Advanced Modeling of Charge Trapping at Oxide Defects

F. Schanovsky, W. Goes, and T. Grasser
Institute for Microelectronics, TU Wien
Gußhausstraße 27-29/E360, 1040 Wien
Fax: +43/1/58801-35099

Abstract—Several reliability issues in MOS transistors - such as the bias temperature instability, hot-carrier degradation, and gate leakage - have been indicated to involve the capture and emission of carriers at point-defects in the oxide. The trapping behavior of these defects depends on the device temperature and the oxide field in a highly non-trivial manner. Detailed capture and emission time constants of single defects have recently been obtained from time-dependent defect spectroscopy (TDDS) measurements. The complex behavior of these time constants is most accurately explained using a multi-state multi-phonon model. In this model, the defects capture and emit carriers through a non-radiative multi-phonon process. Additionally, each defect has (at least) two internal states where each of them gives rise to different trapping dynamics. We give a brief and hopefully intuitive introduction to the theory of non-radiative multi-phonon capture and emission and to the concept of multi-state defects. The relation to the commonly used Shockley-Read-Hall defect description in semiconductor device modeling is discussed.

I. INTRODUCTION

The miniaturization of MOS transistors in the past decade has led to an increase in both the oxide field and the operation temperature, increasing the demands for stability of the insulating oxide. Additionally, the introduction of new materials into the production process results in more complex gate stacks. This increased complexity has brought along an increase in the number of defects in the insulating region, which severely affects the transfer characteristics of the transistor. In consequence, oxide-defect-based reliability issues such as the bias temperature instability (BTI), hot-carrier degradation (HCD), and stress-induced leakage currents (SILC) have climbed to the top of the list of reliability concerns for current technology nodes.

The explanation of the observed degradation effects and accurate life-time projections pose new challenges for semiconductor device simulators. Semiconductor device simulation is usually focused on defects in the semiconductor, which influence the charge transport in the device through scattering or recombination. The capture and emission of carriers at these defects is usually quite fast and the dependence of the capture rates on the applied voltage is largely determined by the carrier concentrations. Additionally, as the semiconducting material is usually crystalline, the defect sites are quite similar, which leads to a negligible variation in the capture and emission rates for different defects. Thus, a defect model based on average transition rates often gives a good description of the behavior of the defect ensemble.

For oxide defects, many of these simplifying conditions are not fulfilled. As the capture and emission processes are slower than in the bulk semiconductor, the charging and discharging kinetics need to be modeled more accurately. Additionally, the usually amorphous oxide makes every defect site unique and average descriptions give poor results. In the last years, a lot of information on the behavior of defects in the MOS oxide has been obtained from bias-temperature stress experiments [1]–[10]. The temperature and field activation observed in these experiments cannot be explained with standard defect models. Quite recently, the behavior of small-area transistors moved into the focus of the scientific attention. In these transistors, single charging and discharging events are visible as distinct steps in the drain current [7], [11]–[17]. The response of these small-area transistors to BT stress has revealed a quite complex behavior of the oxide defects, including a highly non-linear bias dependence of the time constants [18] and correlated gate and drain current fluctuations [19]–[22]. In the following we will focus on this recoverable component. The more permanent component is a topic of current research and the different explanations in the literature are still controversial [6], [8], [23], [24].

Thorough investigations of the experimental data have led to the development of a sophisticated defect model, which is able to explain most of the observed behavior. This defect model describes the carrier capture and emission as non-radiative multi-phonon transitions and also accounts for different internal states of the defect. This document outlines the basic properties of this multi-state multi-phonon defect
model, starting from the commonly employed Shockley-Read-Hall description of defects in semiconductors.

II. DEFECTS IN SEMICONDUCTOR DEVICES

The Shockley-Read-Hall (SRH) theory put forward in 1952 [25] is the standard model for phonon-assisted recombination at defects in semiconductor devices. In the SRH model, the recombination center is described by a capture cross section \( \sigma \) and a trap level \( E_T \). The defect is treated as a localized state that can either be occupied by an electron or unoccupied. The time evolution of the occupancy of the defect \( f \) is described as

\[
\frac{\partial f}{\partial t} = -(k_{dn} + k_{pd})f + (k_{nd} + k_{dp})(1-f),
\]

where \( k_{nd} \) and \( k_{pd} \) are the rates for electron and hole capture (transition from the free state \( n \) or \( p \) to the localized state \( d \)), and \( k_{dn} \) and \( k_{dp} \) are the rates for electron and hole emission (transition from the localized state \( d \) to the free state \( n \) or \( p \)).

The capture process is modeled as a constant flux of the carrier gas moving at the thermal velocity \( v_{th} \) through an opening of area \( \sigma \), see Fig. 1. The emission rates are derived from the principle of detailed balance [25], which ensures that in thermal equilibrium, the occupancy of the trap level follows Fermi-Dirac statistics for a state of energy \( E_T \). The capture and emission rates in the SRH model read

\[
\begin{align*}
  k_{nd} &= \sigma_n v_{th} n(x), \\
  k_{pd} &= \sigma_p v_{th} p(x), \\
  k_{dn} &= k_{nd} \exp\left(-\frac{E_c - E_T}{k_B T}\right), \\
  k_{dp} &= k_{pd} \exp\left(-\frac{E_T - E_c}{k_B T}\right),
\end{align*}
\]

see also Fig. 2.

In semiconductor device simulators, the SRH model is combined with a drift-diffusion based description of the carrier gas and an empirical parametrization for the defect centers using constant capture cross-sections and trap levels. While this description gives good results for the recombination at defects in the semiconductor bulk several works have shown that the detailed description of trapping kinetics requires a more sophisticated theory [26]–[28].

III. PHONON-INDUCED TRANSITIONS

The physical process underlying thermally induced trapping is the non-radiative multi-phonon (NMP) transition. A quantum mechanical theory of NMP transitions was first published by Huang and Rhys in 1952 [29], which has since been broadly applied for the interpretation of measurement data of kinetic trapping experiments [30]–[33]. Introductions to multi-phonon transitions usually involve elaborate quantum mechanics and require a broad knowledge of physical chemistry and mathematical physics. In the present document we draw an intuitive picture for the special case of oxide defects, with the purpose of introducing the reader to the relevant concepts without going too much into the physical details. More detailed discussions can be found in the original papers [34], [35].

From the electronic structure point-of-view, an oxide defect gives rise to a localized state within the forbidden gap of the insulator, i.e. a state with limited spatial extent compared to the quasi-free states in the valence and the conduction band of the semiconductor. The charge carriers can enter or leave this state by means of quantum-mechanical tunneling. These tunneling transitions proceed elastically, i.e. the carrier neither loses nor gains energy in the process. As the energy of the orbit is sharply defined, only carriers which have the exact energy of the trap level can be trapped. Thus, as illustrated in Fig. 3, the trap level acts as an energetic selector and the capture rate will be

\[
k_{nd} \propto T(E_d) D_n(x, E_d) f_n(x, E_d),
\]

where \( T \) is a tunneling coefficient, and \( D_n \) and \( f_n \) are the local density of states and the occupation function for electrons. Similarly, the rate of emission from the localized state reads

\[
k_{dn} \propto T(E_d) D_n(x, E_d)(1 - f_n(x, E_d)).
\]

For the sake of simplicity, we introduce the density of occupied electron states \( n(x, E) = D_n(x, E)f_n(x, E) \) and write the
capture and emission rates as

\[ k_{nd} \propto T(E_d)n(x_i, E_d), \quad (8) \]
\[ k_{dn} \propto T(E_d)(D_n(x_i, E_d) - n(x_i, E_d)). \quad (9) \]

This description assumes a classical carrier model, which does not account for quantum mechanical tunneling. In a quantum-mechanical carrier model, the tunneling coefficient is included in the densities which can then be evaluated at the defect site. The energetic position of the localized electronic orbit depends on the electronic structure of the defect. According to the Born-Oppenheimer approximation, the electronic structure of any molecule depends on the position of its constituent atoms. For the trapping process, this implies that a well-defined level of the localized state can only be given at zero temperature, when the atoms are in their equilibrium position. At finite temperature, the atoms of the MOS device will exhibit a Brownian motion, randomly oscillating around their equilibrium position. This influences the defect level, which will also oscillate in an unpredictable fashion [36]. In consequence, at finite temperatures the sharp energy level for the capture or emission transition is replaced by a probability density \( f(E) \) which gives the probability to find the level around a certain energy. This probability density is called line shape function, due to its origin in optical spectroscopy, where it describes the thermal broadening of absorption lines [29], [30]. Generally speaking, it needs to be considered that a change in the charge state of a defect changes its electronic structure and in consequence the equilibrium positions and oscillation frequencies. For this reason, the line shapes for capture and emission will be different. In consequence every charge transition has its own associated line shape. For this reason we indicate the initial charge state \( q_i \) and final charge state \( q_f \) for a line shape as \( f^{q_i/q_f} \). For an acceptor-like defect, which is neutral when unoccupied and negatively charged when occupied by an electron, the NMP transition rates read

\[ k_{nd} \propto \int_{-\infty}^{\infty} T(E)f^{0/0}(E)n(x_i, E)dE, \quad (10) \]
\[ k_{dn} \propto \int_{-\infty}^{\infty} T(E)f^{-/0}(E)(D_n(x_i, E) - n(x_i, E)). \quad (11) \]

The line shapes of the defect may also reach out to the valence band of the semiconductor, making charge transitions by hole capture and emission possible. The rates for these transitions are analogous to their electron counterparts

\[ k_{pd} \propto \int_{-\infty}^{\infty} T(E)f^{-/0}(E)p(x_i, E)dE, \quad (12) \]
\[ k_{dp} \propto \int_{-\infty}^{\infty} T(E)f^{0/-}(E)(D_p(x_i, E) - p(x_i, E)). \quad (13) \]

The line shape functions is determined by the Born-Oppenheimer potentials of the defect in its various charge states [37]. The calculation of the line shapes requires to go into the details of the quantum-mechanical multi-phonon
theory, which goes beyond the scope of this document. For all systems of practical interest the quantum-mechanical line shape cannot be calculated exactly due to the high dimensionality of the problem and the complexity of the Born-Oppenheimer potentials. It is thus common to assume parabolic potentials for the atoms in the different charge states of the defect. In this approximation, the atomic motion is described as a superposition of harmonic oscillations, which are called (local) eigenmodes of the vibration spectrum. Commonly it is further assumed that only a limited number (usually one) of these vibrational modes has an influence on the transition while the contribution of the other modes can be neglected. Analytical formulae are available for quantum-mechanical line shapes if the frequency of the relevant mode is not modified by the electronic transition [29], [30]. If the frequency changes, the quantum mechanical line shapes can still be calculated numerically [37], simple line shapes can also be calculated using classical statistical mechanics. These expressions neglect the quantum mechanical nature of the vibrating atoms, which is a reasonable for room temperature and above due to the large atomic mass [38].

A. The Trap Level in NMP Theory

While the capture cross-section of the SRH theory is clearly a construct of limited physical relevance, the trap level $E_T$ can also be defined from the NMP theory. To do this, it is necessary to realize that the trap level of SRH theory does not refer to a quantum-mechanical level but a thermodynamical potential. This “thermodynamic” trap level [26], [32], [39], [40] is defined as the value of the electronic chemical potential (Fermi level) at which the defect changes its charge state in thermal equilibrium. For a donor-like trap this means if the Fermi level is below the trap level, the defect is dominantly in the positive state, while if the Fermi level is above the trap level, the defect is dominantly in the neutral state. Due to the differences in the electronic structure, the vibrational properties of the defect may vary strongly between the charge states. These differences add a change in entropy to the trap level, which is consequently slightly temperature dependent [26], [32].

B. Temperature and Field Activation

The NMP theory explains both the temperature and the field activation of the transition process which is experimentally observed. An increase in the temperature results in a broadening of the line shape, as shown in Fig. 5, which increases the capture probability the more, the further away the carrier energy is from the line shape maximum.

The field activation is stronger in oxide defects than in bulk defects due to the absence of free carriers. Just as for the valence and conduction band edges, the energetic position of the line shapes depends on the electrostatic potential in the device. Depending on the applied bias voltage, the electrostatic potential shifts the maximum of the line shape closer to or further away from the occupied states and thus increases or decreases the capture probability, see Fig. 6.

IV. Multi-State Defects

Several experimental observations indicate the existence of internal defect states:
The transitions between different charge states are non-radiative multi-phonon processes 
and the puckered configurations, which have two stable structural configurations. Structures of the dimer 
dimerin a bistable defect, 
The structural reconfigurations are barrier-hopping transitions (Fig. 9), the model assumes that 
which are denoted as neutral and a positive charge state. The structural reconfigurations are barrier-hopping transitions (→) and 
and the transitions between different charge states are non-radiative multi-phonon transitions (→→). 
In the multi-state multi-phonon model, an oxide defect has two stable structural configurations, which can assume a neutral and a positive charge state. The structural reconfigurations are barrier-hopping transitions (→) and the transitions between different charge states are non-radiative multi-phonon transitions (→→).

Decorrelation and bias frequency dependence of the capture and emission time constants. A simple defect model which only switches between two charge states through an NMP process has an inherent correlation of the capture and emission time constants, which is in contrast to the experimental observations [18]. In addition, BT stress experiments with an AC stress voltage have found a frequency dependence of the capture time constant [41]. This frequency dependence is not explainable using a two-state defect but arises naturally from a multi-state defect model.

Curvature in the bias dependence. The gate bias dependence of the capture time constants obtained from time-dependent defect spectroscopy (TDDS) measurements shows a curvature on the log scale in deep inversion, while NMP theory predicts an almost linear behavior in this regime, see Fig. 7 [7, 18].

Anomalous and temporary RTN. In small-area transistors, the charging and discharging of defects is visible as steps in the drain current. The response of some of these defects following bias-temperature stress shows a temporary random-telegraph-noise signal which disappears after a while, see Fig. 8. This change in the trapping behavior of the defect is easily explained by different internal defect states with different capture and emission time constants. 

All this behavior is well-explained by the multi-state multi-phonon model [7, 18, 42]. This model has been developed to explain the degradation and recovery behavior observed in bias temperature stress experiments, but has also been applied to random telegraph noise (RTN) and trap-assisted tunneling scenarios. As illustrated in Fig. 9, the model assumes that the oxide defects can exist in two charge states 1 and 2. In each charge state, the defect can assume one of two internal — or structural — states, which are denoted as 1 and 1′ in
charge state 1, and as 2 and 2' in charge state 2. It is generally assumed that 1 corresponds to a neutral charge state, and 2 to a positive charge state.

While the transitions between the charge states $1 \leftrightarrow 2$ and $2 \leftrightarrow 2'$ are described using the NMP capture and emission rates (10)-(13), the transitions between the internal states $1 \leftrightarrow 1'$ and $2 \leftrightarrow 2'$ are described as an energetic barrier hopping transition using an Arrhenius law

$$k_{\alpha\beta} \equiv \nu \exp \left( -\frac{E_{\alpha\beta}}{k_B T} \right),$$  \hspace{1cm} (14)$$

with $(\alpha, \beta) \in \{(1, 1'), (1', 1), (2, 2'), (2', 2)\}$ where $\nu$ is the attempt frequency and $E_{\alpha\beta}$ is the activation energy for the transition. The existence of internal defect states is a well-studied fact for different defects in SiO$_2$ [40], [44]–[49]. A classical example for a bistable defect as in the multi-state multi-phonon model is the oxygen vacancy in silicon dioxide [40], [50]. As shown in Fig. 10, this defect can exist in a dimer state, where the silicon atoms adjacent to the vacancy form a bond. By overcoming an energetic barrier, which happens by chance through the random motion, one silicon atom can relax back and bond to a back-oxygen. This second configuration has different equilibrium energetics and oscillation frequencies and consequently different trapping behavior. Both of these bonding configurations can exist in both the neutral and the positive state. While the oxygen vacancy is usually taken as an example to illustrate the concept of a multi-state defect, the actual defect structure responsible for the experimentally observed degradation is still unknown and a topic of current research [50]–[55].

Once the states of the defect and the transition rates between those states are defined, the transient behavior of the defect can be modeled. While a two-state defect is fully described by a single occupancy function as in (1), a multi-state defect requires to assign a probability $p_\alpha(t)$ for every state $\alpha$ of the defect. These probabilities have to fulfill the condition $\sum_\alpha p_\alpha(t) = 1$. The time evolution of these probabilities is described by the master equation

$$\frac{\partial p_\alpha}{\partial t} = \sum_{\beta \neq \alpha} k_{\beta\alpha} p_\beta(t) - k_{\alpha\beta} p_\alpha(t).$$  \hspace{1cm} (15)$$

As shown in Fig. 11, this multi-state model drastically improves the reproduction of the experimentally observed capture and emission time constants [7], [22], [42], [56]. Especially for the case of TDDS extracted effective capture and emission time constants, approximate expressions have been derived which represent the multi-state defect as an effective two-state defect.

For the description of the recoverable component of the BTI in large-area MOS devices an ensemble of defects with a random distribution of parameters for both the structural reconfiguration and the NMP transitions is generated. This defect ensemble must be large enough to capture the large variation of energy landscapes for the defects that arises from the amorphous nature of the MOS oxide. The degradation and recovery transients are then obtained from the integration of (15) for all the defects in the ensemble [5], [43], see Fig. 12.

V. CONCLUSION

The modeling of the oxide defects involved in the reliability issues of modern MOS transistors such as the bias temperature instability, hot-carrier degradation and gate leakage requires a detailed description of the electrochemical reactions at the defect site. The recently developed multi-state multi-phonon model successfully reproduces a wide range of experimental data. We have given a brief introduction into the physical theory behind this defect model. The non-radiative multi-phonon theory, which describes the exchange of carriers between the defect and the rest of the device, as well as the concept of internal defect states have been introduced and compared to the standard Shockley-Read-Hall defect model. Examples have been given for the successful application of the model to experimental data.

VI. ACKNOWLEDGEMENT

This work has received funding from the EC’s FP7 grant agreement NMP2010.2.5-1 (MORDRED).
REFERENCES


