Modeling of trap-assisted tunneling on performance of charge trapping memory with consideration of trap position and energy level*

Zhi-Yuan Lun(伦志远)[1], Yun Li(李云)[1], Kai Zhao(赵凯)[2], Gang Du(杜刚)[1,†], Xiao-Yan Liu(刘晓彦)[1], and Yi Wang(王漪)[1]

[1] *Institute of Microelectronics, Peking University, Beijing* 100871, *China*

[2] *School of Information and Communication, Beijing Information Science and Technology University, Beijing* 100101, *China*

(Received 5 March 2016; revised manuscript received 12 April 2016; published online 25 June 2016)

In this work, the trap-assisted tunneling (TAT) mechanism is modeled as a two-step physical process for charge trapping memory (CTM). The influence of the TAT mechanism on CTM performance is investigated in consideration of various trap positions and energy levels. For the simulated CTM structure, simulation results indicate that the positions of oxide traps related to the maximum TAT current contribution shift towards the substrate interface and charge storage layer interface during time evolutions in programming and retention operations, respectively. Lower programming voltage and retention operations under higher temperature are found to be more sensitive to tunneling oxide degradation.

Keywords: trap assisted tunneling, charge trapping memory, tunneling oxide degradation

PACS: 85.30.De, 85.30.–z **DOI:** 10.1088/1674-1056/25/8/088502

1. Introduction

In the past few years we have seen the explosive growth of the non-volatile semiconductor memory market because of the unceasing demand for higher memory density and lower bit cost.[1] During the aggressive scaling of flash memory technology, the floating gate structure has faced great challenges from both technological and physical limitations, such as reducing coupling ratio, cell-to-cell interference, and increasing gate leakage current.[2] In order to continue the scaling trend, great effort has been devoted to developing next-generation memory structures.[3–6] Among these structures, the charge trapping memory (CTM) shows the most excellent potential as one of the most promising candidates for future memory solutions.[7–9] Furthermore, in recent years, high-density CTM in three-dimensional (3D) architectures for the NAND applications has drawn a great deal of attention.[10–12]

Under this circumstance, understanding the physical mechanisms in CTM has growing importance in terms of designing and optimizing CTM devices. The trap-assisted tunneling (TAT) is one of the most important mechanisms influencing the memory performance. On the one hand, TAT is an indispensable part of the tunneling current due to the existence of intrinsic defects in oxide layers. An accurate calculation of TAT current is crucial for achieving high accuracy in modeling CTM performance. On the other hand, it is largely accepted that electrical stress in the tunneling oxide during program/erase (P/E) cycle leads to the generation of interface defects and bulk traps in oxide layers.[13,14] In this regard, the TAT mechanism is related to reliability issues of the memory cells after P/E cycle, leading to a variation of memory characteristics. Especially, degradation in tunneling oxide can cause the increase in charge loss during retention operation. However, previous studies on modeling the TAT mechanism mainly focused on the gate leakage current of the MOS structure.[15–17] Only little work has addressed the TAT effect in the context of flash memory cells.[18,19]

In this paper, the TAT mechanism in CTM cells is carefully modeled and implemented in the developed simulator[20–22] for NAND flash memory. The influence of the TAT mechanism on CTM performance is investigated in consideration of various oxide trap positions and energy levels. The programming and retention performances of CTM after P/E cycle are also evaluated.

2. Physical model and simulation approach

The device structure we used to evaluate the influence of TAT on CTM performance is a standard CTM cell with 4.5-nm/6-nm/10-nm ($SiO_2/Si_3N_4/Al_2O_3$) gate stacks as shown in Fig. 1.

The TAT mechanism is modeled as a two-step process,[23] i.e., the electron capture and emission of the oxide traps. The TAT current driven through each oxide trap is given by the capture and emission times for each trap, $\tau_{c,ch/csl}$ and $\tau_{e/csl}$ respectively, considered from the charge reservoir, i.e., channel or charge storage layer as schematically illustrated in Fig. 2.

*Project supported by the National Natural Science Foundation of China (Grant Nos. 61404005, 61421005, and 91434201).

†Corresponding author. E-mail: gangdu@pku.edu.cn

© 2016 Chinese Physical Society and IOP Publishing Ltd

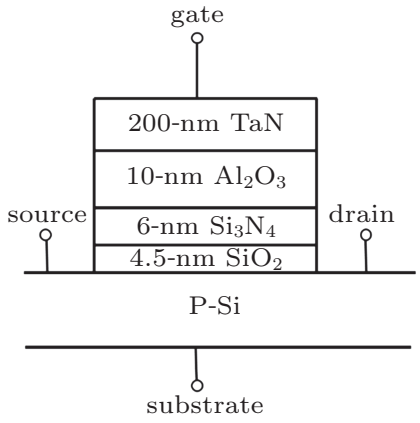


Fig. 1. Schematic illustration of the simulated standard CTM cell with a 4.5/6/10 nm ($SiO_2/Si_3N_4/Al_2O_3$) gate stack.

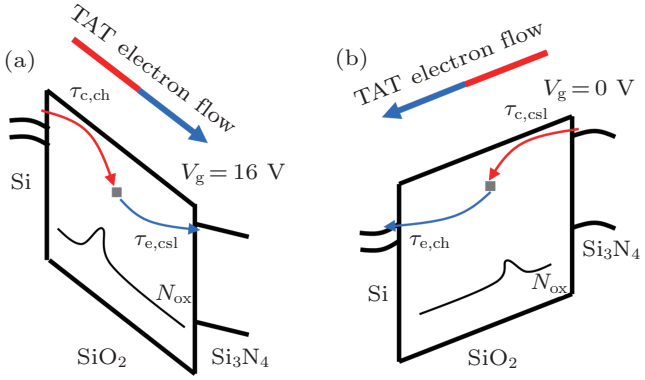


Fig. 2. (color online) Schematics of the capture and emission time constants for calculating TAT current in (a) programming and (b) retention operations.

Accumulating the current-density contributions from the traps in a small interval dx at each distance x, the total TAT current density through the entire oxide layer is expressed as

$$J_{\mathrm{TAT}} = q\int_0^{t_{\mathrm{ox}}} \frac{N_{\mathrm{ox}}(x)}{\tau_{\mathrm{c}}(x)+\tau_{\mathrm{e}}(x)}\,\mathrm{d}x, \tag{1}$$

where t_{ox} is the thickness of oxide layer and N_{ox} is the volume density of oxide traps with respect to distance in tunneling oxide. Solving device electrostatics and current continuity equations in our CTM memory simulator allows the calculation of these time constants, which is based on an SRH-like model[24] with considering the spatial and energetic locations of the oxide traps and the charge reservoir. Here we calculate the time constants for a single acceptor-like oxide trap and assume that it only interacts with the conduction band of the channel and the charge storage layer. The capture and emission time constants are described by the following equations:

$$\tau_{\mathrm{c}}(x)^{-1} = \int_{E_{\mathrm{c}}}^{\infty} c_n(E,x) f_n(E) g_n(E)\,\mathrm{d}E, \tag{2}$$

$$\tau_{\mathrm{e}}(x) = \int_{E_{\mathrm{c}}}^{\infty} e_n(E,x)\,[1-f_n(E)]\,g_n(E)\,\mathrm{d}E, \tag{3}$$

where f_n is the electron energy distribution function in the reservoir and g_n is the density of states. The processes of electron capture and emission are described by energy-dependent capture and emission coefficients c_n and e_n. With regard to the distribution function, the Fermi–Dirac distribution is assumed, and has the useful property,

$$\frac{f_n(E)}{1-f_n(E)} = \mathrm{e}^{-\beta(E-E_{\mathrm{F}})}, \tag{4}$$

where

$$\beta = (k_{\mathrm{b}}T)^{-1}.$$

In addition, according to the SRH theory,[25] the correlation between capture and emission coefficients is

$$\frac{e_n(E)}{c_n(E)} = \mathrm{e}^{-\beta(E-E_{\mathrm{T}})}, \tag{5}$$

with E_{T} being the energy level of the oxide traps.

It should be mentioned that the calculations of capture and emission coefficients are related to the energetic position of E_{T} with respect to the energy level E in the integration. When E_{T} is lower than E, the capture coefficient is energy-independent and can be written as

$$c_n = \upsilon_{\mathrm{th}}\sigma_0\theta, \tag{6}$$

where υ_{th} is the thermal velocity of electron, σ_0 is the capture cross section, and θ is the elastic tunneling coefficient between the reservoir and the oxide trap calculated in the WKB approximation. When the E_{T} is higher than E, the emission coefficient rather than the capture coefficient should be energy-independent.

In order to improve calculation efficiency, an approximation is made to solve the above complicated integral by assuming $E = E_{\mathrm{c}}$ for calculating the time constants.[24] Overall, using Eqs. (1)–(6), we can rewrite the capture and emission time constants as follows:

$$\tau_{\mathrm{c}}^{-1} = n\upsilon_{\mathrm{th}}\sigma_0\theta\,\mathrm{e}^{-\beta\varepsilon_{\mathrm{c}}}, \tag{7}$$

$$\tau_{\mathrm{e}}^{-1} = N_C\upsilon_{\mathrm{th}}\sigma_0\theta\,\mathrm{e}^{-\beta\varepsilon_{\mathrm{e}}}, \tag{8}$$

where N_{c} is the density of states in the conduction band and ε_{c} and ε_{e} are the capture and emission barriers which are determined by

$$\varepsilon_{\mathrm{c}} = \max(0, E_{\mathrm{T}} - E_{\mathrm{C}}), \text{ and } \varepsilon_{\mathrm{e}} = \max(0, E_{\mathrm{C}} - E_{\mathrm{T}}). \tag{9}$$

The abovementioned physical model accounting for the TAT mechanism has been incorporated into our calibrated two-dimensional (2D) self-consistent CTM simulator,[21] which carefully takes into consideration the DT/FN tunneling process, charge trapped/de-trapping mechanisms, and drift-diffusion transport in the gate stack of memory cells. In addition, oxide traps with arbitrary energy level and spatial distribution are considered in the simulation. The main parameters adopted in the simulation are summarized in Table 1.[26,27]

Table 1. Main parameters for the simulation.

Parameter	Symbol	Value
SiO_2 effective mass	m_{ox}	$0.42m_0$
SiO_2 trap cross-section	σ_n	5×10^{-14} cm^2
SiO_2 trap energy	E_T	2.0 eV
Si_3N_4effective mass	m_{SiN}	$0.5m_0$
Si_3N_4 trap energy	$E_{T,SiN}$	1.5 eV

3. Results and discussion

In order to obtain an overview of the influences of the TAT mechanism on programming and retention characteristics of CTM cells, we first perform the simulations in the cases with and without considering the TAT effect. In addition to the TAT mechanism, in the simulation the direct and FN tunneling process within the tunneling oxide are taken into consideration. Comparisons between simulation prediction and experimental measurement are also made. Figure 3 shows a good reproduction of experimental data[27] for all programming voltages considered. A uniform spatial distribution of traps is assumed for the charge storage layer, under a trap density of $N_t = 7 \times 10^{19}$ cm^{-3} and a trap cross section $\sigma_n = 1 \times 10^{-14}$ cm^2. The trap density is an important factor in determining the memory window,[28] and hence is obtained from fitting the experimental characteristics. For the simulation of TAT contribution, we consider typical parameters for oxide defects in accordance with those reported previously[29] for thermally grown oxides. The intrinsic trap distribution across the tunneling oxide is assumed to be uniform under an oxide defect density of $N_{ox} = 3 \times 10^{16}$ cm^{-3}, a trap energy level around 2 eV, and a trap cross section of $\sigma_T = 8 \times 10^{-14}$ cm^2. For the simulated memory cell, the interface traps located at tunneling oxide borders are not included in the simulation. It can be seen that without considering the TAT current, the programming speed is underestimated and simulations and measurements are not in agreement with each other. In addition, the role played by TAT current is found to be non-negligible at lower oxide fields: lower programming voltages, and larger occupation of the traps in the charge storage layer (longer programming time). Using the same material parameters for the charge storage layer and slightly modified parameters for intrinsic oxide defects ($N_{ox} = 2 \times 10^{16}$ cm^{-3} and $\sigma_T = 5 \times 10^{-14}$ cm^2), good agreement between experimental retention measurements[30] and simulations under a wide variety of temperatures can be achieved as shown in Fig. 4. The result indicates that modeling the TAT mechanism is also crucial in predicting retention characteristics, which can lead to extra charge loss from the charge storage layer.

The defect traps allow an extra tunneling path in the oxides, but not all traps have equivalent contribution to TAT current. From Eq. (1), it is known that capture and emission time constants are used to calculate TAT current density of each oxide trap. Figure 5 shows the variations of capture and emission time constants and normalized TAT current contributions with position along tunneling oxide under 16-V programming voltage at 1 μs for the simulated device. It is found that only the oxide traps in a small distance (approximately 1 nm) at the position of 3.5 nm from the substrate interface make the main contribution to the overall TAT current. Similarly, figure 6 shows the variations of capture and emission time constants and normalized TAT contribution with position along tunneling oxide for 375 K retention simulation at 10^7 s. Oxide traps located within a length of 0.2 nm at the position of 3 nm from the substrate interface are related to major TAT current.

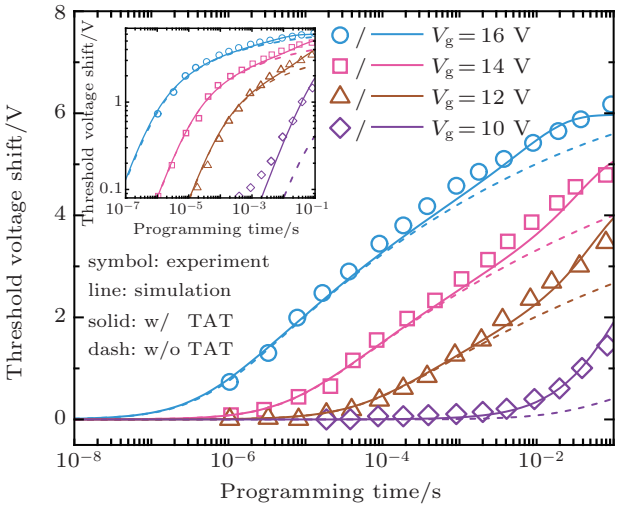


Fig. 3. (color online) Comparison between experimental programming data and simulation results of CTM cells and influence of TAT on programming characteristics.

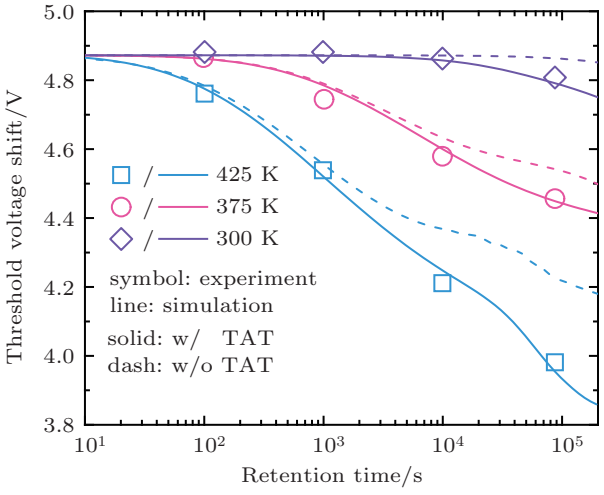


Fig. 4. (color online) Comparison between experimental retention and simulation predictions of CTM cells under different temperatures and influence of TAT on retention characteristics.

However, the positions of peak contribution to the overall TAT current are variable with time evolution under programming and retention operations as shown in Fig. 7. In the programming process, the shape of the curve for normalized TAT contribution becomes narrower with time evolution. Simultaneously, the position of the maximum contribution shifts

towards the substrate interface. In comparison, the shape of the curve keeps unchanged with the peak shifting towards the charge storage layer interface during retention. The variation in positions of the maximum contribution can be ascribed to the movement of band edge in different phases of the programming and retention operations.

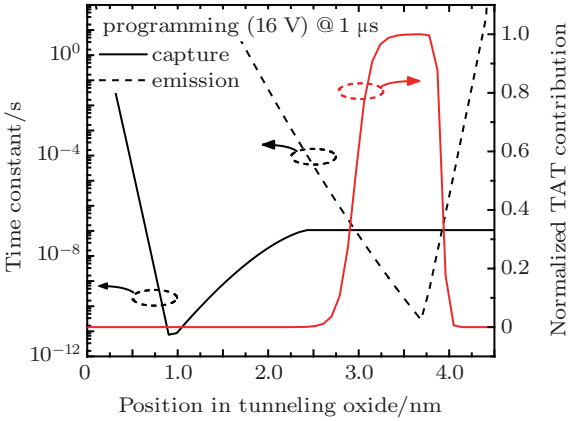


Fig. 5. (color online) Capture and emission time constants used to calculate TAT current density along tunneling oxide and the related normalized TAT current contribution under 16-V programming voltage at 1 μs.

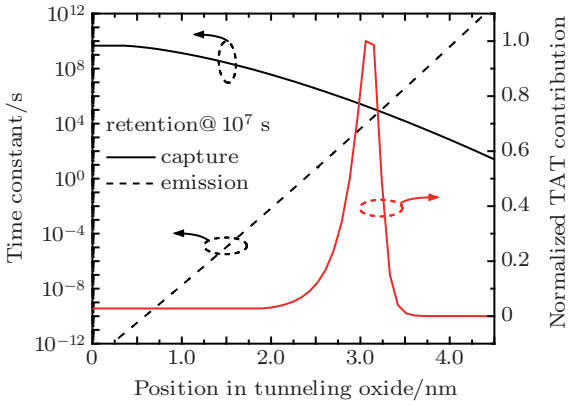


Fig. 6. (color online) Capture and emission time constants used to calculate TAT current density along tunneling oxide and the related normalized TAT current contribution for 375-K retention simulation at 10^7 s.

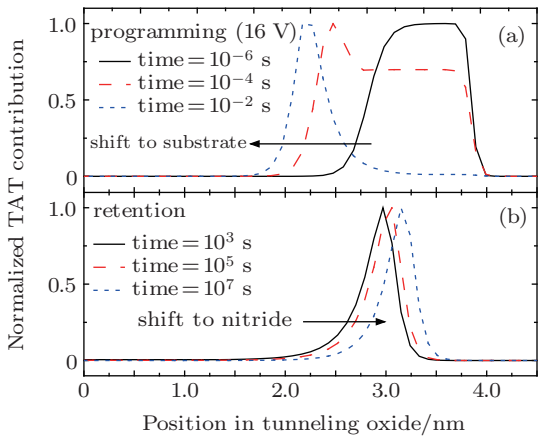


Fig. 7. (color online) The variations of normalized TAT contribution with position in tunneling oxide with time evolution in (a) programming and (b) retention operations.

Apart from the position of oxide defects, trap energy level is another major factor to determine TAT current. To investigate the influences of the trap energy level and spatial position on TAT current density, we assume that the trap distribution in tunneling oxide has a Gaussian distribution with a peak value of 7×10^{18} cm^{-3} and a small standard deviation of 0.2 nm. Simulations of devices with various peak positions and trap energy levels are performed. Figure 8 shows the TAT current density under 16-V programming voltage at 1 μs of a device with a 4.5/6/10 nm gate stack ($SiO_2/Si_3N_4/Al_2O_3$). The oxide traps located around 1 nm from the oxide/nitride interface with lower energy level are related to larger TAT current density. Figure 9 shows the TAT current densities of devices with various peak oxide trap positions and energy levels at 10^7 s of retention under 360 K temperature. It is shown that oxide traps located at 1.4 nm from the oxide/nitride interface are related to the maximum TAT current density, which has little dependence on the trap energy level ranging from 0.6 eV to 4.0 eV.

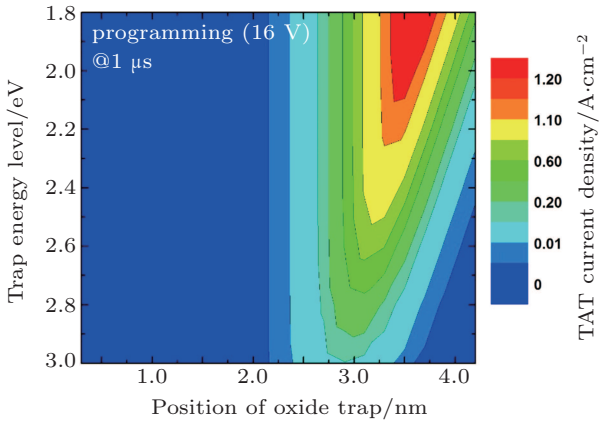


Fig. 8. (color online) TAT current densities under 16-V programming voltage at 1 μs of devices with various peak positions and trap energy levels.

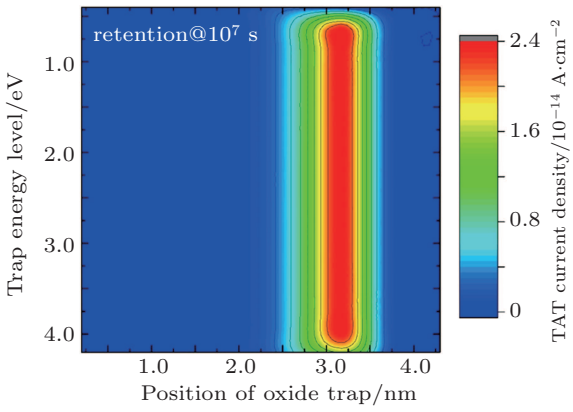


Fig. 9. (color online) TAT current densities at 10^7 s of retention operation under 360-K temperature of devices with various peak positions and trap energy levels.

During P/E operations, the electric field in tunneling oxide can typically exceed 10 MV/cm.[31] A Large number of P/E cycles leads to the generation of oxide traps and thus causes reliability issues of CTM cells. The influences of TAT on programming and retention operations are strongly correlated with trap density in tunneling oxide. Figure 10 shows the programming speeds under different programming voltages with increasing the level of tunneling oxide degradation.

A unanimous programming window of 5 V is used for all cases. It can be seen that programming under lower voltage is more sensitive to oxide degradation. Figure 11 shows the threshold voltage shift after 10^8 s under various retention temperatures with increasing the level of tunneling oxide degradation. Higher temperature is found to be more sensitive to oxide degradation.

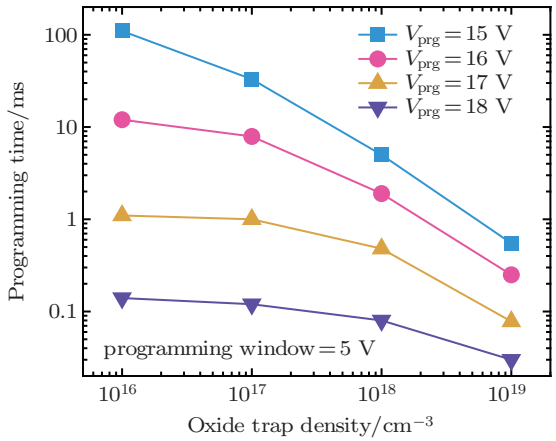


Fig. 10. (color online) Programming speeds under different programming voltages with increasing the level of tunneling oxide degradation.

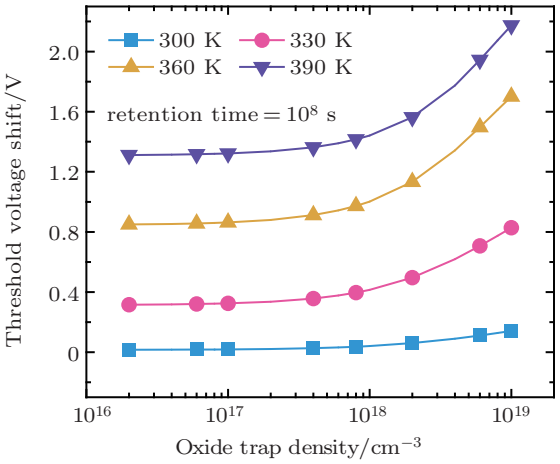


Fig. 11. (color online) Threshold voltage shifts after 10^8 s under various retention temperatures with increasing the levels of tunneling oxide degradation.

4. Conclusions

In this paper, we carefully model TAT mechanism in CTM cells and evaluate the influences of TAT on programming and retention characteristics. Our results indicate that the TAT mechanism is a critical factor for CTM operations, especially at low tunneling oxide fields. The influence of the TAT mechanism on CTM performance is investigated in consideration of various trap positions and energy levels in the tunneling oxide.

The results show that only the oxide traps at a small distance are related to a major contribution to TAT current. In addition, the positions corresponding to the maximum contribution shift towards the substrate interface and the charge storage layer interface during programming and retention operations, respectively. Lower programming voltage and retention operations under higher temperature are more sensitive to tunneling oxide degradation.

References

[1] White M H, Wang Y, Wrazien S J and Zhao Y 2006 *Int. J. High Speed Electron. Syst.* **16** 479
[2] Ielmini D, Spinelli A S, Lacaita A L and Modelli A 2001 *Microelectron. Eng.* **59** 189
[3] Meena J S, Sze S M, Chand U and Tseng T Y 2014 *Nanoscale Res. Lett.* **9** 1
[4] Hou Y, Celano U, Goux L, Liu L, Fantini A, Degraeve R, Youssef A, Xu Z, Cheng Y, Kang J, Jurczak M and Vandervorst W 2016 *Appl. Phys. Lett.* **108** 123106
[5] Chen J, Du G and Liu X Y 2015 *Chin. Phys. B* **24** 057702
[6] Fan X, Chen H P, Wang Q, Wang Y Q, Lv S L, Liu Y, Song Z T, Feng G M and Liu B 2015 *Chin. Phys. Lett.* **32** 068301
[7] Park Y, Choi J, Kang C, Lee C, Shin Y, Choi B, Kim J, Jeon S, Sel J, Park J, Choi K, Yoo T, Sim J and Kim K 2006 *Proceedings of International Electron Devices Meeting*, pp. 1–4
[8] Gilmer D C, Goel N, Park H, Park C, Verma S, Bersuker G, Lysaght P, Tseng H H, Kirsch P D, Saraswat K C and Jammy R 2009 *Proceedings of International Electron Devices Meeting*, pp. 1–4
[9] Liu L F, Pan L Y, Zhang Z G and Xu J 2015 *Chin. Phys. Lett.* **32** 088501
[10] Tanaka H, Kido M, Yahashi K, Oomura M, Katsumata R, Kito M, Fukuzumi Y, Sato M, Nagata Y, Matsuoka Y, Iwata Y, Aochi H and Nitayama A 2007 *Proceedings of Symposium on VLSI Technology*, pp. 14–15
[11] Jang J, Kim H S, Cho W, Cho H, Kim J, Shim S I, Jang Y, Jeong J H, Son B K, Kim D W, Kihyun, Shim J J, Lim J S, Kim K H, Yi S Y, Lim J Y, Chung D, Moon H C, Hwang S, Lee J W, Son Y H, Chung U and Lee W S 2009 *Proceedings of Symposium on VLSI Technology*, pp. 192–193
[12] Park K T, Byeon D S and Kim D H 2014 *Proceedings of the 14th Non-Volatile Memory Technology Symposium*, pp. 1–5
[13] Lun Z, Wang T, Zeng L, Zhao K, Liu X, Wang Y, Kang J and Du G 2013 *Proceedings of International Conference on Simulation of Semiconductor Processes and Devices*, pp. 292–295
[14] Fayrushin A, Lee C H, Park Y, Choi J H and Chung C 2013 *IEEE Trans. Electron Dev.* **60** 2031
[15] Larcher L 2003 *IEEE Trans. Electron Dev.* **50** 1246
[16] Vianello E, Driussi F, Esseni D, Selmi L, van Duuren M J and Widdershoven F 2006 *Proceeding of the 36th European Solid-State Device Research Conference*, pp. 403–406
[17] Zhang M, Huo Z, Yu Z, Liu J and Liu M 2011 *J. Appl. Phys.* **110** 114108
[18] Lee K, Kang M, Seo S, Kang D, Li D H, Hwang Y and Shin H 2014 *Electron Dev. Lett.* **35** 51
[19] Amoroso S M, Gerrer L, Adamu-Lema F, Markov S and Asenov A 2013 *Proceedings of International Reliability Physics Symposium*, pp. 3B.4.1–3B.4.6
[20] Peng Y, Liu X, Du G, Liu F, Jin R and Kang J 2012 *Chin. Phys. B* **21** 078501
[21] Lun Z, Du G, Zhao K, Liu X and Wang Y 2016 *Sci. China: Inf. Sci.* DOI: 10.1007s11432-015-5475-7
[22] Song Y, Liu X, Du G, Kang J and Han R 2008 *Chin. Phys. B* **17** 2678
[23] Gerrer L, Markov S, Amoroso S M, Adamu-Lema F and Asenov A 2012 *Microelectron. Reliab.* **52** 1918
[24] Grasser T 2012 *Microelectron. Reliab.* **52** 39
[25] Shockley W and Read W T 1952 *Phys. Rev.* **87** 835
[26] Vandelli L, Padovani A, Larcher L, Southwick R G, Knowlton W B and Bersuker G 2011 *IEEE Trans. Electron Dev.* **58** 2878
[27] Padovani A, Larcher L, Heh D and Bersuker G 2009 *Electron Dev. Lett.* **30** 882
[28] Seo Y J, Kim K C, Kim H D, Joo M S, An H M and Kim T G 2008 *Appl. Phys. Lett.* **93** 063508
[29] Padovani A, Larcher L, Verma S, Pavan P, Majhi P, Kapur P, Parat K, Bersuker G and Saraswat K 2008 *Proceedins of International Reliability Physics Symposium*, pp. 616–620
[30] Padovani A, Larcher L, Heh D, Bersuker G, Della M V and Pavan P 2010 *Appl. Phys. Lett.* **96** 223505
[31] Moon P, Lim J Y, Youn T U, Park S K and Yun I 2014 *Solid-State Electron.* **94** 51