Dielectric Breakdown Characteristics of Stacked High-k Dielectrics

Byoung Hun Lee\textsuperscript{1,2} and Rino Choi\textsuperscript{3}

\textsuperscript{1} Department of Material Science and Engineering, \\
\textsuperscript{2} Department of Nanosystems Engineering, \\
Gwangju Institute of Science and Technology (GIST) \\
Oryong-dong 1, Buk-gu, Gwangju, Korea 500-712 \\
\textsuperscript{3}Material Science and Engineering Department, Inha University, \\
253 Yonghyun-Dong, Nam-Ku, Incheon, Korea 402-751

Dielectric breakdown characteristics of high-k dielectric have been intensively studied to develop a lifetime extrapolation model for device with metal/high-k gate stacks. Majority of prior works treated the high-k dielectric as a single layer dielectric like thermally grown SiO\textsubscript{2} while the actual structure of high-k dielectric consists of two layers (high-k layer stacked on interfacial SiO\textsubscript{2} like interfacial layer). And, the results of reliability test were interpreted using the model developed for SiO\textsubscript{2}. In the previous works, the potential limit of such approach has been pointed out; 1) severe electric field distortion due to transient charging in high-k dielectric during the stress, 2) excessive electrical stress applied to the interfacial layer. Thus, it was suggested that the conventional lifetime extrapolation method using the data obtained from a high field stress may not be correct for a high-k dielectric. In this paper, several breakdown models suggested for high-k dielectric will be reviewed and compared to provide a comprehensive understanding on the current knowledge and challenges in this area.

Introduction

After more than a decade of intensive research, high-k dielectrics have been implemented into the gate stack of 32 nm or 45 nm node high performance CMOS devices [1]. As a result, the power consumption in recent CPU from Intel dropped by close to an order of magnitude from the previous generation. Also, more device makers are implementing high-k dielectric into their roadmap even for less aggressively scaled products such as low standby power applications and memory devices [2].

However, the reliability evaluation methods for metal electrode/high-k dielectric stack have not been extensively discussed even after the high-k dielectrics are adopted as a part of main stream technology. The study on the negative bias temperature instability (NBTI) and positive bias temperature instability (PBTI mostly due to charge trapping) of high-k dielectric has been very active [3,4] and it is agreed that the transient charge trapping in the bulk high-k layer and a SiO\textsubscript{2} like interface state generation at the interfacial layer co-exist with a different degree of mixture depending on the stress methods. In the studies on hot carrier injection (HCI) stresses, there is a consensus that the device degradation after the HCI stress consists of both a permanent degradation at
the interfacial layer and a transient charge trapping at high-k layer even though how the resulting device shift after HCI can be de-convoluted into two components is still a challenging subject [5,6]. Detrapping bias was used to eliminate the portion of transient charge trapping from the total device shift, but the selection of detrapping bias and detrapping time is not a simple task because the detrapping bias itself can be a source of device shift. The mechanism of time dependent dielectric breakdown (TDDB) has been studied extensively, but there is no clear consensus yet.

There are several reasons for this controversy. First, two different integration schemes (Gate first scheme and replacement gate scheme) have been used to implement new gate stack technology [7]. Intel uses a replacement gate scheme and companies in IBM alliance are using a gate first integration scheme. There are many significant differences between these two schemes in terms of gate stack processing and material selection [8,9]. The difference in heat cycle between the two schemes changes the oxygen distribution and inter-diffusion within the gate stack [10]. The degree of flat band voltage ($V_{FB}$) roll off problem due to oxygen redistribution is more severe for the gate first integration while the gate last integration scheme is relatively free from it [7,11,12]. To mitigate the $V_{FB}$ roll off problem, the gate first integration scheme uses more complex material system.

Second, there are no unified gate stack materials or processes for the metal electrode/high-k dielectric gate stack. To mitigate $V_{FB}$ roll off problem, the gate first integration scheme needed to introduce additional dipole materials such as La-O and Al-O into HfO$_2$ using capping or alloying method [13,14]. Doping in HfO$_2$ changes the reliability characteristics significantly [15]. Also, the choice of materials should be different to optimize the device performance even within the same integration scheme. For example, a channel doping can be slightly reduced to improve the channel mobility when band edge workfunction metal electrodes are not needed to get a low threshold ($V_{th}$) values [16]. Or, a simple and low cost processing scheme such as single metal gate/single high-k dielectric stack can be used to reduce the process cost while limiting the device design window [17].

All of these complexities are due to the fact that the gate stack materials and its processing method itself are important factors in the control of the effective workfunction (EWF) or threshold voltage unlike the case of poly silicon/SiO$_2$ stack where poly silicon processing or SiO$_2$ processing has only minor impacts on the workfunction or threshold voltage. As a result, the reliability of gate stack no longer means a simplified modeling of the reliability of gate stack itself. It should include the effect of other processes including overall gate stack integration scheme, heat cycle and gate patterning process.

Thus, a study on the intrinsic properties of gate stack became very complicated and it is very difficult to delineate a unique degradation mechanism intrinsic to high-k dielectrics. This paper will summarize the existing approaches presented in the literatures to provide insights on the diversity of reliability study for alternative gate stack and suggest a model to controversies among different models.

Key models for the intrinsic reliability
In this section, reliability models used in recent publications (implicitly and explicitly) will be discussed. Fig. 1 describes a model which is similar to the percolation model for SiO₂. According to this model, defects generated by energetic electrons are gradually accumulated within the high-k dielectric throughout the stress period and lead to the failure when a breakdown path is formed [18]. Supporters of this approach present a gradual degradation of gate leakage current (stress induced leakage current) as an evidence for this model [19,20].

Also, indirectly, higher charge pumping current after a constant voltage stress, especially at a low frequency, is used as another strong evidence of gradual defect creation and accumulation in high-k dielectric layer. Based on this model, the defects generated in high-k dielectric was separated from those of SiO₂ by subtracting a charge pumping current at high frequency region from that of low frequency region.

The basic assumption of this approach is that the charge pumping measurement can detect the defects in high-k dielectric layer through the interfacial layer at low frequency due to longer pulse time to access trap sites [21]. However, there are opposite opinions about the penetration depth of charge pumping. Heh et al proposed that the penetration depth of charge pumping is only around 1nm even at 1 KHz, which is typically near to the interface between SiO₂ interfacial oxide and high-k dielectric [22,23]. Numerical simulation of Masuduzzaman et al also indicated that the conventional charge exchange can only happen up to ~ 1nm, but they proposed the charge exchange between high-k layer and silicon substrate through a direct tunneling may happen to a little deeper distance up to ~1.5 nm [24].

![Fig. 1. Schematic band diagrams showing the mechanism of progressive breakdown of high-k dielectric](image)

According to these works, the increased charge pumping current at low frequency is more likely a signature of defect generation within the interfacial SiO₂ layer or at the internal interface between interfacial SiO₂ and high-k dielectric. Then, the charge pumping current at low frequency after a stress can be a evidence corroborating with the other works claiming that the defect generation rate at the interfacial SiO₂ layer which is in contact with high-k layer is higher than the rate at a pure SiO₂ layer [25]. This tentative conclusion might match with other physical analysis indicating that this
interfacial SiO$_2$ layer is oxygen deficient and most likely intermixed with some of metal atoms from high-k layer such as Hf, Al or La.

Fig. 2 is a schematic band diagram of generated subordinate carrier injection (GSCI) model [26,27]. Key concept of this model is that the TDDB lifetime is mainly controlled by the subordinate carrier component in the gate current. For example, in case of substrate injection stress (+$V_G$), electrons are the main carriers and holes are the subordinate carriers controlling the breakdown process. In case of the gate injection stress (-$V_G$), a hole current is the main carrier and TDDB characteristics are governed by an electron current. The data presented for this model are very convincing, especially at high field stress regime. Total amount of subordinate carrier to the breakdown ($Q_{BD}$) is almost constant in high field stress region ($V_G > \pm 4.5$ V). However, an electron $Q_{BD}$ for HfAlO$_x$/SiO$_2 = 5.71\text{nm}/1.3\text{nm}$ stack start to increase at a stress bias lower than $V_G<-4.5$ V, indicating that the QBD model valid for high field stress region may not be accurate at low field stress region [27]. This is an important clue to resolve the controversies in TDDB model.

Despite of clear experimental evidences, some aspects of this model need more clarification. This model doesn’t have a physical model explaining why the carrier dominating the breakdown process should be different for different stress polarities. By using $Q_{BD}$ as an indicator for the breakdown model, this model implicitly assumes a gradual defect accumulation in a high-k dielectric like the model shown in Fig. 1. Then, why the defect generation mechanism within the same high-k dielectric layer should be different for different stress polarities? A significant difference between high field stress case and low field stress case under a gate injection stress cannot be easily explained. Also, most of data to support GSCI model were collected at $\pm 4V-6V$, which is relatively high field stress for HfAlO$_x$/SiO$_2 = 5.71\text{nm}/1.3\text{nm}$. Electric field calculation shows that each layers are subject to the stress field higher than their intrinsic breakdown field. Thus, this model may be suitable for the TDDB characteristics under a high field stress, but the applicability for low field stress region is not fully examined yet.

$$+V_G: \text{Substrate injection} \quad -V_G: \text{Gate injection}$$

Fig. 2. Schematic band diagrams showing the Generated Subordinate carrier injection induced breakdown mechanism (GSCI) of high-k dielectric.
Fig. 3 is a schematic band diagram of thermo-chemical breakdown model, which emphasizes the electric field driven breakdown mechanism due to a atomic bond stretching beyond its critical value [28]. In this model, the breakdown process of high-k dielectric layer is initiated when the electric field applied to a high-k dielectric is higher than its intrinsic bulk breakdown field [29]. Experimental evidence supporting this model was shown by Lee et al [29]. In their work, a metal-insulator-metal (MIM) structure was used to study the breakdown mechanism of a single layer of HfO₂ and Hf-silicate without interfering with an interfacial SiO₂ layer. The typical field dependence of thick SiO₂ gate dielectric was ~0.93 decade/(MV/cm) while that of MIM Hf-silicate was 2.04 decade/(MV/cm) [29]. The electric field dependence in MIM HfSiOₓ capacitor more than two times stronger than that of SiO₂ indicates that the breakdown process of high-k dielectric layer itself is strongly influenced by the electric field rather than the charge fluence like in SiO₂ layer. According to this model, the TDDB lifetime of high-k layer under the stress below the intrinsic breakdown field should have a much longer than the high field stress case and a gradual degradation the in high-k layer should be minimal at a low field stress. This prediction is contradictory to the experimental observation supporting the GSCI model at high field stress region (Fig. 2). Thus, a unified TDDB model that can explain experimental observations for both models: strong field dependence of TDDB of MIM capacitor and strong QBD dependence of MIS capacitor.

![Fig. 3. Schematic band diagrams showing the field dependent breakdown mechanism of high-k dielectric. Dielectric breakdown happens when the electric field in HfO₂ layer is higher than the critical field.](image)

**Unified model for the intrinsic reliability**

As mentioned above, one of the major factors that differentiate high-k dielectrics from SiO₂ in term of reliability characteristics is that the high-k dielectric is actually a stacked film consists of high-k dielectric layer and SiO₂-like interfacial layer [30]. When one of the stacked layers broke down first, the other layer will be subject to a strongly localized stress at the breakdown point, which will result in a hard breakdown. This process can happen as a step wise breakdown process or the stress current crowding can
cause an immediate breakdown of two stacked layers. In either case, depending on the process and stress condition, the weak layer of the stacked layers will dominate the reliability characteristics of the whole stack. Several publications on the reliability characteristics of high-k dielectrics already observed that the interfacial SiO$_2$ layer will be degraded first under a constant voltage stress [31,32]. Yet, a complete understanding on the degradation mechanism has not been established.

How the breakdown at an interfacial SiO$_2$ layer and the gradual degradation in a high-k dielectric can be combined to provide a universal TDDB model for a stacked high-k dielectric? An analysis on the electric field distribution within metal electrode/high-k dielectric stack showed that the electric field applied to SiO$_2$ layer always reaches the intrinsic breakdown field (~15 MV/cm) well before the electric field applied to high-k layer reaches its intrinsic breakdown field (3.9 MV/cm for HfO$_2$) if the stress bias is gradually increased [33]. More detail reasoning for this conclusion is as following.

Fig. 4 shows a thickness map of HfO$_2$/SiO$_2$ stack. Two lines shown in Fig. 4 represent the boundary of breakdown field at the SiO$_2$ and HfO$_2$ layer respectively. The line separating the region 1 and region 2 is calculated using the equation (1) [34].

$$T_{phy,ox} = \frac{\varepsilon_{ox}}{\varepsilon_{hk}} \left( T_{phy,hk} - \frac{\varepsilon_{hk}}{\varepsilon_{ox}} \frac{V_{stress}}{0.01 \cdot E_{bd,ox}} \right)$$ (1)

At region 1, both layers are below the intrinsic breakdown field when the gate bias is 3 V (for SiO$_2$ (15 MV/cm) and HfO$_2$ (4 MV/cm)). However, as the thickness of gate stack is reduced, the stress field within the gate stack is changed to region 2 where the stress field applied to the interfacial SiO$_2$ layer is higher than 15 MV/cm. A similar transition happens if the stress bias is increased at a given stack (fix the thickness for high-k layer and SiO$_2$ layer and change the stress bias). As shown in Fig. 4, the interfacial layer always reaches the breakdown field first.

![Fig. 4. Thickness map showing three different stress regions at a stress bias of 3 V (presented at Int. Workshop. on Dielectric Thin Film, 2009) [34].](image_url)
Lee et al also showed that many studies on the reliability of high-k dielectric used an excessively high electric field stress because it is usual to choose a stress value close to a hard breakdown point to accelerate the reliability test [34]. However, in stacked structure, even when the electric field applied to SiO$_2$ layer reached an intrinsic breakdown limit, an actual hard breakdown does not occur because the leakage current through the SiO$_2$ layer is reduced by the additional tunneling barrier due to the overlying high-k dielectric layer. Thus, by choosing a stress bias that is close to an extrinsic hard breakdown bias, the interfacial layer will be subjected to a seriously high electric field [33].

Furthermore, if the stress bias is high enough to increase the electric field for both high-k dielectric and interfacial layer beyond their intrinsic limit, it is likely that the breakdown process can be influenced by the total charge fluence, especially the one directly affecting the breakdown process of weak link, which is typically the interfacial layer. Since many prior works indicated that the interfacial SiO$_2$ layer is the weak link in the breakdown process under both polarities, GSCI model needs to explain how the breakdown process at the interfacial SiO$_2$ layer can be governed by hole current under a substrate injection and electron current under a gate injection. Alternative possibility is that the weak layer of high-k dielectric stack should be different for different polarities, which is less likely to.

Fig. 5 summarizes the idea to resolve this dilemma. This model has not been experimentally examined, but each physical process used in this model has been studied in the literatures. Thus, it will be still meaningful to start the discussion on TDDB process with this kind of modeling. First assumption of this model is that the electron current in the substrate injection and the hole current in the gate injection don’t degrade the interfacial SiO$_2$ layer significantly. This is reasonable assumption because both currents are mostly due to a direct tunneling and these are the main carriers in GSCI model, which are not the source of dielectric degradation.

Second, under a substrate injection stress shown in Fig. 5, the electrons injected from a substrate are accelerated towards a gate electrode and generate the electron-hole pair through the emission of surface plasmons at metal electrode/high-k dielectric interface [35]. Even though the electron-hole pair generation at the metal electrode/high-k dielectric interface is less intuitive, Fisheetti et al showed that the plasmon mediated electron-hole pair can be generated even at metal electrode if the energy injecting electrons are high enough. The hot holes generated at the top interface move back to the interfacial SiO$_2$ layer through the stress field direction and get trapped or generate another electron-hole pairs at the channel. Okada et al measured the constant hole Q$_{_{BD}}$ with a poly silicon gate at the beginning and their works are extended to metal gate case later. This model explains the source of subordinate hole carriers in GSCI model using electron-hole pair generation at metal/high-k dielectric interface or poly silicon/high-k dielectric interface. Then, the Q$_{_{BD}}$ controlled by hot hole current under a substrate injection stress can be explained by GSCI model. This process should be same at low field stress and high field stress even though the amount of hot hole generation at the top interface can be changed.

Third, under a gate injection stress, the electrons injected from the electrode are accelerated toward the substrate as shown in Fig. 5 and the accelerated electrons directly degrade the interfacial SiO$_2$ layer through an electron-hole pair generation at interfacial
layer and/or Si channel [35]. In this process, the electron-hole pair generation and hole trapping at the interfacial SiO₂ layer dominates the degradation of SiO₂ layer and the gradual degradation in high-k dielectric is ignored if the stress field applied to the high-k layer is below the intrinsic breakdown field of high-k layer. This model suggests that the process generating hot holes which is buried in the major hole current can be a real source of dielectric degradation even though the electron current appears to dominate the charge to breakdown process. Indirect evidence for this assumption is that the gradual electron \( Q_{BD} \) increase observed at the low field stress region of gate injection stress of GSCI experiment. As the stress bias is reduced, the energy of electrons injected from the gate is reduced below the critical value to generate an electron-hole pair. Then, hot hole generation process will be slowed down and total amount of electron current has to increase to generate more hot holes. This conjecture matches with the observation made for GSCI model.

**Gate injection**

- Electron injection from gate
- Electron acceleration
- Electron-hole pair generation at interfacial SiO₂ or substrate
- Hot hole induced defect generation in SiO₂
- Leakage current increase
- Destructive feedback

**Substrate injection**

- Electron injection from substrate
- Electron acceleration
- Electron-hole pair generation at electrode
- Hole acceleration backward to channel
- Hot hole induced defect generation in SiO₂
- Leakage current increase
- Destructive feedback

Fig. 5. Schematic band diagrams and sequences explaining the dielectric degradation process for the proposed unified reliability model at a low field stress.
In summary, the model described above suggests that the constant electron $Q_{BD}$ observed in GSCI experiment can be correlated with hot hole $Q_{BD}$ of interfacial SiO$_2$ layer, which is the major factor affecting the TDDB characteristics of stacked high-k dielectric. Merit of this model is that the degradation of SiO$_2$ is only controlled by the hot hole generation-trapping process and the breakdown is always initiated from the interfacial SiO$_2$ layer. Thus, the controversy over GSCI model about the different degradation mechanisms for the breakdown of same layer can be resolved. As mentioned above, this is only a theoretical reasoning to come up with a unified TDDB model. More quantitative study on the electron-hole pair generation mechanism at metal electrode/high-k dielectric is necessary to support this explanation.

Summary

This paper reviewed key models to explain the TDDB characteristics of high-k dielectrics. Main controversy is 1) whether there is a significant gradual defect generation in high-k dielectric and 2) whether it plays an important role in the TDDB process. The electric field used in the reliability study of high-k dielectric is relatively high to study the intrinsic TDDB characteristics of high-k dielectrics. The results obtained from the high field accelerated stress might not be directly applicable to lower field stress case due to the strong field dependence of TDDB characteristics. Direct study on the field dependence of TDDB of high-k dielectric layer alone is necessary to resolve this controversy.

A unified model (hot hole induced interface degradation) that can explain the characteristics of TDDB characteristics of high-k dielectric under both stress polarities is proposed. One unclear aspect of this model is how holes are generated at the metal/high-k dielectric interface when an electron injected from a substrate side is accelerated under a positive gate bias and reached the metal/high-k dielectric interface. More quantitative study on this aspect with different electrode metals is necessary.

Acknowledgments

This works was partially supported by DASAN project through a grant provided by Gwangju Institute of Science and Technology in 2008 and World Class University (WCU) program at the Gwangju Institute of Science and Technology (GIST) through a grant provided by Ministry of Education, Science and Technology (MEST) of Korea in 2009.

References


