to the top electrode and so on.](#)

![Fig. 3. Evolution of the filament resistance with time during retention loss in the LRS state of HfO₂ RRAM. A quasi-linear trend of resistance increase towards rupture (big jump in \( R_{fil} \)) due to preferential passivation of vacancy defects closer to the metal–electrode interface is observed.](#)

![Fig. 4. Lognormal plot of the filament resistance \( (R_{fil}) \) after retention failure occurred (filament rupture). Note the widening spread in the resistance distribution with reducing SET compliance level. Significant variability in the filament resistance, shape and size is inherent for ultra-low power switching applications. The inset shows the zoomed-in version of the distribution spread at the low percentile region.](#)
well on the Weibull scale, upon closer observation, deviations in the low percentile trends exist. This brings to question the validity of Weibull stochastics for retention failure time analysis. It is key to mention here that from our percolation based representation of the filament, we realize that almost all of the retention failures occurred in Row I, closest to the top electrode (Row I). The GEV works especially well when the number of random variables being analyzed for maxima is large enough. The cumulative density function for GEV is given by Eq. (9), where \( \mu \) is the

Fig. 5. Time dependent shift and evolution of the filament resistance distribution in the LRS state for 500 simulation trials considering \( E_{ad} = 1.4 \text{ eV} \) and \( T = 398 \text{ K} \). Note the initial shifts being parallel to each other and well represented by the Lognormal distribution. For baking time \( >800 \text{ s} \), the high percentile region starts to bend indicating onset of rupture and subsequently, the distributions become bimodal with an increasing percentage of population of devices undergoing filament rupture.

Fig. 6. Weibull plot of the retention failure time distribution at the low percentile regime for a wide range of \( I_{comp} \) ranging from 1–100 \( \mu \text{A} \). Upon close observation, it is more evident that the Weibull behavior is not very well obeyed at lower percentiles, which questions the validity of this assumption. Possibility of other suitable distributions to better model retention failure cannot be ignored.

Fig. 7. Comparison of the fitting of the GEV, Weibull, Lognormal, Gamma and Exponential failure distributions to the (a) simulated LRS retention loss data (\( I_{comp} = 100 \mu \text{A} \)) as well as the (b) experimental data extracted from Ref. [14] for Ti-doped HfO\(_x\) (\( U_{comp} = 50 \mu \text{A} \)). The maximum likelihood estimate (MLE) analysis shows that the GEV distribution best fits the set of data in (a), while the limited set of data in (b) does not provide a clear conclusion.

Please cite this article as: N. Raghavan, et al., Statistics of retention failure in the low resistance state for hafnium oxide RRAM using a Kinetic Monte Carlo approach, Microelectronics Reliability (2015), http://dx.doi.org/10.1016/j.microrel.2015.06.090
These asymmetric structures, the failure time distribution shifts significantly. When a similar retention loss simulation is carried out on the TE or BE interfaces to realize inverted cone and conical percolation theory, Markov chain, ionic transport and KMC algorithm. The model helped in quantifying the trends of resistance variation and retention loss time in terms of the shape and mean values of the distribution. The role of different shapes of the filament was investigated and the analysis reveals that retention loss is predominantly governed by the number of defects in the percolation cell matrix row closest to the TE. The GEV distribution was identified to be the best representative for retention failure data. Further study is required to extend the model to account for the possible lateral diffusion of vacancies from the filament as proposed recently in Ref. [8]. The lateral diffusion of vacancies could be an additional significant factor other than the vertical back diffusion of oxygen ions that is detrimental in lowering the retention lifetime of the device. The possible existence of a locally high interfacial diffusion barrier at the TE-oxide junction (Ref. [16]) may also be an important aspect to consider in the model.

**Acknowledgment**

The authors would like to thank the International Design Center (IDC) at the Singapore University of Technology and Design (SUTD) for the technical resource support. The first author would like to acknowledge the travel funding support provided through the Start-Up Research Grant (SRG) by SUTD and the SUTD-MIT Joint Post-Doctoral Fellowship fund.

**References**


Fig. 8. Weibull plot of the retention loss time for three different assumed shapes of the filament—a symmetrical filament in black, inverted cone in blue and upright cone in violet, given \(I_{\text{comp}} = 100 \mu \text{A}\) and \(T = 398 \text{ K}\). The distributions tend to deviate quite a lot from the standard Weibull model, more so at the low percentile regime.

\[
F(t; \mu, \sigma, \xi) = \exp \left\{ - \left[ 1 + \xi \frac{t - \mu}{\sigma} \right]^{-1/\xi} \right\}
\]