ELectronics EXpress

LETTER

IEICE Electronics Express, Vol.14, No.16, 1–11

A physical model of electron trapping/detrapping in electrically stressed oxide

Hongyi Wang, Cong Li[a)], Bingbing Zhang, Shunqiang Xu, Liming Zheng, and Jianfei Wu

School of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, China

a) *cong.li.nudt@aliyun.com*

Abstract: A physical model of electron trapping/detrapping in electrically stressed oxide has been proposed in this paper. The new model is based on both inelastic multi-phonon trap-assisted tunneling and thermal emission, and also considers the capture effect of oxide bulk traps. It handles every trap separately, and establishes the dynamic procedure of traps capture and emission of electrons. Finally, through the proposed model we may accurately and effectively obtain the filling state of all the oxide traps at any stress and any time, which is very useful for the modeling of the endurance and data retention characteristics of floating gate nonvolatile memories.

Keywords: trapping, detrapping, oxide trap, electron, endurance, retention

Classification: Electron devices, circuits and modules

References

[1] N. Mielke, *et al.*: "Flash EEPROM threshold instabilities due to charge trapping during program/erase cycling," IEEE Trans. Device Mater. Rel. **4** (2004) 335 (DOI: 10.1109/TDMR.2004.836721).

[2] J. B. Velamala, *et al.*: "Compact modeling of statistical BTI under trapping/detrapping," IEEE Trans. Electron Devices **60** (2013) 3645 (DOI: 10.1109/TED.2013.2281986).

[3] K. B. Sutaria, *et al.*: "Aging statistics based on trapping/detrapping: Compact modeling and silicon validation," IEEE Trans. Device Mater. Rel. **14** (2014) 607 (DOI: 10.1109/TDMR.2014.2308140).

[4] Y. Higashi, *et al.*: "Unified transient and frequency domain noise simulation for random telegraph noise and flicker noise using a physics-based model," IEEE Trans. Electron Devices **61** (2014) 4197 (DOI: 10.1109/TED.2014.2365015).

[5] M. Banaszeski da Silva, *et al.*: "A physics-based statistical RTN model for the low frequency noise in MOSFETs," IEEE Trans. Electron Devices **63** (2016) 3683 (DOI: 10.1109/TED.2016.2593916).

[6] T. R. Oldham, *et al.*: "Spatial dependence of trapped holes determined from tunneling analysis and measured annealing," IEEE Trans. Nucl. Sci. **33** (1986) 1203 (DOI: 10.1109/TNS.1986.4334579).

[7] R. Yamada, *et al.*: "Analysis of detrap current due to oxide traps to improve Flash memory retention," Proc. IEEE IRPS (2000) 200 (DOI: 10.1109/RELPHY.2000.843915).

[8] G. I. Wirth, *et al.*: "Statistical model for MOSFET bias temperature instability component due to charge trapping," IEEE Trans. Electron Devices **58** (2011)

DOI: 10.1587/elex.14.20170565
Received June 1, 2017
Accepted July 18, 2017
Publicized August 2, 2017
Copyedited August 25, 2017

2743 (DOI: 10.1109/TED.2011.2157828).
[9] D. Garetto, *et al.*: "Small signal analysis of electrically-stressed oxides with Poisson-Schroedinger based multiphonon capture model," Proc. Int. Workshop on Computational Electronics (2010) 1 (DOI: 10.1109/IWCE.2010.5677950).
[10] L. Larcher: "Statistical simulation of leakage currents in MOS and Flash memory devices with a new multiphonon trap-assisted tunneling model," IEEE Trans. Electron Devices **50** (2003) 1246 (DOI: 10.1109/TED.2003.813236).
[11] F. Jimenez-Molinos, *et al.*: "Physical model for trap-assisted inelastic tunneling in metal-oxide-semiconductor structures," J. Appl. Phys. **90** (2001) 3396 (DOI: 10.1063/1.1398603).
[12] M. Herrmann and A. Schenk: "Field and high-temperature dependence of the long term charge loss in erasable programmable read only memories: Measurements and modeling," J. Appl. Phys. **77** (1995) 4522 (DOI: 10.1063/1.359414).
[13] S. M. Amoroso, *et al.*: "Impact of statistical variability and 3D electrostatics on post-cycling anomalous charge loss in nanoscale Flash memories," Proc. IEEE IRPS (2013) 3B.4.1 (DOI: 10.1109/IRPS.2013.6531980).
[14] L. Vandelli, *et al.*: "A physical model of the temperature dependence of the current through SiO_2/HfO_2 stacks," IEEE Trans. Electron Devices **58** (2011) 2878 (DOI: 10.1109/TED.2011.2158825).
[15] P. J. McWhorter, *et al.*: "Modeling the memory retention characteristics of silicon-nitride-oxide-silicon nonvolatile transistors in a varying thermal environment," J. Appl. Phys. **68** (1990) 1902 (DOI: 10.1063/1.346580).
[16] S. M. Sze: *Physics of Semiconductor Devices* (Wiley, New York, 2007) 3rd ed. 228.

1 Introduction

As the process node scales down, the reliability issues of floating gate nonvolatile memories become severe, such as endurance and data retention. The program/erase cycling will generate oxide traps, which are usually partly occupied by electrons or holes, then the trapped charge can affect the device's tunneling current and threshold voltage, and finally degrade the memory's endurance capability. Furthermore, the trapped charge and the floating gate charge together determine the memory cell's threshold voltage, thus electron detrapping from oxide bulk traps would also affect the cell's threshold voltage [1], even though the floating gate charge keeps unchanged. Therefore, to accurately model the memory cell's endurance and retention characteristics, the filling state of all the bulk traps must be obtained, which requires a comprehensive electron trapping-detrapping (T-D) model. In addition, the electron T-D phenomenon has also great impact on the device's bias temperature instability (BTI) [2, 3] and random telegraph noise (RTN) [4, 5] as the size scales.

The classical method of simulating electron T-D is based on the tunneling front (TF) model [6], as shown in Fig. 1. This model assumes that electron emission from bulk traps only by tunneling, the tunneling front $x_m(t)$ gradually transits from the interface to the inside of the oxide, and $x_m(t)$ is proportional to the logarithm of time. The TF model is a semi-empirical model, and it does not consider the cases of electron detrapping through thermal emission and electron capture by traps, thereby it cannot accurately predict the filling state of the bulk traps. R. Yamada

DOI: 10.1587/elex.14.20170565
Received June 1, 2017
Accepted July 18, 2017
Publicized August 2, 2017
Copyedited August 25, 2017

et al. extended the TF model by taking detrapping through thermal emission into account [7], which is believed to be the main way for shallow level traps. In recent years, there are many researches on the T-D model, but most of them are based on the statistical empirical equations [2, 3, 8], thus cannot give the filling state of bulk traps. D. Garetto *et al.* proposed a T-D model based on multi-phonon trap-assisted-tunneling (TAT) mechanism [9], but its calculation is too complicated since it needs to solve the Possion–Schroedinger equations. Besides, it does not include the detrapping way of thermal emission.

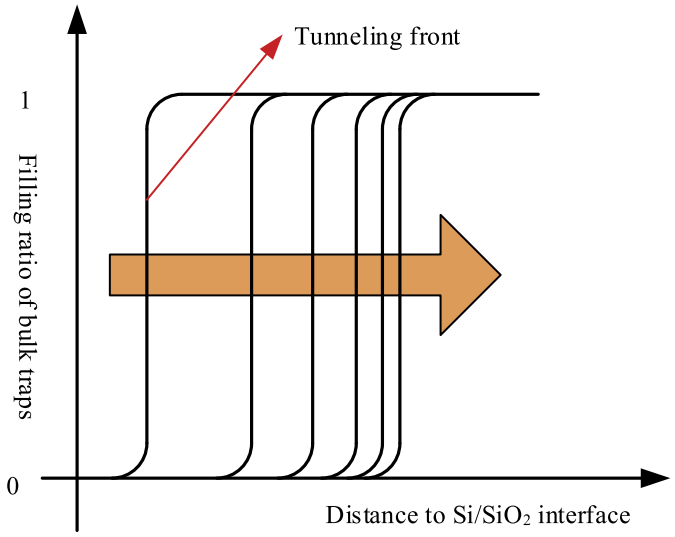


Fig. 1. Schematic diagram of TF model

In this paper, a comprehensive dynamic electron T-D model is proposed based on multi-phonon TAT and thermal emission, the new model calculates for every single bulk trap according to the trap's capture time and emission time, and finally establish the dynamic process of traps capturing and emitting electron under any given stress condition.

2 Proposed electron T-D model

The schematic diagram of the proposed electron T-D model is shown in Fig. 2, which mainly consists of three processes: 1) trap captures electron through tunneling, 2) trapped electron escapes from trap through tunneling, and 3) trapped electron passes through the potential barrier by thermal emission. The first two processes can be described by multi-phonon TAT mechanism in 2.1, and the last one can be modeled by the method introduced in 2.2.

2.1 Inelastic multi-phonon TAT

We adopt inelastic multi-phonon TAT mechanism [4, 10, 11, 12, 13] to model the processes of capture and emission of electrons. Due to the coupling of electron with oxide phonons, a series of virtual trap energy levels are generated in the oxide band gap, thereby enlarges the tunneling probability. The time constants of capture and emission of electron by and from the j^{th} trap, denoted as $\tau_{c,j,TAT}$ and $\tau_{e,j,TAT}$ respectively, can be expressed as:

DOI: 10.1587/elex.14.20170565
Received June 1, 2017
Accepted July 18, 2017
Publicized August 2, 2017
Copyedited August 25, 2017

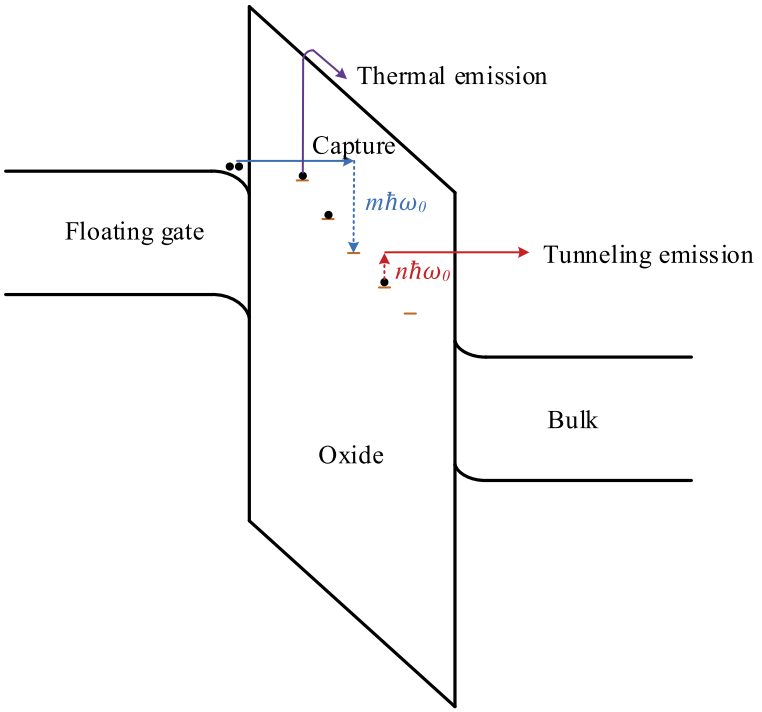


Fig. 2. Schematic diagram of the proposed electron T-D model

$$\tau_{c,j,TAT}{}^{-1} = \sum_{m} (\tau_{c,j,m})^{-1} \tag{1}$$

$$\tau_{e,j,TAT}{}^{-1} = \sum_{n} (\tau_{e,j,n})^{-1} \tag{2}$$

where $\tau_{c,j,m}$ ($\tau_{e,j,n}$) is the time required for the electron capture (emission) by (from) the j^{th} trap while releasing (absorbing) m (n) phonons as shown in Fig. 2, and can be expressed as [10, 12]:

$$(\tau_{c,j,m})^{-1} = N_{j-1}(E_{j,m}) \cdot f_{j-1}(E_{j,m}) \cdot P_{T,j-1,j}(E_{j,m}) \cdot Ca_{j,m} \tag{3}$$

$$(\tau_{e,j,n})^{-1} = N_{j+1}(E_{j,n}) \cdot (1 - f_{j+1}(E_{j,n})) \cdot P_{T,j,j+1}(E_{j,n}) \cdot Em_{j,n} \tag{4}$$

here $E_{j,k}$ is the generated virtual energy levels, N_j is the density of states, f_j is the Fermi–Dirac occupation probability, $P_{T,i,j}(E)$ is the tunneling probability of electron from the i^{th} trap to the j^{th} trap, $Ca_{j,m}$ and $Em_{j,n}$ are the trap capture and emission rates respectively.

$$Ca_{j,m} = c_{0,j} L(m) \tag{5}$$

$$Em_{j,n} = c_{0,j} L(n) exp\left(\frac{-n\hbar\omega_0}{kT}\right) \tag{6}$$

here $c_{0,j} = \frac{(4\pi)^2 r_{T,j}{}^3}{\hbar E_{g,ox}} (\hbar\Theta_0)^3$, $r_{T,j}$ is the *j-th* trap's capture radius, $r_{T,j} = \sqrt{\sigma_{T,j}/\pi}$, $\sigma_{T,j}$ is the trap's cross section, and $E_{g,ox}$ is the oxide bandgap. Besides, $\hbar\Theta_0$ is the oxide's electro-optical energy, $\hbar\Theta_0 = \left(\frac{q^2\hbar^2F^2}{2m_{ox}}\right)^{1/3}$, F is the effective electric field.

The multi-phonon transition probability $L(m)$ can be given by:

$$L(m) = \left(\frac{f_B + 1}{f_B}\right)^{m/2} exp[-S(2f_B + 1)] I_m(2S\sqrt{f_B(f_B + 1)}) \tag{7}$$

where S is the Huang-Rhys factor [14], f_B is the Bose function, $f_B = \frac{1}{exp(\hbar\omega_0/kT)-1}$, $I_m(\xi)$ is the modified Bessel function of order m.

DOI: 10.1587/elex.14.20170565
Received June 1, 2017
Accepted July 18, 2017
Publicized August 2, 2017
Copyedited August 25, 2017

2.2 Thermal emission

The time constant of electron detrapping from the j-th trap through thermal emission, $\tau_{e,j,TE}$, has been derived by P. J. McWhorter *et al.* based on Shockey–Read–Hall (SRH) statistical method [15], and is given by

$$\tau_{e,j,TE} = (AT^2)^{-1} exp\left(\frac{E_t}{kT}\right) \tag{8}$$

here $A = \frac{4\sqrt{6}\pi^{\frac{3}{2}}\sigma_t k^2 m_{ox}}{h^3}$, m_{ox} is the effective electron mass in oxide, E_t is the energy depth of the j-th trap, σ_t is the trap's cross section.

Considering the potential barrier lowering effect by the image force and the immobility of the positive charge [16], we have the barrier reduction of $\Delta\phi = \sqrt{q|F_{ox}|/\pi\varepsilon_{ox}}$, where q is the elementary charge, ε_{ox} is the permittivity in oxide. Thereby we have

$$\tau_{e,j,TE} = (AT^2)^{-1} exp\left(\frac{E_t - q \cdot \Delta\phi}{kT}\right) \tag{9}$$

Finally, for the j-th trap, the time constant of it capturing an electron $\tau_{c,j}$, and that of it emitting an electron $\tau_{e,j}$, can be described as

$$\tau_{c,j} = \tau_{c,j,TAT} \tag{10}$$

$$\tau_{e,j} = (\tau_{e,j,TAT}{}^{-1} + \tau_{e,j,TE}{}^{-1})^{-1} \tag{11}$$

2.3 Simulation procedure

To model the electron T-D process of all the bulk traps in the oxide, we should first create the following arrays: 1) *StateArray*, to record the state information of every trap; 2) *TauCaArray* and *TauEmArray*, to record the time constants $\tau_{c,j}$ and $\tau_{e,j}$ of each trap respectively; 3) *StateChangeArray*, to record when will each trap changes its state. The simulation procedure of the proposed electron T-D model is shown in Fig. 3, and the details are given below.

1) Input the simulation conditions, including the device's structural information, the trap parameters and the bias condition.

2) Generate the bulk traps according to the trap parameters, note that the number of traps N_{ot} follows the Poisson distribution, the position coordinates obey the uniform distribution, and the energy level E_t obeys Gauss distribution.

3) Initialize the bulk traps' *StateArray*.

4) Calculate the electric field in the oxide F_{ox}.

5) According to eq. (10) and eq. (11), calculate $\tau_{c,j}$ and $\tau_{e,j}$ for each bulk trap, and then create the time constant arrays *TauCaArray* and *TauEmArray*.

6) Combining *StateArray*, *TauCaArray* and *TauEmArray*, we may simply deduce the *StateChangeArray*.

7) Find the minimum value of *StateChangeArray* as *minValue* and its corresponding trap as *trapID*, which means that trap *trapID* will change its state at time *minValue*, and the corresponding value in *StateArray* should be updated.

8) According to the new *StateArray*, update the corresponding value in *StateChangeArray*.

9) Repeat step 7) to 8) until the simulation time reaches the preset value. If the change of bias exceeds 10%, return to step 4).

DOI: 10.1587/elex.14.20170565
Received June 1, 2017
Accepted July 18, 2017
Publicized August 2, 2017
Copyedited August 25, 2017

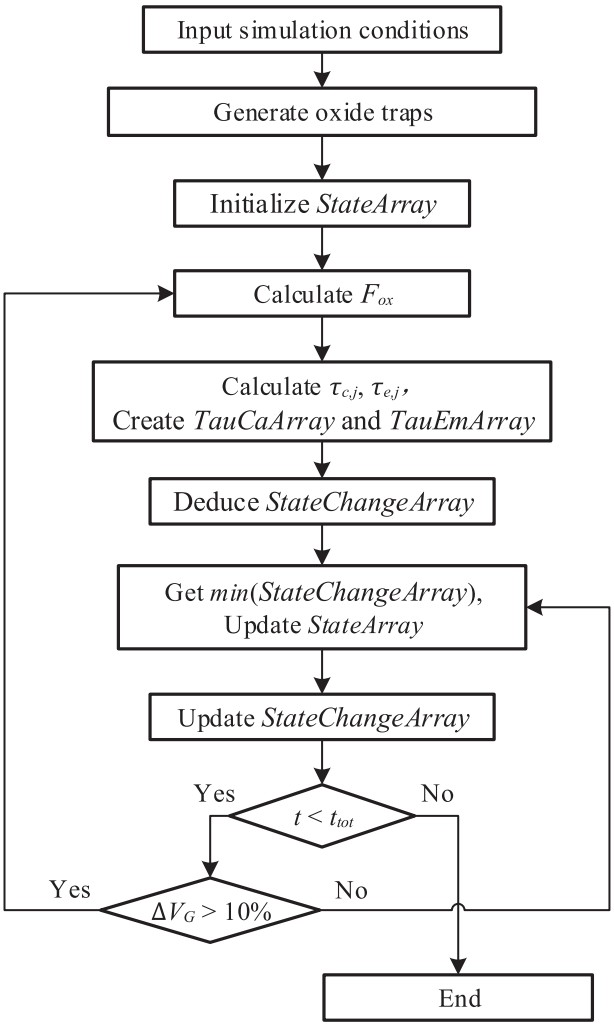


Fig. 3. The simulation flowchart of the proposed electron T-D model

3 Experimental results and discussion

Table I. Trap parameters of different devices. These parameters are obtained by fitting the TAT model with the measured stress induced leakage current (SILC).

Device Type	E_t (eV)	ΔE_t (eV)	m_{ox}	S	$\hbar\omega_0$ (meV)
NMOS	2.55	0.1	$0.284m_0$	25	60
PMOS	2.5	0.1	$0.282m_0$	22	60

Fig. 4 gives the electron T-D simulation results of NMOS transistor under low stress (gate voltage $V_G = 2$ V). Fig. 4a presents the filling state of the bulk traps at the end of the simulation, here '0' represents empty, '1' means filled by an electron, and the horizontal coordinate z is the vertical distance between the trap and the interface of Si/SiO_2. Fig. 4b shows the change of oxide charge per unit area Q_{ot}', here $Q_{ot}' = \sum_N \frac{q}{A}\frac{t_{ox}-z_i}{t_{ox}}$, A is the oxide area, and t_{ox} is the oxide thickness. Fig. 4c and Fig. 4d show the changes of gate charge per unit area Q_G' and bulk charge per unit area Q_B' respectively. From the simulation results we can see that, electron detrapping is proportional to the logarithm of time, and this is fully consistent with

DOI: 10.1587/elex.14.20170565
Received June 1, 2017
Accepted July 18, 2017
Publicized August 2, 2017
Copyedited August 25, 2017

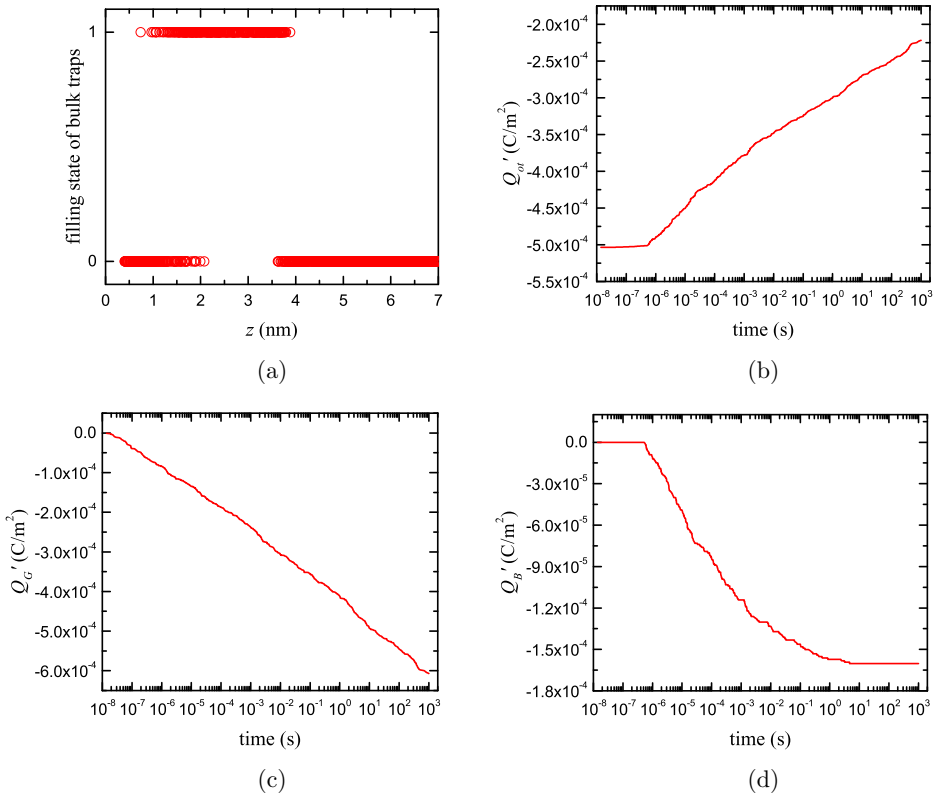


Fig. 4. The electron T-D simulation results of NMOS transistor under low stress. $W = L = 0.4\,\mu\text{m}$, $N_{ot} = 1 \times 10^{18}\,\text{cm}^{-3}$, the trap parameters are given in Table I, $V_G = 2\,\text{V}$, the simulation time is 1000 s, and all the bulk traps are assumed to be filled by electrons at the initial moment.

the measured results [1, 7]. Furthermore, both Q_G' and Q_B' decrease with time, which means that electrons detrap from bulk trap to both gate and bulk electrodes. Therefore, through this model, the electron trapping/detrapping behavior can be simulated accurately.

Fig. 5 gives the electron T-D simulation results of NMOS transistor under high stress ($V_G = 8\,\text{V}$). As we can see, the device enters equilibrium state at about 1 ms, then Q_{ot}' remains basically unchanged, electrons enter the oxide from the bulk and then go to the gate. The trap filling state under equilibrium state is shown in Fig. 5a, and the vertical position range of the filled traps is marked as [x_1 x_2]. Therefore, the filling state of the all the traps under any stress may be obtained by the proposed model, and this is essential for the modeling of the endurance and data retention characteristics of floating gate memories.

Fig. 6 shows the relationship between x_1, x_2 of different devices and stress, and Fig. 7 shows that between the time required to reach the equilibrium state t_s and stress. As we can see, x_1 and x_2 are nearly proportional to the logarithm of F_{ox}, and as the stress increases t_s decreases rapidly. It should be noted that, for negative biased PMOS transistor, its bulk traps are all empty in equilibrium state as shown in Fig. 6d, this is because $R_{c,j,TAT}$ is far less than $R_{e,j,TAT}$ in negative biased case, i.e., the rate of trapping electrons by bulk traps is much less than the rate of detrapping electrons.

DOI: 10.1587/elex.14.20170565
Received June 1, 2017
Accepted July 18, 2017
Publicized August 2, 2017
Copyedited August 25, 2017

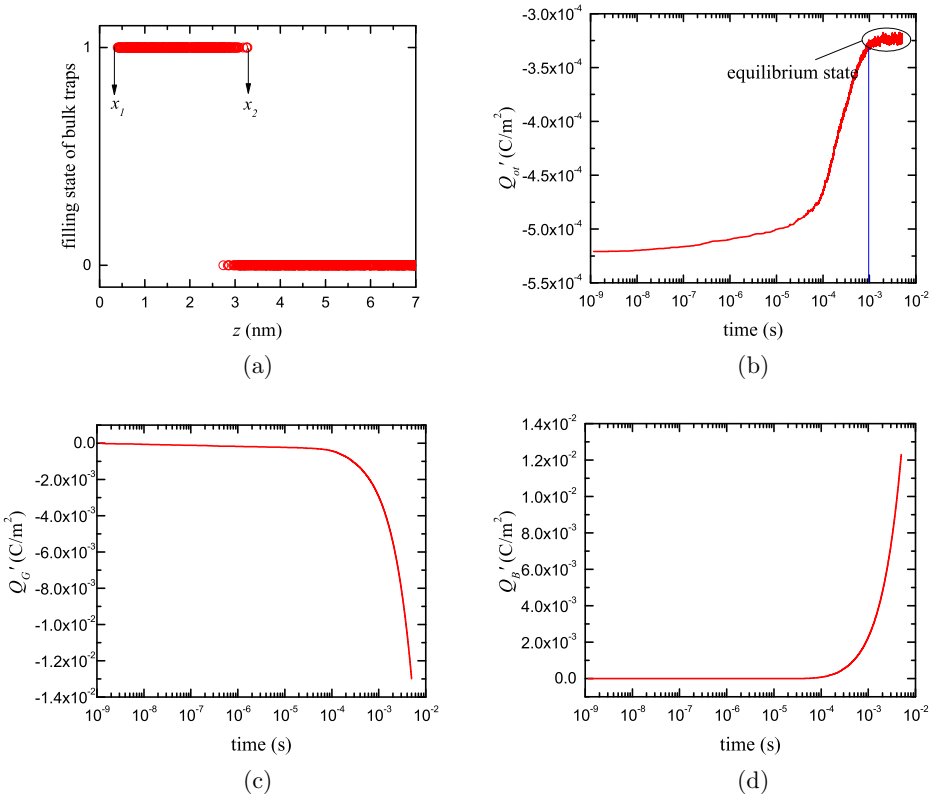
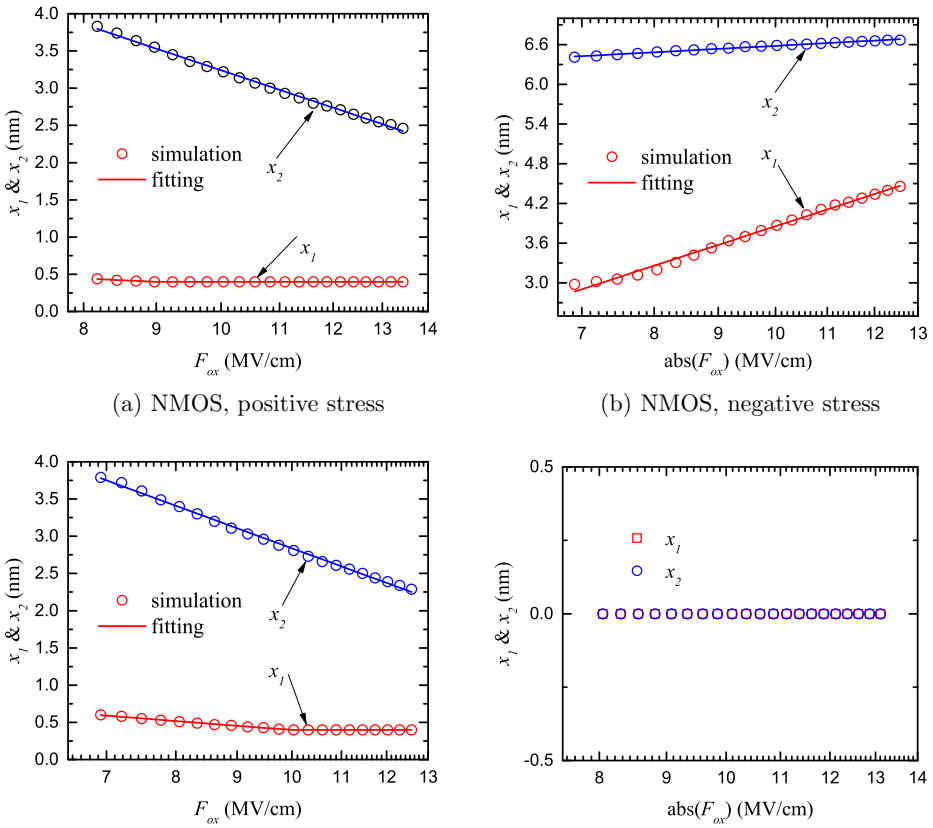


Fig. 5. The electron T-D simulation results of NMOS transistor under high stress. $W = L = 0.4\,\mu m$, $N_{ot} = 1 \times 10^{18}\,cm^{-3}$, the trap parameters are given in Table I, $V_G = 8\,V$, the simulation time is 5 ms, and all the bulk traps are assumed to be filled by electrons at the initial moment.

DOI: 10.1587/elex.14.20170565
Received June 1, 2017
Accepted July 18, 2017
Publicized August 2, 2017
Copyedited August 25, 2017



(a) NMOS, positive stress

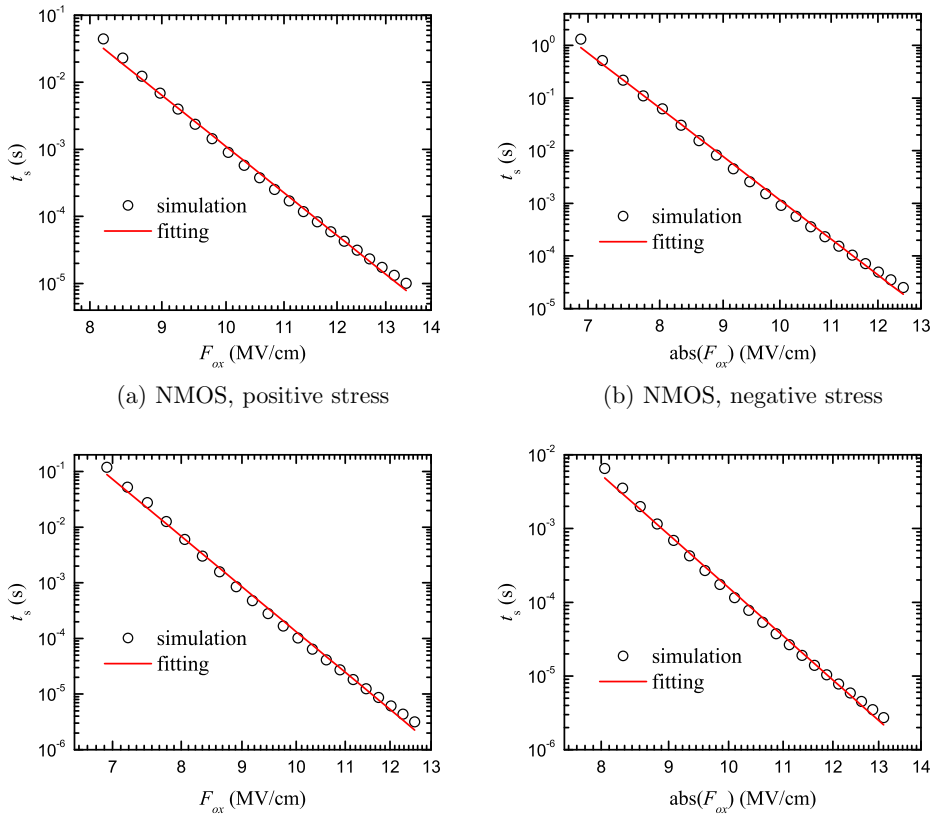
(b) NMOS, negative stress

(c) PMOS, positive stress

(d) PMOS, negative stress

Fig. 6. The relationship between x_1, x_2 of different devices and stress, the trap parameters are given in Table I.

DOI: 10.1587/elex.14.20170565
Received June 1, 2017
Accepted July 18, 2017
Publicized August 2, 2017
Copyedited August 25, 2017



(a) NMOS, positive stress (b) NMOS, negative stress

(c) PMOS, positive stress (d) PMOS, negative stress

Fig. 7. The relationship between t_s of different devices and stress, the trap parameters are given in Table I.

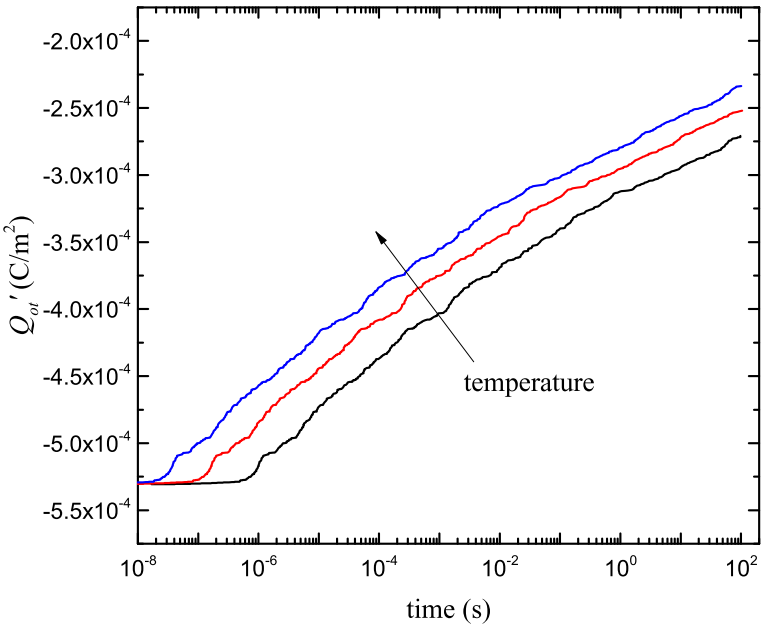


Fig. 8. The trends of electron detrapping with temperature. NMOS, $W = L = 0.4\,\mu m$, $N_{ot} = 1 \times 10^{18}\,cm^{-3}$, the trap parameters are given in Table I, $V_G = 2\,V$, all the bulk traps are assumed to be filled by electrons at the initial moment, the simulation temperatures are 25 °C, 85 °C and 150 °C respectively.

DOI: 10.1587/elex.14.20170565
Received June 1, 2017
Accepted July 18, 2017
Publicized August 2, 2017
Copyedited August 25, 2017

Fig. 8 also gives the trends of electron detrapping with temperature. The higher the temperature, the faster the electron detrapping, since the high temperature accelerates the tunneling and thermal emission mechanisms shown in Fig. 2. Furthermore, the proposed model is calculated much more effectively than the model in [9], since it does not need to solve the Possion–Schroedinger equations.

4 Conclusion

In this paper a physical comprehensive model of electron trapping/detrapping in electrically stressed oxide is presented. The proposed model takes both inelastic multi-phonon trap-assisted tunneling and thermal emission into account for the first time, and thanks to the new dynamic simulation framework which calculates for every single bulk trap separately, we may precisely simulate the dynamic procedure of traps capture and emission of electrons. Furthermore, the new model is computed effectively. Simulations under both conditions of low stress and high stress are proceeded to validate the proposed model, and the simulation results are fully consistent with the measured results and other empirical equations. In conclusion, the new model enables us to accurately and effectively obtain the filling state of all the oxide traps at any given stress and any time, which would be very helpful for the modeling of the reliability issues of nonvolatile memories and other semiconductor devices.

Acknowledgments

This work was supported by the China Postdoctoral Science Foundation funded project under Grant 2015M572794.

DOI: 10.1587/elex.14.20170565
Received June 1, 2017
Accepted July 18, 2017
Publicized August 2, 2017
Copyedited August 25, 2017