



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

M.S. THESIS

Three dimensional simulation of retention
characteristics of Charge Trap (CT) NAND
Flash memory

(Charge Trap (CT) NAND Flash memory 의
리텐션 특성에 관한 3 차원 시뮬레이션)

By

Sangyong Park

February 2013

SCHOOL OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE
COLLEGE OF ENGINEERING
SEOUL NATIONAL UNIVERSITY

Three dimensional simulation of retention characteristics
of Charge Trap (CT) NAND Flash memory

(Charge Trap (CT) NAND Flash memory 의
리텐션 특성에 관한 3 차원 시뮬레이션)

지도교수 박 영 준

이 논문을 공학석사 학위논문으로 제출함

2013 年 2 月

서울대학교 대학원

전기-컴퓨터 공학부

박 상 용

박상용의 공학석사 학위논문을 인준함

2013 年 2 月

위 원 장 이 종 호 (인)

부위원장 박 영 준 (인)

위 원 박 병 국 (인)

Abstract

To investigate the retention characteristics of the CT (Charge Trap) Flash memory, the simulation study with various NAND Flash structures has been conducted using the three dimensional self-consistent simulator [1], which is the 3D simulator developed in house, considering the vertical and lateral charge transport in the conduction band of the nitride layer. All the simulations are based on the SANOS (Si /Al₂O₃/ Si₃N₄/ SiO₂/Si) device with the SA-STI (Self-Aligned Shallow Trench Isolation) structure. Prior to retention simulations, the programming simulation has been conducted under the temperature 300K by enforcing the 18V to the gate electrode to extract the initial charge density distribution of the retention situation. The programming is stopped when the threshold voltage shift is 4V, and the charge distribution is settled as the initial charge density distribution of the retention simulation.

First of all, to ascertain the dependency on the extended nitride length over the gate region, several structures (0nm, 15nm, 30nm, 150nm) are simulated. Since the emitted electrons cannot transport over the nitride region, the extended nitride length decides the saturation point of the charge loss. Because NAND Flash memory has an array structure, the charge transport to the extended nitride region could affect the neighbor cells. To demonstrate this phenomenon, two programmed patterns (check-board pattern and solid pattern) are simulated. In the check-board pattern, the threshold voltage of the eased cell has increased. The charge loss speed of the programmed cell, in the solid pattern, has decreased until the programmed electrons

spread and stop at the lateral edge. The retention cycle (Program- Retention-Erase) behavior has been also simulated. The charges which have been trapped in the extended region in the previous retention-term disturb the additional transport to the extended nitride region in the next retention mode. This causes a change in the slope of the retention curve.

The activation energy of the CT Flash memory has been also studied by simulating the various tunneling oxide thickness, considering charge transport along the vertical and lateral directions. And then, the various forms of the Arrhenius plot have been modeled by separating the temperature ranges into three regions. At high temperature (Region-1), the charge loss is mainly affected by thermal emission and electron transport, so that the activation energy is same as the the tunneling free case. The temperature region showing reduced activation energy is called Region-2. Since the impact of the tunneling current become larger, the Arrhenius plot under high temperature region is depressed. In the low temperature (Region-3), thermal energy is needed to move the electrons to the tunneling interface or outside of the gate region until the charge loss reaches the criterion. Thus, the activation slope is increased up to that of the device with thick tunneling oxide.

Consequently, in this thesis, the dependency of the extended nitride length, the programmed pattern and the retention cycle is confirmed as an important factor in analyzing the retention characteristics of the SANOS cells. Various forms of the Arrhenius plot are understood by the charge transport model.

Keywords: CT(Charges Trap) Flash memory, Retention, Charge loss, Lateral Spreading, Activation energy, 3D simulation.

Student Number: 2011-20837

Contents

Abstract.....	i
Contents	iv
List of Figures	vi
Chapter 1	1
Introduction.....	1
1.1 Motivation	1
1.2 CT NAND Flash memory	4
1.2.1 Basic Flash memory cell operation	5
1.2.2 CT (Charge Trap) Flash memory	7
1.2.3 3D Vertical NAND Flash memory	8
1.3 Issues of CT NAND Flash memory	13
1.3.1 Lateral charge spreading	13
1.3.2 Activation energy	13
1.4 Outline of thesis	14
Chapter 2	16
Physics and simulation structures.....	16
2.1 Physical models	16
2.2 Simulated structures	18
Chapter 3	20
Simulation of the charge transport effects.....	20
3.1 The initial charge distribution.....	20
3.2 The dependency of the extended nitride length	23
3.3 The dependency of and on the surround cells	29

3.3.1 Check-Board pattern.....	29
3.3.2 Solid pattern.....	34
3.4 The dependency of the retention cycle.....	38
3.4.1 Infinitely extended nitride structure.....	38
3.4.2 30nm extended nitride structure.....	42
3.5 Conclusion.....	42
Chapter 4	46
Simulation of the activation energy	46
4.1 Conventional expectation method of the data life time.....	46
4.2 Initial electron distributions.....	50
4.3 Modeling of the multiple activation energy	52
4.3.1 Criterion above the dent point (15% charge loss criterion)	54
4.3.2 Criterion below the dent point (20% charge loss criterion)	61
4.4 Conclusion.....	63
Chapter 5	69
Conclusions	69
5.1 Summary.....	69
5.2 Future work.....	70
5.2.1 Charge loss path through the blocking layer.....	70
5.2.2 Modeling for more accurate expectation of the life time.....	71
5.2.3 3D Monte-Carlo Simulation.....	71
References	74

List of Figures

Fig. 1-1 (a) The trend of the bits per chip (left : linear scale, right log scale), (b) The half pitch size of the NAND Flash and the step of the 3D vertical structure.	3
Fig. 1-2 (a) The first Flash memory structure (b) Floating Gate type Flash structure (c) CT type Flash structure.	4
Fig. 1-3 Array structures of (a) NOR Flash structure (b) NAND Flash structure.	5
Fig. 1-4 Band diagrams and schematics of FG NAND Flash memory operation in (a) Programming operation (b) Retention situation (c) Erasing operation.	9
Fig. 1-5 V_{th} shifts of NAND Flash memory after (a) Programming operation (b) Erasing operation.	10
Fig. 1-6 The concept of SLC and MLC (a) 1 bit/cell (SLC) (b) 2bit/cell (MLC) (c) 3bit/cell (MLC).	10
Fig. 1-7 TEM images of Floating Gate type cell and CT type cell [24].	11
Fig. 1-8 TEM images of 3D Vertical structures (a) TCAT structure [6], (b) BiCs structure [25].	12
Fig. 2-1 Three-Dimensional SANOS band diagrams and charge flux diagrams in retention mode.	17
Fig. 2-2. Simulated structures of the SA-STI SANOS device with (a) No extended nitride, (b) finitely extended nitride, (c) infinitely extended nitride, and (d) three gate string.	19
Fig. 3-1. The transient distribution of (a) the conduction electrons, (b) the trapped electrons during the program mode of the finitely-extended nitride structure at the middle of the channel.	21
Fig. 3-2. Transient trapped electron distributions of the infinitely-extended structures from 10^{-6} sec to 10^{-1} sec during the programing.	22
Fig. 3-3. Threshold voltage shifts of the devices. 6nm tunnel oxide structure is also simulated as reference device to evaluate the vertical charge loss.	23
Fig. 3-4. Transient conduction and trapped electron distributions along the vertical direction	

(a) no-extended, (b) finitely-extended (30nm), and (c) infinitely-extended structures.	25
Fig. 3-5. Transient trapped electron distributions along the channel direction at the blocking layer interface (Left) and the tunneling layer interface (Right) of the (a) no-extended, (b) finitely-extended (30nm), and (c) infinitely-extended structures.	26
Fig. 3-6. Transient conduction electron distributions of the infinitely-extended structures from 10^{-2} sec to 10^5 sec.	27
Fig. 3-7. Transient trapped electron distributions of the infinitely-extended structures from 10^{-2} sec to 10^5 sec.	28
Fig. 3-8. Check-board pattern structure and retention curves. Dashed line is the situation when Gate0 is programmed and Gate1 and Gate2 are erased.	30
Fig. 3-9. Transient trapped electron distributions along the channel direction at the blocking layer interface (Top) and the tunneling layer interface (Bottom) of the check-board pattern.	31
Fig. 3-10. Transient trapped electron distributions along the vertical direction at the middle of (a) the programmed Gate0 and (b) the erased Gate1.	32
Fig. 3-11. Transient trapped electron distributions of the check-board pattern from 10^{-2} sec to 10^5 sec.	33
Fig. 3-12 Solid pattern structure and retention curves. Dashed line is the situation when Gate0 is programmed and Gate1 and Gate2 are erased.	35
Fig. 3-13. Transient trapped electron distributions along the channel direction at the blocking layer interface (Top) and the tunneling layer interface (Bottom) of the check-board pattern.	36
Fig. 3-14. Transient trapped electron distributions of the check-board pattern from 10^{-2} sec to 10^5 sec.	37
Fig. 3-15 Delta threshold voltage shift curves of the three gates string structure as the cycle is repeating.	39
Fig. 3-16 Trapped electron density of the three gates string structure at the interface of the tunneling layer when the each retention cycle step is stopped.	40
Fig. 3-17 Top is retention curves of three gates string structure after the cycle is repeated. The dashed lines indicate the threshold voltage shift of the nearby cell. Bottom is the trapped electron density profiles after each retention step is finished.	41

Fig. 3-18 Delta threshold voltage shift curves of the 30nm extended nitride structure as the cycle is repeating.	43
Fig. 3-19 Trapped electron density of the 30nm extended nitride structure at the interface of the tunneling layer when the each retention cycle step is stopped.	44
Fig. 3-20 Top is retention curves of 30nm extended nitride structure after the cycle is repeated. Bottom is the trapped electron density profiles after each retention step is finished.	45
Fig. 4-1 Arrhenius plots of FG Flash memory which are extracted from reference [43][44][45].	48
Fig. 4-2 Arrhenius plots of CT Flash memory which are extracted from reference [47][48][49][50].	49
Fig. 4-3. (a) The initial trapped electron distributions along the vertical direction of the various tunnel oxide cases. (b) The tunneling current of the four cases during the bake.	51
Fig. 4-4 Simulated retention curves of the infinitely extended structure with (a) 4nm tunnel oxide, (b) 2.8nm tunnel oxide, (c) 2.6nm tunnel oxide and (d) 2.4nm tunnel oxide.	53
Fig. 4-5. Simulated Arrhenius plots when criterion is 15% shift of the threshold voltage.	56
Fig. 4-6. Transient trapped electron distributions along the vertical direction of the various tunnel oxide cases during the retention situation at 500K temperature.	57
Fig. 4-7. Transient trapped electron distributions along the channel direction during the retention situation at 500K temperature at the tunneling layer interface of the various tunnel oxide cases.	58
Fig. 4-8. Transient trapped electron distributions along the vertical direction of the various tunnel oxide cases during the retention situation at 400K temperature.	59
Fig. 4-9. Transient trapped electron distributions along the channel direction during the retention situation at 400K temperature at the tunneling layer interface of the various tunnel oxide cases.	60
Fig. 4-10 Simulated Arrhenius plots when criterion is 20% shift of the threshold voltage.	64
Fig. 4-11. Transient trapped electron distributions along the vertical direction of the various tunnel oxide cases during the retention situation at 300K temperature.	65
Fig. 4-12. Transient trapped electron distributions along the channel direction during the	

retention situation at 300K temperature at the tunneling layer interface of the various tunnel oxide cases.	66
Fig. 4-13. Comparisons of the data life times and activation energies between 15% and 20% criteria.	67
Fig. 4-14. Transient trapped electron distributions of the 2.4nm tunneling layer as baking in various temperatures.....	68
Fig. 5-1 The Monte-Carlo simulation sequence for the retention characteristic of the CT Flash memory.....	73

Chapter 1

Introduction

1.1 Motivation

During the past ten years, the explosive improvement and popularization of the mobile devices, such as digital camera, smart phone, tablet device and SSD, have brought the demands for the high density non-volatile memory, particularly NAND Flash memory. Many memory companies have accelerated the developing speed to satisfy the customers.

According to the report of ITRS (International Technology Roadmap of Semiconductor), the density of the Flash memory product in market has been higher than the prediction of ITRS [2]. This means that the scaling down speed of the critical dimension is faster than the expectation. As a result, the ITRS for the half-pitch size has been updated every year. The trend of the bits per chip and the half-pitch size are shown in Fig. 1-1.

So far, the conventional NAND flash memory has been the FG (Floating Gate) structure. In the FG structure, it is difficult to reduce the thickness of the floating poly silicon because of the coupling ratio between the control gate- floating gate. Therefore, as the critical dimension of the gate size shrinks, the aspect ratio of the floating poly has been increased. The high aspect ratio makes critical problems in maintaining both the cell characteristics and reliable fabrication. The narrow spaces between patterns increase capacitances and cause the cell to cell interference [3]. Since the etch depth is getting deeper while the pattern size is getting smaller, the process also becomes increasingly difficult. Because of these problems, the conventional FG structure is facing the limitation in scale down to the under 30nm generation.

To overcome the scaling limitations, the CT (Charge Trap) structure have been suggested and considered as a promising alternative structure because of the thin trap layer. Even though the CT structure has many merits, there are some critical reliability issues which have not been clearly understood yet. One of the major issues is the retention characteristic. This is the reason why the FG structure has been mainly used in the NAND Flash memory industry until now. Another idea to increase the bit density in a same chip is a 3D structure. In recent year, various 3D vertical structures, such as Stacked Flash memory, VSAT, VG, BiCs and TCAT, has been also attempted to keep the scaling trend [4][5][6][7]. It should be noted that the most 3D structures use the CT type cell from the stand point of ease of the fabrication processes. Consequently, to move foreawrd to the next generation in the

scaling trend, we have to understand the retention mechanisms and characteristics of the CT type NAND Flash memory cells.

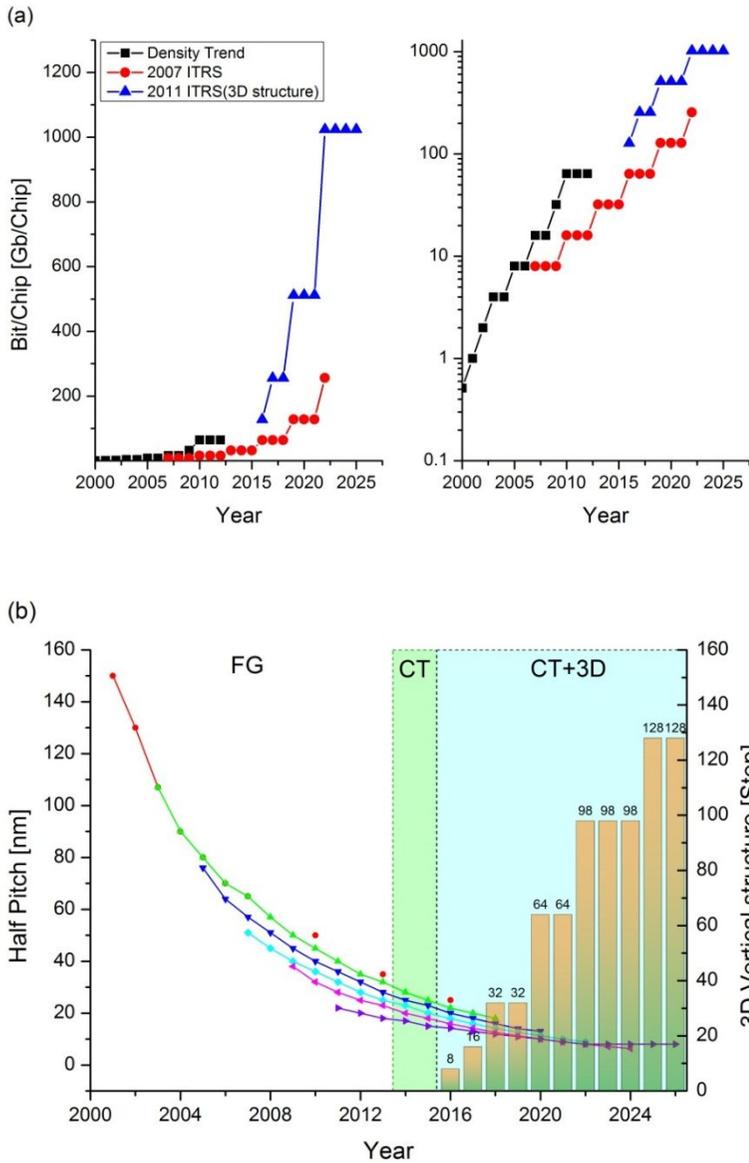


Fig. 1-1 (a) The trend of the bits per chip (left : linear scale, right log scale), (b) The half pitch size of the NAND Flash and the step of the 3D vertical structure.

1.2 CT NAND Flash memory

The Flash memory is one of the non-volatile memories invented by Dr. Fujio Masuoka while working for Toshiba in 1984 [8]. The ‘Flash’ is originated by the camera flash because the data of the Flash memory is erased in a flash mode. The structure of the Flash memory is shown in Fig. 1-2. The conventional FG Flash memory has two gates. One is the floating gate which stores charges. The other is the control gate which controls the transistor. In the case of the CT Flash memory, the thick floating poly is replaced by thin nitride trap layer. Flash memory has two kinds, NAND and NOR as show in Fig. 1-3. The two Flash memories are distinguished by the cell array structure. The memory transistors are connected in series in the NAND Flash but, in the NOR Flash, each of transistors has Bit-line and Ground line individually. Therefore, the NAND Flash is used for a high density memory and NOR Flash is used for a high speed memory. In this section, the basic NAND Flash memory operation will be explained and the features of CT NAND Flash memory with 3D Vertical NAND Flash will be introduced.

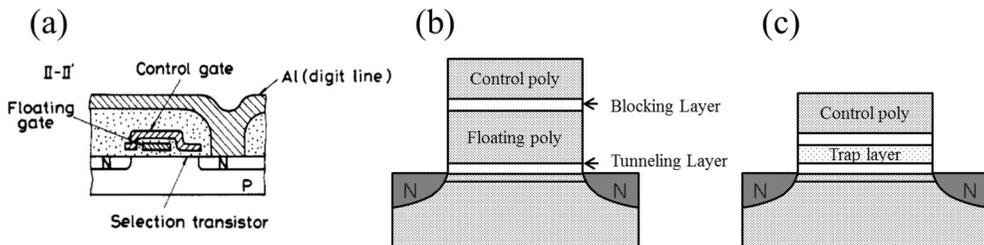


Fig. 1-2 (a) The first Flash memory structure (b) Floating Gate type Flash structure (c) CT type Flash structure.

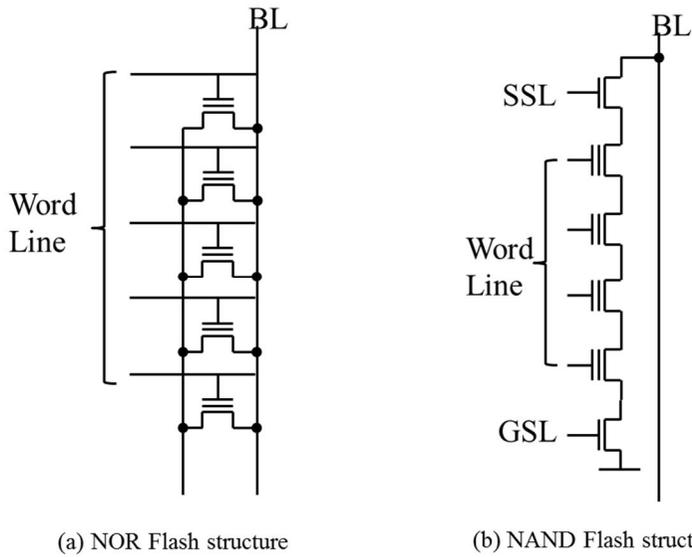


Fig. 1-3 Array structures of (a) NOR Flash structure (b) NAND Flash structure.

1.2.1 Basic Flash memory cell operation

The main operations of the NAND Flash memory are Program, Erase and Read. Programming means the injecting the electrons into the floating poly when a positive voltage is applied to the control gate. Since the NAND Flash memory has three more layers below the control gate, the applied voltage delivered to the floating gate is reduced by the coupling ratio [9]. The floating gate is usually n+ poly silicon so that there is no voltage drop. The main mechanism of the charge injection is the Fowler-Nordheim tunneling [10]. The energy band diagram and a schematic of the programming situation is shown in Fig. 1-4(a). After programming operation, the threshold voltage is shifted to the positive direction as depicted in Fig. 1-5(a).

The programmed charges are confined in a floating poly layer because the floating poly is isolated by the dielectric material. This situation is the retention mode as shown in Fig. 1-4(b). The tunneling layer and blocking layer prevent the programmed charges from escaping. Because the NAND Flash memory is a Nonvolatile memory, this retention characteristic is the most important factor for the memory reliability. The nonvolatile memory needs at least ten years of charge retention. The floating poly is conductive material. Therefore, when the tunneling oxide or the blocking oxide has some leakage path, the programmed charges are disappearing simultaneously. SILC(Stress Induced Leakage Current) is the represented example [11][12] As the memory device shrinks, the tunneling layer is vulnerable to a stress and direct tunneling of the programmed charges. In scaling down the device, the reliability of the tunneling oxide is one of the most important issues.

To erase the cell, the programmed electrons have to be ejected from the floating poly to substrate. The main mechanism of the erase of NAND Flash is also the FN tunneling. FN tunneling occurs when high positive voltage is applied to the substrate or high negative voltage is applied to the gate. The band diagram and schematic are shown in Fig. 1-4(c). After erasing operation, the threshold voltage of the cell shifts to the negative voltage as shown in Fig. 1-5(b). NOR flash memory can erase specific cell because the cells are separated each other. By contrast, NAND Flash cannot erase only a selected cell but erases by a unit block.

Read operation is verified whether the memory cell is programmed or not. If the

drain current doesn't flow with applying 0V to the gate, the cell is decided to be successfully programmed. When the cell has been erased, the current would flow. Recently, the MLC (Multi-Level-Cell) could classify the amount of the charges in the floating poly into 4 or 8 states [13][14][15][16]. In other words, the cell stores two or three bits in a cell to increase data storage density. Since the MLC needs more steps to verify, read operation for MLC is slower than the SLC (Single-Level-Cell). The concept of SLC and MLC is demonstrated in Fig.1-6. Basic operation of the CT Flash memory is same as the FG Flash memory except that the charges are trapped in trap layer instead of the floating poly.

1.2.2 CT (Charge Trap) Flash memory

The CT type memory device was suggested in the last 1960s [17][18]. The first structure of CT memory is MNOS (Metal/Si₃N₄/SiO₂/Si) structure [19]. It is conceptually the MOS structure but the oxide layer is replaced with nitride layer and oxide layer. Since the nitride layer has a lot of traps, injected electrons or holes are captured by traps and the threshold voltage shifted accordingly. The injected electrons or captured electrons, however, can be leaked to the metal gate region in programming operation. So, the first MNOS has the thick nitride thickness to prevent the leakage current to the metal gate.

Until now, there are many efforts to improve the CT type memory. To prevent the injected electron from tunneling out to the metal gate, MONOS (Metal/SiO₂/Si₃N₄/SiO₂/Si) structure was suggested [20][21]. The SONOS (Si/SiO₂/Si₃N₄/SiO₂/Si) mitigated the instability of the MONOS structure by replacing the Metal

gate with the Silicon gate [22]. TANOS (TaN/Al₂O₃/Si₃N₄/SiO₂/Si) and WANOS (W/Al₂O₃/Si₃N₄/SiO₂/Si) structures have been adopted to improve the program/erase speed by using the high-k dielectric, Al₂O₃, as a blocking layer [23].

The CT type memory cells have many merits. One of the major advantages is thinner thickness of the trap layer than the floating gate. The comparison of the two devices using TEM images is shown in Fig. 1-7. The thin trap layer of the CT reduces the cell to cell interference and makes the fabrication easy. Another merit is that the CT type cell has discrete traps. All injected electrons in the floating poly could be free when a leakage path over the tunneling oxide and blocking oxide exists. The CT type cell, however, loses only an electron which is trapped in the region connected to a leakage path. This feature of CT type cell is immune to local degradation of the tunneling oxide and thinner tunneling oxide. As a result, the Flash memory scaling trend encourages the adoption of the the CT Flash memory.

1.2.3 3D Vertical NAND Flash memory

Even if the size of pattern could be reduced to the several nano-meters, there is a limitation of the density of the bit. The simplest method to overcome this limitation is adoption of the vertically stacked structure. As the stacked layers increase, the density of bit increases. There have been many kinds of proposals for the 3D vertical NAND Flash, such as the stacked Flash memory, VSAT, VG, BiCs, TCAT and so on [4][5][6][7]. Among these 3D structures, BiCs and TCAT have been actively developed by semiconductor companies recently. Fig.1-8 is the TEM images of TCAT and BiCs [6][25]. ITRS in 2011 expects the 3D vertical NAND Flash will be

a main stream in the cell structure for the NAND Flash memory above the 128 Gb as shown in Fig. 1-1.

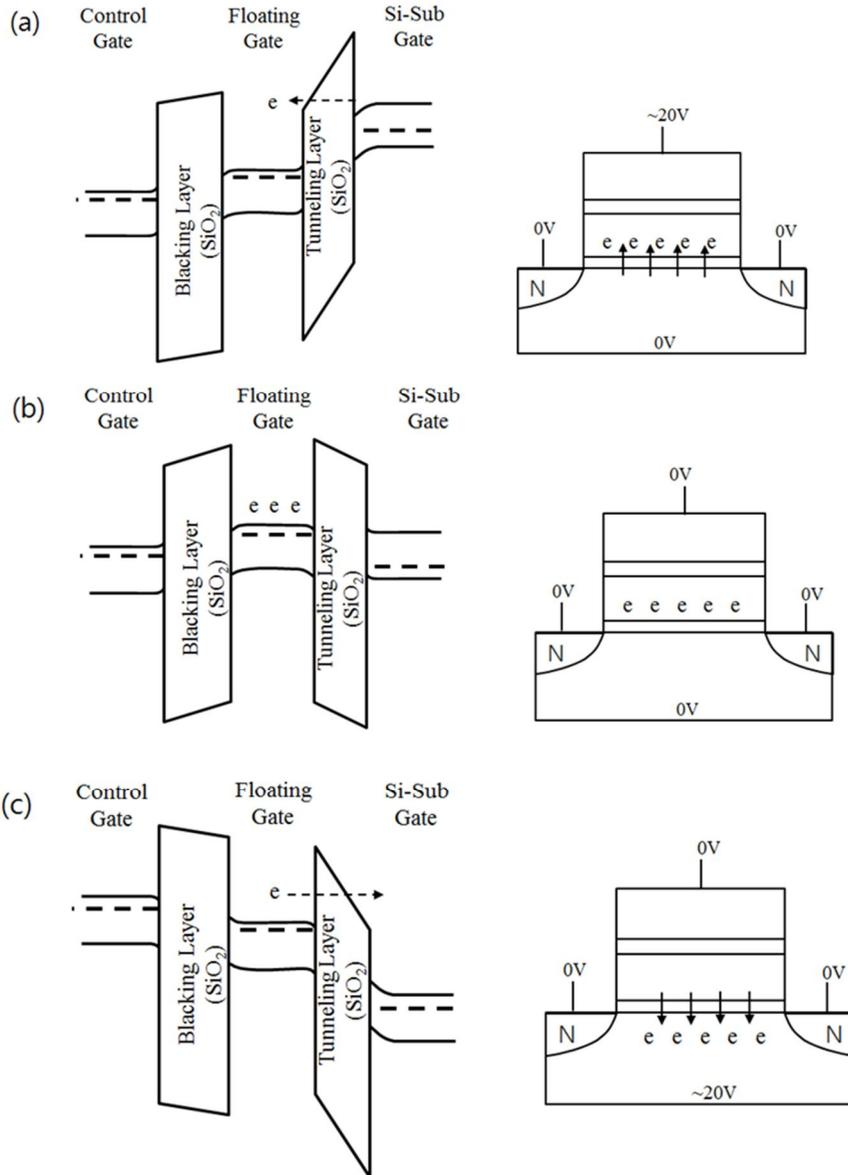


Fig. 1-4 Band diagrams and schematics of FG NAND Flash memory operation in (a) Programming operation (b) Retention situation (c) Erasing operation.

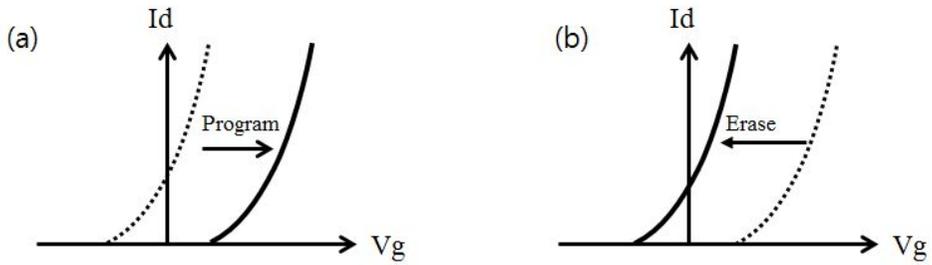


Fig. 1-5 V_{th} shifts of NAND Flash memory after (a) Programming operation (b) Erasing operation.

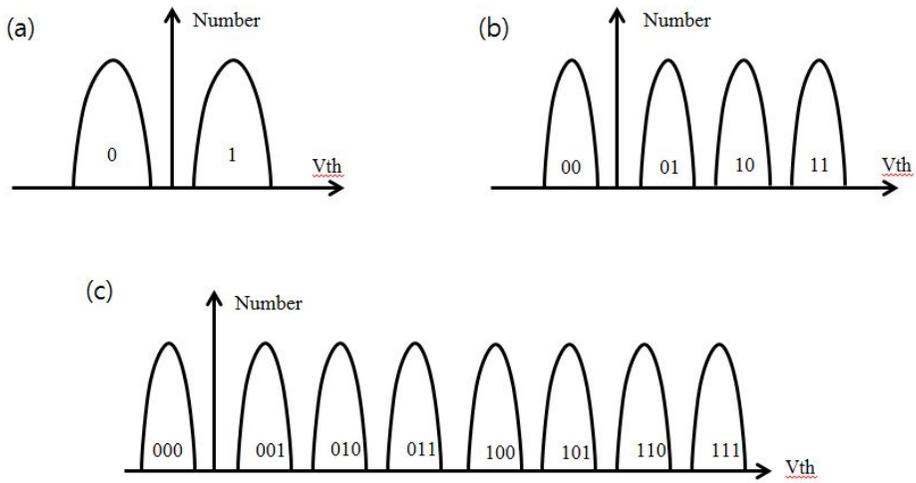


Fig. 1-6 The concept of SLC and MLC (a) 1 bit/cell (SLC) (b) 2bit/cell (MLC) (c) 3bit/cell (MLC).

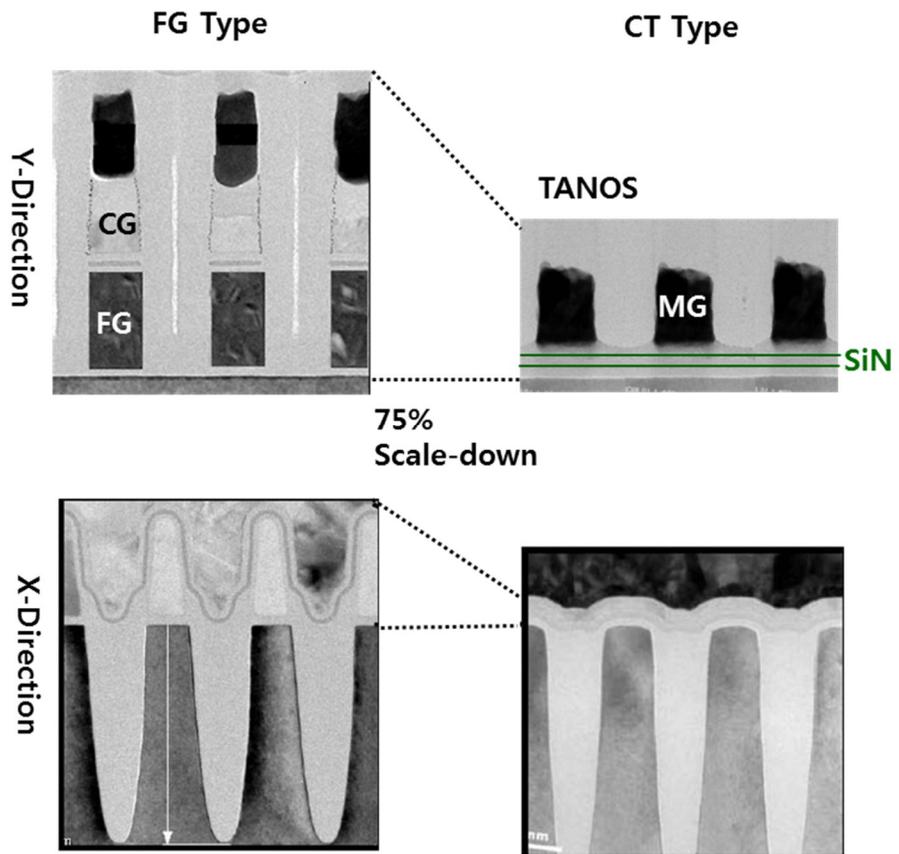
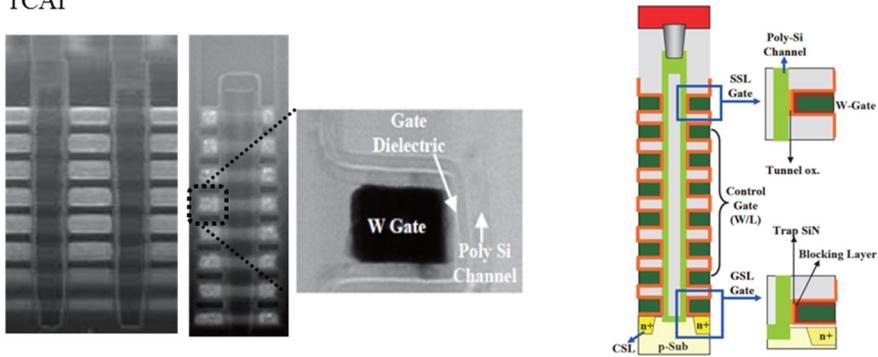


Fig. 1-7 TEM images of Floating Gate type cell and CT type cell [24].

(a) TCAT



(b) BiCs

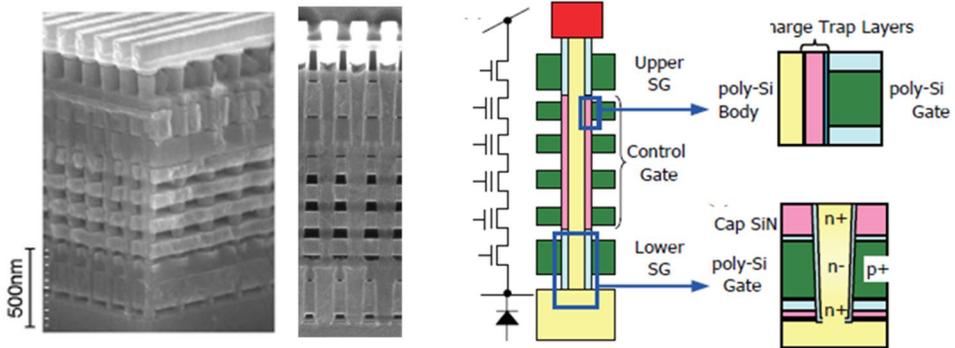


Fig. 1-8 TEM images of 3D Vertical structures (a) TCAT structure [6], (b) BiCs structure [25].

1.3 Issues of CT NAND Flash memory

1.3.1 Lateral charge spreading

As mentioned in previous section, in addition to many merits of the CT Flash memory over the conventional NAND flash memory with the floating gate the use of the CT type cell in 3D NAND Flash is unavoidable because of the vertical scalability of the trap layer.

Even though the CTF type cell has many merits, the reliability issues, especially the retention characteristics degraded by the lateral charge spreading, have not been clearly understood yet. Many previous researchers on modeling and simulation of the charge loss mechanism have been reported but limited to the 1D effect considering only along the vertical direction [26][27][28][29][30]. The experimental data, however, reveals that the charge transport in the lateral direction should be taken into account [31][32]. Moreover, it is difficult to define the trap layer only within the cell region both in the planar and the vertical structure so that the lateral charge spreading cannot be prevented [33]. Therefore, the investigation on the influence of the lateral charge spreading is vital to design the device and predict their characteristics. Some authors reported 2D simulations considering the lateral charge spreading but they assume the initial charge distribution and neglect the vertical charge redistribution [34][35].

1.3.2 Activation energy

Nonvolatile memory has to guarantee the retention characteristics as long as 10

years. To expect the retention time in room temperature from the stressed measurement, usually performed at the elevated temperature, an extrapolation method has been used. The activation energy is extracted from the Arrhenius curve which is plotted by the acceleration test in high temperature.

The Arrhenius equation, however, does not include the independent factor on the temperature, such as FN tunneling. And the measured data in accelerating condition is limited in a high temperature. If the activation energy in the high temperature region is not correct, the retention time extrapolated at the room temperature may introduce error. Incorrectly predicted performance of the device could make vital errors both the makers and the consumers. Previous research tried to simulate the effect of the tunneling on the activation energy. However, the simulation are conducted in a one dimension without considering the charge redistribution along the vertical and lateral direction [36][37].

1.4 Outline of thesis

The first object of this thesis is simulating the charge transport effect in the nitride layer on the retention characteristics in three dimension. Second object is to study the dependency of the activation energy on the tunneling oxide and charge transport in the nitride layer. All the simulation results come from in-house 3D simulator [1]. In chapter 2, for the physical understand of the results, the physics, models and structures used in this simulator are introduced. The retention characteristics are shown in Chapter 3. The dependency on the laterally extended nitride length is

simulated in various extended structures. The dependency of the programmed pattern of the surround cell is simulated in the solid pattern and the check-board pattern and the threshold voltage shift of the target cell and the surrounded cell are monitored. Finally, the retention cycle is conducted by repeating the program/retention/erase operation. All of the results are explained by the electron density distributions. The consideration of the activation energy to predict the data life time is described in Chapter 4. The conventional methods of expectation of the data retention time in room temperature will be explained and the vulnerable points of the methods will be brought up. The impact of the tunneling oxide thickness on the activation energy is simulated by varying the tunneling oxide. The activation energy is extracted at two criterions; threshold shift time for 15% and 20%. And the discussions about the main factor determining the activation energy will be made as the temperature varies. As a future work, the scheme and object of the Monte-Carlo simulation of the discrete trap and discrete electron in retention mode are planned. The conclusion and summary are included in the final Chapter.

Chapter 2

Physics and simulation structures

2.1 Physical models

Three-dimensional band diagram together with the charge fluxes in the nitride layer during the retention mode are depicted in Fig. 2-1. The x-direction and y, z-direction indicates the vertical direction and the lateral direction, respectively. In this simulation, the major charge fluxes are classified into four main mechanisms in retention mode as shown in Fig.2-1. The two fluxes are tunneling fluxes which are toward to the substrate conduction band from the traps (flux-1 in Fig. 2-1) and conduction band of nitride (flux-2 in Fig. 2-1). These currents are expressed in (1) and (2), respectively.

$$J_{TC} = qn_r T_n v \Delta l \quad (1)$$

$$J_{CC} = qn_c T_n v_T \quad (2)$$

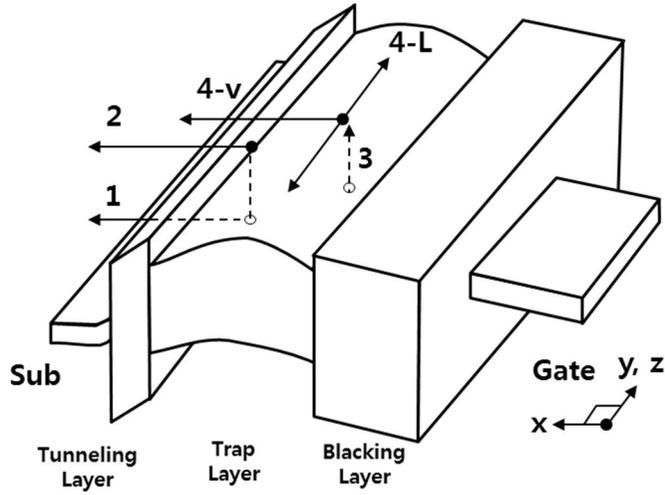


Fig. 2-1 Three-Dimensional SANOS band diagrams and charge flux diagrams in retention mode.

where n_T is the trapped electron density and n_C is the electron density of the conduction band at the interface, ν is a hitting frequency to escape the trap, T_n is tunneling probability based on the WKB approximation, Δl is the distance between nodes and v_T is the average tunneling velocity of the electrons [29][30]. Assuming that the blocking layer is thick enough to block the back tunneling in the retention mode, the tunneling current of the nitride to gate electrode is ignored.

The electron capture and emission between the trap and conduction band of the nitride are depicted in Fig. 2-1 (flux-3). The capture is model by the SRH statistics and the emission process is represented by the Poole-Frenkel model as in (3) and (4) [38][39]

$$\frac{\partial n_c}{\partial t} = \nabla \cdot J_n - c_n(N_T - n_T)n_c + v \exp\left(-\frac{E_t - \beta\sqrt{\varepsilon}}{k_B T}\right)n_T \quad (3)$$

$$\frac{\partial n_T}{\partial t} = c_n(N_T - n_T)n_c - v \exp\left(-\frac{E_t - \beta\sqrt{\varepsilon}}{k_B T}\right)n_T \quad (4)$$

where c_n is capture coefficient, N_T is the total trap density in the nitride, E_t is the trap energy, β is the Frenkel constant, ε is the electric field and J_n is the flux-4 in Fig. 2-1. The flux-4 is the charge transport in the conduction band of the nitride which can be written as

$$J_n = q\mu_n n_c \varepsilon + qD_n \nabla n_c = q\mu \left[n_c (-\nabla \phi) + \frac{k_B T}{q} \nabla n_c \right] \quad (5)$$

It may be noted here the charge transport along the vertical and lateral direction in both the program and the retention mode are considered so that the current density is calculated three-dimensionally. Since this thesis focused on the retention characteristics, the hole contribution is neglected [26][27][36]. More details of the physics of the simulation are described in ref.[0].

2.2 Simulated structures

The 3D SANOS (Si/Al₂O₃/Si₃N₄/SiO₂/Si) devices with 4nm tunneling oxide, 8nm Si₃N₄, 14nm Al₂O₃, and 30nm channel length are considered in this study. The dependency of the extended nitride length is analyzed by comparing the four structures. Fig. 2-2(a) is the control device without lateral spreading along the

channel direction. From 1nm to 30nm extended nitride structure are selected to represent the finitely extended nitride case (Fig. 2-2(b)). To reflect the infinitely extended nitride structure, we defined the extended nitride to the 150nm as shown Fig. 2-2(c).

In fact, the actual NAND flash memory has the array structure. Thus, the three gate string structure is used to show the interference with the neighbor cells (Fig. 2-2(d)). Because all of these structures are the SA-STI structure, the changes of the charge distribution along the z-direction in the nitride are neglected [40].

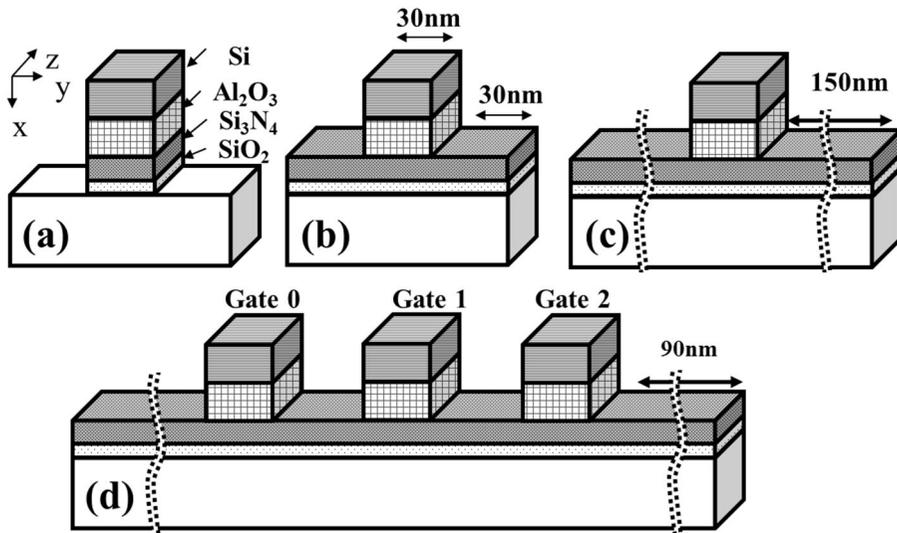


Fig. 2-2. Simulated structures of the SA-STI SANOS device with (a) No extended nitride, (b) finitely extended nitride, (c) infinitely extended nitride, and (d) three gate string.

Chapter 3

Simulation of the charge transport effects

3.1 The initial charge distribution

Many previous retention simulation are started with assumed the initial charge distribution [34][35]. The initial charge distribution, however, is an important factor of the retention characteristics as shown in ref [29][41]. Thus, all of the retention simulations in this thesis are conducted after programming the devices by enforcing 18V to the gate electrode under the temperature 300K. The programming simulation is considering the charge transport in conduction band of nitride layer three dimensionally and self-consistently. The transient charge distribution of the conduction electron density and the trapped electron density are shown in Fig. 3-1. The programming is stopped when the threshold voltage shift is 4V.

As time went on, the injected electrons fill the traps from the interface of the tunneling layer to the interface of the blocking layer. Since it is assumed that the

blocking layer is sufficiently thick to prevent the back tunneling to the gate, the transported electrons are piled up at the blocking layer interface. The trapped electron density does not exceed the trap density which is fixed with $1e20/cm^3$. The transient 3D profile of the trapped electron distributions is shown in Fig. 3-2.

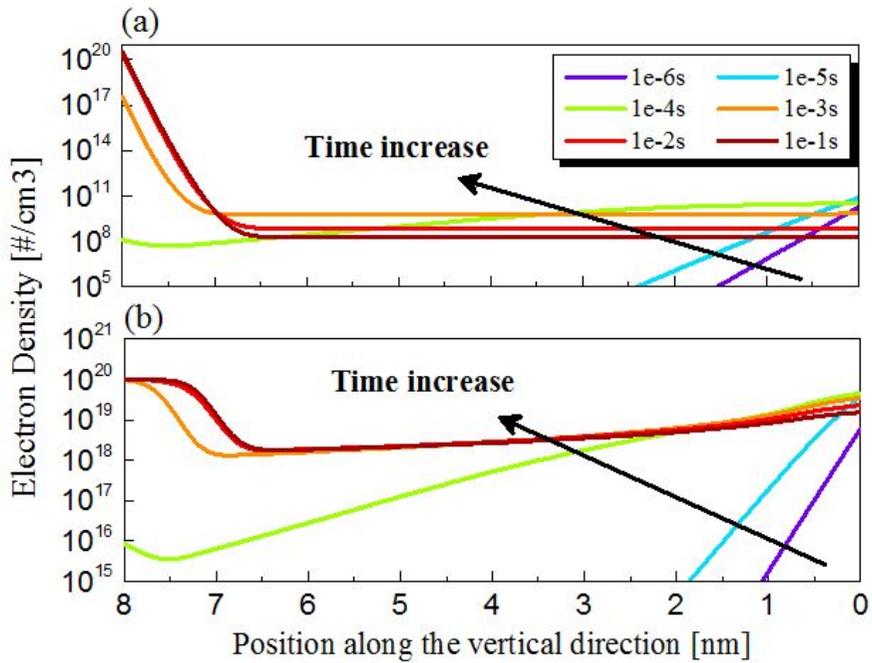


Fig. 3-1. The transient distribution of (a) the conduction electrons, (b) the trapped electrons during the program mode of the finitely-extended nitride structure at the middle of the channel.

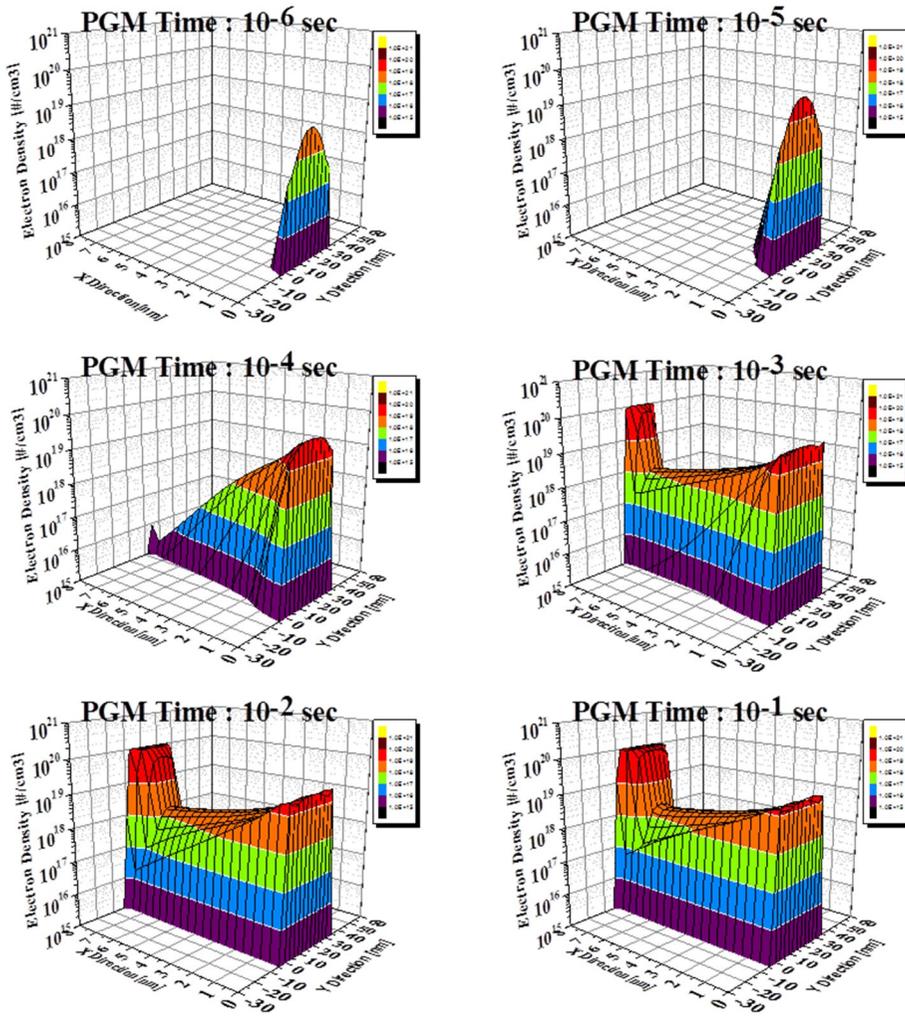


Fig. 3-2. Transient trapped electron distributions of the infinitely-extended structures from 10⁻⁶ sec to 10⁻¹ sec during the programing.

3.2 The dependency of the extended nitride length

Fig. 3-3 is the retention curves of the devices which have different extended nitride length each other. These structures are already shown in Chapter 2.

First of all, to distinguish the vertical charge loss from the lateral charge loss, we simulated a structure which has the 6nm tunneling oxide while the others have the 4nm tunneling oxide as shown in Fig. 3-3. The thick (Ext_0nm TunOx 6nm) tunneling oxide structure has no vertical and lateral charge loss except for the charge redistribution along the vertical direction [29]. Thus, the difference between the no-extended nitride device (Ext_0nm) and the 6nm tunnel oxide device (Ext_0nm TunOx 6nm) indicates the vertical charge loss only.

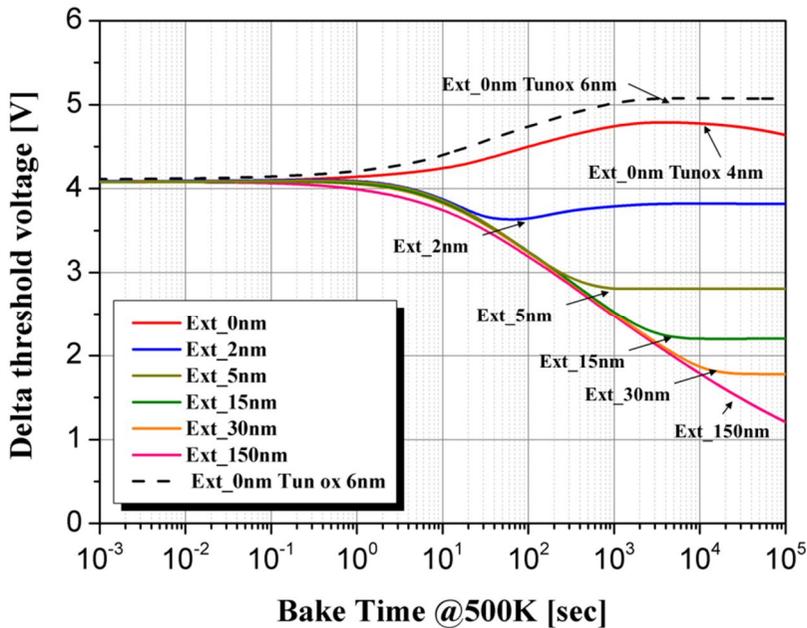


Fig. 3-3. Threshold voltage shifts of the devices. 6nm tunnel oxide structure is also simulated as reference device to evaluate the vertical charge loss.

The increment of the threshold voltage of the no-extended nitride structure is due to the charge transport along the vertical direction from the gate side to the tunneling oxide side as shown in Fig. 3-4 (a).

Since the trapped electrons are transported along the vertical and the lateral direction in the cases of the finitely and the infinitely extended structure as shown in Fig. 3-5 (b)(c) ,Fig. 3-6 and Fig. 3-7, the significant threshold voltage shift is observed, unlike the no-extended nitride structure in Fig. 3-3 (Ext_0nm). Interestingly, the saturation of the threshold voltage occurs and the saturation point depends on the extended nitride length in Fig. 3-3. The saturation can be explained by the transient distribution of the trapped electron density as shown in Fig. 3-5. When the extended nitride region is filled with the transported charges up to the equal level of the gate region, the finitely-extended nitride structure has no lateral charge loss but has the vertical charge loss. Therefore, the retention curves are saturated earlier as the extended nitride length becomes shorter. In this simulation, the side surface of the gates are regarded as a damage free surface. If the surfaces has a damages or any other leakage path, there wouldn't be a saturation point [33].

From this result, the extended nitride length is confirmed as an important parameter in predicting the retention time and deciding the specification of the structure, such as a side-wall spacer size [31].

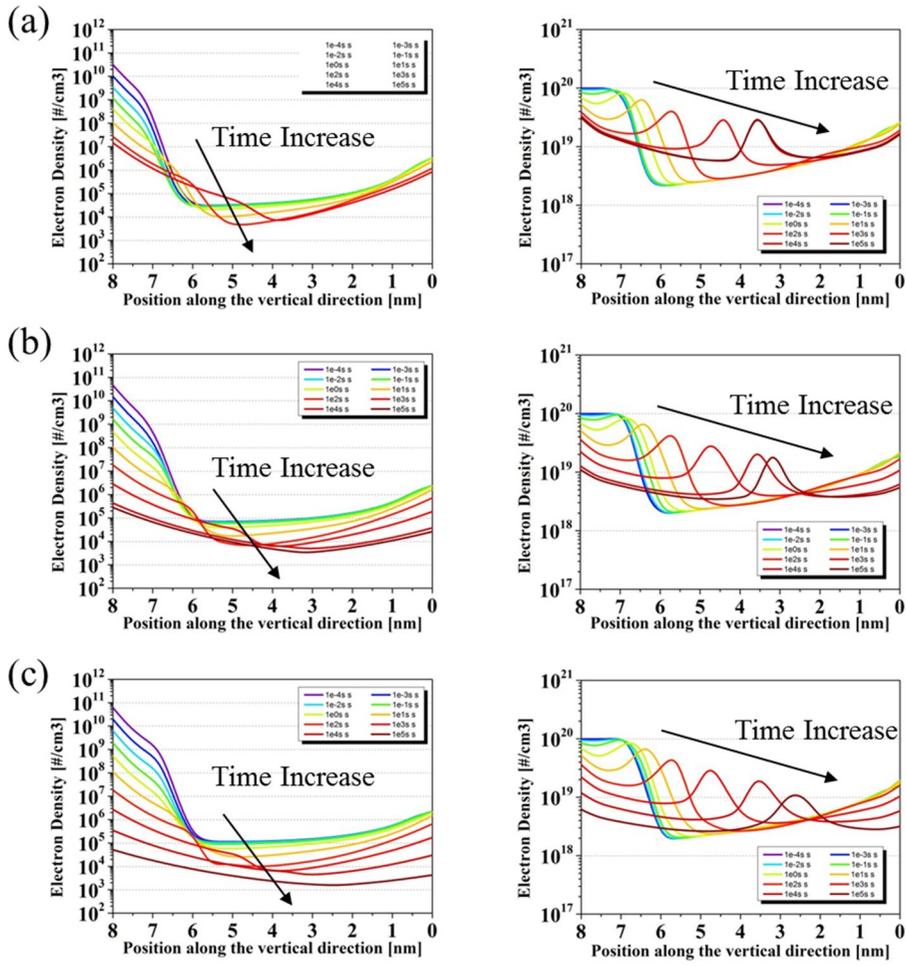


Fig. 3-4. Transient conduction and trapped electron distributions along the vertical direction (a) no-extended, (b) finitely-extended (30nm), and (c) infinitely-extended structures.

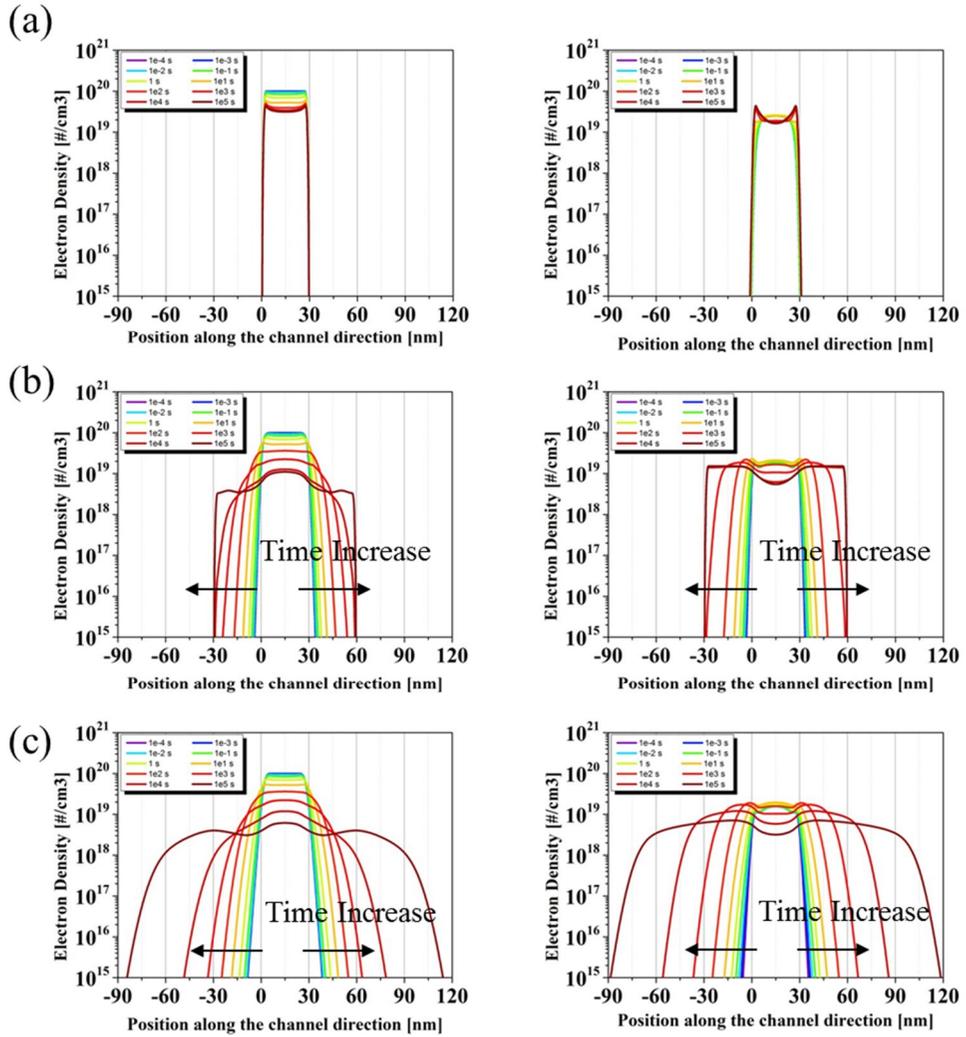


Fig. 3-5. Transient trapped electron distributions along the channel direction at the blocking layer interface (Left) and the tunneling layer interface (Right) of the (a) no-extended, (b) finitely-extended (30nm), and (c) infinitely-extended structures.

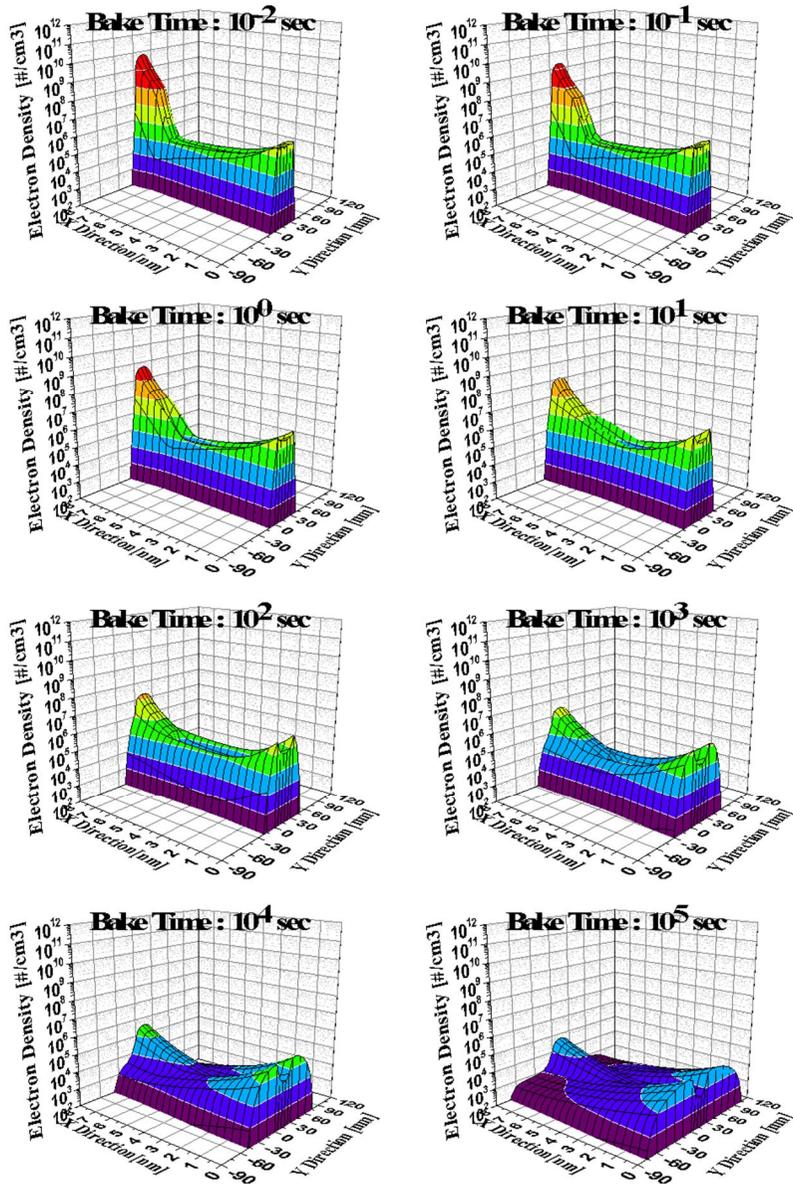


Fig. 3-6. Transient conduction electron distributions of the infinitely-extended structures from 10^{-2} sec to 10^5 sec.

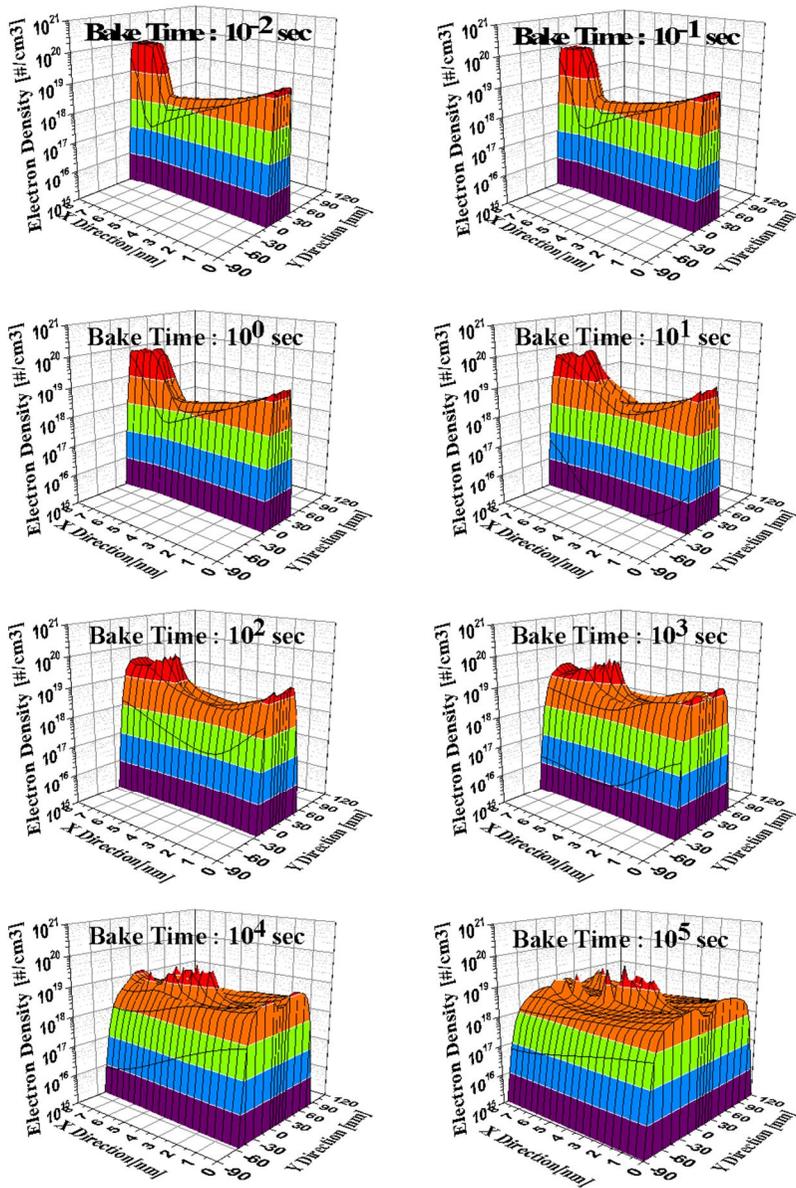


Fig. 3-7. Transient trapped electron distributions of the infinitely-extended structures from 10^{-2} sec to 10^5 sec.

3.3 The dependency of and on the surround cells

To understand the influence of the lateral charge spreading on the neighbor cell, the check-board and the solid program patterns are simulated with the infinitely-extended nitride structure. Because we assumed that the devices have the SA-STI structure, the nearby cells in the other string is neglected and the only two nearby cells in a same string line affect the monitoring cells.

3.3.1 Check-Board pattern

For the check-board (CKBD) pattern simulation, the string structure is simulated with a programmed cells and an erased nearby cell. The simulated structure, the programmed pattern and delta threshold voltage shift as baking are drawn in Fig. 3-8. As the programmed electrons are transport to the lateral direction, the threshold voltage of the Gate1 is increased. The threshold voltage shift of the erased cell is shown in the Fig. 3-8 (Single – Dashed line) when the only one side cell is programmed. In the check-board pattern, the threshold voltage of the erased cell is increased two times than the one side programmed situation, because the cell is affected by the two programmed cells surrounding the erase cell. The electron transport is illustrated in Fig. 3-9, Fig. 3-10 and Fig. 3-11 with the trapped electron density distribution. Since the transported electrons from the programmed cells reach the erased cell (Gate1) region, the threshold shift is increased until the threshold voltage of the erased cell is same as that of programmed cells. The threshold voltage shift of the programmed cells is not different from the single transistor retention characteristics shown in Fig. 3-3.

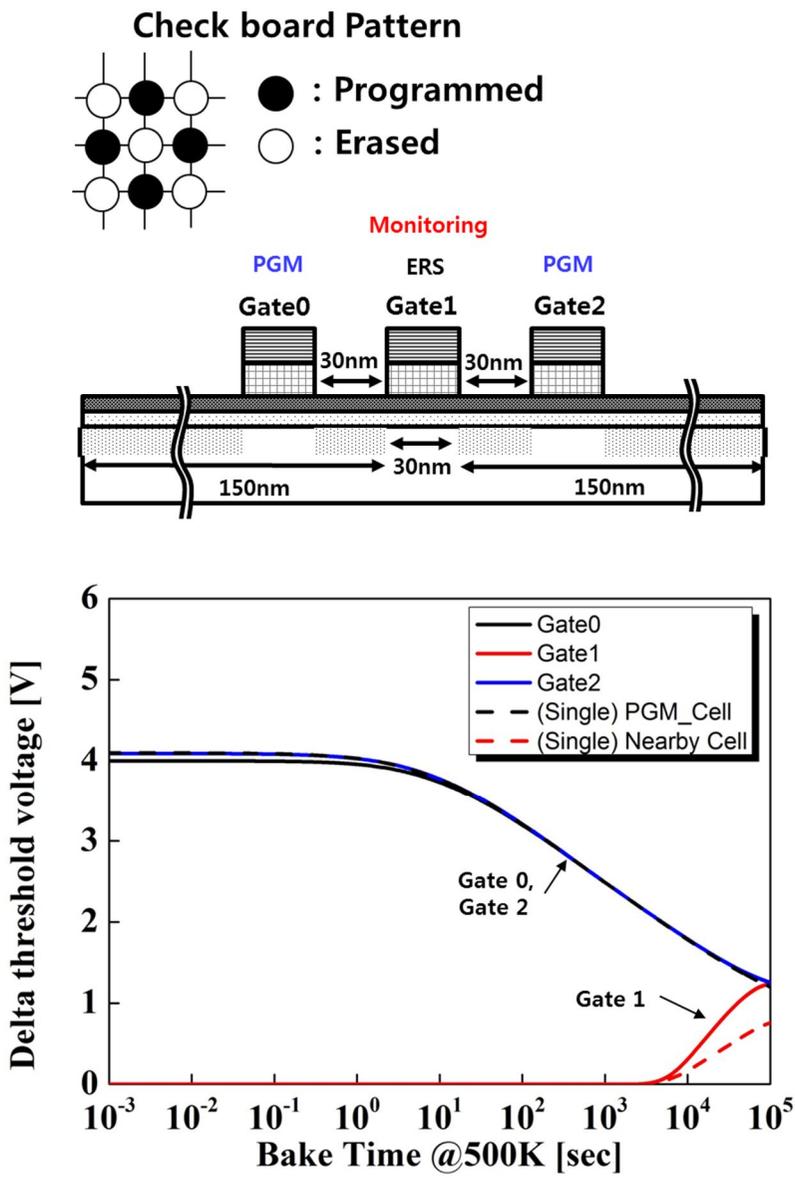


Fig. 3-8. Check-board pattern structure and retention curves. Dashed line is the situation when Gate0 is programmed and Gate1 and Gate2 are erased.

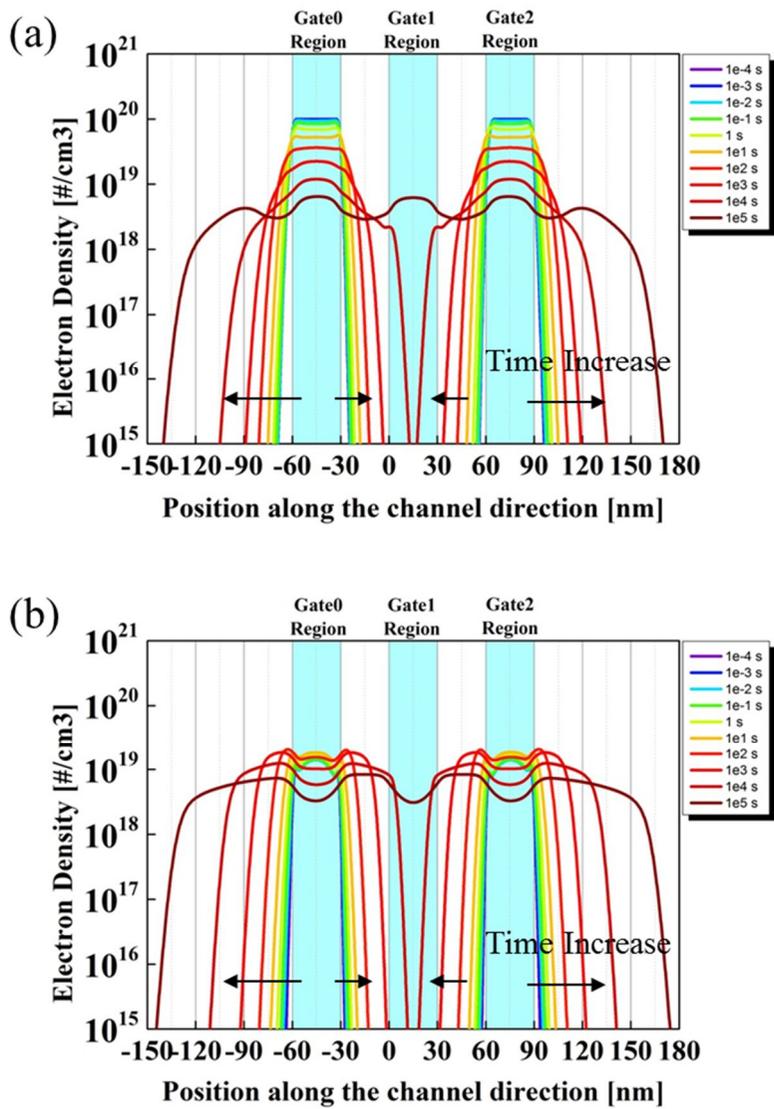


Fig. 3-9. Transient trapped electron distributions along the channel direction at the blocking layer interface (Top) and the tunneling layer interface (Bottom) of the check-board pattern.

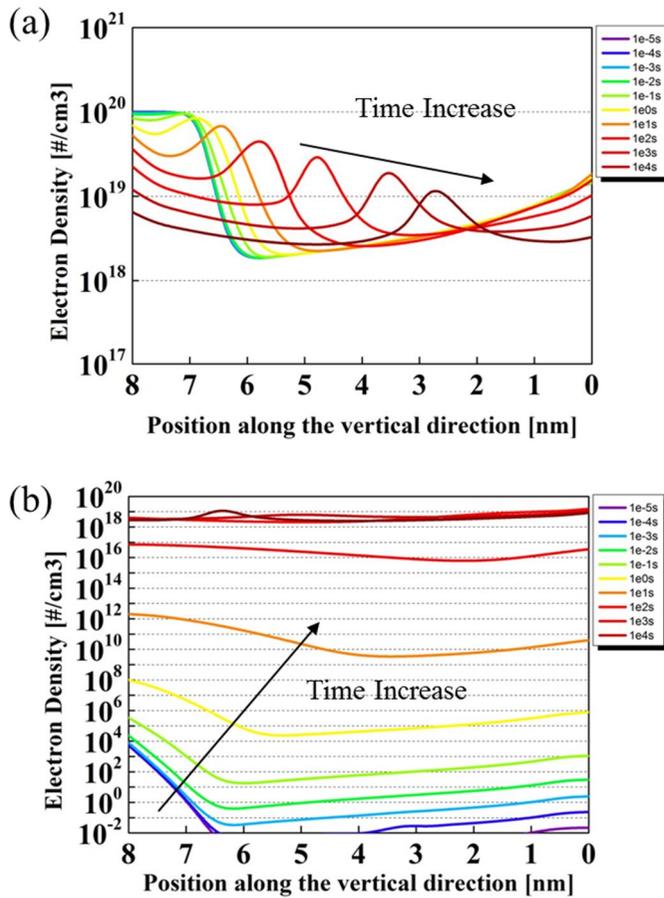
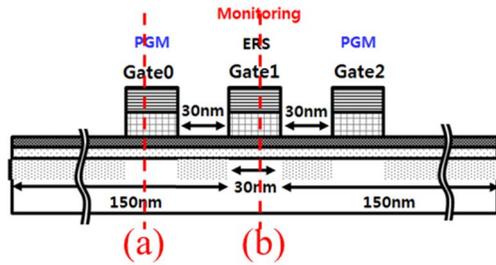


Fig. 3-10. Transient trapped electron distributions along the vertical direction at the middle of (a) the programmed Gate0 and (b) the erased Gate1.

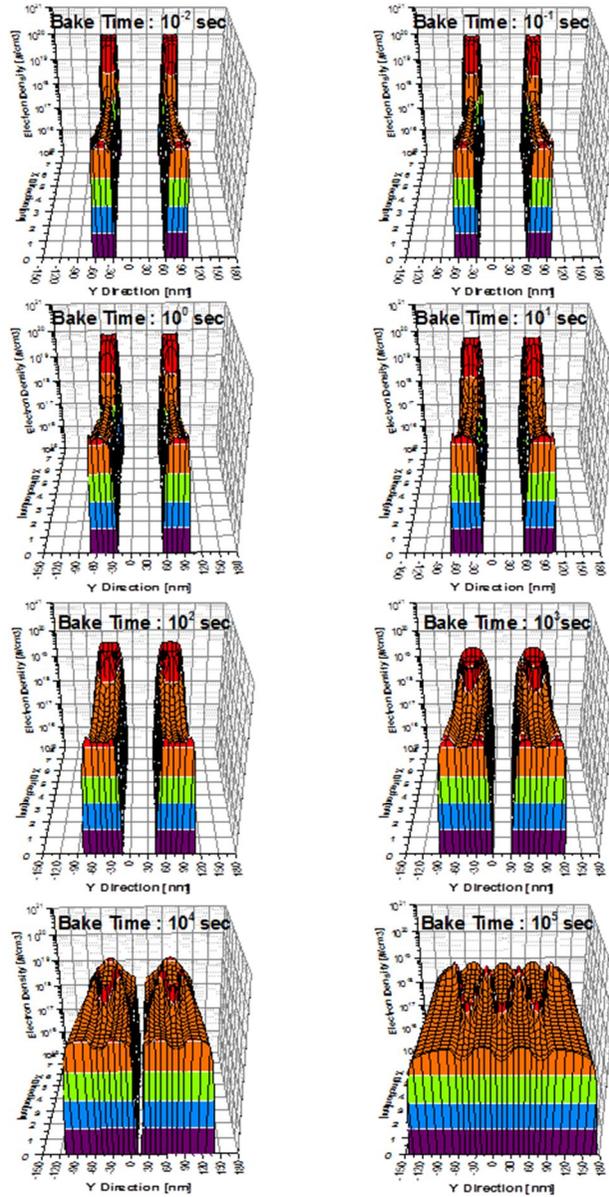


Fig. 3-11. Transient trapped electron distributions of the check-board pattern from 10⁻² sec to 10⁵ sec.

3.3.2 Solid pattern

In the solid pattern, since all cells are programmed, the space region between the gates is the only way for the transport. Hence, the uncovered region with the gate is filled by the electrons which come from the neighboring cell until the potential of the uncovered region is equal to the cell region. As the space region is filled with the electrons, the lateral charge transport is interrupted so that the decrease rate of the threshold voltage is reduced as shown in Fig. 3-12. The retention curves, however, are not saturated because the electrons are transported to the outside of the solid pattern array. The charge density in the space region also decreases after the space region is filled as shown in Fig. 8(b). In addition, Fig. 8 (a) and Fig. 8 (b) show that the retention characteristics of the edge cell is not same as the inner cells in the solid pattern. These pattern dependency results are consistent with the experimental data in [31]

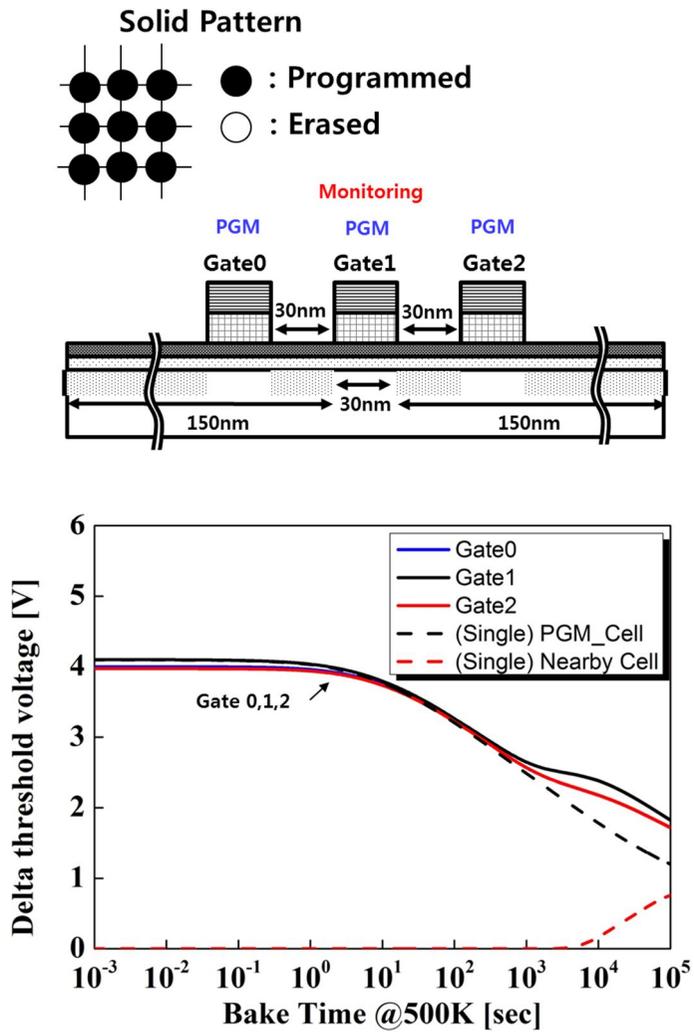


Fig. 3-12 Solid pattern structure and retention curves. Dashed line is the situation when Gate0 is programmed and Gate1 and Gate2 are erased.

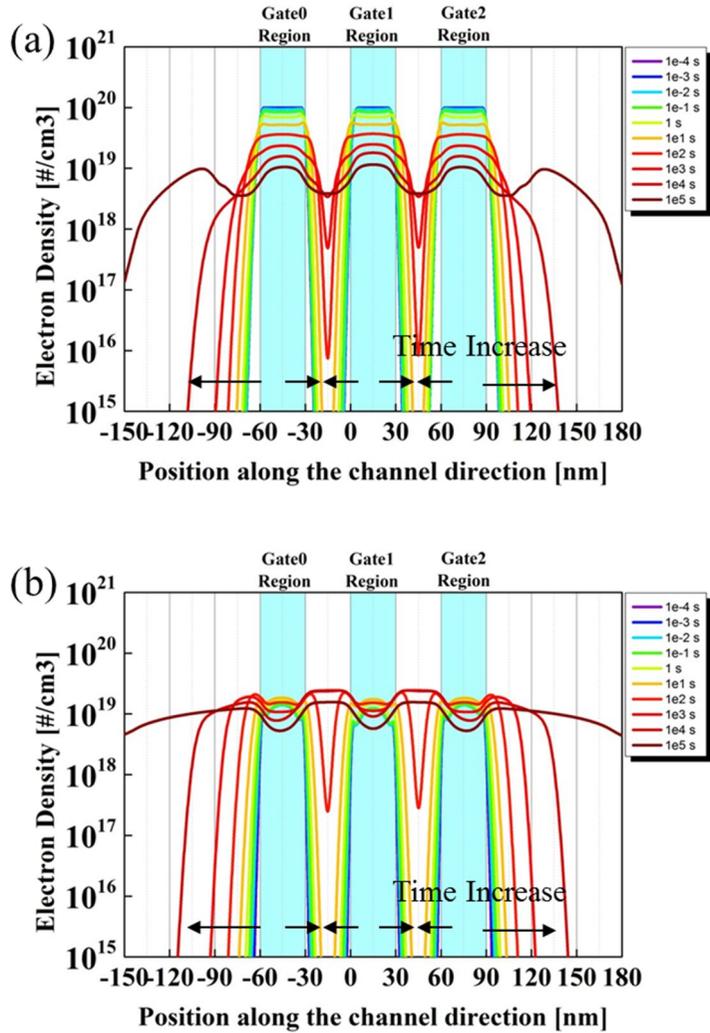


Fig. 3-13. Transient trapped electron distributions along the channel direction at the blocking layer interface (Top) and the tunneling layer interface (Bottom) of the check-board pattern.

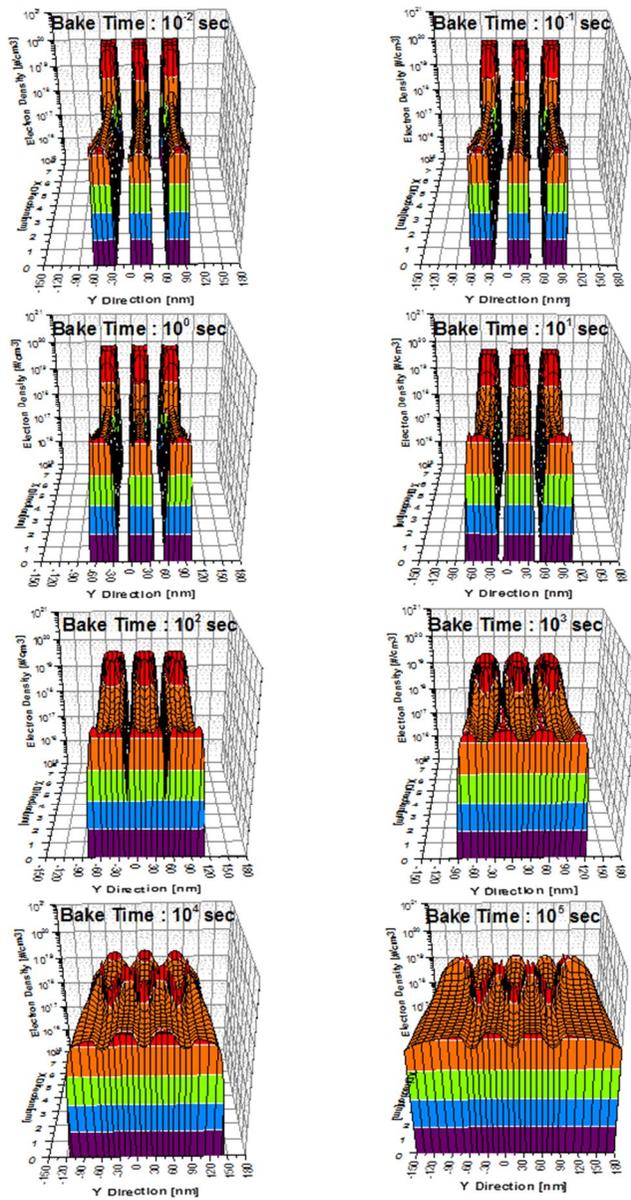


Fig. 3-14. Transient trapped electron distributions of the check-board pattern from 10^{-2} sec to 10^3 sec.

3.4 The dependency of the retention cycle

The remaining electrons after the retention mode could affect to the second retention characteristic. To reflect this situation, we simulated the retention cycle (programming – retention – erase) using the three gate structure and finitely (30nm) extended nitride structure.

3.4.1 Infinitely extended nitride structure

In the string structure with continuous nitride layer, the centered cell and the nearby cell are monitored to inspect the charge transport effect on the nearby cell as the cycle is performing. Fig. 3-15 shows the threshold voltage shift and Fig. 3-16 depicts the trapped electron density of the each step. During the retention steps, the threshold voltage of the nearby cell is changed because of the electron transport. For the program and erase steps, the nearby cell is not affected. After a retention step, the remaining electrons in the uncovered region disturb the charge spreading at the next retention tern. Therefore, the slope of the threshold voltage shift is decreased as the retention cycle is repeated as shown in Fig. 3-17.

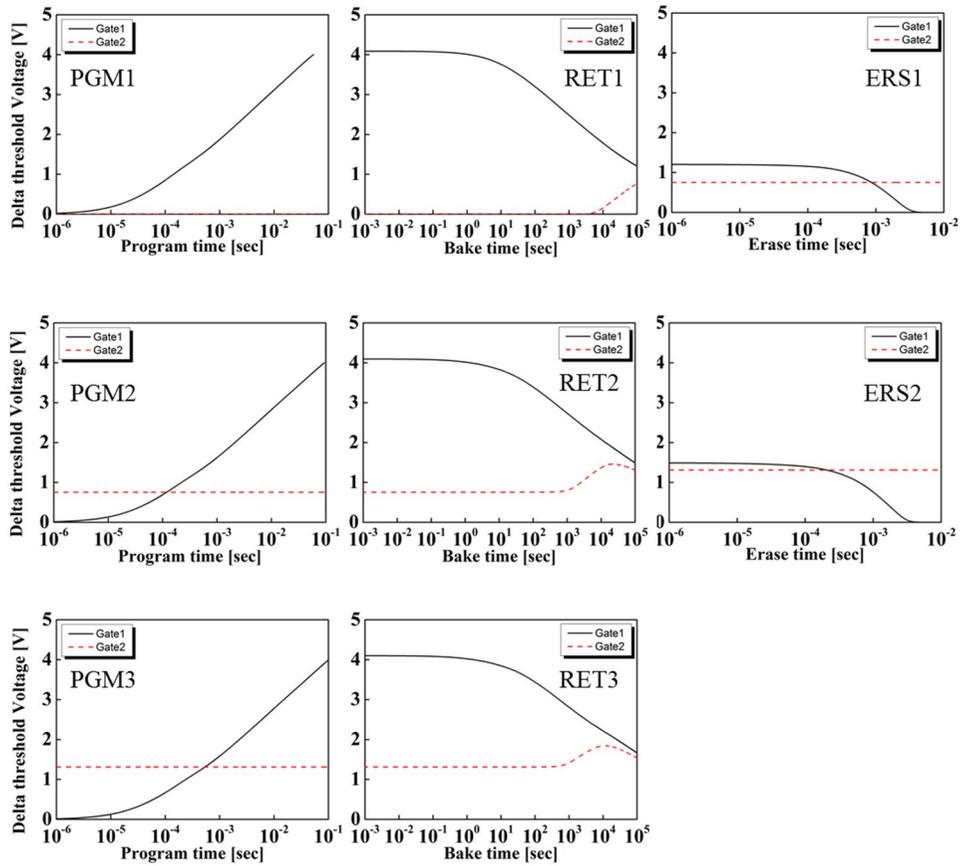


Fig. 3-15 Delta threshold voltage shift curves of the three gates string structure as the cycle is repeating.

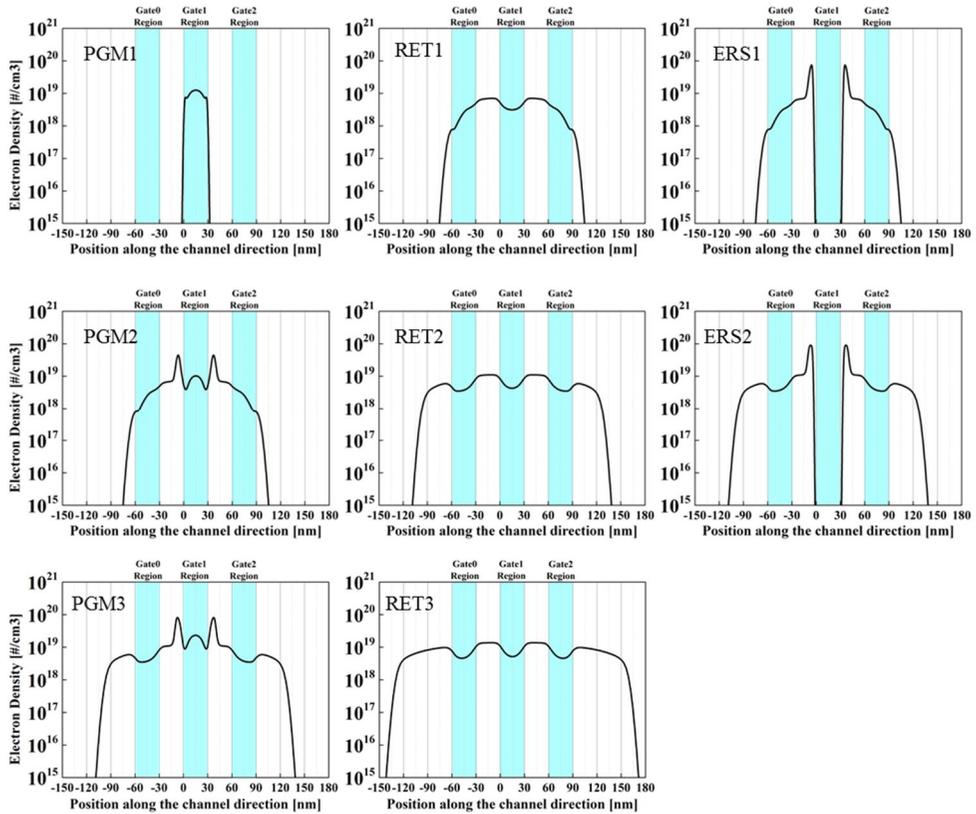


Fig. 3-16 Trapped electron density of the three gates string structure at the interface of the tunneling layer when the each retention cycle step is stopped.

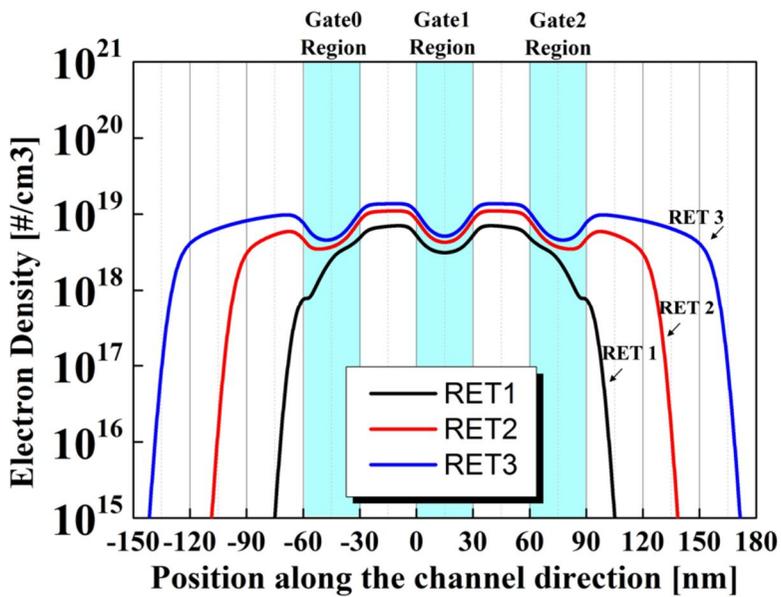
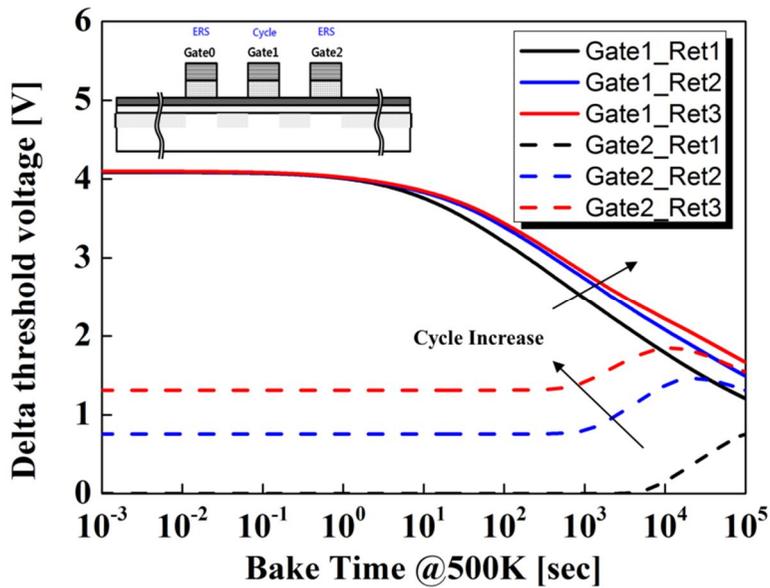


Fig. 3-17 Top is retention curves of three gates string structure after the cycle is repeated. The dashed lines indicate the threshold voltage shift of the nearby cell. Bottom is the trapped electron density profiles after each retention step is finished.

3.4.2 30nm extended nitride structure

In the case of the 30nm extended nitride structure, the programmed electrons couldn't transport over the nitride region. Since the electrons in the extended region are piled up as the retention step repeating, the charge loss to the lateral direction is degraded and the threshold voltage saturation is increased. That is shown in Fig. 3-18, Fig. 3-19 and Fig. 3-20.

The experimental data of the retention cycle are reported in [35]. This Simulation results indicates that measurement sequence impacts on the retention performance of the CT type Flash memory.

3.5 Conclusion

Using the 3D self-consistent simulator, we simulated the effects of the lateral charge spreading in the CT Flash memory. The initial charge distribution has been obtained from the same simulator after the programming considering the charge transport in the conduction band of the silicon nitride self-consistently. Considering device structures with various lateral extended nitride, it has been concluded that the extended nitride length is an important parameter to determine the retention characteristics. Since that the neighbor cells affect the monitoring cell of interest, the programmed pattern has to be considered in analyzing the retention characteristics. Finally, we showed that the charge loss characteristic in the retention cycle is also strongly affected by the lateral charge spreading.

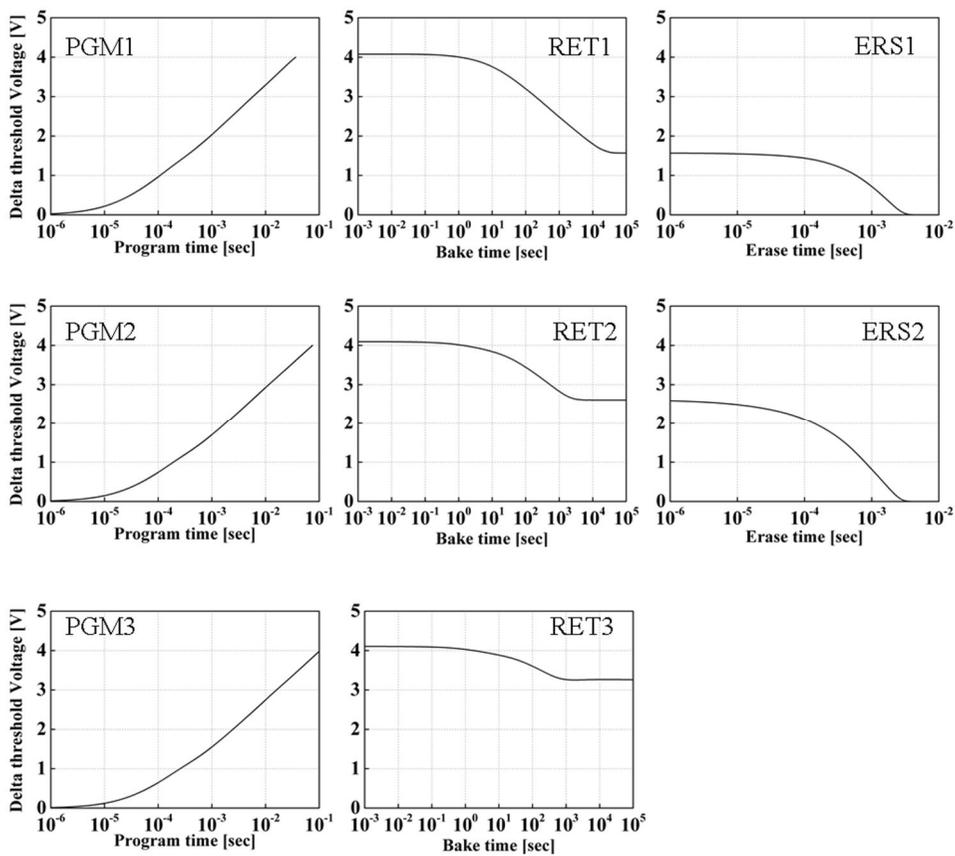


Fig. 3-18 Delta threshold voltage shift curves of the 30nm extended nitride structure as the cycle is repeating.

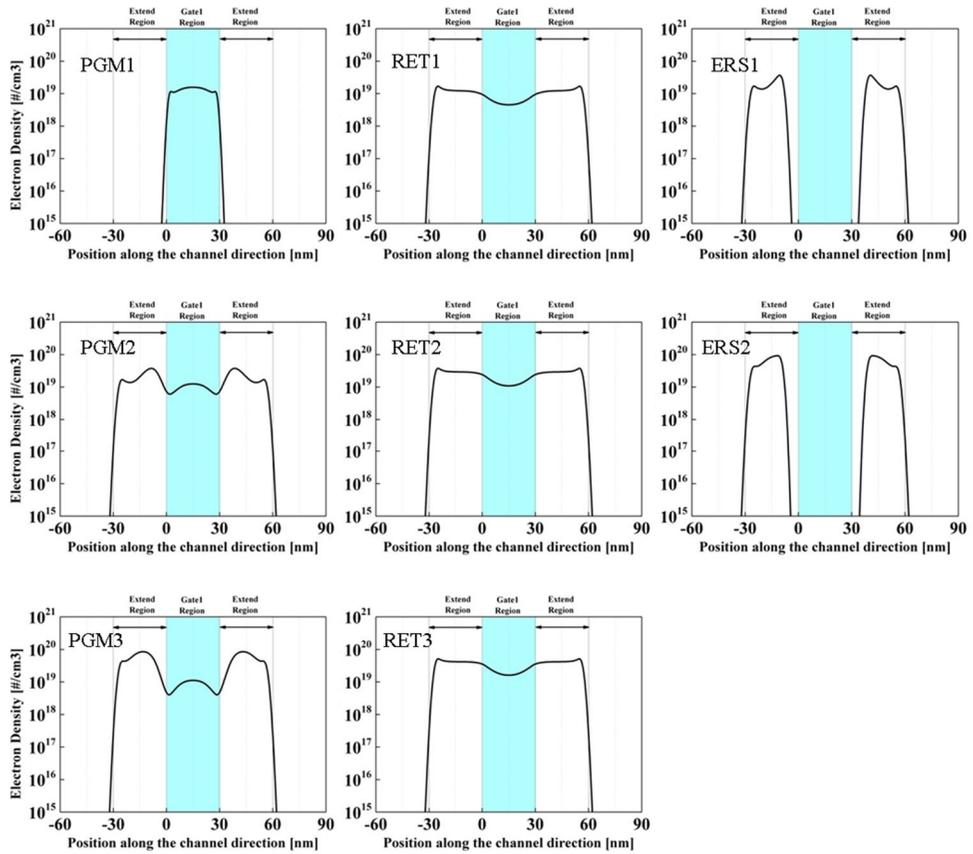


Fig. 3-19 Trapped electron density of the 30nm extended nitride structure at the interface of the tunneling layer when the each retention cycle step is stopped.

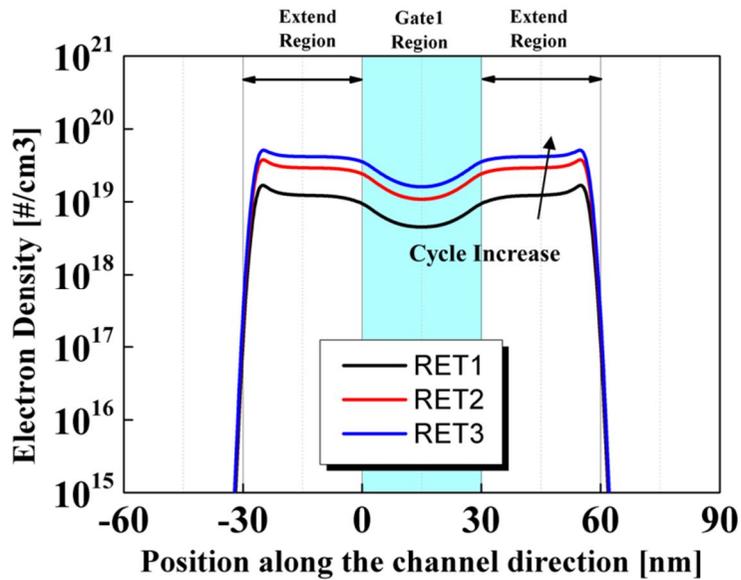
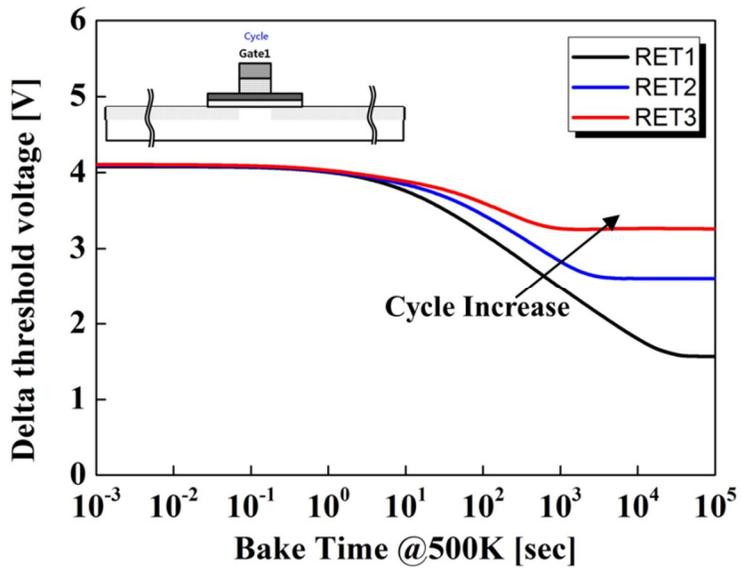


Fig. 3-20 Top is retention curves of 30nm extended nitride structure after the cycle is repeated. Bottom is the trapped electron density profiles after each retention step is finished.

Chapter 4

Simulation of the activation energy

4.1 Conventional expectation method of the data life time

Flash memory requires data retention time for ten years at operating temperature. Since the retention test of the requirement takes much long time, extrapolation method in Arrhenius plot has been used for the conventional FG Flash memory after testing the life times at the various high temperature [41][42]. As many experimental studies reported, however, the retention characteristic is affected not only the thermal emission but also the field enhanced tunneling leakage [42]. Fig.4-1 shows various tendencies of the experimental data [43][44][45]. Fig. 4-1(a) is a standard case of the Arrhenius plot which has constant activation energy. The degradation of

activation energy in the Fig. 4-1(b) and Fig. 4-1(c) is modeled by the effect of the tunneling leakage current through the IPD layer [43][44]. Fig. 4-1(d) shows very strange curve which is bent line. This would be modeled in section 4.3. As the Fig. 4-1 shows, the activation energy is varied according to the temperature, so that forecasting the life time by extrapolating the Arrhenius plot is difficult.

For the CT type Flash memory, the conventional Arrhenius plot has been used to predict the data retention time just like the FG Flash memory [46]. CT Flash memory also has diverse activation energies which depend on the temperature. The experimental data extracted from reported papers are shown in Fig. 4-2. The linear lines of the the Fig. 4-2 (a) and Fig. 4-2 (b) mean the dominant charge loss mechanism of the device is the thermal effect. The degradation of the Fig. 4-2(c) is explained by the tunneling current [36]. Fig. 4-2(d) has not been clearly understood yet. Interestingly, the bent shape is similar to the Fig. 4-1(d).

As a result, extrapolation method using activation energy in predicting the retention time is valid only if the charge loss is strongly dominated by the thermal emission. The modification of the method in anticipating the life time is needed with an in-depth consideration about the charge loss mechanism.

In the following sections, the ambiguous Arrhenius plots of the CT Flash memory are investigated using the 3D simulator. And the bent curves of the Fig.4-1 and Fig. 4-2 are modeled by the charge transport.

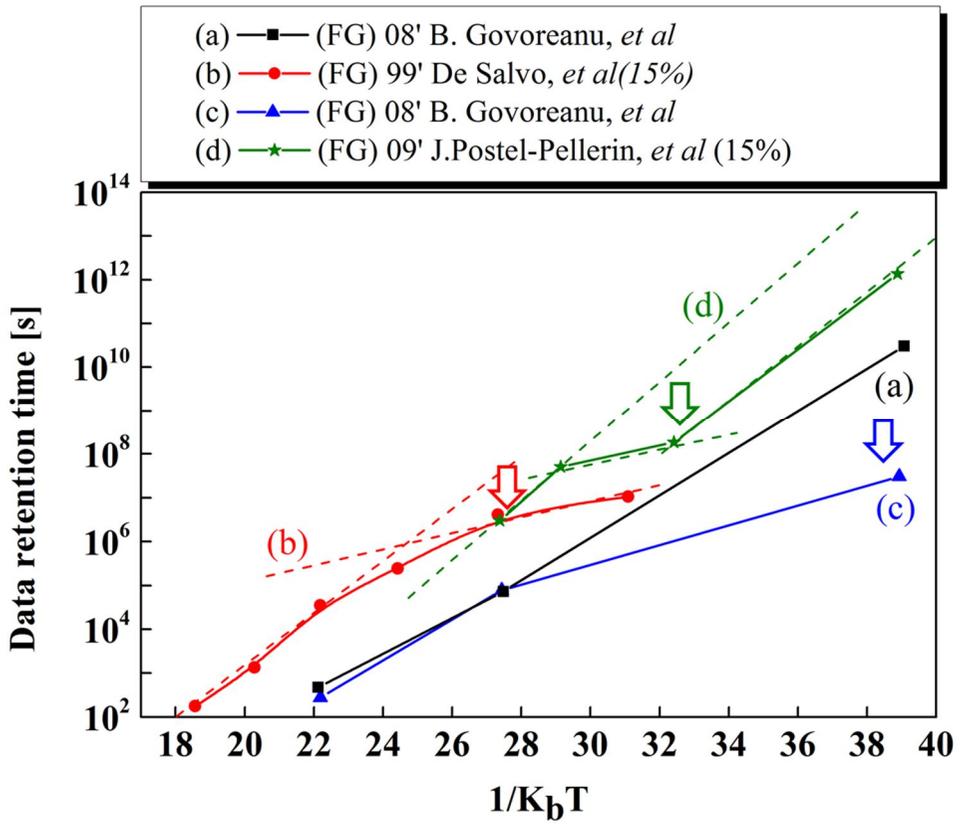


Fig. 4-1 Arrhenius plots of FG Flash memory which are extracted from reference [43][44][45].

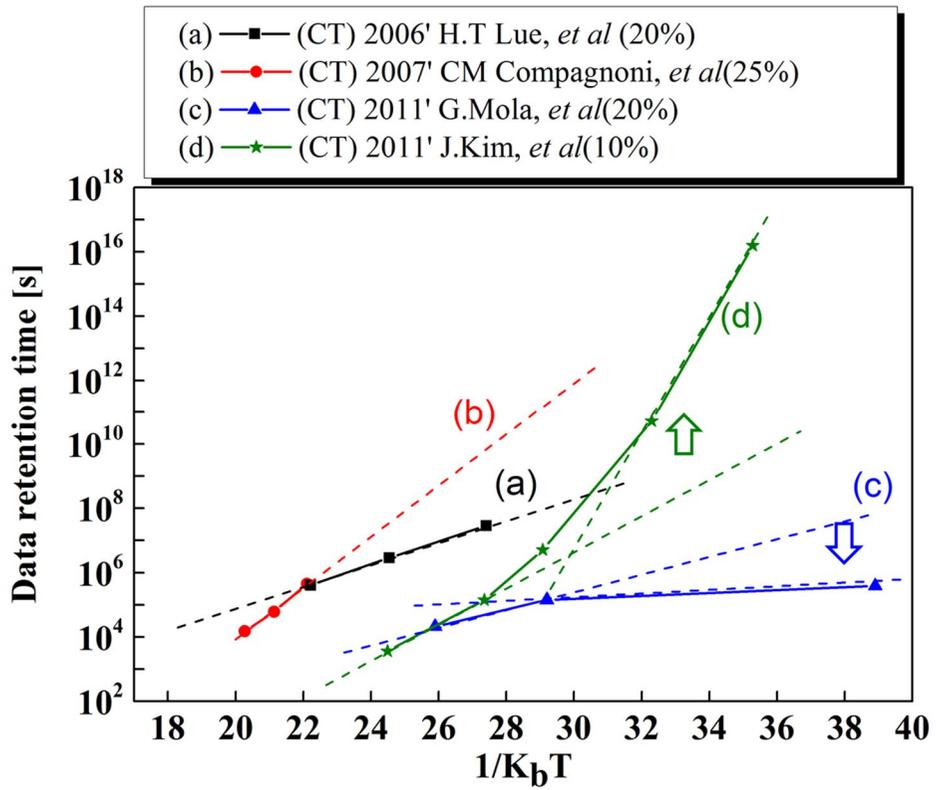


Fig. 4-2 Arrhenius plots of CT Flash memory which are extracted from reference [47][48][49][50].

4.2 Initial electron distributions

To understand the degradation of the activation energy in low temperature region ($< 400\text{K}$), the SANOS devices with various tunneling layer thickness are simulated. Before conducting the retention simulation, all of the cases are separately programmed to reflect the initial electron distribution of each case on the threshold voltage shift. Fig. 4-3 (a) illustrates the initial trapped electron densities. Naturally, the more tunnel layer is thin, the more tunneling current is high at the tunneling layer interface. Therefore, the trapped electron density of the thinner tunneling layer is higher at the tunneling layer interface. The difference of the initial electron distribution affects tunneling current at the retention situation as shown in Fig. 4-3 (b).

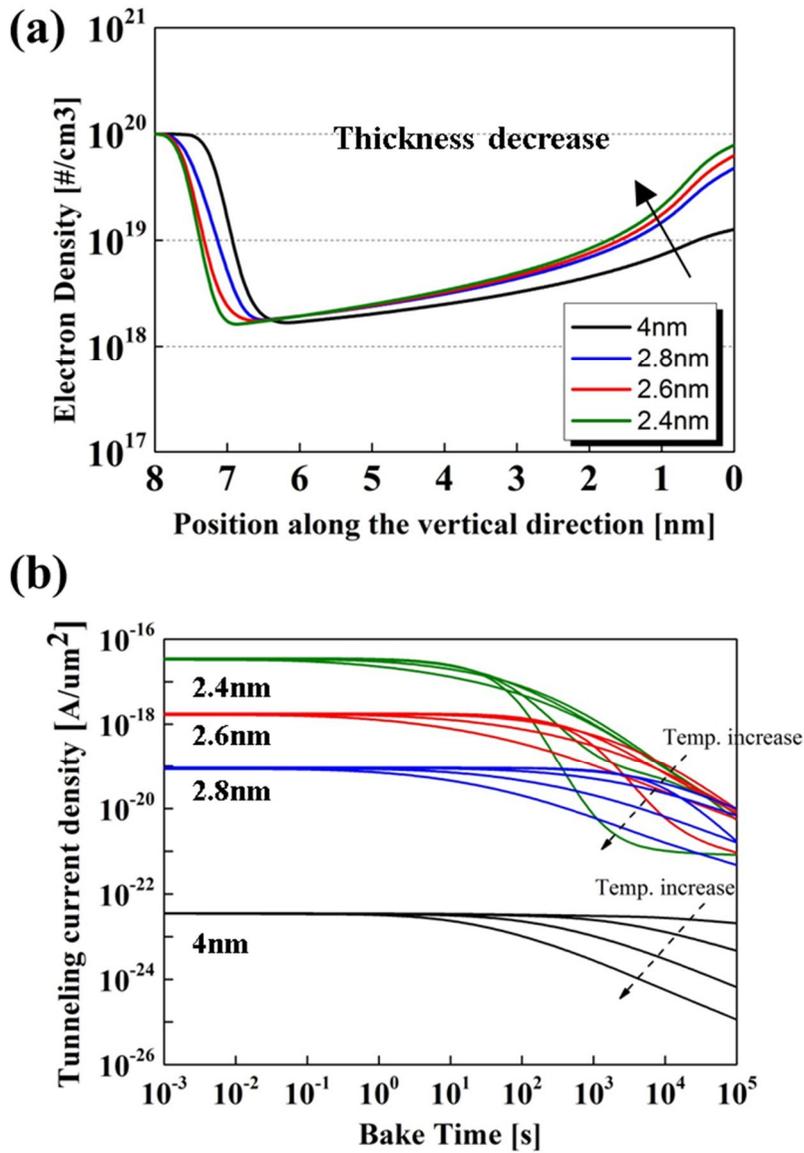


Fig. 4-3. (a) The initial trapped electron distributions along the vertical direction of the various tunnel oxide cases. (b) The tunneling current of the four cases during the bake.

4.3 Modeling of the multiple activation energy

First, we plot the retention vs bake time with the temperature range from 300k to 550k for devices with various tunneling oxide thickness in Fig. 4-4. It is interesting to point that the retention curves for lower temperature are dented for two thin oxide cases. The criteria for the retention time can be defined as the time for 15% charge loss (above the dent) and 20% charge loss (below the dent). For the different criterion, different activation energy can be obtained. It may be also interesting to see the detailed trapped electrons and their loss during the stress at different temperature for various device structures (mainly different tunneling oxide) in Fig. 4-4. We define region-1, 2 and 3 (defined later in Fig. 4-8) according to the different slope in the Arrhenius plot.

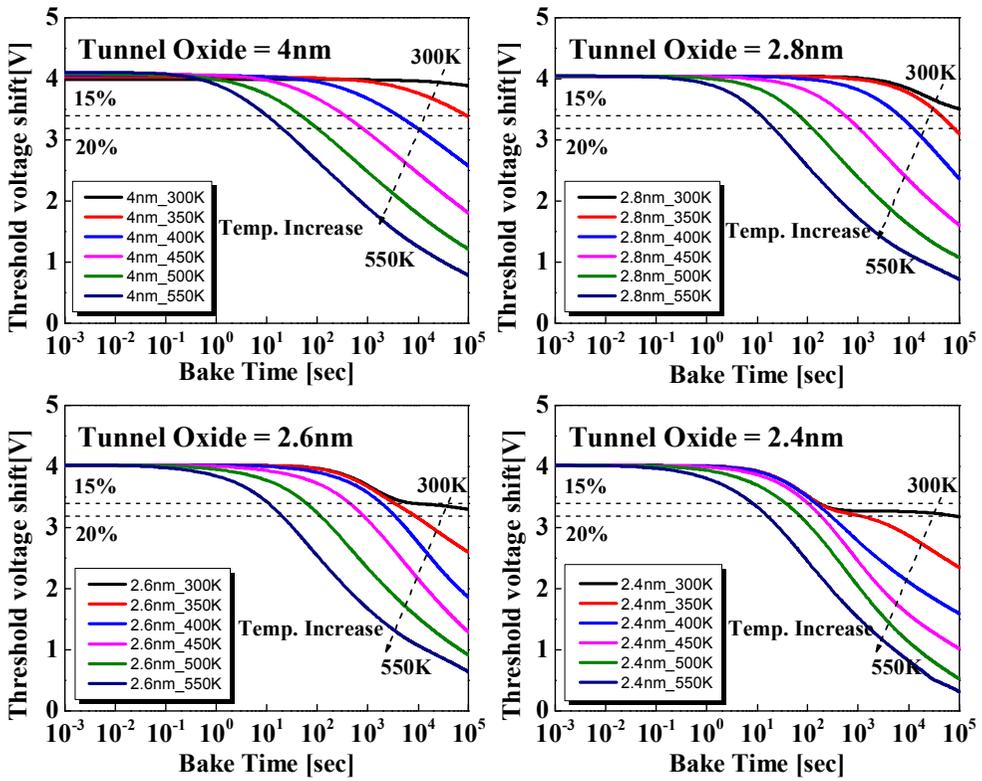


Fig. 4-4 Simulated retention curves of the infinitely extended structure with (a) 4nm tunnel oxide, (b) 2.8nm tunnel oxide, (c) 2.6nm tunnel oxide and (d) 2.4nm tunnel oxide.

4.3.1 Criterion above the dent point (15% charge loss criterion)

The data retention time defined by the 15% charge loss criterion vs. $1/kT$ is depicted in Fig. 4-5. The trend shows that the device with the oxide of 4nm has a unique slope while thinner devices have the varying slope according to the temperature range. It is known that the Arrhenius plot shows a unique activation energy when the thermal emission is a dominant charge loss mechanism [42]. As the tunneling current becomes important for the cases of thinner oxide, the slope of the Arrhenius plots decreases in the low temperature region. Thus, the different activation energy may be defined according to the different temperature regions.

The high temperature region which has the same activation energy as the tunneling free case (4nm oxide) is defined as Region-1. Since the thermal emission and lateral transport is faster than the tunneling charge loss, the data retention time is determined by the temperature in this region. The transient trapped electrons distribution in Region-1 (500K for 2.4nm oxide) is shown in Fig. 4-6 and Fig. 4-7. The temperature region showing reduced activation energy is called Region-2. Fig. 4-8 and Fig. 4-9 show that the lateral charge transport and the tunneling charge loss in the Region-2. In this region, the direct tunneling, which is insensitive to the temperature, becomes more dominant than the thermal emission and lateral transport. Therefore, the thinner tunneling oxide causes more charge loss and reduces more data retention times. The independency of the regions on the temperature is also more obvious in the thinner tunneling oxide because the impact of the tunneling current become larger. The distribution of the trapped electrons along the vertical

direction is shown in Fig. 4-7 and Fig. 4-9.

This tunneling effect on the activation energy agrees with the trend reported in [43] for the FG structure where the charge loss through the IPD layer is the dominant charge loss mechanism. Thus, the degradation of the activation energy reported in [49] can be explained by the tunneling leakage. The effect of the tunneling oxide layer thickness on the activation energy is already reported in ref [35]. However, the initial charge distribution depending on the tunneling layer thickness and charge redistribution during the bake were neglected.

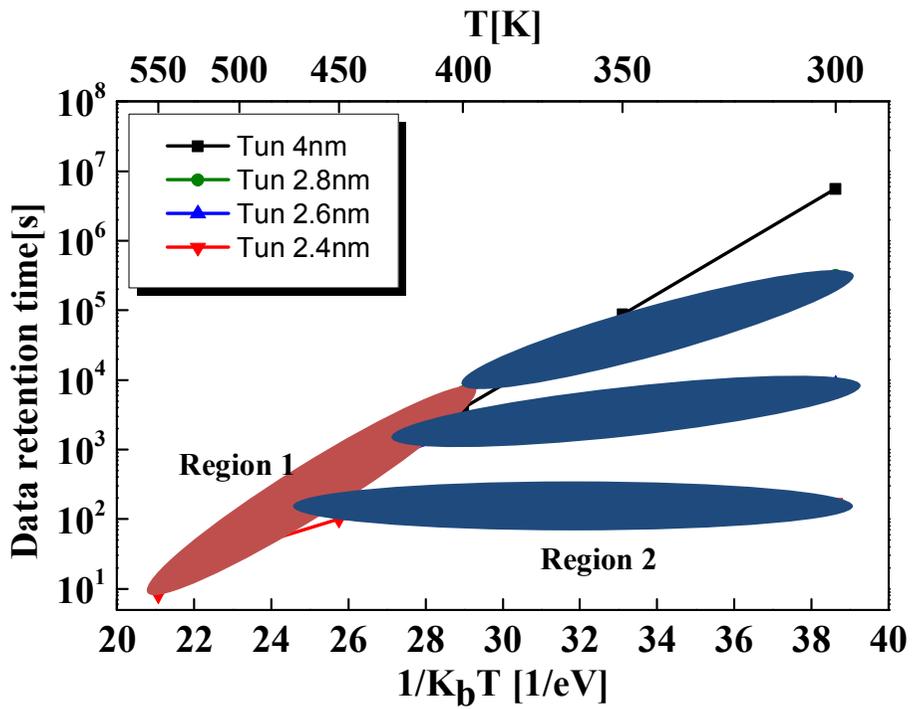


Fig. 4-5. Simulated Arrhenius plots when criterion is 15% shift of the threshold voltage.

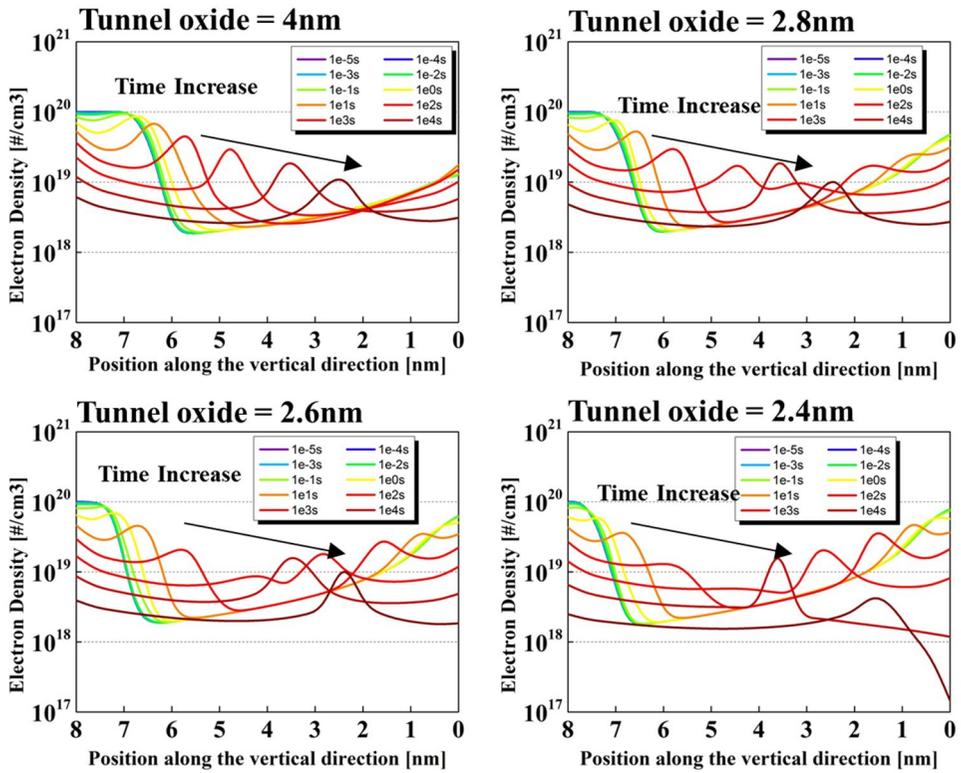


Fig. 4-6. Transient trapped electron distributions along the vertical direction of the various tunnel oxide cases during the retention situation at 500K temperature.

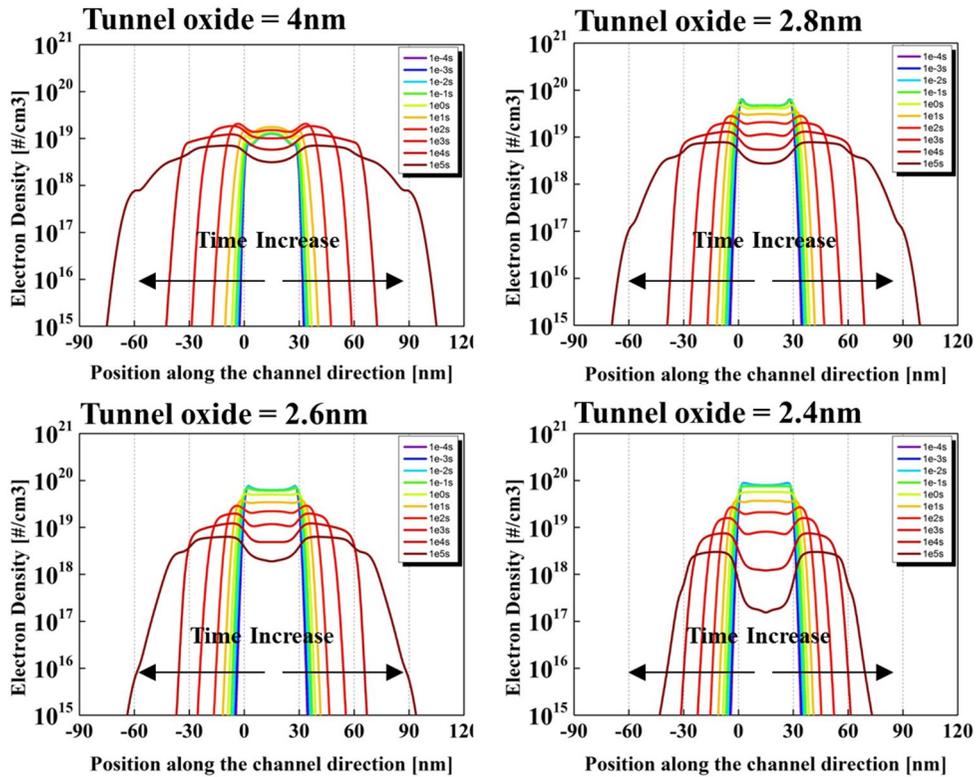


Fig. 4-7. Transient trapped electron distributions along the channel direction during the retention situation at 500K temperature at the tunneling layer interface of the various tunnel oxide cases.

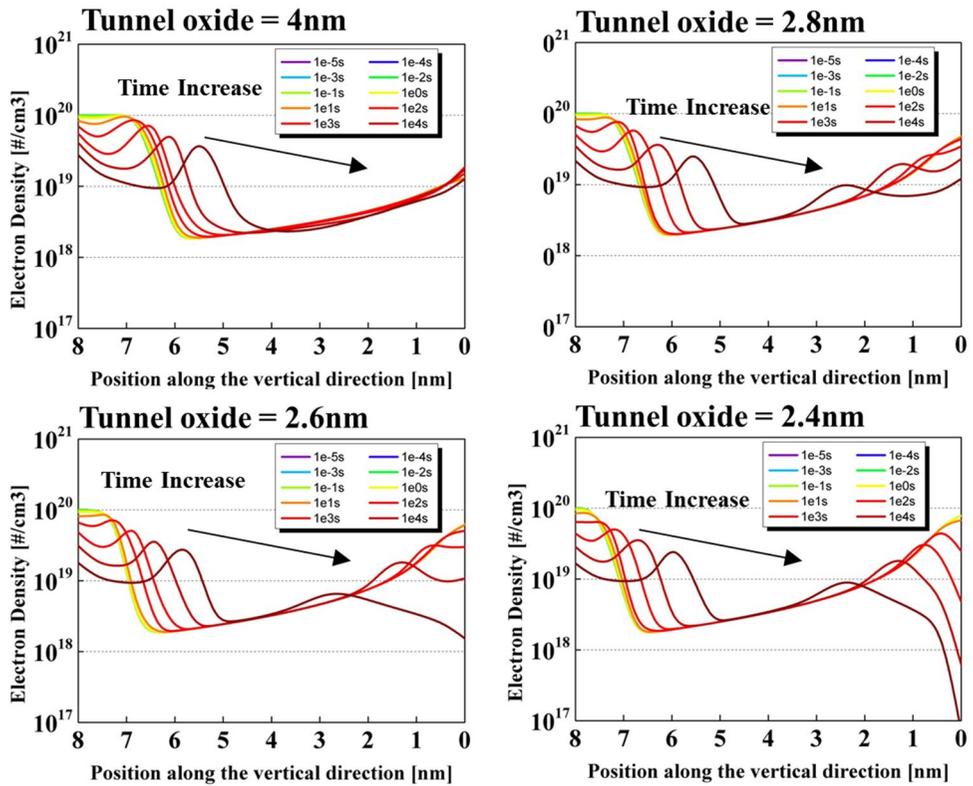


Fig. 4-8. Transient trapped electron distributions along the vertical direction of the various tunnel oxide cases during the retention situation at 400K temperature.

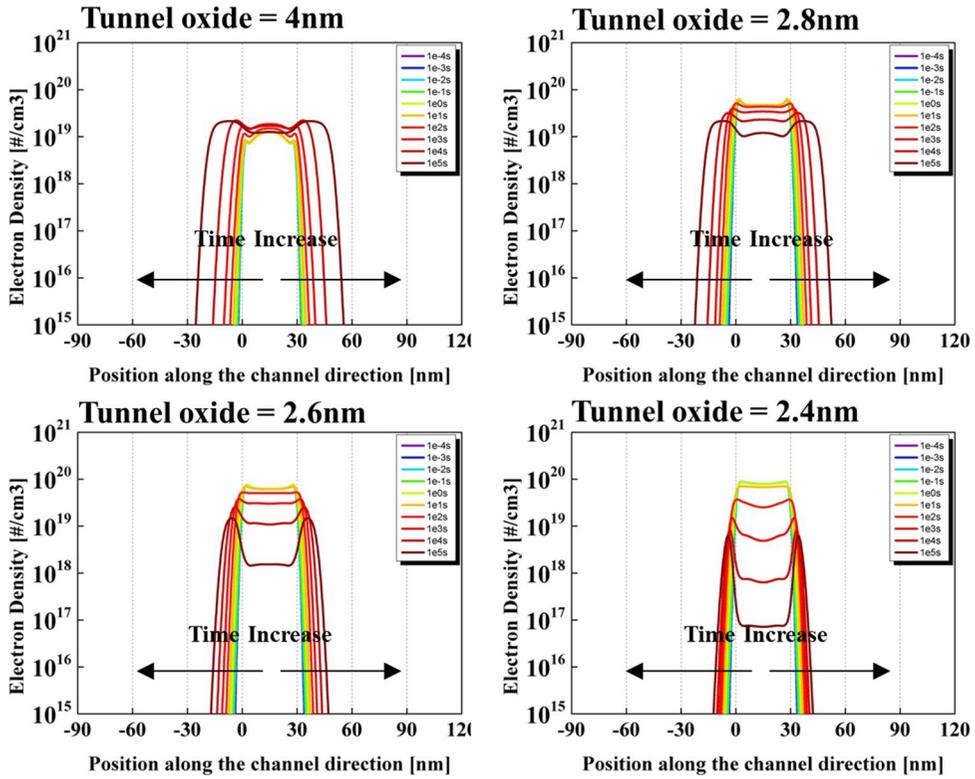


Fig. 4-9. Transient trapped electron distributions along the channel direction during the retention situation at 400K temperature at the tunneling layer interface of the various tunnel oxide cases.

4.3.2 Criterion below the dent point (20% charge loss criterion)

The Arrhenius plots with the 20% charge loss criterion are depicted in Fig. 4-10. Unlike the case of the 15% charge loss criterion, the Arrhenius plots can be divided into three regions according to the slope (the activation energy). Region-1 and Region-2 have the same meaning as the case of 15% criterion. Interestingly, the activation energy in Region-3 restores back to that in Region-1. This is one of the effects caused by the electron transport on the activation energy. The electron distributions shown in Fig. 4-11 and Fig. 4-12 clearly explain the raising of the activation energy in the Region-3. As the trapped electrons tunnel out, the electrons near the interface are depleted and eventually become negligible. This is shown as the dent point in Fig. 4-4. Neither the electron transport to the interface nor charge loss to the uncovered gate region is suppressed at the low temperature. Consequently, the additional charge loss is determined by the electron transport only, and the activation energy increases up to the tunneling free activation energy in Region-1. The electron density distributions of the three regions are summarized in Fig. 4-14.

The transition temperature between regions varies with the life time criterion and the tunneling oxide thickness. In the 15% criterion, since the trapped electrons are able to afford the threshold voltage shift, Region-3 is not shown. The increase of the activation energy in [50] can be explained by the electron transport model. As the result, a prediction of the lifetime in the room temperature changes depending on the stress temperature because of the multiple activation energies. The comparisons of the data life times and activation energies between 15% and 20% criterions is shown

in Fig. 4-13. If the activation energy is extracted from the high temperature stress condition even though the tunneling leakage exists, the prediction of the lifetime would be overestimated. On the other hand, if the lifetime is extrapolated from Region-2, the retention time could be seriously underestimated. Therefore, the consideration of the multiple activation energies is essential for an accurate prediction of the reliability of the CT Flash memory.

4.4 Conclusion

To investigate the activation energy of the CT Flash memory, the SANOS structures with various tunneling oxide thickness are simulated in several temperatures considering vertical and lateral charge transport.. The Arrhenius plot which has a tunneling leakage path is divided into three regions, resulting in multiple activation energies. In high temperature region (Region-1), the slope of the plot is same with the structure without leakage path. In the middle temperature region (Region-2), the data life time rarely depends on the temperature due to the tunneling leakage. In the low temperature region (Region-3), the slope is increased up to the activation energy of the leakage free case. The regions are varied with the life time criterion and the tunneling oxide thickness.

Using this model, the decrease and increase of the activation energy according to the stress temperature condition can be understood with the tunneling leakage and electron transport effect, respectively. Therefore, understanding of the multiple activation energy is necessary for an accurate prediction of the retention time.

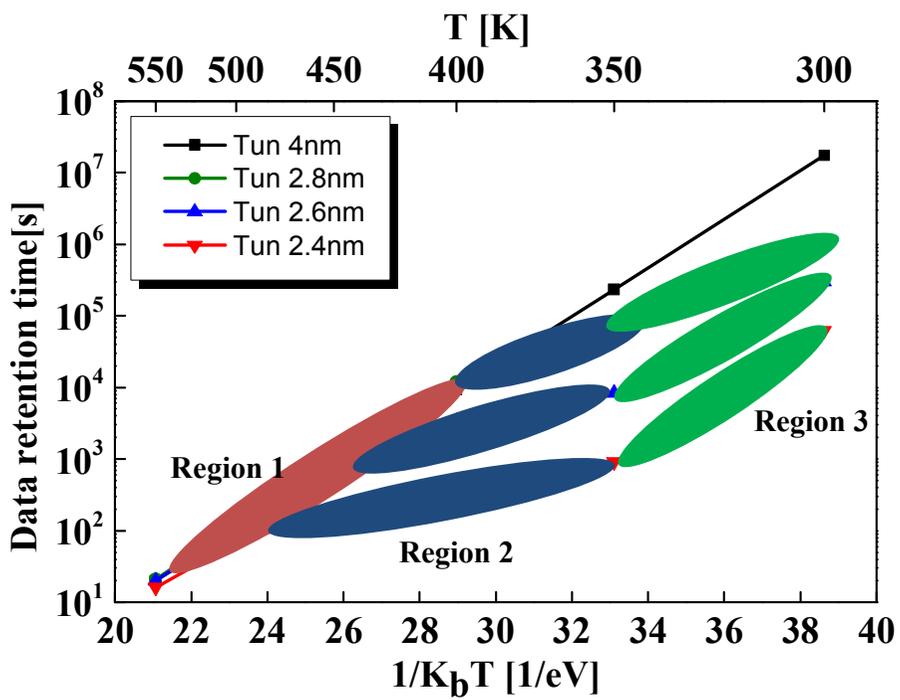


Fig. 4-10 Simulated Arrhenius plots when criterion is 20% shift of the threshold voltage.

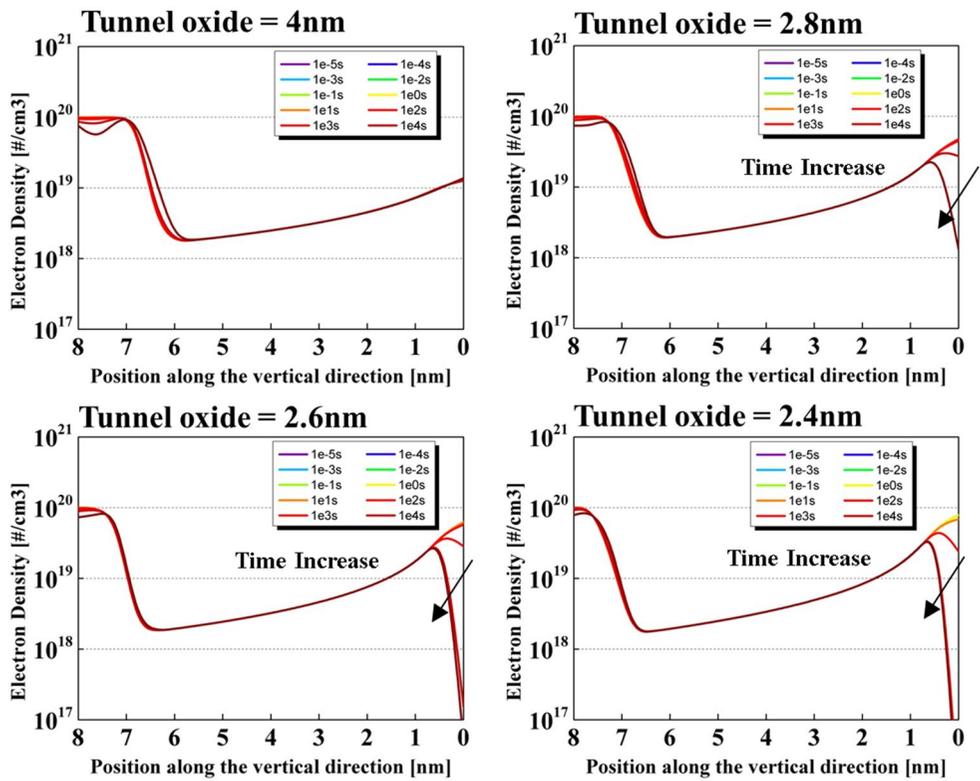


Fig. 4-11. Transient trapped electron distributions along the vertical direction of the various tunnel oxide cases during the retention situation at 300K temperature.

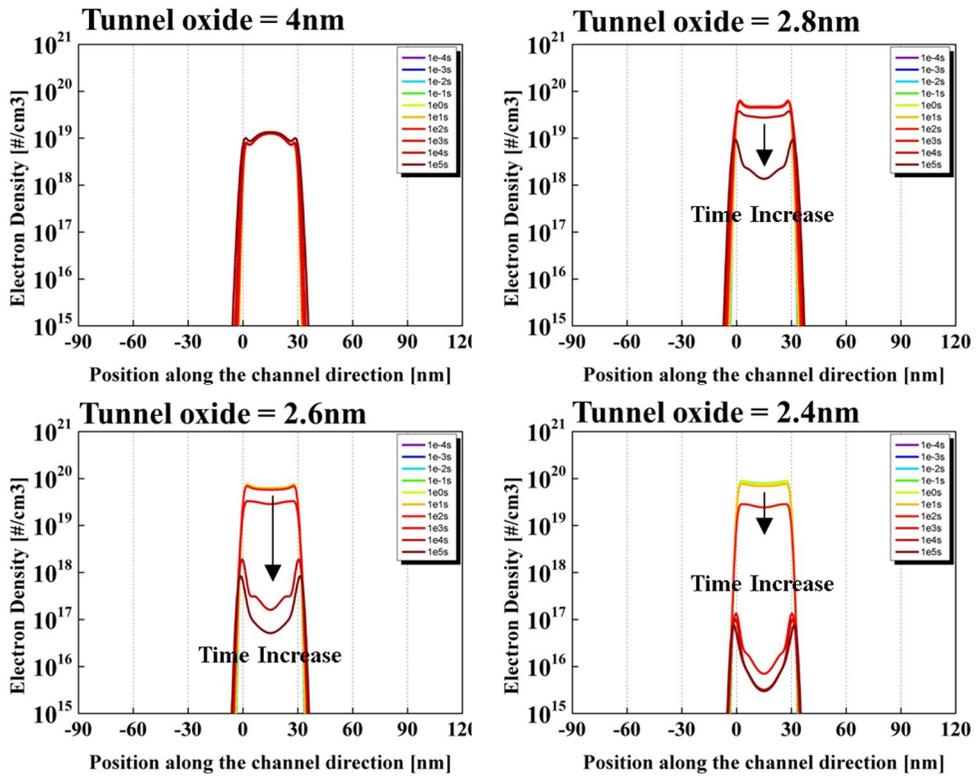


Fig. 4-12. Transient trapped electron distributions along the channel direction during the retention situation at 300K temperature at the tunneling layer interface of the various tunnel oxide cases.

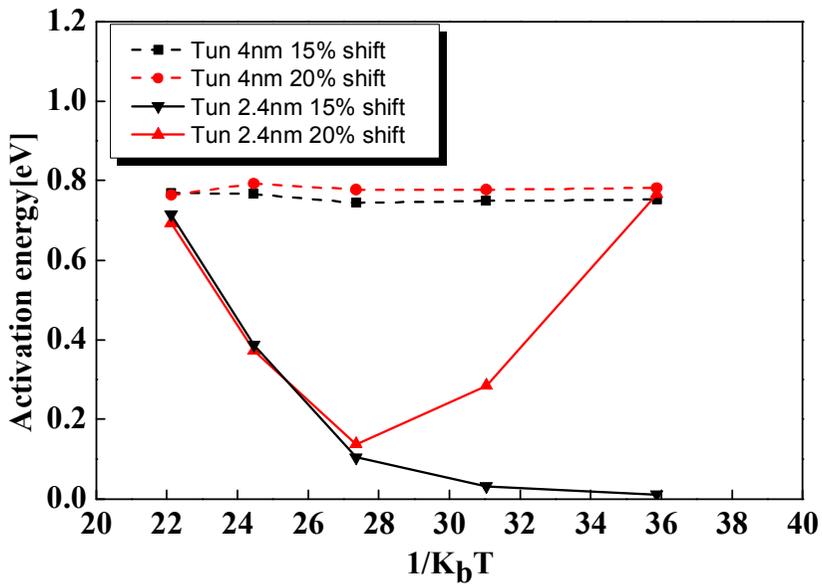
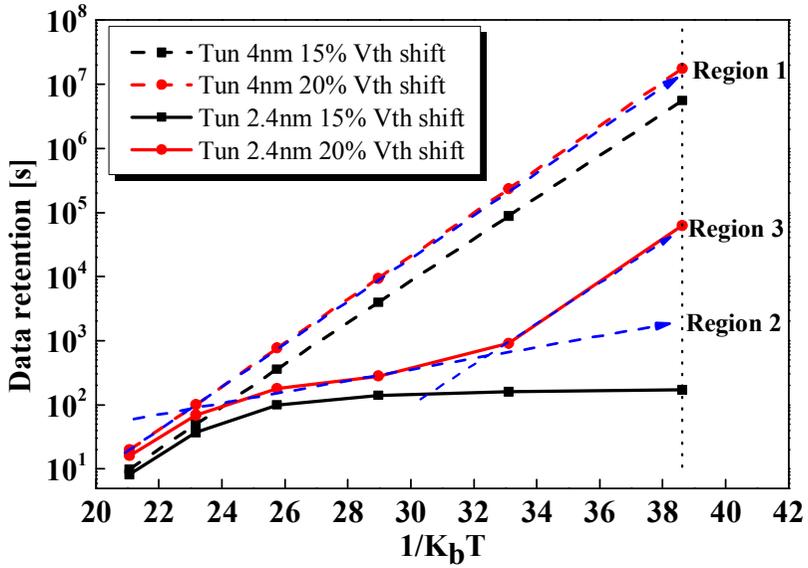


Fig. 4-13. Comparisons of the data life times and activation energies between 15% and 20% criterions.

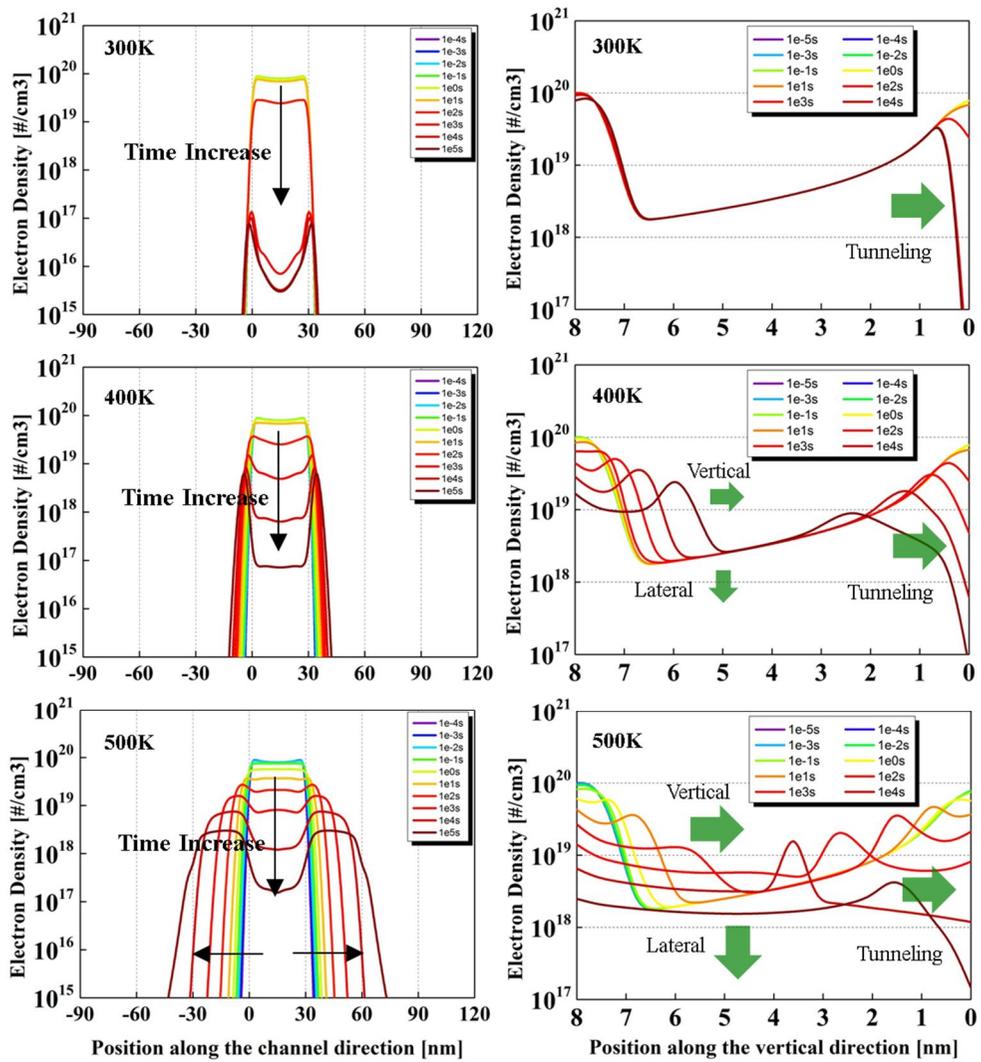


Fig. 4-14. Transient trapped electron distributions of the 2.4nm tunneling layer as baking in various temperatures.

Chapter 5

Conclusions

5.1 Summary

Using an in-house developed self-consistent 3D simulator, the charge transport effects on the retention characteristics of the CT flash memory have been studied. For the first time, the dependency of the extended nitride length, programmed pattern of the neighbor cells and retention cycle are studied for the SANOS device structure. From the simulation results, the extended nitride length is confirmed as an important factor in designing the memory cell device. It is also clearly shown that the surround cells have to be considered in analyzing the retention characteristics because the status of the nearby cells affects the transport property of the monitoring cell. Finally, it has been found that the transported electrons to the extended nitride region affect the retention characteristics in the neighbor cell.

The activation energy of the CT Flash memory associated with the retention

characteristics is also investigated to understand various activation energies in different temperature ranges in the Arrhenius plots. Four thicknesses of the tunneling oxide are simulated in various temperature ranges and the data retention times are plotted against the $1/kT$. Arrhenius plots are divided into three temperature regions, resulting in multiple activation energies. In Region-1, because the thermal emission and lateral charge loss is the main mechanism of the threshold voltage, the slope of the plot is the same as the tunneling free case. Region-2 is defined as the temperature region in which direct tunneling is dominant. Since tunneling current is insensitive to the temperature, the plot is suppressed in this region. The slope of Region-3 returns to that of Region-1. When the trapped electrons near the tunneling layer are not enough to make the threshold voltage shift up to the charge loss criterion for the retention failure, additional charge loss occurs after electrons are transported to the tunneling interface or the outside of the gate region. This is the reason why the slope of the Arrhenius plot follows that of Region-1.

Consequently, in analyzing the retention characteristics of the CT Flash memory, charge loss taken place in three-dimensional transport must be considered for more accurate analysis.

5.2 Future work

5.2.1 Charge loss path through the blocking layer

In this thesis, it is assumed that the blocking layer is thick enough to prevent the back tunneling to the gate. When the blocking layer is made by the high- k material,

there are considerable trap sites and it may contribute to the a leakage path. Therefore, for more accurate simulation, the back tunneling toward the gate region has to be considered for in the programming and reteintion mode operations.

5.2.2 Modeling for more accurate expectation of the life time

As shown in Chapter 4, the prediction of the retention time by extrapolating the life time from the measment data only in a limited high temperature is not accurate for most of the situation. Therefore, the new model which considers the tunneling effect and the electron transport has to be developed.

5.2.3 3D Monte-Carlo Simulation

Many previous retention modeling works including this thesis are based on a continuum simulations regarding the electrons and traps as continuous density parameters. For a large scale device, it is reasonable to use the average parameter. In the case of the small device, however, one electron or one trap may give appreciable impacts to device characteristics. If the trap density is $10^{19}/\text{cm}^3$, for example, there would be 0.01 trap in 1nm^3 cube. The nitride layer with the $10 \times 10 \times 5\text{nm}$ volume has only five traps. As a result, consideration of the discretized trap and electron is inevitable for the simulation of the next generation memory devices. In simulating the retention characteristics considering the discretized electron and trap, the Monte-Carlo simulation is one of the powerful methods to model the electron motion in a discrete manner.

For these reasons, it may be recomandable to develop the 3D Monte-Carlo

framework and compare with the simulation results based on the continuum assumption such as the one in this thesis. The comparison for the dependency of the extended nitride, programmed pattern and retention cycle would become important in the future scaled memory cell. The recommended simulation sequence of the 3D Monte-Carlo simulation for the CT memory cell is shown in Fig. 5-1. Using the 3D Monte-Carlo framework, more reliable expectation of the retention characteristics would be possible for the very small size CT device. Comparison between the simulation data in this thesis and the Monte-Carlo simulation is a future work for the more accurate prediction of the CT memory cell characteristics.

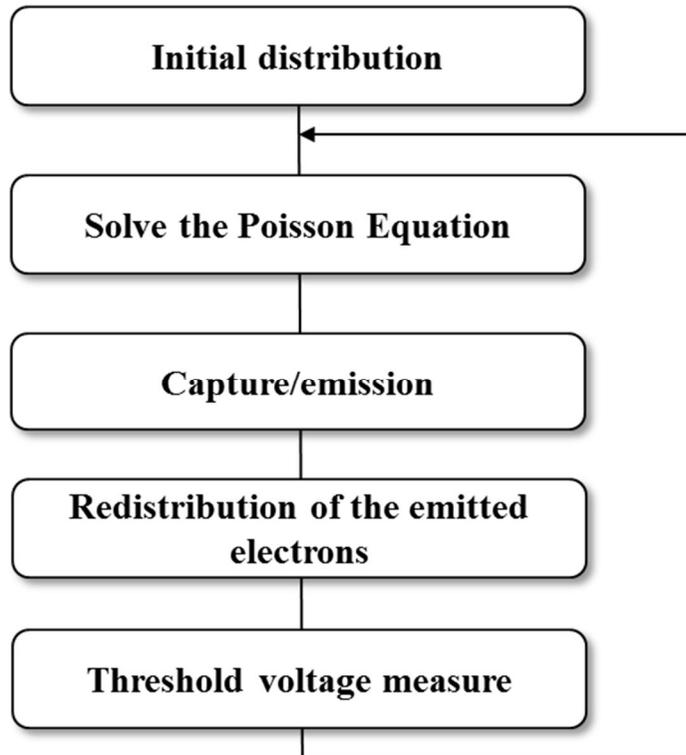


Fig. 5-1 The Monte-Carlo simulation sequence for the retention characteristic of the CT Flash memory.

References

- [1] Kwang Sun Jeon, “A Study on the Spatial Trap Profile in the TANOS Device using the Numerical Simulation and Charge Pumping Method”, *Ph.D Dissertation, School of Electrical Engineering, Seoul National University, Korea, 2011.*
- [2] *International Technology Roadmap for Semiconductors*. Semiconductor Industry Association, [Online] Available: <http://www.itrs.net/>
- [3] Jae-Duk Lee, Sung-Hoi Hur, Jung-Dal Choi "Effects of floating-gate interference on NAND flash memory cell operation", *IEEE Electron Device Letters*, Vol. 23(5), pp. 264–266, 2002
- [4] S.-M. Jung, J. Jang, W. Cho, H. Cho, J. Jeong, Y. Chang, J. Kim, Y. Rah, Y. Son, J. Park, M.-S. Song, K.-H. Kim, J.-S. Lim, and K. Kim. “Three dimensionally stacked nand flash memory technology using stacking single crystal si layers on ild and tanos structure for beyond 30nm node”. *Electron Devices Meeting, 2006. IEDM '06. International*, pages 1–4, 11-13 Dec. 2006.
- [5] T. Endoh, K. Kinoshita, T. Tanigami, Y. Wada, K. Sato, and K. Yamada, "Novel ultrahigh density Flash memory with a stacked-surrounding gate transistor (S-SGT) structural cell", *IEDM Tech. Dig.*, pp.33 -36 2001
- [6] H. Tanaka , M. Kido , K. Yahashi , M. Oomura , R. Katsumata , M. Kito , Y. Fukuzumi , M. Sato , Y. Nagata , Y. Matsuoka , Y. Iwata , H. Aochi

- and A. Nitayama "Bit cost scalable technology with punch and plug process for ultra high density Flash memory", *VLSI Symp. Tech. Dig.*, pp.14 2007
- [7] J. Jang , H. S. Kim , W. Cho , H. Cho , J. Kim , S. I. Shim , Y. Jang , J. H. Jeong , B. K. Son , D. W. Kim , K. Kim , J. J. Shim , J. S. Lim , K. H. Kim , S. Y. Yi , J. Y. Lim , D. Chung , H. C. Moon , S. Hwang , J. W. Lee , Y. H. Son , U. I. Chung and W. S. Lee "Vertical cell array using TCAT (Terabit Cell Array Transistor) technology for ultra high density nand flash memory", *VLSI Symp. Tech. Dig.*, pp.192 -193 2009
- [8] F. Masuoka, M. Asano, H. Iwahashi, T. Komuro, S. Tanaka, "A new flash E2PROM cell using triple polysilicon technology," *Electron Devices Meeting*, 1984 International , vol.30, no., pp. 464- 467, 1984
- [9] P. Pavan, R. Bez, P. Olivo, and E. Zanoni, "Flash memory cells-an overview," *Proceedings of the IEEE* , vol.85, no.8, pp.1248-1271, Aug 1997
- [10]M. Lenzlinger, "Fowler-Nordheim tunneling into thermally grown SiO₂", *Journal of Applied Physics*, vol.40, No., 1pp278-283, Jan. 1969.
- [11] N.Naruke, S.Taguchi, and M.Wada, "Stress Induced Leakage Current Limiting to Scale Down EEPROM Tunnel Oxide," *IEDM Technical Digest*, p. 424, 1988
- [12] D. J. DiMaria and E. Cartier, "Mechanism for stress-induced leakage currents in thin silicon dioxide films" *Journal of Applied Physics*, vol. 78, no. 6, pp. 3883-3894, 1995
- [13] M. Bauer, R. Alexis, G. Atwood, B. Baltar, A. Fazio, K. Frary, M. Hensel, M. Ishac, J. Javanifard, M. Landgraf, D. Leak, K. Loe, D. Mills, P. Ruby, R. Rozman, S. Sweha, S. Talreja, and K. Wojciechowski, "A multilevel-cell 32 Mb Flash memory", *IEEE ISSCC Dig. Tech. Pap.*, vol.

351, pp.132 -133, 1995

- [14] Tae-Kyung Kim, Sungnam Chang, Jeong-Hyuk Choi, "Floating gate technology for high performance 8-level 3-bit NAND flash memory", *Solid-State Electronics*, Volume 53, Issue 7, Pages 792–797, July 2009
- [15] Y. Li, *et al*, "A 16 Gb 3-bit per cell (X3) nand flash memory on 56 nm technology with 8 MB/s write rate", *IEEE J. Solid-State Circuits*, vol. 44, no. 1, pp.195 -207 2009
- [16] H. Nobukata, S. Takagi, K. Hiraga, T. Ohgishi, M. Miyashita, K. Kamimura, S. Hiramatsu, K. Sakai, T. Ishida, H. Arakawa, M. Itoh, I. Naiki, Noda, M, "A 144 Mb 8-level NAND flash memory with optimized pulse width programming", *VLSI Circuits. Digest of Technical Papers. 1999 Symposium*, 1999
- [17] Henry C. Pao and M. O'Connell, "Memory behavior of an MNS capacitor", *Appl. Phys. Lett.* 12, 260, 1968
- [18] H. A. R. Wegener, A.J. Lincoln, H.C. Pao, M.R. O'Connell, R.E. Oleksiak, H. Lawrence, "The variable-threshold transistor, a new electrically-alterable, nondestructive read-only storage device", *IEEE IEDM Tech. Dig*, 1967.
- [19] T.L. Chu, J.R. Szedon, C.H. Lee, "The preparation and C-V characteristics of Si Si₃N₄ and Si SiO₂ Si₃N₄ structures", *Solid-State Electronics*, Volume 10, Issue 9, Pages 897–905, September 1967
- [20] B. V. Keshevan and H. C. Lin, "MONOS memory element," *IEEE IEDM Tech. Dig*, p. 140, 1968.
- [21] E. Suzuki, H. Hiraishi, K. Ishii, Y. Hayashi, "A low-voltage alterable EEPROM with metal-oxidenitride-oxide-semiconductor (MONOS) structures", *IEEE Trans. Electron Devices*, vol. ED-30, p. 122, 1983.

- [22] P. Chen, "Threshold-alterable Si-gate MOS devices", *IEEE Trans. Electron Devices*, vol. ED-24, p. 584, 1977
- [23] C.H. Lee, K.I Choi, M.K. Cho, Y.H. Song, K.C. Park, K. Kim, "A Novel SONOS Structure of SiO₂/SiN/Al₂O₃ with TaN metal gate for multi-giga bit flash memories," *IEDM Tech. Dig.*, pp. 613-616, 2003.
- [24] JungDal Choi, "Advanced Flash Technology", *Lecture note of the SNU-SEC Lectures on Advanced Semiconductor Device and Technology*, 2011 P54~55,
- [25] Y. Fukuzumi, R. Katsumata, M. Kito, M. Kido, M. Sato, H. Tanaka, Y. Nagata, Y. Matsuoka, Y. Iwata, H. Aochi and A. Nitayama, "Optimal Integration and Characteristics of Vertical Array Devices for Ultra-High Density, Bit-Cost Scalable Flash Memory", *IEDM Technical Digest*, pp. 449-452, 2007.
- [26] Yu Wang; Marvin H. White, "An analytical retention model for SONOS nonvolatile memory devices in the excess electron state", *Solid-State Electronics*, vol. 49, pp. 97-107, 2005.
- [27] Y. Yang, Marvin H. White, "Charge retention of scaled SONOS nonvolatile memory devices at elevated temperatures", *Solid-State Electronics*, vol. 44, pp. 949-958, 2000.
- [28] A. Furnemont , M. Rosmeulen , A. Cacciato , L. Breuil , K. De Meyer , H. Maes and J. Van Houdt "A consistent model for the SANOS programming operation", *Proc. Non-Volatile Semicond. Memory Workshop*, pp.96 2007
- [29] E. Vianello , F. Driussi , P. Palestri , A. Arreghini , D. Esseni , L. Selmi , N. Akil , M. van Duuren and D. S. Golubovi, "Impact of the Charge Transport in the Conduction Band on the Retention of Si–Nitride Based Memories", *ESSDERC*, pp. 107-110, 2008.

- [30] E. Vianello, F. Driussi, A. Arreghini, P. Palestri, D. Esseni, L. Selmi, N. Akil, M. J. van Duuren, and D. S. Golubovic, "Experimental and simulation analysis of program/retention transients in silicon nitride-based NVM cells," *IEEE Trans. Electron Devices*, vol. 56, no. 9, pp. 1980–1990, Sep. 2009."
- [31] C.S. Kang, J.D. Choi, J.S. Sim, C.H. Lee, Y.C. Shin, J.T. Park, J.S. Sel, S.H. Jeon, Y.W. Park, and K.N. Kim, "Effects of lateral charge spreading on the reliability of TANOS (TaN/AlO/SiN/Oxide/Si) NAND flash memory," *IRPS*, pp. 167-170, 2007."
- [32] E. Vianello, E. Nowak, D. Mariolle, N. Chevalier, L. Perniola, G. Molas, J. P. Colonna, F. Driussi, and L. Selmi, "Direct probing of trapped charge dynamics in SiN by Kelvin force microscopy", 2010 *IEEE International Conference on Microelectronic Test Structures*, pp. 94-97, 2010"
- [33] M. Florian Beug, Thomas Melde, Jan Paul, Roman Knoefer, "TaN and Al₂O₃ Sidewall Gate-Etch Damage Influence on Program, Erase, and Retention of Sub-50nm TANOS NAND Flash Memory Cell", *Electron Devices, IEEE Transactions on*, vol. 58, no.6, 2011.
- [34] A. Furnémont, A. Cacciato, L. Breuil, M. Rosmeulen, H. Maes, K. De Meyer, and J. Van Houdt, "Physical understanding and modeling of SANOS retention in programmed state," *Solid-State Electronics*, vol. 52, pp. 577–583, 2008."
- [35] A. Maconi, A. Arreghini, C.M. Compagnoni, G. Van den bosch, Spinelli, "Impact of lateral charge migration on the retention performance of planar and 3D SONOS devices", *proc. ESSDERC* , pp. 195-198., 2011
- [36] A. Arreghini, N. Akil, F. Driussi, D. Esseni, L. Selmi, and M.J. Van Duuren, "Long term charge retention dynamics of SONOS cells", *Solid-*

- State Electronics*, vol. 52, pp. 1460-1466, 2008.
- [37] B. DeSalvo, G. Ghibaudo, G. Pananakakis, B. Guillaumot, P. Candelier, F. Mondon, and G. Reimbold, "Experimental and theoretical investigation of nonvolatile memory data-retention", *IEEE Trans. Electron Devices*, vol. 46, pp.1518 1999
- [38] R.M.Hill, "Poole-Frenkel Conduction in Amorphous Solids", *Philos. Mag.*, vol. 23, pp. 59-86,1971
- [39] W.Shockely and W.T.Read Jr,"Statistics of the Recombinations of Holes and Electronics," *Phys.Rev.*,vol.87, no.5,pp 835-842,Sep.1952.
- [40] J. S. Sim , J. Park , C. Kang , W. Jung , Y. Shin , J. Kim , J. Sel , C. Lee , S. Jeon , Y. Jeong , Y. Park , J. Choi and W.-S. Lee "Self aligned trap-shallow trench isolation scheme for the reliability of TANOS (TaN/AlO/SiN/Oxide/Si)", *Proc. Non-Volatile Semicon. Memory Workshop*, pp.110, 2007
- [41] Bocquet M, Vianello E, Molas G, Perniola L, Grampeix H, Martin F et al. An indepth investigation of physical mechanisms governing SANOS memories characteristics. *In: Proc. of IMW* , Monterey (USA), p. 60–3., 2009
- [42] H.Nozawa, S. Kohyama, "A thermionic electron emission model for charg retention in samos structures", *Jpn. J. Appl. Phys.*, vol. 21, no. 2, pp. L111, 1982.
- [43] B. D. Salvo, G. Ghibaudo, G. Pananakakis, G. Reimbold, F. Mondond, B. Guillaumot, and P. Candelier, "Experimental and theoretical investigation of nonvolatile memory data-retention", *IEEE Trans. Electron Devices*, vol. 46, pp.1518 -1524 1999
- [44] B. Govoreanu, J. Van Houdt, "On the Roll-Off of the Activation Energy Plot in High-Temperature Flash Memory Retention Tests and its Impact

on the Reliability Assessment", *IEEE Electron device letter*, VOL. 29, NO. 2, FEBRUARY 2008

- [45] J. Postel-Pellerin, F. Lalande, P. Canet, R. Bouchakour, F. Jeuland, L. Morancho, "Modeling charge variation during data retention of MLC Flash memories", *Microelectronics Reliability*, 49, 1060–1063, 2009
- [46] S. Kumar, and E.L. Russell, "Model to predict ONO non-volatile memory", *Integrated Reliability Workshop (IRW)*, pp. 34-40, 2001.
- [47] H. T. Lue, S. Y. Wang, Y. H. Hsiao, E. K. Lai, L. W. Yang, T. Yang, K. C. Chen, K. Y. Hsieh, R. Liu, and C. Y. Lu, "Reliability Model of Bandgap Engineered SONOS (BE-SONOS)" *Tech. Digest 2006 International Electron Devices Meeting*, session 18-4, 2006.
- [48] C. Compagnoni, A. S. Spinelli and A. L. Lacaita "Experimental study of data retention in nitride memories by temperature and field acceleration", *IEEE Electron Device Lett.*, vol. 28, pp.628 2007
- [49] G. Molas, J.P. Colonna, R. Kies, D. Belhachemi, M. Bocquet, M. Gely, V. Vidal, P. Brianceau, E. Martinez, A.M. Papon, C. Licitra, L. Vandroux, G. Ghibaudo, B.D. Salvo, "Investigation of charge-trap memories with AlN based band engineered storage layers", *Solid-State Electronics*, 58 (2011) 68–74
- [50] J. Kim, C. Kang, S.-I. Chang, J. Kim, Y. Jeong, C. Park, J.-H. Kang, S.-H. Kim, S. Hwang, B.-I. Choe, J. Park, J. Chung, Y. Park, J. Choi, C. Chung, "New phenomena for the Lifetime Prediction of TANOS based Charge Trap NAND Flash Memory", *Device Research Conference (DRC)*, pp. 99-100, 2010.

초 록

본 논문에서는 자체적으로 개발된 삼차원 소자 시뮬레이터를 이용하여 전자의 수직, 수평적 이동을 고려한 소자의 보유특성에 대한 연구를 기술하였다. 본 시뮬레이션에서는 SA-STI(Self-Aligned Shallow Trench Isolation) 구조의 SANOS(Si/Al₂O₃/Si₃N₄/SiO₂/Si) 구조를 기본으로 시뮬레이션을 진행하였다. 초기의 전하의 분포 상태를 가정하지 않고 시뮬레이션을 이용하여 변화된 문턱 전압이 4V 가 될때까지 자체적으로 일관성있게 전자의 이동이 고려된 쓰기동작을 하였다. 이로부터 초기 전하의 분포상태를 얻은 후 모든 보유특성에 대한 시뮬레이션이 진행되었다.

첫번째로 실리콘 질화막의 게이트 영역 밖으로 확장된 영역의 길이에 대한 의존도를 확인하기 위해서 여러길이(0nm, 15nm, 30nm, 150nm)의 소자를 시뮬레이션 하였다. 둘째로 주변 메모리 셀들의 쓰여진 상태에 의해서 어떠한 영향을 받는지에 대한 연구를 위해서 체크보드 패턴으로 주변 셀을 쓰기 상태로 만든 상황과 모든 셀들이 모두 쓰기 상태로 된 상황에 대해서 보유특성 시뮬레이션을 진행하였다. 세번째로 여러 번의 보유 동작을 반복할경우 보유특성의 변화정도를 확인하기 위하여 쓰기/보유/지우기를 연속적으로 시뮬레이션 하였다.

마지막으로 실온에서의 보유가능 기간을 예측하기 위한 활성화

에너지를 측정함에 있어서 터널링 산화막의 두께에 따른 수직방향으로의 전하 손실이 활성화 에너지에 주는 영향을 시뮬레이션 하였다. 또한, 문헌에 보고된 여러가지 형태의 아레니우스 그래프를 삼차원 시뮬레이션을 통해서 나온 전자의 밀도분포를 기반으로 3 가지 구간으로 나누어 모델링하여 설명하였다.

요약하자면 본 연구에서는 삼차원 SANOS 시뮬레이션을 통해서 실리콘 질화막의 확장 길이에 따른 보유특성의 차이를 확인하여 소자를 디자인함에 있어서 확장길이의 중요성을 강조하였고 주변 셀들의 쓰여진 상태 그리고 쓰기/보유/지움의 반복행위가 보유특성에 주는 영향을 확인하였다. 또한 활성화 에너지를 계산함에 있어서 터널링 누설전류로 인한 손실의 영향성 및 실리콘질화막내에서의 전자의 이동을 고려하여야 정확한 보유기간예측을 할 수 있음을 연구하였고 이를 토대로 여러가지 모양의 아레니우스 그래프를 해석하였다..

주요어 : 낸드 플래시 메모리, 삼차원 시뮬레이션, 전하 포획형 메모리, CT, 리텐션, 활성화 에너지

학 번 : 2011-20837

감사의 글

우선 삼성전자에 들어가면서부터 계속 가지고 있던 '학업을 지속하게 해달라'는 기도제목을 들어주신 하나님께 감사드립니다. 또한, 2년이라는 시간 동안 건강하게 많은 것을 배우고 무사히 졸업할 수 있도록 지켜주셔서 감사합니다.

떨리는 심정으로 물리전자연구실을 찾아왔던 게 엇그제 같은데 벌써 2년이라는 시간이 흘렀습니다. 그동안 세상도 많이 바뀌고 사람들도 바뀌고 저 또한 많이 바뀐 것 같습니다. 예전부터 많이 꿈꾸던 대학원 생활이기 때문에 많은 기대와 설렘을 가지고 시작했었는데 이제는 편안함과 익숙함만 남았습니다. 무언가 새롭고 좀 더 깊은 것에 대한 갈증을 조금은 해결한 것 같아서 한편으로 뿌듯하면서도 이제 시작인 것 같은데 떠나야 해서 아쉬움이 너무나 많이 남습니다. 몸과 마음이 너무 지쳐 있을 때 좋은 기회가 주어졌고 뒤도 안 돌아보고 도망치듯 학교로 나왔던 것 같습니다. 이제 다시 돌아가야 한다는 생각에 많은 걱정이 몰려오지만, 내가 가장 잘 할 수 있고, 나를 가장 필요로 하는 곳에 일을 하러 간다는 생각에 신입사원으로 입사할 때와 같은 떨림을 가지고 씩씩하게 돌아가려 합니다.

많은 분들이 지난 2년동안 물심양면으로 저를 도와주셨습니다. 학업뿐 아니라, 인생에 대해서도 많은 지도를 해주신 박영준 교수님께 먼저 감사드립니다. 삶을 대하는 자세에 대해서 많은 것을 배우고 갑니다. 학형(우) 분들로는 회사에서 먼저 나와서 이런저런 조언을 해주신 혁제형, 석하형, 입학 할 때부터 동고동락한 희중이형, 한 살 터울이라고 잘해주고 가장 힘들 때 큰 힘을 준 진정 고마운 준명이형, 여러 가지 도움을 준 해외여행의 동반자 성욱이, 학교생활 동안 너무나도 가깝게 지냈던 일오와 승만이, 39동의 해결사 재흥이, 부러운 몸매의 석향이, 노벨상 탈호석이, 신비한 인태, 위치 선정에 탁월한 상우, 집까지 데려다 주는 수고를 마다하지 않으시는 주성이형, 짧은 시간이 아쉬운 준현이, 언제나

성실히 열심히 일하는 친한 친구 예진이, 그리고 짧은 시간동안 같이했던 천준호박사님, 박수영박사님, 경석이형, 광선이형, 영준이에게 감사의 마음을 전합니다.

학교를 간다고 했을 때 누구보다 좋아하셨던 아버지, 어머니, 건강하고 올바르게 잘 키워주셔서 너무나 감사합니다. 멀리 남해에서 저를 언제나 자랑스러워 해주시는 장인 장모님! 언제나 힘주셔서 감사합니다. 학교로 나올 때 여러 조언을 해주신 자형, 누나, 효빈이네 가족, 멀리 있어서 자주 못 찾아가서 죄송한 처형, 형님, 은세, 리인이네 가족, 항상 잘 못해줘서 미안한 성욱이 모두 감사합니다. 마지막으로 지난 2년동안 신희임에도 불구하고 항상 늦게 들어오는 나를 이해해주고 언제나 내 건강 걱정을 해주며 내가 하고 싶은 일을 하도록 지원해주는 평생의 나의 동반자, 보보, 귀님 심지현에게 너무나 고맙고 너무나 사랑한다는 말을 전하고 싶습니다.

2013년 2월 301동 1018호에서