國立交通大學

電子工程學系 電子研究所碩士班

碩士論文

非揮發性記憶體的儲存電荷的空間分佈對元件特性的影響

Effects of Spatial Distribution of the Stored-Charge on Device Characteristics of Non-volatile Memory

研 究 生：余昆武

指導教授：崔秉鉞 教授

中華民國九十八年十月

非揮發性記憶體的儲存電荷的空間分佈對元件特性的影響

Effects of Spatial Distribution of the Stored-Charge on Device Characteristics of Non-volatile Memory

研究生：余昆武　　Student : Kun-Wu YU

指導教授：崔秉鉞　　Advisor : Bing-Yue Tsui

國立交通大學

電子工程學系　電子研究所

碩士論文

A thesis

Submitted to Department of Electronics Engineering & Institute of Electronics

College of Electrical and Computer Engineering

National Chiao Tung University

in Partial Fulfillment of the Requirement

for the Degree of Master

in

Electronic Engineering

2009

Hsinchu, Taiwan, Republic of China

中華民國九十八年十月

非揮發性記憶體的儲存電荷的空間分佈對元件特性的影響

研究生：余昆武　　　　　　　　**指導教授：崔秉鉞**

國立交通大學電子工程系　　電子研究所碩士班

摘　要

在本論文中，探討了非揮發性記憶體的儲存電荷的空間分佈對元件特性的影響。根據模擬的結果，調變臨界電壓(V_t)和調變閘極引發汲極漏電流(GIDL)是與電荷儲存的位置有關。當電荷儲存在通道中央的上方位置時，整個次臨界曲線會向右移，這是由於負電荷在捕陷電荷層的正中央會造成通道區域的電子位障上升。當電荷儲存在通道的上方並且靠近汲極的接面時，只有上半部份的次臨界曲線會向右移。這是因為通道區域中的電子位障被汲極電壓稍微地拉低了，所以下半部份的次臨界曲線不會移動，但是電子位障仍然不夠低，不足以讓電子完全地導通，所以上半部份的次臨界曲線會向右移。再來，當電荷儲存的位置正好在汲極接面的正上方時，閘極引發汲極漏電流會大量地增加。這是因為在閘極和汲極間的垂直電場變強了，所以閘極引發汲極漏電流會上升。長通道元件和短通道元件有相似的儲存電荷的空間分佈對元件特性的影響。

此外，我們證實在 N 型通道的多晶矽/氧化鋁/氧化鉿/氧化矽/矽(SAHOS)記憶體元件上單一位元胞中可以有三位元的記憶體特性。為了增加水平方向上的電荷儲存空間，我們把捕陷電荷層延伸到側壁空間層的底下。結合調變臨界電壓、

調變正向讀取的閘極引發汲極漏電流以及調變反向讀取的閘極引發汲極漏電流，這些記憶體元件可以有三位元的操作。這些元件中，源極/汲極與閘極重疊的結構顯示出比非重疊的結構有較好的記憶體性能。在源極/汲極與閘極重疊的元件上，臨界電壓可以位移產生 7V 的記憶窗口，並且在外插到十年線後仍擁有良好的儲存資料持久性。此外，在 10^5 次寫入/抹除之後，此記憶體元件在調變臨界電壓上仍維持良好的性能。當此記憶體元件被運用在 NOR 型的非揮發性記憶體的陣列結構中，干擾效應對於臨界電壓來說是可以忽略的。閘極引發汲極漏電流可以在調變後產生約 100 倍大小的差異，但是沒有良好的儲存資料持久性，也沒有良好的耐操度。因為電荷是儲存在靠近側壁空間的角落，所以「水平電荷遷移」和增加電荷流失速率的缺陷都是造成沒有良好的儲存資料持久性的原因。而沒有良好的耐操度的可能原因是在每一次的寫入/抹除之後，儲存電荷的空間分佈改變了。對元件的可靠度來說，汲極的干擾效應對於閘極引發汲極漏電流是個問題。

Effects of Spatial Distribution of the Stored-Charge on Device Characteristics of Non-volatile Memory

Student: Kun-Wu Yu Advisor: Bing-Yue Tsui

Department of Electronics Engineering
Institute of Electronics
National Chiao Tung University

Abstract

In this thesis, the effects of spatial distribution of the stored-charge on device characteristics of non-volatile memory are evaluated. According to the simulated results, the modulation of the threshold voltage (V_t) and the modulation of the Gate-Induced-Drain-Leakage (GIDL) are related to the stored-charge positions. When the charges are stored at the position which is above the channel center, the whole subthresthold curve is moved to the right because of the negative charges stored at the center of the trapping layer resulting in a raise of the electron barrier of the channel region. When the charges are stored at the position which is above the channel and near the drain junction, only the upper half subthresthold curve is moved to the right. Because the electron barrier of the channel region is slightly dragged down by the drain voltage, the lower half subthreshold curve keeps. However, the electron barrier is not low enough to allow complete electron conduction, so the upper half subthresthold curve is moved to the right. Next, when the charges are stored just

above the drain junction, the GIDL current increase largely. Because the vertical electric field between the gate and the drain is enhanced and then results in larger GIDL current. Both the long channel device and the short channel device have the similar effects of stored-charge distribution on device characteristics.

Moreover, the 3-bit per cell memory characteristics are demonstrated on the n-channel poly-Si/Al_2O_3/HfO_2/SiO_2/Si (SAHOS) memory device. In order to increase lateral charge storage space, the HfO_2 charge trapping layer extends to the underneath of the spacer. The 3-bit operations of these memory devices are demonstrated by combining the V_t modulation, the GIDL current modulation on forward read, and the GIDL current modulation on reverse read. The devices with the S/D-to-gate overlap structure show better memory performances than those with S/D-to-gate non-overlap structure. For the devices with the S/D-to-gate overlap structure, the V_t can shift with large memory window of 7V, and shows good 10-year extrapolated charge retention. Moreover, high endurance after 10^5 P/E cycles is exhibited on the V_t modulation. The disturbance effects of the V_t are negligible when this memory device is implemented by the NOR array architecture. The GIDL current can be modulated by about two orders of magnitude, but it shows poor retention and poor endurance. Because the stored charges are near the corner of the spacer, poor retention is due to “lateral charge migration” and the defects which enhance charge loss rate. The poor endurance is possible as a result of the change of the stored charges distribution after every P/E cycle. In addition, the drain disturbance on the GIDL is also an issue for device reliability.

誌謝

隨著論文的完成，碩士生涯也到了尾聲。在寫論文的這一年多裡，我受到許許許多多的人幫助，使我能夠順利地完成我的碩士論文，謹以此文表達我的感謝。

首先，我要感謝我的指導教授　崔秉鉞老師。不管是做研究的態度還是做研究的方法，老師都給我很大的助益。老師在研究上非常正直，對於是非對錯總是秉持著自己的原則，不會得過且過，這樣的態度給了我很多正面的思考。另外，老師在訓練學生培養研究能力上，更是仔細謹慎，不管是做實驗的技巧或是實驗現象背後的學理，老師都會抽絲剝繭地與學生們討論，並且傾囊相授。然後，我還要特別感謝老師非常有耐心地逐字逐句修改我的論文，給予我許許多多的建議，老師您辛苦了！

其次，我要感謝財團法人工業技術研究院提供沉積氧化鋁和氧化鉿的服務，感謝工研院蔡銘進組長和韋佩儀工程師在元件的製作上給予許多協助。另外，我也要感謝國家奈米元件實驗室與國立交通大學奈米中心提供半導體製程的機台設備，還有許多機台工程師與小姐們的幫忙。

感謝實驗室的謝志民、盧季霈、羅正愷學長給予許多研究上的建議，感謝李振銘學長教我實驗機台並且給予很多實驗上的協助。謝謝同學瑞堯、俊凱、依成在實驗與修課上的協助與砥礪，也謝謝學弟們這一年的陪伴。另外還要特別感謝另一個實驗室的陳建華學長，每次教我機台都非常地仔細。還要感謝凱瑜常常在我研究的低潮時勉勵我。

最後，我要感謝我的家人們！謝謝爸爸媽媽一直以來的栽培與付出，對於我求學的路上總是給予許許多多的支持與勉勵，對於我的生活起居更是照顧的無微不至。還有我親愛的弟弟，一直扮演我的家人、朋友與玩伴三種角色，在我人生的路上帶給我很多的快樂。另外，還有很多一直為我打氣的親朋好友們，謝謝你們！

Contents

Table Captions

Chapter 2

Chapter 3

Chapter 4

Figure Captions

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 1

Introduction

1-1 Evolution of Non-volatile Memory

Since the millennium, people's daily life habit has been changed by various kinds of portable electronic products, such as notebook computer, digital camera, MP3 player, personal digital assistant(PDA), USB, iPod…and so on [1]. The memory devices need to be adopted into all of these electronic products to make them work for different functions. These memories can be divided into volatile memory and non-volatile memory. The use of non-volatile memory (NVM) is to remain the storage data for a long time without power supply, and then portable electronic products can work just by battery due to low power consumption of non-volatile memory.

Non-volatile memories are mainly classified into non-charge-based memory and charge-based memory. The typical charge-based memory is the so called flash memory. There are three types of flash memory including the floating gate (FG) type, SONOS (Silicon/Oxide/Nitride/Oxide/Silicon) type, and nano-crystal type. The mainstream of NVM nowadays is floating gate (FG) type. Some non-charge-based memories have been in small volume production, such as magnetroresistive random access memory (MRAM) [2-3]. The production technologies of ferroelectric random access memory (FeRAM) [4-5] and phase change random access memory (PCRAM) [6-7] are still under development. A novel resistive random access memory (RRAM) attracts much more attention recently [8-12]. All types of flash memories and non-charge-based memories are introduced in detail in the following section.

Smart system must have larger brains which include both sophisticated functions (code storage) and more memory capacity (data storage). These characteristics can be served by two types of flash memories. The NOR-type memory has fast and random access capability for the code storage, and the NAND-type memory has the page access architecture for date storage [13].

1-1-1 Floating Gate Non-volatile Memory

Before the floating gate (FG) non-volatile memory, the magnetic-core memory [14] has a lot of issues, such as large volume, high power consumption, and high cost. Therefore, new kind of memory needs to be invented to replace the magnetic-core memory. In 1967, the first floating gate non-volatile memory was invented by D. Kahng and S. M Sze at Bell Labs [15]. As shown in Fig. 1-1, the stacked-gate FG non-volatile memory structure is mainly used as embedded memories for portable electronics. Nowadays, the main products of the non-volatile memory market still adopt the poly-silicon floating gate structure.

However, as scaling down the device size, the conventional FG memories encounter several limitations [16-18]. First, it is difficult to reduce the operation voltage of FG NVMs because the read and program/erase (P/E) speed in the FG non-volatile memory is related to the operation voltage. Unfortunately, the high operation voltage would result in high power consumption. Second, the thickness of tunneling oxide would be thinner than 4~5 nm while scaling down the channel length. Then, the charges stored in the FG would be easy to leak through any defect in the tunneling oxide to the silicon substrate, and the degradation of retention characteristic of FG memory occurs. Moreover, endurance and disturbance characteristics of the FG memory will degrade by some reliability issues such as stress-induced leakage current improvement and tunneling oxide breakdown while scaling the tunneling oxide thickness. However, thick tunneling oxide will slow down the operation speed, so there is a trade-off between reliability and speed for designing the thickness of tunneling oxide.

Therefore, to overcome the scaling limits of the conventional FG structure, some structures have been suggested by the International Technology Roadmap for Semiconductors (ITRS) including SONOS (Silicon/Oxide/Nitride/Oxide/ Silicon) type and nano-crystal type [19].

1-1-2 SONOS Non-volatile Memory

In 1969, the predecessor of SONOS type non-volatile memory were invented which is p-channel metal-nitride-oxide-silicon (MNOS) structures with 5nm thick silicon dioxide as tunneling layer, 50 nm silicon nitride as charge trapping layer, and a gate electrode sputtered by Al [20-21]. Because charges stored in the nitride layer of the MNOS memory are easy to leak into metal gate, the retention characteristic is poor. The solution to improve the retention characteristic is adding a blocking oxide between the metal gate and the nitride trapping layer. Therefore, the SONOS type non-volatile memory was invented.

In the recent 10 years, a lot of SONOS type flash memory papers have been published [18-19, 22-23]. As shown in Fig. 1-2, the basic structure of SONOS non-volatile memory which stores charges in discrete traps of the silicon nitride layer show better retention and endurance characteristics. Today, it’s the most promising candidate to replace the traditional FG non-volatile memory and have compatibility with standard CMOS technology. The tunneling oxide thickness can be decreased because the stored charges are localized in nitride trapping layer. When defect is generated in the tunneling oxide, only the charges stored near the defect can leak through the tunneling oxide. Therefore, the reliability of SONOS type memory can be improved while comparing with the FG memory. Moreover, the dielectric constant of silicon nitride is double higher than silicon dioxide, so the vertical electric field will increase in tunneling oxide. The higher vertical electric field results in the improvement of Fowler-Nordheim (FN) P/E speed and the reduction of operation voltage.

However, there are many issues exhibited in SONOS type memory. First, to achieve lower power consumption, the operation voltages need to be reduced. Second, the “charge migration” is a big issue [24]. The charges stored in trapping size may move to the neighboring trapping site. For example, the vertical migration of trapping electrons occurs as Si_3N_4 is used as trapping layer, and the lateral migration of trapping electrons occurs as HfO_2 is used as trapping layer [24]. Third, tunneling oxide will be degraded by Channel-Hot-Electron (CHE) programming which accompanies the degradation of endurance performance. The fourth one, gate and drain disturbance may cause movement of stored charge. Last, erase saturation must be concerned [25].

Recently, as a result of those issues existed in the simple SONOS memory, some advanced SONOS type memories such as TaN/AlO/SiN/Oxide/Si (TANOS) [26] and bandgap engineered SONOS (BE-SONOS) [27-28] have been proposed to improve the P/E speed and/or the retention characteristics. First, take a look at TANOS structure, aluminum oxide as blocking oxide layer is a high-κ material. Due to the low gate coupling rate of the conventional SONOS memories, thin tunneling oxide is required to improve P/E speed, but thin tunneling oxide will lead to poor retention characteristic because of charge leakage path formed in tunneling oxide. Using aluminum oxide to replace silicon dioxide as blocking layer, the high gate coupling rate results in higher P/E speed, and maintains the thicker tunneling oxide to preserve good retention characteristics. On the other hand, the traditional n^+ poly-Si gate is replaced by TaN metal gate. Due to the higher work function of TaN metal, the unwanted backward FN tunneling current of electron from top gate can be reduced [29]. For the BE-SONOS, the tunneling oxide is replaced by an ONO (oxide/nitride/oxide) structure which brings about more holes tunneling current at high electric field during erase operation due to the band offset [27]. On the other hand, the thick ONO layer can prevent direct tunneling of holes from substrate at low electric field, so the retention characteristic can be improved by suppressing charge loss [27].

1-1-3 Nano-crystal Non-volatile Memory

Recently, non-volatile memory with nano-crystals has attracted a lot of attention due to its potential for next generation non-volatile memory. As shown in Fig. 1-3, charges are stored in isolated nano-crystals or nano-dots embedded in the trapping layer. Each nano-crystal can store few electrons and then affect the conductivity of channel to cause threshold voltage shift. There are many advantages of nano-crystal NVMs. First, compared to the structure of the SONOS type memory, the lateral migration of stored charges can be suppressed by the dielectric isolation between nano-crystals. It represents better retention characteristics than the SONOS type memory. Second, just few stored charges will leak through the leaky paths formed in the tunneling layer nearby the nano-crystal, and most of the stored charges can remain in the isolated nano-crystals. Therefore, compared to conventional FG memory devices, nano-crystals NVMs can decrease the thickness of the tunneling layer, and then improve the P/E speed or suppress the operation voltage to obtain low power consumption. Third, the vertical electric field between gate and substrate across nano-crystals will increase, and the work function of nano-crystal can be designed for optimizing the device characteristics [30-33]. Last, nano-crystal devices have the capability of storing two bits of information.

Some possible scaling limits of nano-crystal device have been proposed. When the nano-crystal non-volatile memory device scales down to sub-65nm node, uniform distribution, high density, and small size of nano-crystal is needed. However, the small nano-crystal (< 5nm in diameter) will decrease the trapping efficiency during program operation due to coulomb blockade and quantum confinement effects [34-35]. Consequently, small nano-crystal results in the small memory window. Therefore, the scaling limit for nano-crystal device depends on how small nano-crystals can be made as well as the means to uniformly deposit them.

1-1-4 Non-charge-based Non-volatile Memory

While scaling down the flash memory device, there are some issues need to be conquered, such as low program/erase speed, poor retention time, and high operation voltage [36]. Therefore, non-charge-based NVMs attract significant attention, especial the RRAMs. The resistive switching phenomenon is the change of the resistance of the metal-insulator-metal (MIM) memory cell [37]. Generally, an initial electroforming step is required to apply at fresh samples, and then the system can be switched between a conductive ON-state and a less conductive OFF-state [37].

Resistance switching phenomenon could be divided into unipolar and bipolar systems. Unipolar means the operation of switching is independent on the polarity of the voltage and current signal. In unipolar system, the high-resistance state (OFF) is switched (“ON”) into the low-resistance state (ON) by a threshold voltage and the compliance current. On the other hand, the unipolar system in its low-resistance state (ON) is switched (“OFF”) into the high-resistance state (OFF) by a higher current and a voltage below the set voltage. In contrast, the bipolar system is which the set to the ON state occurs at one voltage polarity, and the reset to the OFF state take place on reversed voltage polarity [37].

Although RRAM has potential to further scale down than flash memory, RRAM also has a lot of issues required to be resolved. The current issues for RRAM are operation variation, current reduction and device yield. Operation variation issues are about the deviation of operation voltage and values of high and low resistance states.

Nowadays, the only method to increase storage information density is to continuously scale-down the flash memory device and realize multi-bit and/or multi-level per memory cell.

1-2 Motivation

Because it is not easy to further scale down memory device size, multi-bit per memory cell becomes the only method to increase storage information capacity. Recently, some researches show that the SONOS type memories have the possibility to realize two bits per cell, and the device characteristics have been demonstrated [38-41]. Because the charge trapping layer of the SONOS type memory can store charges locally, more than one bit data information can be achieved in one memory device. In addition, modulation of the Gate-Induced-Drain-Leakage (GIDL) current has been proposed to increase bit information in a memory cell, and the simulation results are shown in Fig. 1-4 [38]. When the charges are stored in bit 2, the GIDL current significantly increase by six orders on forward read, but the V_t shift is very small. On the other hand, when the charges are stored in bit 1, the GIDL current doesn’t increase on forward read, but the V_t shift is remarkably large [38].

Therefore, in this thesis, we use Sentaurus-TCAD simulation software to simulate the I_d-V_g curves with various charge distributions in the trapping layer. The aim of this study is to illustrate multi-bit data information in one memory cell.

We have mentioned that the SONOS type non-volatile memory has poor retention and scaling issues. The operation voltage and power consumption are still not low enough. High-relative permittivity (high-κ) charge trapping layers in poly-Si/SiO_2/high-κ/SiO_2/Si (SOHOS) memory structure are of increasing attention. High-κ materials as a charge trapping layer have been proposed to replace the nitride layer in the SONOS structure to reduce the total thickness of gate dielectric stack and to improve the P/E speed as a result of high gate coupling to the conductive channel [24, 42–43]. In order to further improve the charge retention characteristics and reduce the thickness of gate dielectric stack, the high-κ materials with a large barrier height have been adopted to replace silicon dioxide as a blocking layer, such as Al_2O_3 films [44–47].

A high work function metal gate electrode in metal/Al_2O_3/HfO_2/SiO_2/Si (MAHOS) memory has been proposed but only capacitor structure was fabricated and evaluated [47]. The calculated energy band diagram on Al/Al_2O_3/HfO_2/SiO_2/p-Si (solid line) and Al/SiO_2/HfO_2/SiO_2/p-Si (dash line) structures under the P/E mode is shown in Fig. 1-5 [47]. Under the program operation, electron current (J_{electron}) from the Si substrate can be increased and hole current (J_{hole}) from the gate electrode can be reduced by using the high-κ Al_2O_3 as a blocking oxide. In contrast, under the erase operation, the J_{hole} from the Si substrate can be increased and the electron current from the gate electrode can be reduced [47].

In this work, poly-Si/Al_2O_3/HfO_2/SiO_2/Si (SAHOS) non-volatile memory devices are fabricated to demonstrate the multi-bit per cell operation predicted by the TCAD simulation. We replace metal gate by poly-Si gate to have better feasibility with standard CMOS integration process. Poly-Si gate has self-alignment feature to form source and drain region, and it is important for scaling down device size and increasing device density. Since the charge trapping layer HfO_2 film of SAHOS memory can store charges locally, this SAHOS memory has potential to realize multi-bit per cell. Combined with the simulation results in this thesis, we demonstrated 8 states of 3 bits information in this SAHOS memory successfully and clearly point out the stored-charge positions.

1-3 Thesis Organization

This thesis is divided into five chapters and the contents of each chapter are described as follows.

In chapter 1, the evolution of non-volatile memory has been reviewed, and various kinds of the memory structure have been compared. Accordingly, we choose SAHOS type memory as the research target to study the changes of I_d-V_g characteristics by varying the stored charge positions.

In chapter 2, experimental procedure will be described. The device structure parameters and physical models used in the numerical simulation are shown clearly. The fabrication process of the 2μm gate length poly-Si/Al_2O_3/HfO_2/SiO_2/Si (SAHOS) non-volatile memory will be illustrated. In addition, the measurement techniques of electrical characteristics are described.

In chapter 3, we use the simulation tool of Sentaurus-TCAD to study the I_d-V_g characteristics to study the effects of the position of the stored charges. The charges are located in trapping layer with different lateral positions such as above channel center, above channel and near drain junction, above channel and next to drain junction, above drain and next to drain junction, and above drain and away drain junction.

In chapter 4, the basic transistor characteristics of the SAHOS non-volatile memory fabricated in chapter 2 will be discussed firstly. We will show that the SAHOS flash memory can store 3 bits data information in one memory cell. The threshold voltage (V_t) can shift with large memory window of 7V. Moreover, the V_t modulation presents good retention of nearly no charge loss after 10^5 sec and high endurance of 22% degradation after 10^5 P/E cycles. The GIDL current can be modulated by about two orders of magnitude, but it shows poor retention and poor endurance. We will discuss the possible reasons for the poor retention and endurance performance. Last, we demonstrate the disturbance effect when this memory device is implemented by the NOR array architecture.

In chapter 5, conclusions and future works are presented.

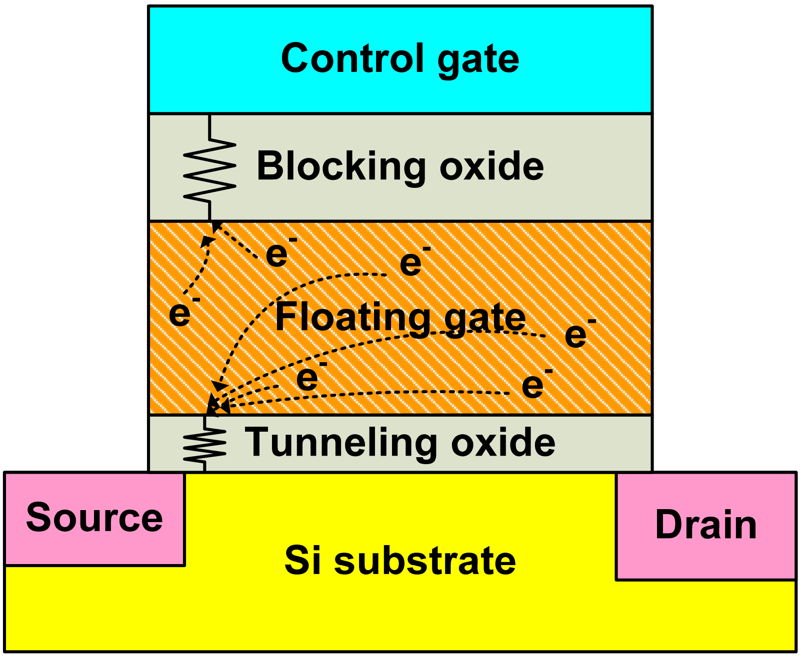


Fig. 1-1: Basic concept of floating gate (FG) non-volatile memory.

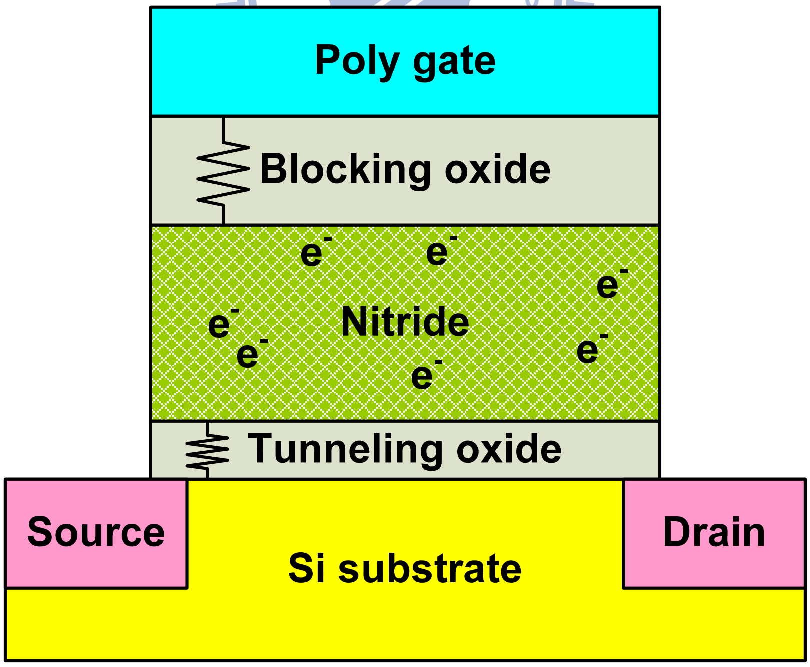


Fig. 1-2: Basic concept of SONOS non-volatile memory

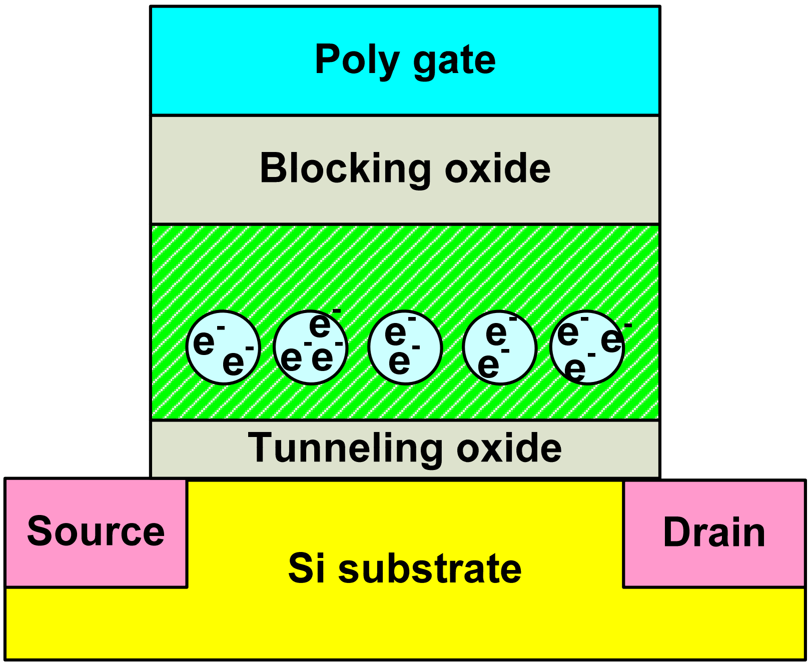
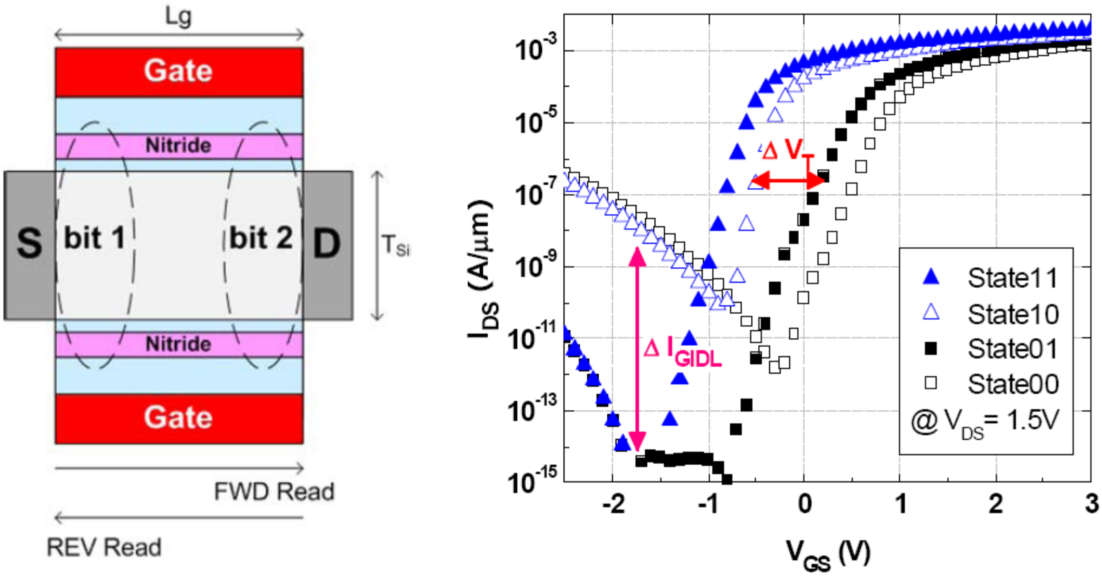


Fig. 1-3: Basic concept of nano-crystal non-volatile memory.



(a) (b)

Fig. 1-4: (a) 2D schematic cross-section of the dual-bit FinFET SONOS NVM cell. (b) Simulated I_{DS}-V_{GS} characteristics show that V_T and GIDL can be used to distinguish the state of Bit 1 and Bit 2. In this figure, ‘0’ (‘1’) refers to the programmed (erased) state [38].

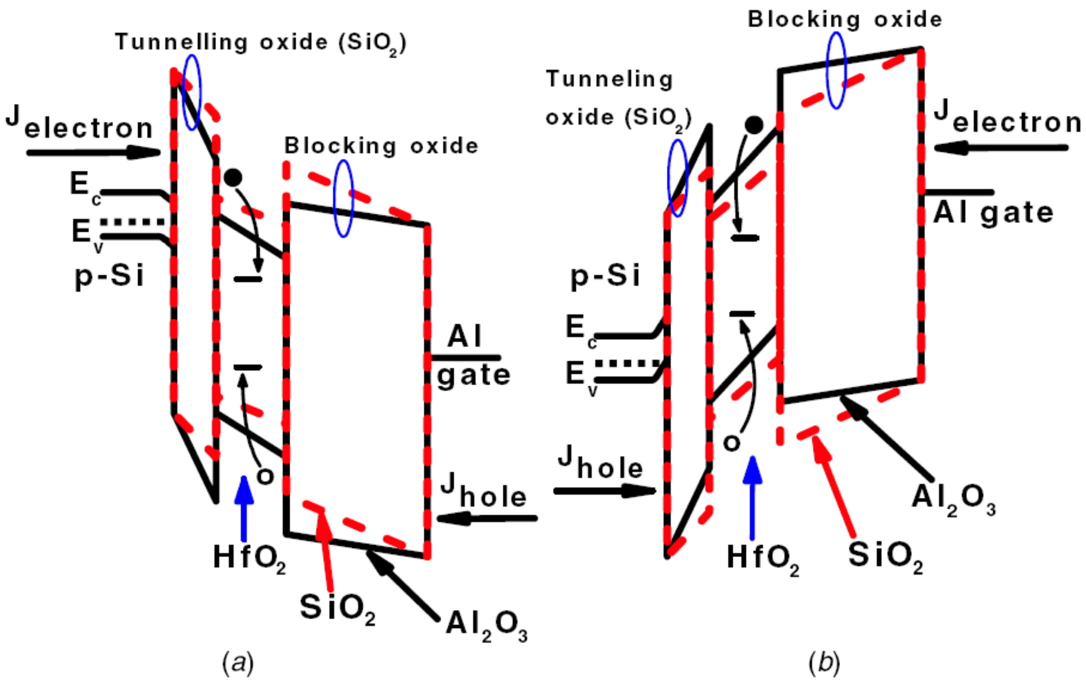


Fig. 1-5: Schematic energy band diagrams of $Al/Al_2O_3/HfO_2/SiO_2$/p-Si (solid line) and $Al/SiO_2/HfO_2/SiO_2$/p-Si (dash line) memory structures under (a) program and (b) erase modes [47].

Chapter 2
Experimental Procedure

2-1 Simulation Method

2-1-1 Device Structure

To study the effects of the stored-charge distribution on device characteristics in non-volatile memory, Sentaurus-TCAD tool was used. The device structure was defined by the device simulator DEVISE. In this work, both long channel and short channel devices were studied. Then, the position of the stored-charge was varied relative to the source and drain metallurgic junction position.

The 2-D long channel non-volatile memory structure is shown in Fig. 2-1. The fixed device parameters includes channel width = 1μm, source and drain doping concentration N_{SD} = $5\text{x}10^{19}\text{cm}^{-3}$, substrate doping concentration N_{sub}= $5\text{x}10^{16}\text{cm}^{-3}$, poly-Si gate doping concentration N_{Gate} = $1\text{x}10^{20}\text{cm}^{-3}$, spacer length L_{spacer} = 50nm, gate oxide thickness t_{ox} = 12nm, length of the charge storage region L_{charge} = 30nm, thickness of the charge storage region t_{charge} = 3nm, tunneling oxide thickness t_{tunnel} =4nm, blocking oxide thickness $t_{blocking}$ = 5nm, and stored-charge quantity = 0 or $-1\text{x}10^{-15}$Coul. The varied parameters are channel length (L_c), gate to S/D overlap length (L_{ol}), and stored-charge position. The gate length is fixed at 0.4μm. Therefore, the overlap length L_{ol} is 50nm as L_c=0.3μm and 0nm as L_c=0.4μm. Five different stored-charge positions were considered. They are above channel center, above channel and 30nm away from drain junction, above channel and next to drain junction, above drain and next to drain junction, and above drain and 30nm away from drain junction.

The 2-D short channel non-volatile memory structure is shown in Fig. 2-2. The stored-charge position is defined as silicon, and the spacer is defined as SiO_2. The fixed device parameters includes channel length = 50nm, channel width = 1μm, source and drain doping concentration N_{SD} = $1x10^{19}cm^{-3}$, substrate doping concentration N_{sub}= $2x10^{18}cm^{-3}$, poly-Si gate doping concentration N_{Gate} = $1x10^{20}cm^{-3}$, spacer length L_{spacer} = 25nm, gate oxide thickness t_{ox} = 10nm, length of the charge storage region L_{charge} = 3nm, thickness of the charge storage region t_{charge} = 2nm, tunneling oxide thickness t_{tunnel} = 4nm, blocking oxide thickness $t_{blocking}$ = 4nm, and stored charge quantity = 0, $-1x10^{-16}$, or $-2x10^{-16}$Coul. The varied device parameter is stored-charge position. Four different stored-charge positions were considered. They are above channel center, above channel and 3nm away from drain junction, above channel and next to drain junction, and above drain and next to drain junction.

2-1-2 Physical Models

DESSIS in Sentaurus-TCAD was used to declare physical models and physical parameters, and to simulate electrical characteristics. Mobility model, bandgap narrowing model, recombination model, and band-to-band tunneling model were included [48]. There are three components used in mobility model. “Doping Dependence” is the first one, and it means the mobility would degrade as doping concentration increase. “High Field Saturation” is the second one, and it means the mobility would be saturated at high electric field. “Enormal” is the last one, and it means the mobility would be affected by the vertical electric field. For recombination model, doping-dependent SRH recombination model is adopted. Recombination through deep levels in the gap is usually labeled as Shockley–Read–Hall (SRH) recombination. Phonon-assisted band-to-band tunneling cannot be neglected in steep pn-junctions or in high normal electric fields of MOS structures [48].

2-2 Device Fabrication

Fig. 2-3 shows the main process flow of the poly-Si/Al_2O_3/HfO_2/SiO_2/Si (SAHOS) non-volatile memory. The devices were fabricated on a 6” p-type Si wafer. At beginning, dry oxide was thermally grown to 35 nm thick in a lateral furnace system. Then, BF_2^+ ions were implanted into wafer backside at 60 KeV to a dose of $5x10^{15}$ cm^{-2}. Zero marks were patterned by optical lithography and plasma etching. In lateral low pressure chemical vapor deposition (LPCVD) system, Si_3N_4 was deposited to 150 nm thick. Then, device active regions were patterned, and the Si_3N_4 and SiO_2 layers were etched by plasma etching system of model TEL5000. BF_2^+ ions were implanted at 60 KeV to a dose of $2x10^{13}$ cm^{-2} to form the p-type channel stop. The wet oxide of 550nm thick was thermally grown in a lateral furnace, and then wafers were dipped in hot H_3PO_4 and dilute HF solutions sequentially to remove Si_3N_4 and SiO_2, respectively. To finish the LOCOS (local oxidation of Si) isolation structure, the final step is that a wet oxide of 30nm thick was grown in a lateral furnace, and then wafers were dipped in dilute HF solutions to remove oxide. Fig. 2-3(a) shows the LOCOS structure of the device.

After LOCOS isolation process, the tunneling oxide was thermally grown to 4 nm thick in a vertical furnace system. Next, a 5nm thick HfO_2 was deposited as trapping layer in a clustered ALD system and a 20nm thick Al_2O_3 was deposited as blocking oxide in the same ALD system followed by a deposition of 150 nm thick LPCVD amorphous-Si gate in a LPCVD system. Then, the amorphous-Si gate was doped with BF_2^+ at 40KeV to a dose of $5x10^{15}cm^{-2}$. After that, gate dopant activation was executed by a RTA annealing at 900℃ for 20 seconds in nitrogen ambient, and the amorphous-Si was turned into poly-Si simultaneously. Then, TEOS oxide with a thickness of 100nm as hard mask was deposited on the poly-Si film to avoid unwanted anti-doping during n^+ S/D ion implantation. The gate pattern was transferred from i-line photo resist to the TEOS layer by dry etching, and then the remaining etching steps were carried out with the TEOS layer as hard mask. The etching process stopped at the trapping layer. After gate patterning, the schematic cross-sectional structure is shown in Fig. 2-3(b).

Afterwards, the spacer composed of 10 nm SiO_X and 50 nm SiN_X was formed as shown in Fig. 2-3(c). Next, P_{31}^+ ions were implanted into the S/D region at 20 KeV to a dose of $5x10^{15}$ cm^{-2} as shown in Fig. 2-4(d). Then the S/D dopants were activated at 900 °C for 20 sec, 60 sec or 180 sec in nitrogen ambient as shown in Fig. 2-3(e). The gate hard mask and native oxide on the S/D region were removed by dipping in dilute HF solution. A 25 nm thick Ni film was deposited by e-gun evaporation followed by a two-step Ni-salicide process [49]. The first step is to form Ni_2Si phase, so the annealing process is executed at 300℃ for 45 minutes in vacuum system. The unreacted Ni was removed by H_2SO_4/H_2O_2 = 3:1 solution. The second step is to transform the Ni_2Si phase to the NiSi phase, and the annealing was executed at 600℃ for 30 seconds. After the two-step Ni-salicide process, the S/D region and gate electrode became silicide structure as shown in Fig. 2-3(f). Finally, the poly-Si at backside of the substrate was removed by the $HF:HNO_3:CH_3COOH$ = 6:20:7 solution and the oxide was removed by dilute HF solution followed by a deposition of 300nm thick Al film as back electrode of the devices which is shown in Fig. 2-3(g). Table 2-1 summarizes the split conditions used in this thesis.

2-3 Electrical Characterization Techniques

To characterize the SAHOS non-volatile flash memory fabricated in this thesis, several measurements were carried out. We used Agilent 4156C to measure the static current-voltage (I-V) characteristics of the memory device. An Agilent 41501A pulse generation expander was utilized to generate pulse signals during program and erase operations. The substrate terminal of devices was always biased at 0V.

There are many ways to determine the threshold voltage (V_t) which is the most important parameter of non-volatile memory devices. In this thesis, the method to determinate the threshold voltage is *constant drain current method,* and this technique is usually adopted in non-volatile memory devices. The gate voltage at a constant drain current value is taken as the threshold voltage. The threshold current value is 100 μA at the drain voltage 1V for sample A, and 1μA at the drain voltage 1V for sample B and sample C. During I-V measurement, the range of gate voltage sweeping should be carefully controlled. If the sweeping range is too large, the unexpected V_t shift may occur during measurement. Before establishing the characteristics of memory device, the I_d-V_g curve should be recorded repeatedly to make sure V_t stability.

In this thesis, gate-induced drain leakage (GIDL) can be modulated, so that the GIDL can be used to store information. The off current (GIDL) is measured at constant gate voltages of -5V, -7V or -8V as the drain voltage is fixed at +1V.

Because the information can be stored by V_t modulation or off current (GIDL) modulation separately in this thesis, various memory characteristics are need to be defined and measured respectively. The definition of memory characteristics and methods to measure these characteristics are described as follows.

2-3-1 P/E speed and memory window

(A) V_t modulation

V_t should be measured as quickly as possible after each program or erase operation at different gate biases and different pulse widths so that we can obtain the characteristics of P/E speed. The "V_t shift" is defined as the difference of V_t after each program or erase operation. Before applying another P/E signal, it is essential to return V_t to the same state. For example, before measuring program speed by applying different pulse widths, the device must be erased to the same V_t. The memory window is defined as the V_t difference between program state and erase state. During program and erase operations, both source and drain terminals were biased at 0V, because the charge injection is performed by Fowler-Nordheim (FN) tunneling in this work.

(B) GIDL modulation

To characterize the P/E speed, off current (GIDL) was measured after each program or erase operation at different P/E conditions. The bias condition which would increase and decrease the GIDL current is defined as program and erase condition, respectively. The device could be programmed by a negative gate bias, and both source and drain terminals were biased at 0V. On the other hand, the device could be erased by applying a high drain voltage, and both gate and source terminals were biased at 0V. When measuring P/E speed, the GIDL currents were recorded as a function of gate bias after applying different P/E pulses. Before applying another P/E pulse, it is essential to return the GIDL current to the same value by applying a suitable bias condition. Then, the memory window is defined as the difference of GIDL current between program state and erase state.

2-3-2 Retention

In any nonvolatile memory technology, it is essential to maintain information for a long time. This means the loss rate of the stored charge need to be as slow as possible.

Generally, retention performance is evaluated as the variation of both program state and erase state as a function of storage time at specified storage temperature. But, in this study, the change of erase state, both V_t and GIDL, as time goes by is so small, that it can be neglected. To evaluate the retention performance of the program state, the V_t shift and the change of GIDL were recorded as a function of time. In addition, a small V_g sweeping range should be used to detect the V_t to prevent unwanted V_t shift during measurement. The ten year residual memory window can be obtained by extrapolation.

2-3-3 Endurance

Endurance is the number of the P/E operation cycles that the memory device can operate without breakdown or vanishing memory window.

(A) V_t modulation

The V_t on program state and erase state were recorded during sequential P/E cycles. The sequential pulse signals with fixed pulse width and rise/fall time were pulsed into memory devices to establish the endurance characteristics. In this work, 10^4 P/E cycles were measured on sample A, and 10^5 P/E cycles were measured on sample B.

(B) GIDL modulation

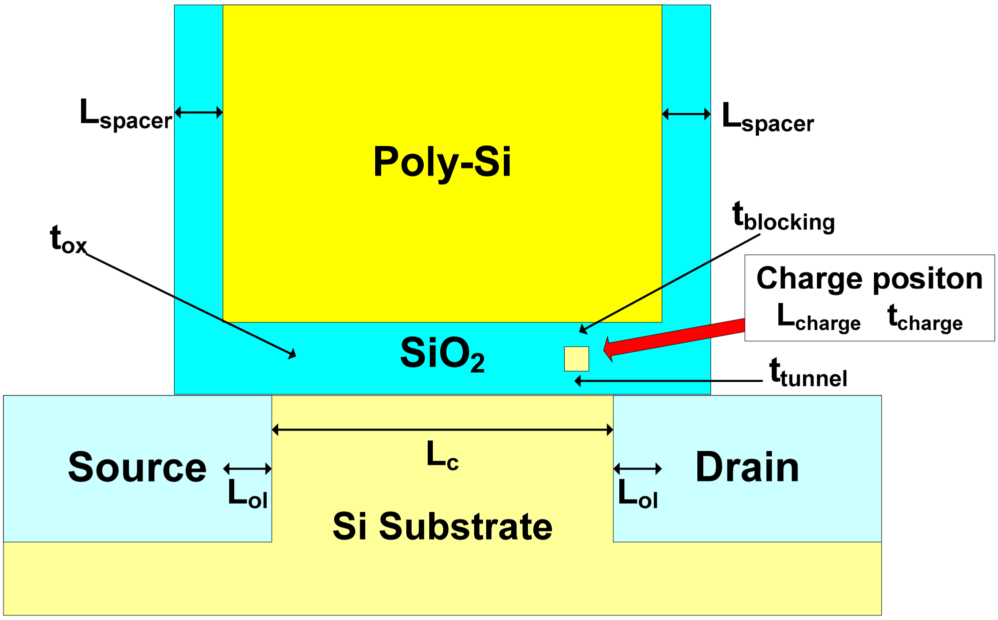
The GIDL on program state and erase state were recorded during sequential P/E cycles. The sequential pulse signals with fixed pulse width and rise/fall time were pulsed into memory devices to establish the endurance characteristics. In this work, 200 P/E cycles were measured on both sample A and B.

2-3-4 Disturbance

As the result of sharing the same word line and bit line in a memory array, the disturbance problems occur. Generally speaking, the main voltage disturbance on flash memory can be divided into “read disturbance”, “gate disturbance”, and “drain disturbance”. First, when a memory cell operates at read procedure, the read disturbance is defined as the change of the values of V_t and GIDL in this work. Second, when the selected memory cell is operating at high gate voltage, the other unselected memory cells sharing the same word line may exhibit the variation of V_t or GIDL, and this phenomenon is called the gate disturbance. Last, when the selected memory cell is operating at high drain voltage bias, the other unselected memory cells sharing the same bit line may exhibit the variations of V_t or GIDL, and this phenomenon is called the drain disturbance. While measuring the disturbance characteristics, we apply the specific stress conditions to the memory device, and the I-V characteristics were measured at specific times.

Table 2-1: Process conditions of the SAHOS non-volatile memory device fabricated in this thesis.

Sample	S/D dopant activation at 900℃	S/D junction overlap with poly-Si gate
A	20s	No
B	60s	Yes
C	180s	Yes

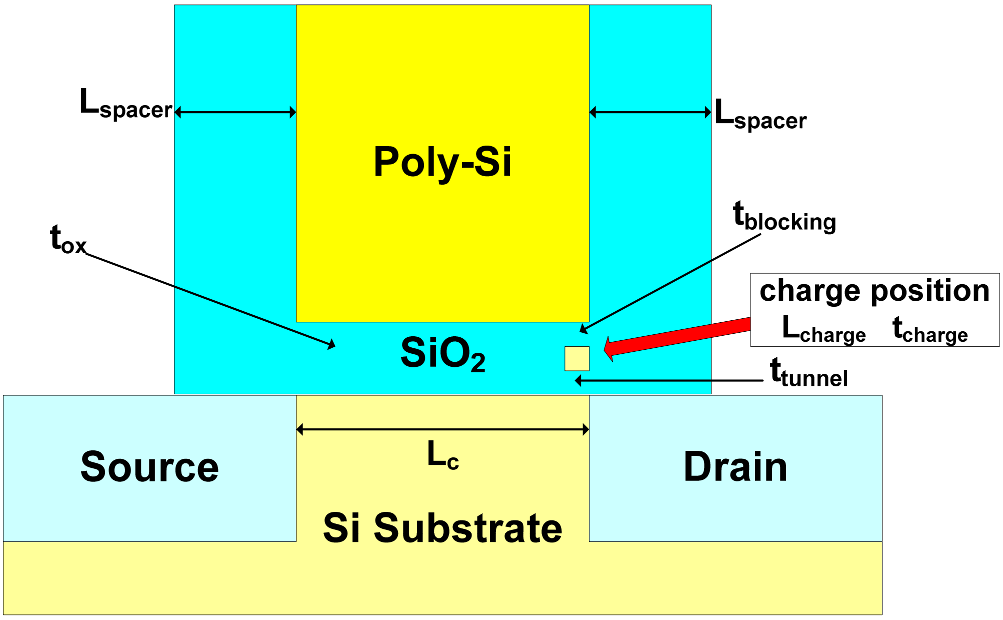


(a)

Channel length L_c (μm)	**0.3**	**0.4**
N_{SD} (cm^{-3})	**$5x10^{19}$**	
N_{sub} (cm^{-3})	**$5x10^{16}$**	
N_{Gate} (cm^{-3})	**$1x10^{20}$**	
L_{spacer} (nm)	**50**	
t_{ox} (nm)	**12**	
L_{charge} (nm)	**30**	
t_{charge} (nm)	**3**	
t_{tunnel} (nm)	**4**	
$t_{blocking}$ (nm)	**5**	
L_{ol} (nm)	**50**	**0**
Stored charge quantity (Coul)	**0 or $-1x10^{-15}$**	

(b)

Fig. 2-1: Long channel device (a) device structure (b) structural parameters

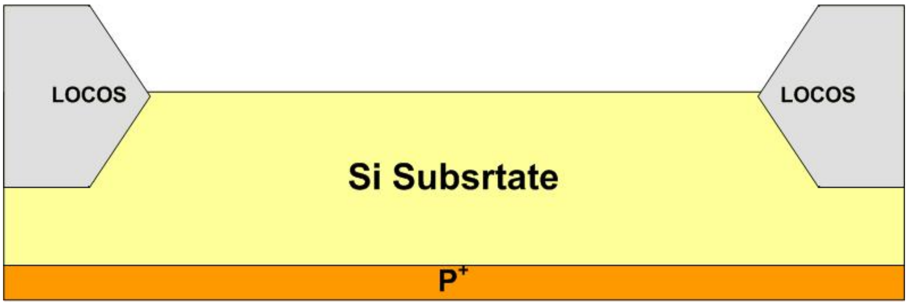


(a)

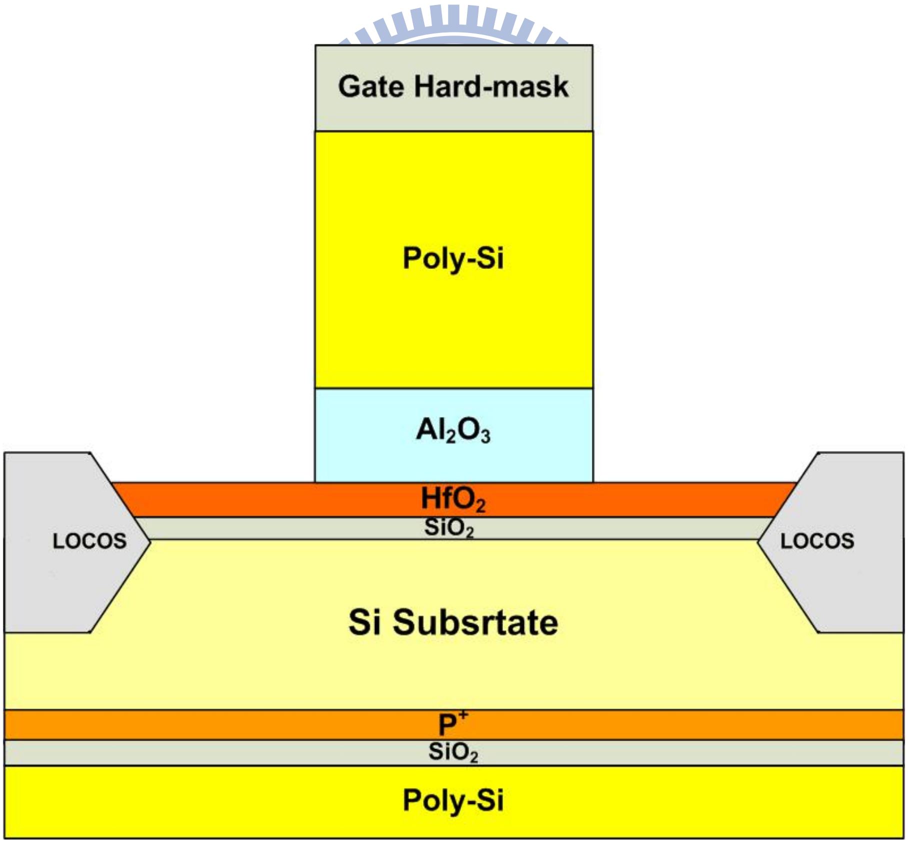
Channel length L_c (nm)	**50**
N_{SD} (cm^{-3})	**$1x10^{19}$**
N_{sub} (cm^{-3})	**$2x10^{18}$**
N_{Gate} (cm^{-3})	**$1x10^{20}$**
L_{spacer} (nm)	**25**
t_{ox} (nm)	**10**
L_{charge} (nm)	**3**
t_{charge} (nm)	**2**
t_{tunnel} (nm)	**4**
$t_{blocking}$ (nm)	**4**
Stored charge quantity (Coul)	**0, $-1x10^{-16}$, $-2x10^{-16}$**

(b)

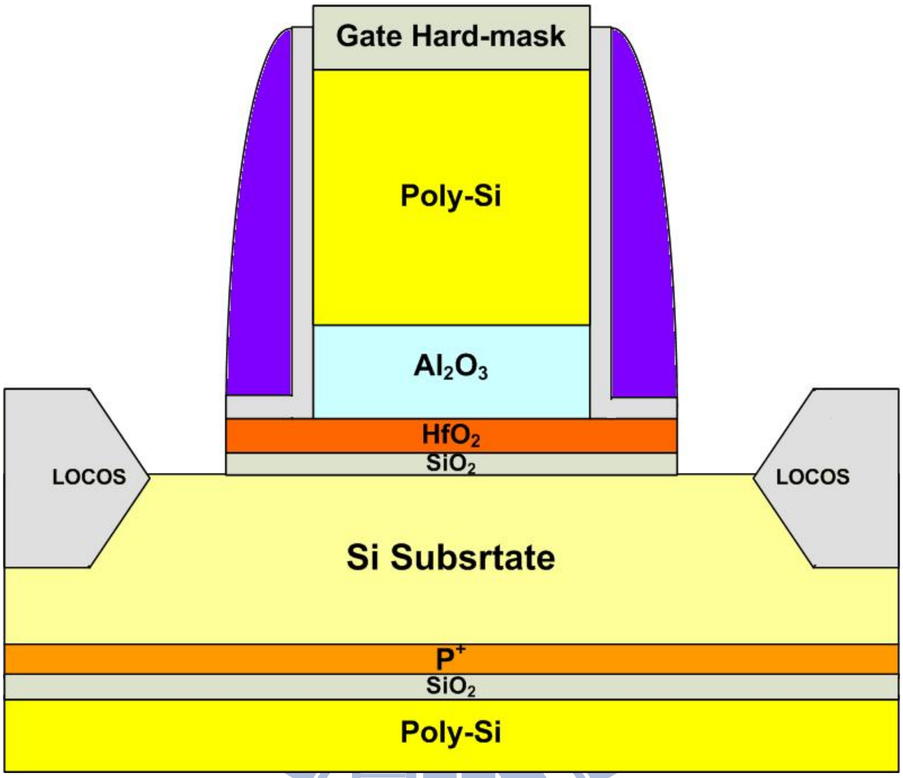
Fig. 2-2: Short channel device (a) device structure (b) structural parameters


LOCOS
LOCOS
Si Subsrtate
P+

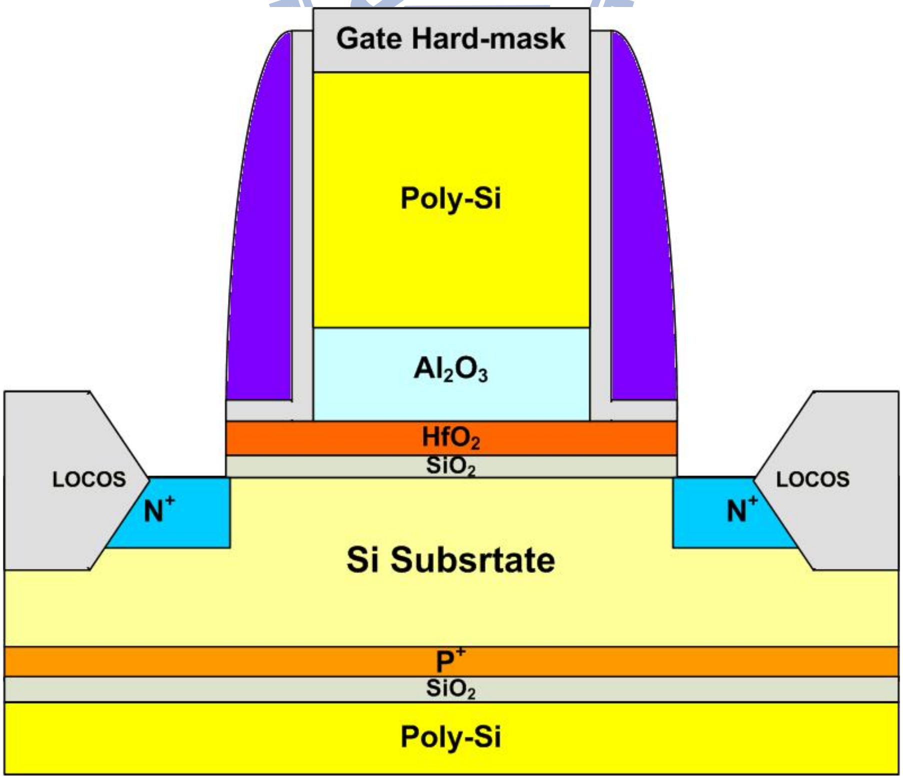
(a)


Gate Hard-mask
Poly-Si
Al2O3
HfO2
SiO2
LOCOS
LOCOS
Si Subsrtate
P+
SiO2
Poly-Si

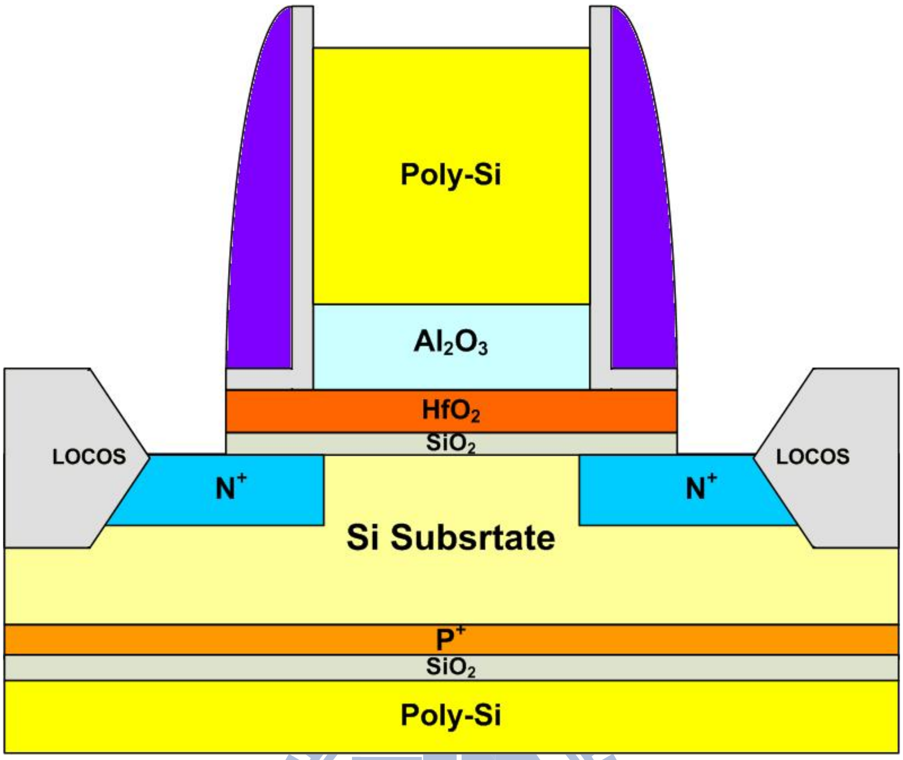
(b)


Gate Hard-mask
Poly-Si
Al2O3
HfO2
SiO2
LOCOS
LOCOS
Si Subsrtate
P+
SiO2
Poly-Si

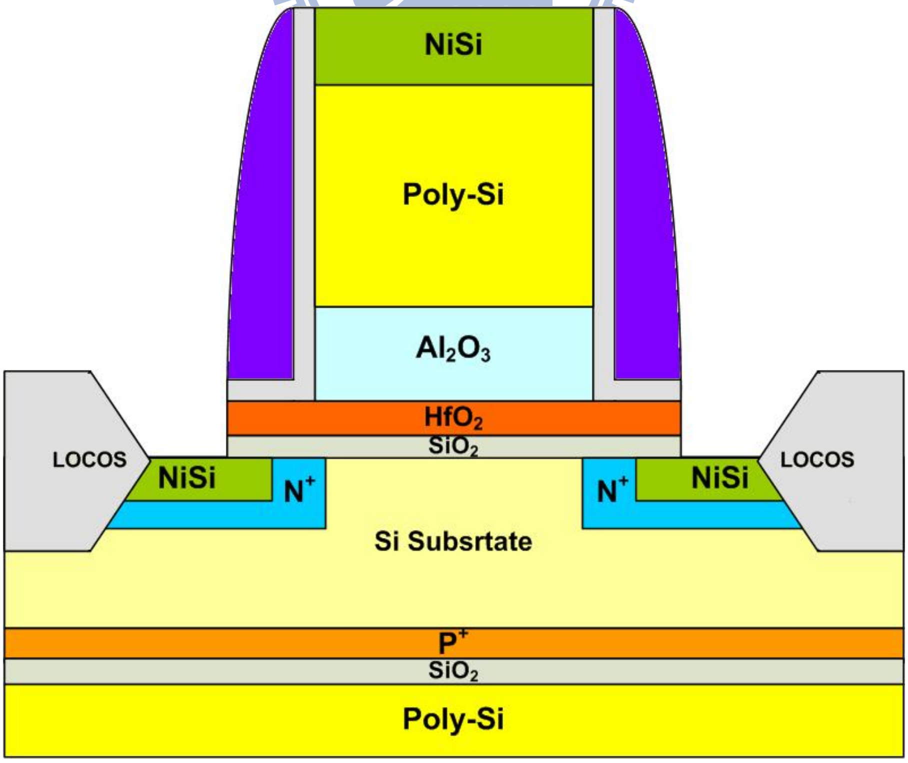
(c)


Gate Hard-mask
Poly-Si
Al2O3
HfO2
SiO2
LOCOS
N+
N+
LOCOS
Si Subsrtate
P+
SiO2
Poly-Si

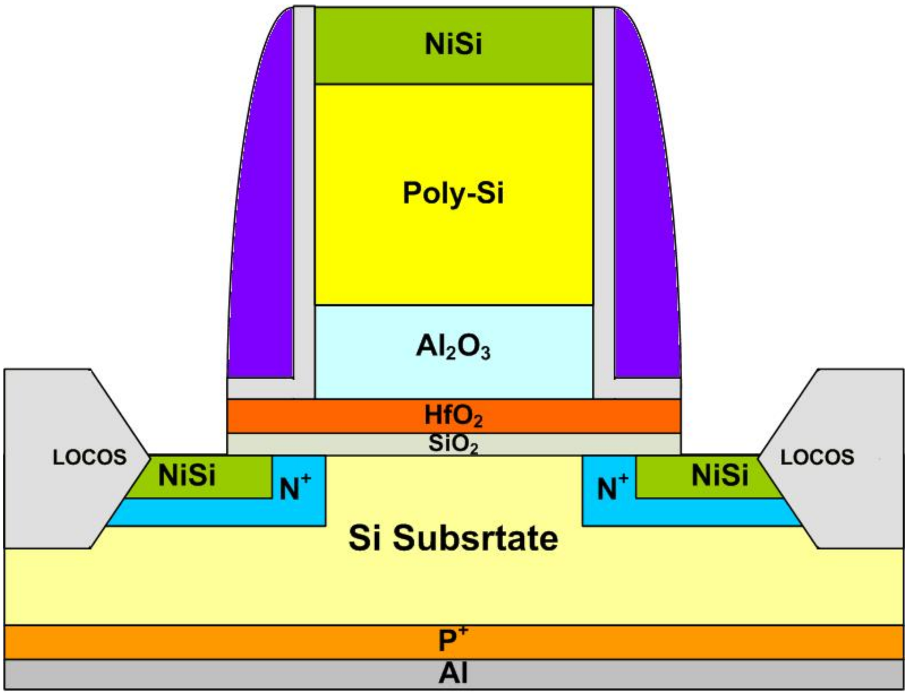
(d)


Poly-Si
Al_2O_3
HfO_2
SiO_2
LOCOS
N^+
Si Subsrtate
N^+
LOCOS
P^+
SiO_2
Poly-Si

(e)


NiSi
Poly-Si
Al_2O_3
HfO_2
SiO_2
LOCOS
NiSi
N^+
N^+
NiSi
LOCOS
Si Subsrtate
P^+
SiO_2
Poly-Si

(f)



(g)

Fig. 2-3: Process flow and cross-sections of the poly-Si/Al2O3/HfO2/SiO2/Si (SAHOS) non-volatile memory. (a) after LOCOS process, (b) after dielectric stack deposition and gate patterning, (c) after spacer formation, (d) after S/D ion implantation, (e) after S/D activation, (f) after gate hard mask removal and silicide formation, (g) after back electrode formation.

Chapter 3

Simulation Results of Stored Charge Distribution on Device Characteristics

3-1 Introduction

As non-volatile memory scaling down, the conventional NVMs encounter many challenges to be overcome. Therefore, the memory device which can store more than one bit per cell becomes the promising device for further improving storage information capacity. The SONOS type memory has potential to realize two bits per cell, and the device characteristics with two bits per cell have been published [38-41]. Some papers have proposed that the modulation of the Gate-Induced-Drain-Leakage (GIDL) current is a potential way to distinguish a bit in the two-bit memory device [38-41]. The GIDL has been investigated since 1987 by J. Chen et al. [50]. Fig. 3-1 shows that the deep-depletion region is formed in the gate-to-drain overlap region, and the band-to-band tunneling mechanism of carrier occurs [50]. On one hand, the electrons at the valence-band tunnel into the conduction band and then are collected at the drain. On the other hand, the holes generated at the deep-depletion region flow to the substrate simultaneously [50].

The modulation of the V_t and the modulation of the GIDL are related to the stored-charge positions. Therefore, in this chapter, the effects of stored-charge distribution on device characteristics will be investigated by simulation tool at first.

3-2 Long Channel SONOS type NVM

3-2-1 Overlap between S/D and gate

The simulation parameters have already been declared in chapter 2. In the structure of the device with channel length L_c = 0.3 μm, the gate overlap with the source and drain (S/D) region by 50nm. The center position of the channel is defined as the origin of the X-axis. The stored-charges at the trapping layer have five different horizontal positions: Position 1 "above channel center", Position 2 "above channel and 30nm away from drain junction", Position 3 "above channel and next to drain junction", Position 4 "above drain and next to drain junction", Position 5 "above drain and 30nm away from drain junction". These five positions are indicated in Fig. 3-2. The I_d-V_g characteristics with stored-charges at various positions were simulated and discussed as following. In order to investigate how various charge storage positions can affect the carrier barrier of the channel, the conduction band energy of the channel is discussed later.

First, compared with the non-charge-stored case, the I_d-V_g characteristics with stored-charges at Position 1 show higher V_t. The whole subthresthold curve was moved to the right as shown in Fig. 3-2 (a). The negative charges stored at the center of the trapping layer result in a raise of the electron barrier of the channel region, and the conduction band energy diagram is shown in Fig. 3-3. The conduction band energy diagram was drawn along the X-axis within the channel which is 1 nm below the interface which is between the Si substrate and the tunneling oxide. Second, compared with the non-charge-stored case, the I_d-V_g characteristics with charges at Position 2 show higher V_t, but only the upper half subthresthold curve was moved to the right as shown in Fig 3-2 (b). Next, Fig. 3-2 (c) shows the I_d-V_g characteristics with charges stored at Position 3 so that the off-current, GIDL, increased largely.

Because the charges storage position is next to the drain junction, it can enhance the vertical electric field between the gate and the drain and then result in larger GIDL. Next, the charges are stored at Position 4 as shown in Fig.3-2 (d). Fig. 3-2 (d) and Fig. 3-2 (c) have similar characteristics as a result of the same physical mechanism. Finally, when the charges are stored at Position 5 as shown in Fig. 3-2 (e), there is no difference compared to the non-charge-stored case.

When the charges are at Position 2, the I_d-V_g characteristics attract additional attention. Fig. 3-4 shows the conduction band energy along the X-axis of the channel with the stored-charges at Position 2 as V_d = +1V and V_s = 0V. From Fig. 3-4, the electron barrier at V_g = 0V with charges at Position 2 is insignificant compared to the electron barrier of whole channel region. The height of the electron barrier with stored-charges is dragged down by the drain voltage so that the lower half subthreshold curve could keep. When the gate voltage increases to +2V, the electron barrier of channel center decreases apparently. However, the electron barrier with stored-charges decreases much slower than that of channel center, so it can't allow complete electron conduction. Therefore, the upper half subthreshold curve was moved to the right side.

In summary, placing charges at the above-mentioned charge storage positions can result in the movement of the whole subthreshold curve, the movement of the upper half subthreshold curve, or the increase of the GIDL current. Table 3-1 summarizes the effects of the stored-charge distribution on the device characteristics for the L_c = 0.3 μm long channel device.

3-2-2 Non-overlap between S/D and gate

As mentioned in chapter 2, in the structure of the device with channel length L_c = 0.4 μm, the gate does not overlap with the S/D region. In the trapping layer, the stored-charges can be placed at four different positions: Position 1 “above channel center”, Position 2 “above channel and 30nm away from drain junction”, Position 3 “above channel and next to drain junction”, Position 4 “above drain and next to drain junction”, as shown in Fig. 3-5. The I_d-V_g characteristics with stored-charges at various positions are also shown in Fig. 3-5. The non-overlap structure and the overlap structure have the similar I_d-V_g characteristics. It confirms that the effects of stored-charge distribution on device characteristics are according to the relative position between the S/D junction and the stored-charge position, but not according to the overlap between the S/D region and the gate.

It is deserved to be mentioned that the GIDL current significantly increase when the charges are stored at Position 4 of this non-overlap structure. This characteristic gives us an idea to extend the trapping layer to the region under the spacer.

3-3 Short Channel SONOS type NVM

From the simulated results of the long channel device, the memory device can store a bit by the modulation of the V_t or the modulation of the GIDL current respectively. Therefore, it has potential to fabricate a multi-bit memory. Next, the I_d-V_g characteristics of the short channel length (50 nm) device with various charge storage positions are shown in Fig. 3-6.

The charges are located in trapping layer with different lateral positions such as Position 1 “above channel center”, Position 2 “above channel and 3nm away from drain junction”, Position 3 “above channel and next to drain junction”, Position 4 “above drain and next to drain junction”. The I_d-V_g characteristics of the short channel device are similar to the long channel device. The only one difference is the I_d-V_g

characteristic with the stored-charges at Position 2. In the long channel device, only the upper half subthresold curve shifted to the right. However, in short channel device, the whole subthresold curve shifted to the right. Moreover, in the long channel device, the GIDL current did not increase when the stored-charge at Position 2, but the GIDL current increased slightly in the short channel device.

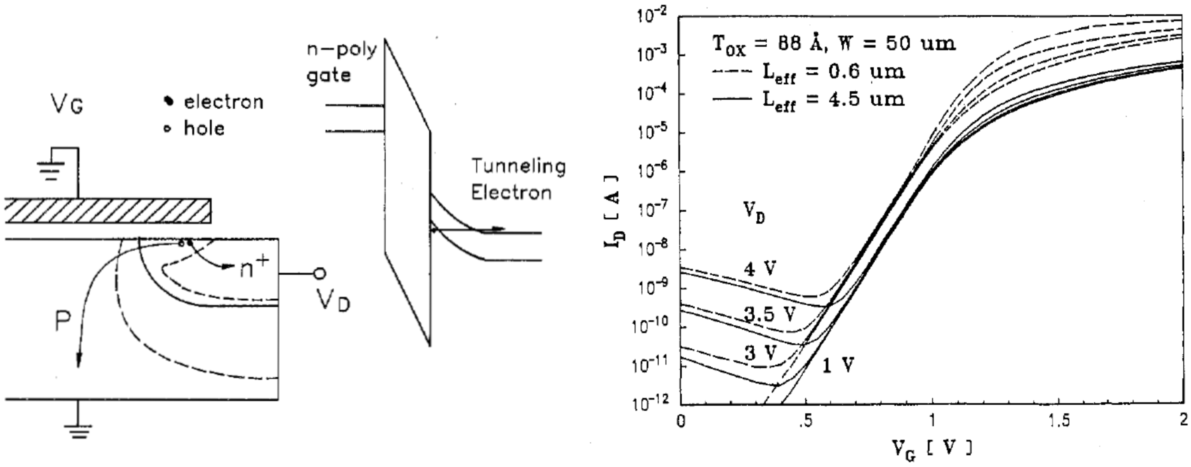
The GIDL current increases when the charges are stored at Position 4 in the short channel device. This phenomenon also occurs in the long channel device. Therefore, it confirms that fabricating multi-bit memory device by extending trapping layer to the region under the spacer is feasible at both long and short channel device. Table 3-2 summarizes the effects of stored-charge distribution on device characteristics for 50 nm short channel device.

Table 3-1: The summary of the effects of stored-charge distribution on device characteristics for channel length 0.3 μm

Position	charge storage	V_t shift	GIDL increase
1	above channel center	Yes	No
2	above channel and 30nm away from drain junction	Yes, only upper half	No
3	above channel and next to drain junction	No	Yes
4	above drain and next to drain junction	No	Yes
5	above drain and 30nm away from drain junction	No	No

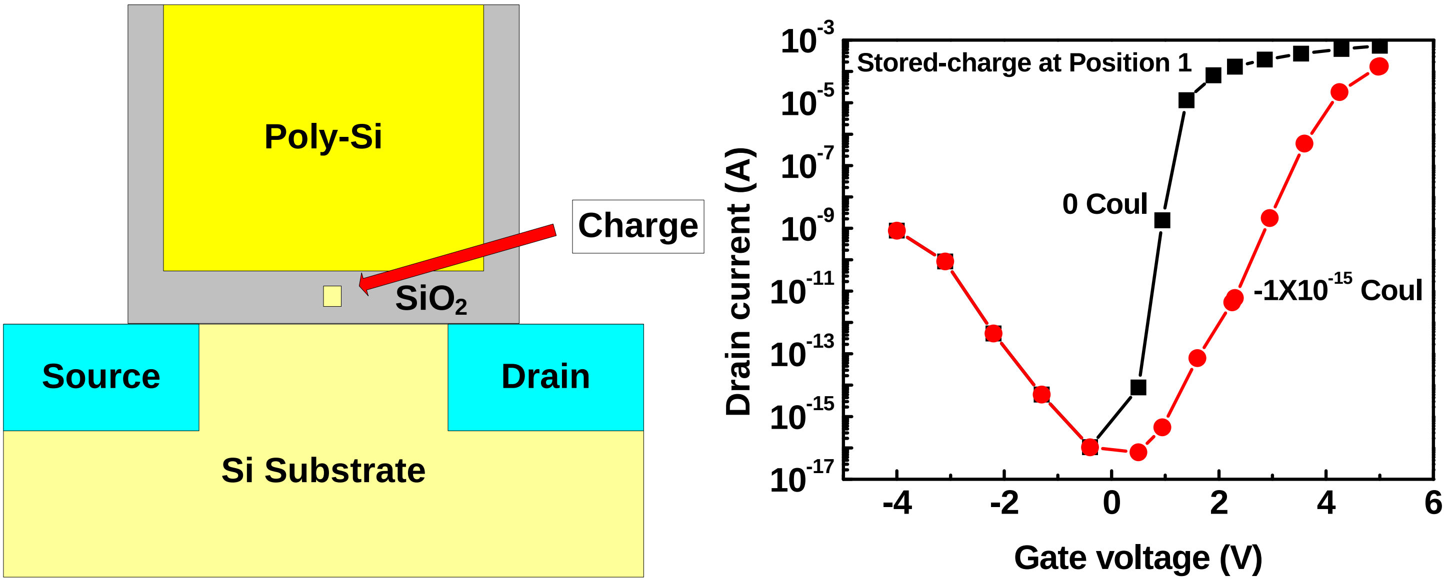
Table 3-2: The summary of the effects of stored charge distribution on device characteristics for channel length 50 nm

Position	Stored charge region	V_t shift	GIDL increase
1	above channel center	Yes	No
2	above channel and 3nm away from drain junction	Yes, slightly	Yes, slightly
3	above channel and next to drain junction	No	Yes
4	above drain and next to drain junction	No	Yes

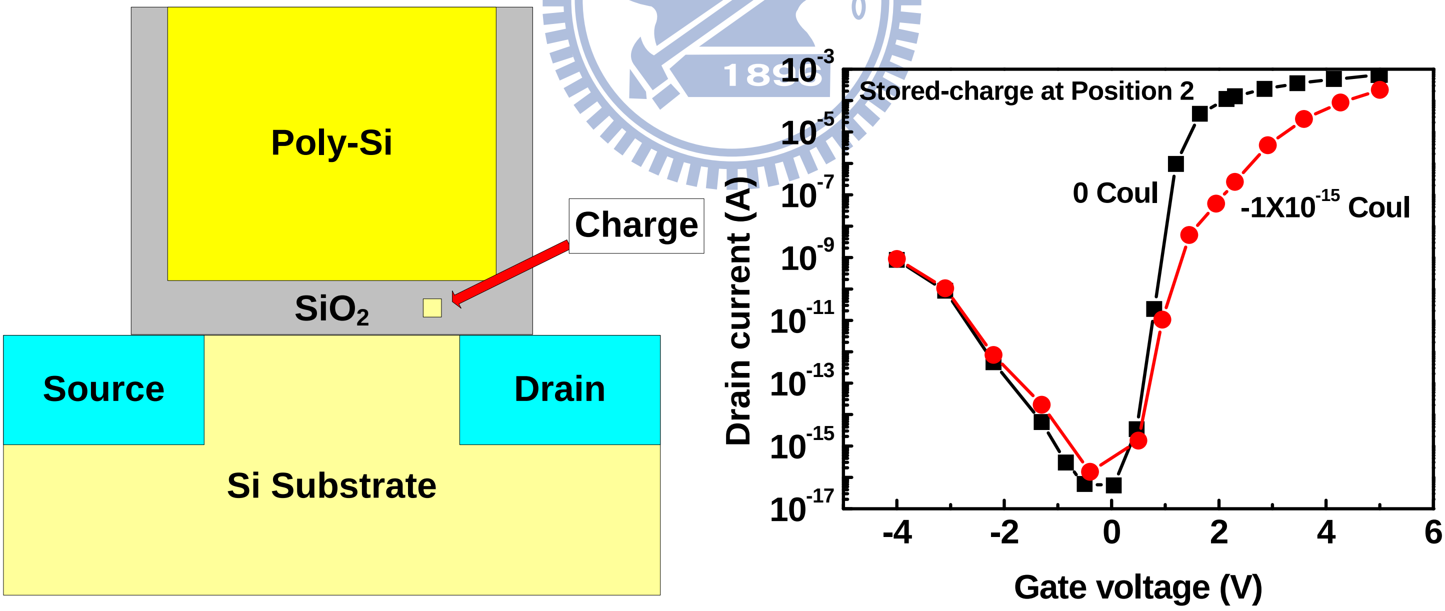


(a) (b)

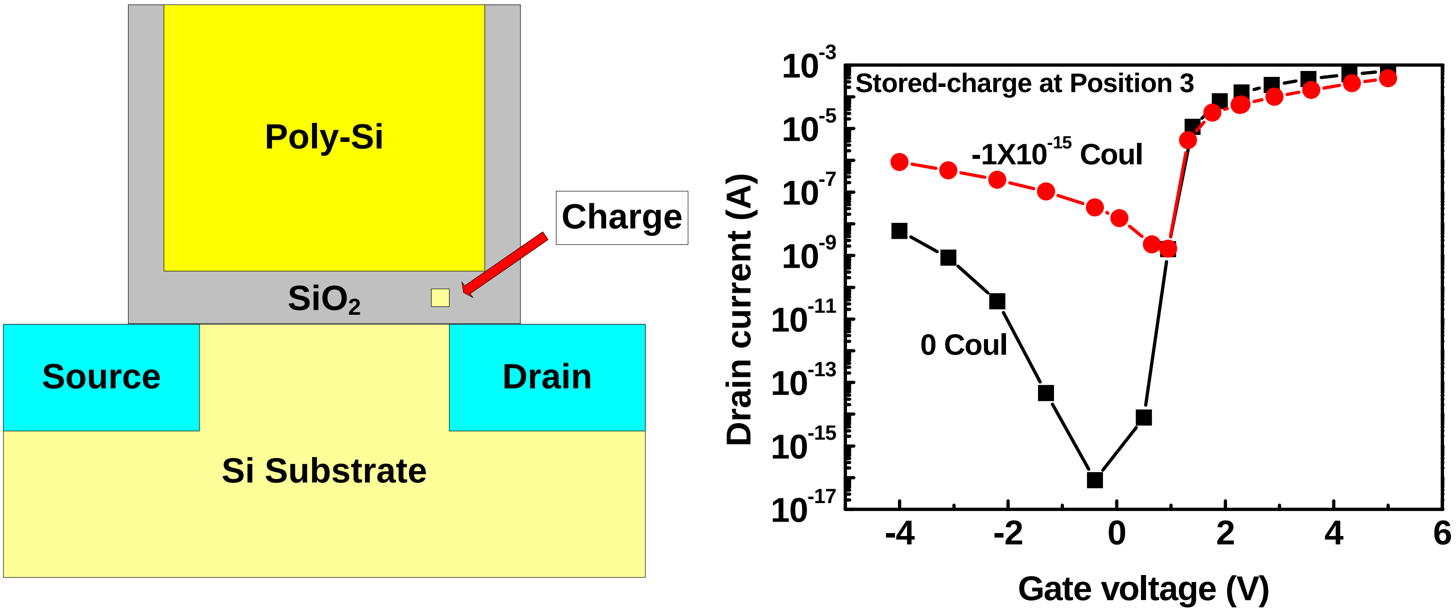
Fig. 3-1: (a) Physical mechanisms of the GIDL (b) current-voltage characteristics of the GIDL phenomenon [50].


Poly-Si
Charge
SiO₂
Source
Drain
Si Substrate
Stored-charge at Position 1
0 Coul
-1X10⁻¹⁵ Coul
Drain current (A)
10⁻³
10⁻⁵
10⁻⁷
10⁻⁹
10⁻¹¹
10⁻¹³
10⁻¹⁵
10⁻¹⁷
-4
-2
0
2
4
6
Gate voltage (V)

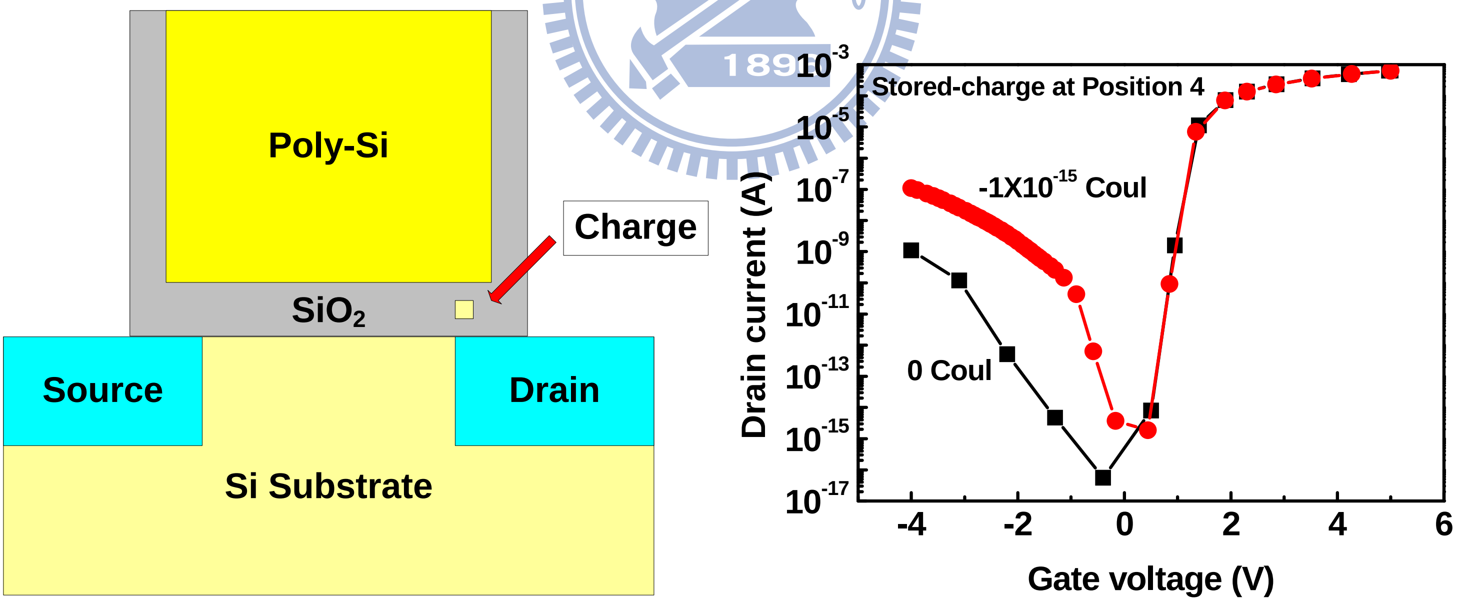
(a)


Poly-Si
Charge
SiO₂
Source
Drain
Si Substrate
Stored-charge at Position 2
0 Coul
-1X10⁻¹⁵ Coul
Drain current (A)
10⁻³
10⁻⁵
10⁻⁷
10⁻⁹
10⁻¹¹
10⁻¹³
10⁻¹⁵
10⁻¹⁷
-4
-2
0
2
4
6
Gate voltage (V)

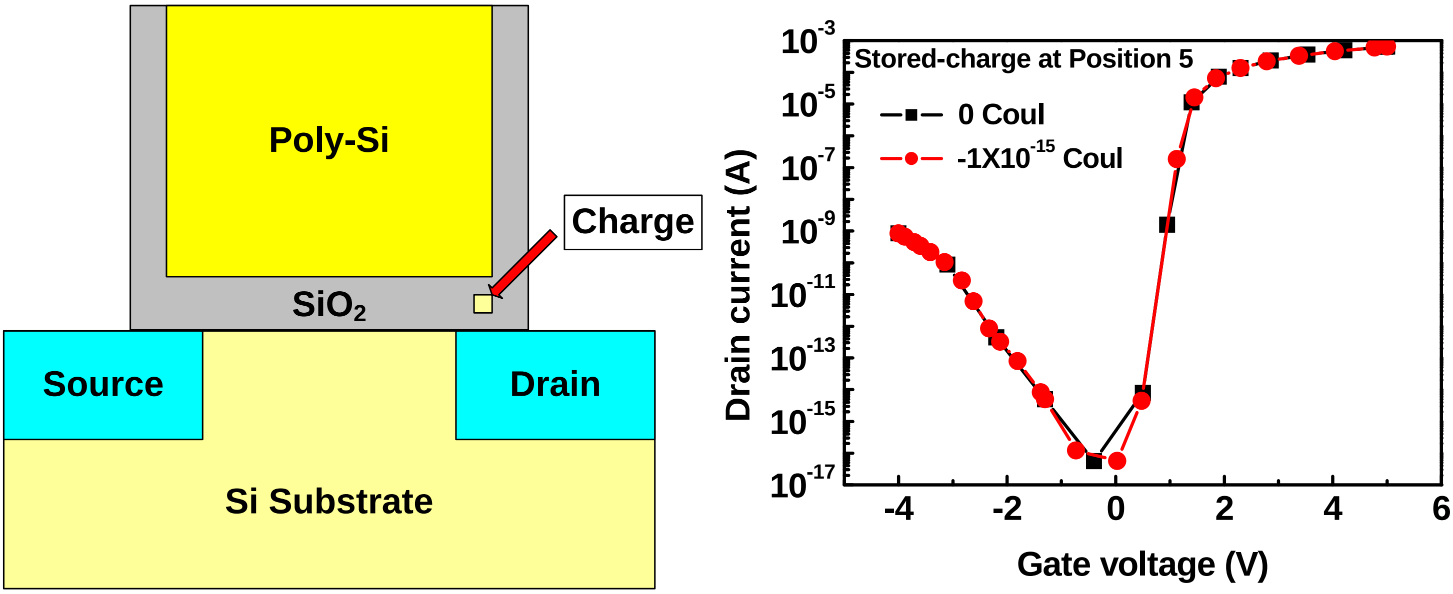
(b)


Poly-Si
Charge
SiO_2
Source
Drain
Si Substrate
Stored-charge at Position 3
$-1X10^{-15}$ Coul
0 Coul
Drain current (A)
10^{-3}
10^{-5}
10^{-7}
10^{-9}
10^{-11}
10^{-13}
10^{-15}
10^{-17}
-4
-2
0
2
4
6
Gate voltage (V)

(c)

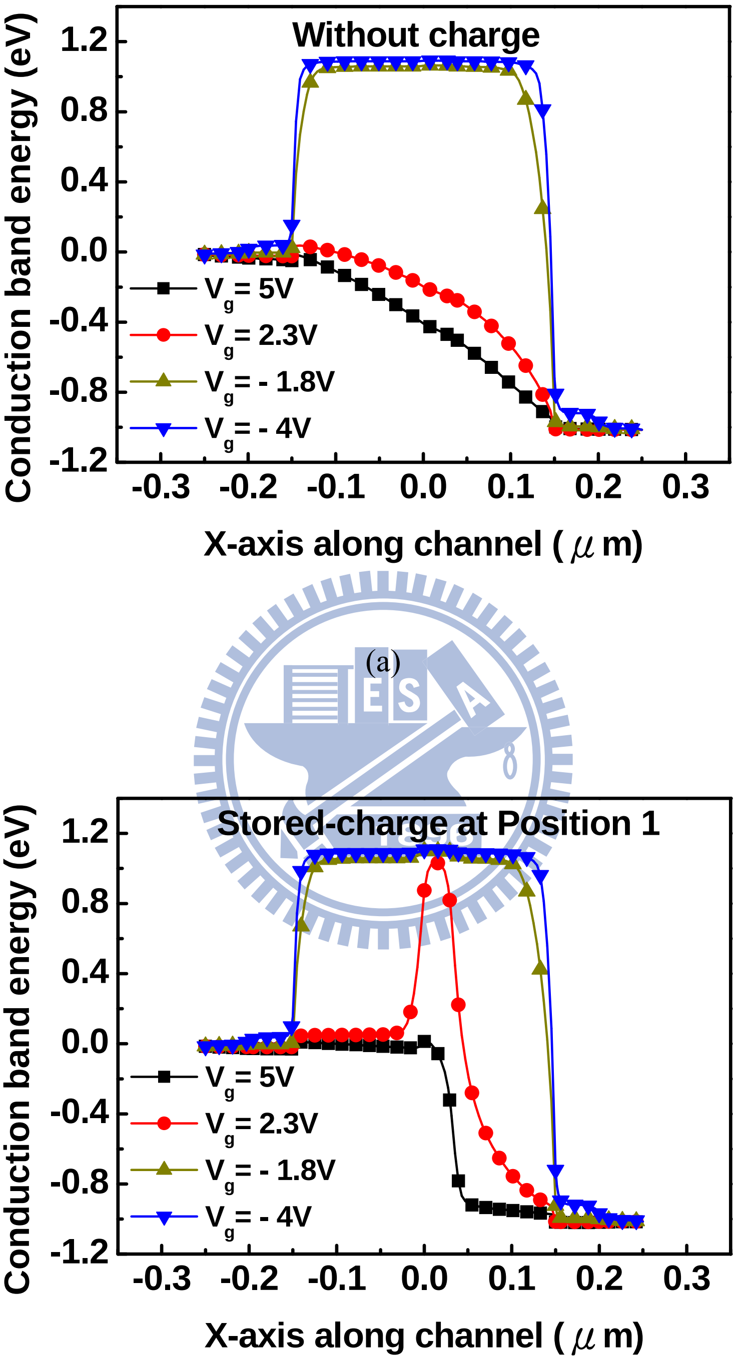

Poly-Si
Charge
SiO_2
Source
Drain
Si Substrate
Stored-charge at Position 4
$-1X10^{-15}$ Coul
0 Coul
Drain current (A)
10^{-3}
10^{-5}
10^{-7}
10^{-9}
10^{-11}
10^{-13}
10^{-15}
10^{-17}
-4
-2
0
2
4
6
Gate voltage (V)

(d)



(e)

Fig. 3-2: The I_d-V_g characteristics with the stored-charges at various positions for the long channel 0.3 μm device. (a) Position 1 “above channel center”, (b) Position 2 “above channel and 30nm away from drain junction”, (c) Position 3 “above channel and next to drain junction”, (d) Position 4 “above drain and next to drain junction”, (e) Position 5 “above drain and 30nm away from drain junction”.



(a)

(b)

Fig. 3-3: Conduction band energy along the X-axis of the channel (a) without charge (b) stored-charge at Position 1.

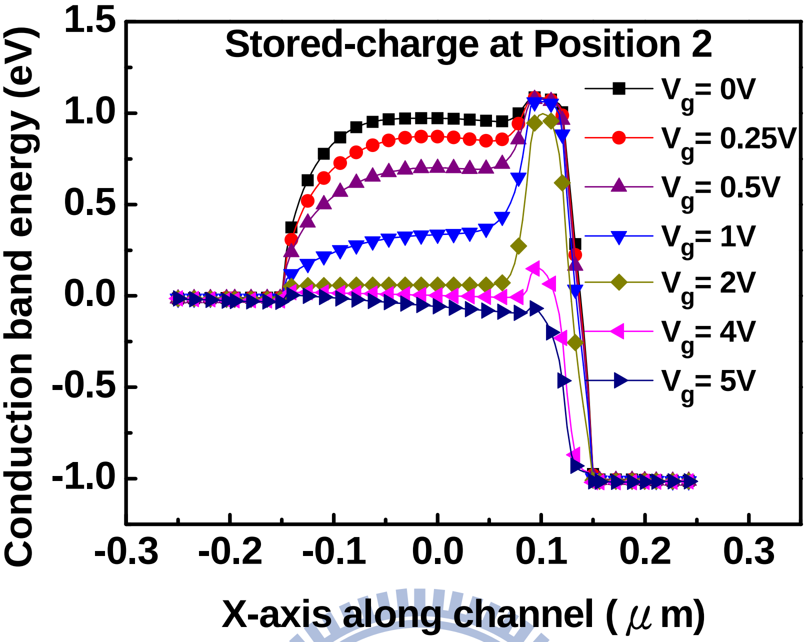
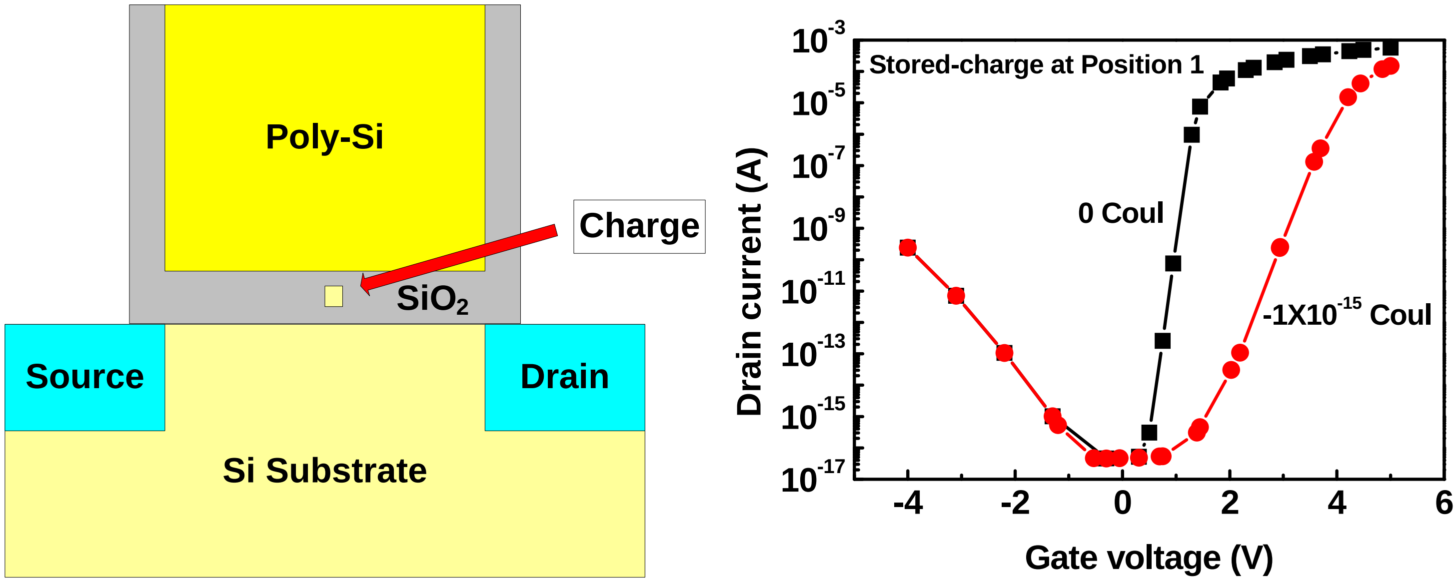
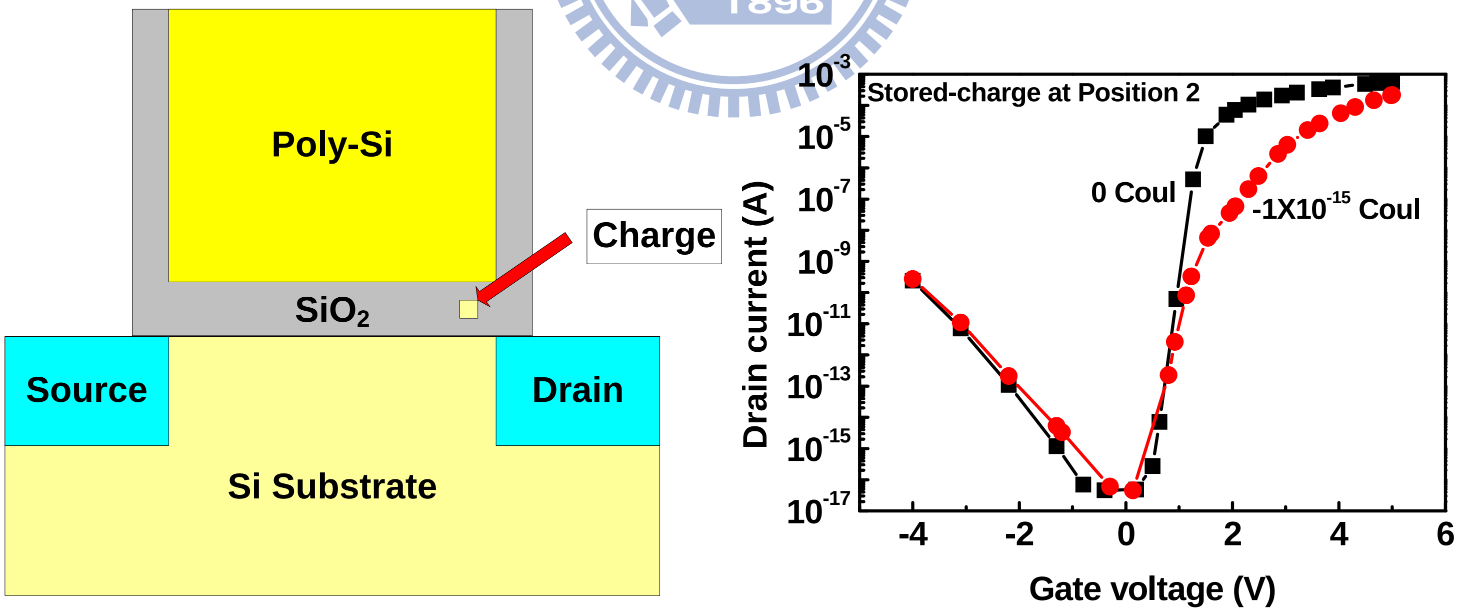


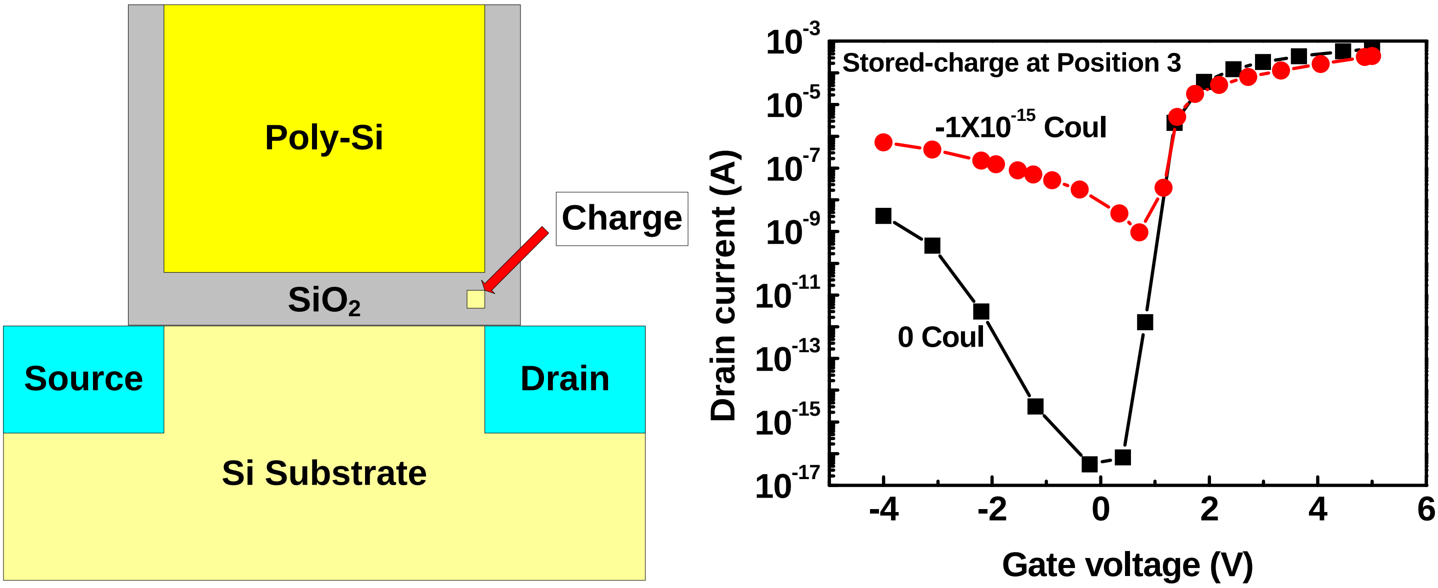
Fig. 3-4: Conduction band energy along the X-axis of the channel with the stored-charges at Position 2 as $V_d = 1V$ and $V_s = 0V$.


Poly-Si
Charge
SiO2
Source
Drain
Si Substrate
Stored-charge at Position 1
0 Coul
-1X10-15 Coul
Drain current (A)
10-3
10-5
10-7
10-9
10-11
10-13
10-15
10-17
-4
-2
0
2
4
6
Gate voltage (V)

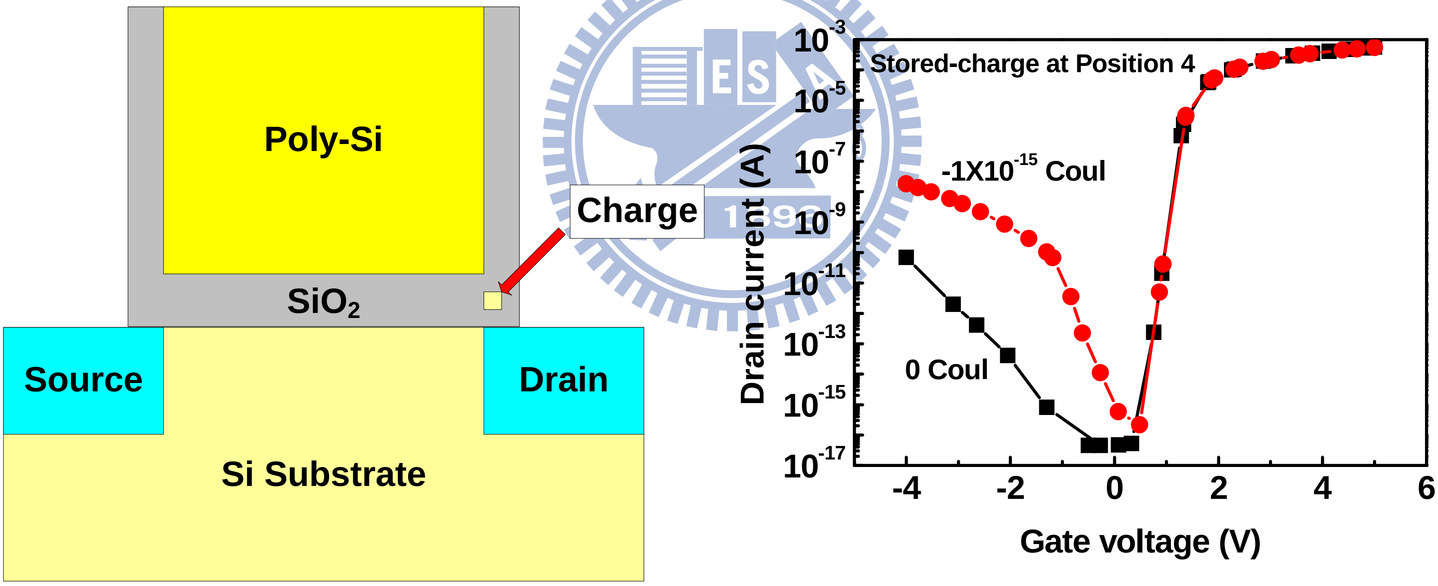
(a)


Poly-Si
Charge
SiO2
Source
Drain
Si Substrate
Stored-charge at Position 2
0 Coul
-1X10-15 Coul
Drain current (A)
10-3
10-5
10-7
10-9
10-11
10-13
10-15
10-17
-4
-2
0
2
4
6
Gate voltage (V)

(b)

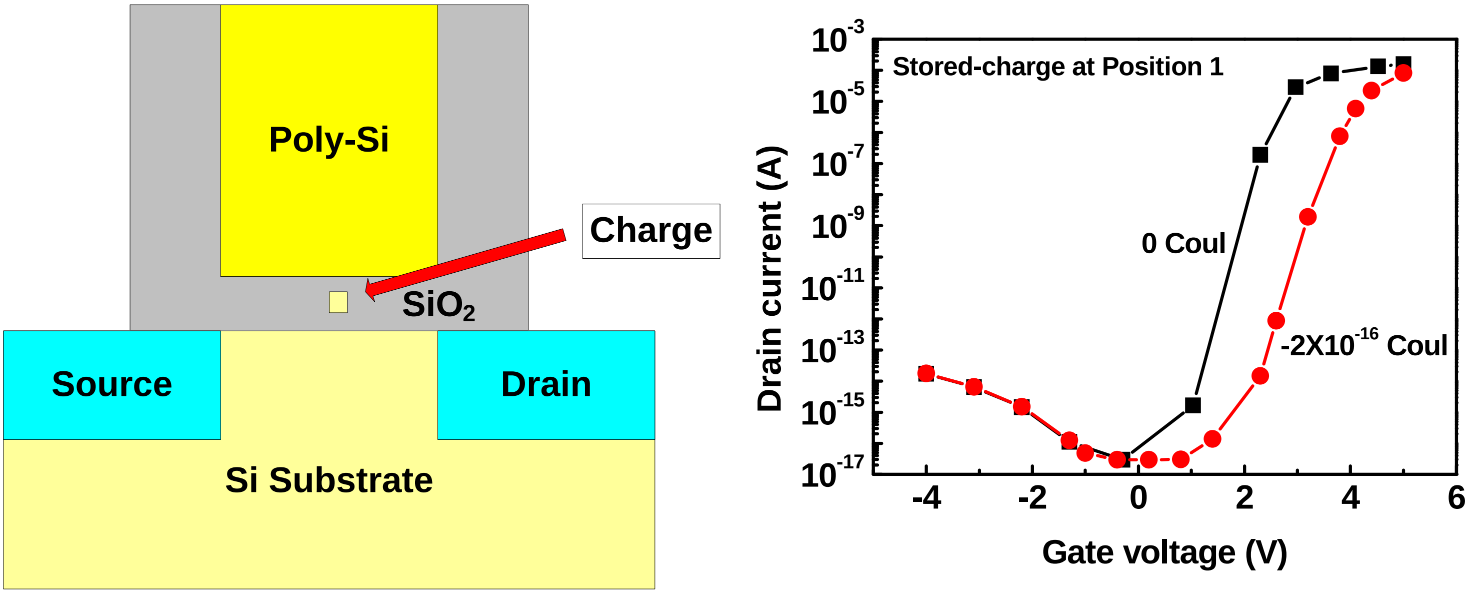


(c)

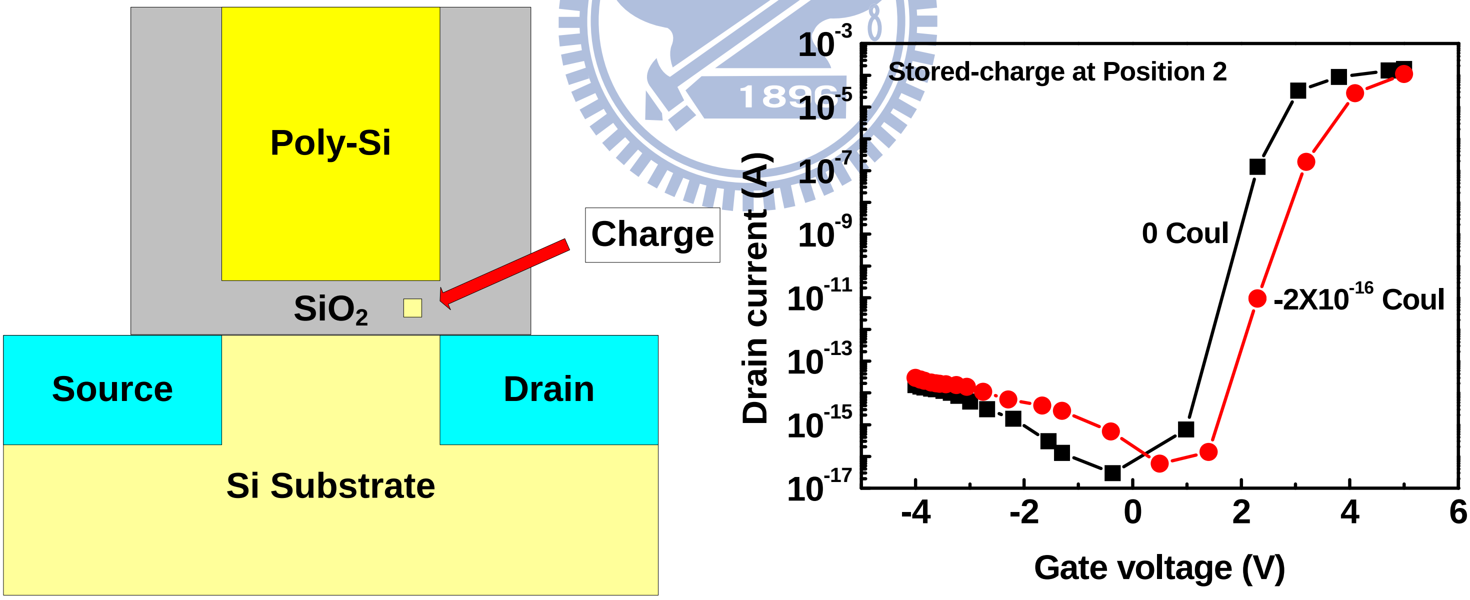


(d)

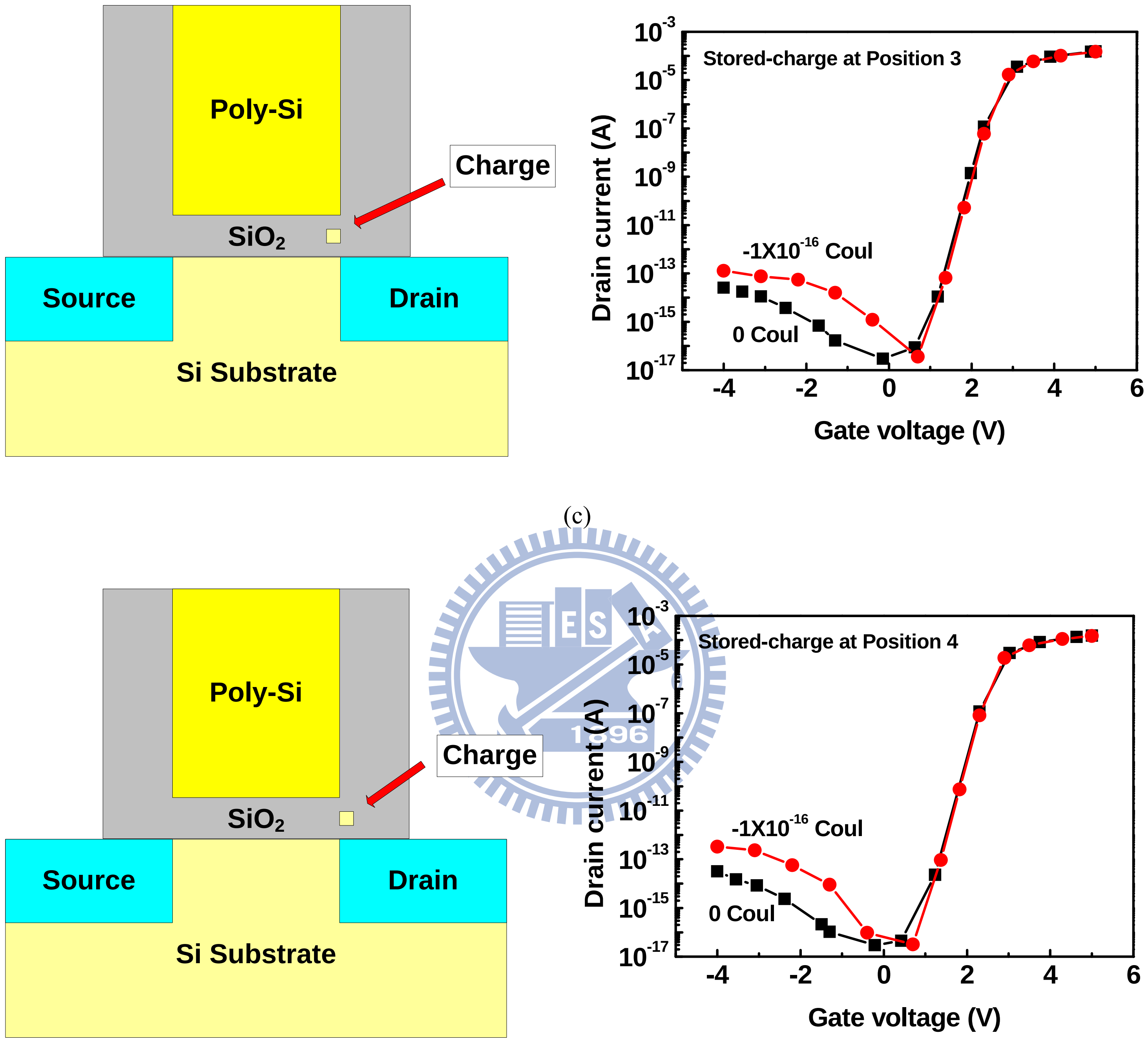
Fig. 3-5: The I_d-V_g characteristics with the stored-charges at various positions for the long channel 0.4 μm device. (a) Position 1 “above channel center”, (b) Position 2 “above channel and 30nm away from drain junction”, (c) Position 3 “above channel and next to drain junction”, (d) Position 4 “above drain and next to drain junction”.


Poly-Si
Charge
SiO$_2$
Source
Drain
Si Substrate
Stored-charge at Position 1
0 Coul
-2X10^{-16} Coul
Drain current (A)
Gate voltage (V)
10^{-3}
10^{-5}
10^{-7}
10^{-9}
10^{-11}
10^{-13}
10^{-15}
10^{-17}
-4
-2
0
2
4
6

(a)


Poly-Si
Charge
SiO$_2$
Source
Drain
Si Substrate
Stored-charge at Position 2
0 Coul
-2X10^{-16} Coul
Drain current (A)
Gate voltage (V)
10^{-3}
10^{-5}
10^{-7}
10^{-9}
10^{-11}
10^{-13}
10^{-15}
10^{-17}
-4
-2
0
2
4
6

(b)



(c)

(d)

Fig. 3-6: The I_d-V_g characteristics with the stored-charges at various positions for the short channel 50 nm device. (a) Position 1 “above channel center”, (b) Position 2 “above channel and 3nm away from drain junction”, (c) Position 3 “above channel and next to drain junction”, (d) Position 4 “above drain and next to drain junction”.

Chapter 4

Characteristics of SAHOS Non-volatile Memory

4-1 Introduction

In this chapter, the 3-bit per cell memory characteristics are demonstrated by the SAHOS memory device. First of all, how to operate multi-bits per memory cell are illustrated, and the possible physical mechanisms of those operations are also discussed. We can store information separately by different methods including the V_t modulation, the GIDL current modulation on forward read and the GIDL current modulation on reverse read. Therefore, memory device with 3-bit per cell can be achieved. Then, the basic memory characteristics including the memory window, the P/E speed, the retention, the endurance and the disturbance of these memory devices are evaluated carefully. The effects of different S/D activation time are also discussed.

All devices discussed in this chapter have dimensions of gate length/gate width (L/W) = 2μm/4μm. The forward read operation means that we apply +1V to the drain and 0V to the source, and the reverse read operation means that we apply 0V to the drain and +1V to the source. The substrate terminal of the devices was always biased at 0V in this chapter.

4-2 Device Structure Inspection

Transmission electron microscopy (TEM) image of the poly-Si/Al_2O_3/HfO_2/

SiO_2/Si (SAHOS) non-volatile memory (sample C) with gate length of 2 μm are shown in Fig. 4-1 and Fig. 4-2. The thickness of the SiO_2 tunneling oxide is about 3.7nm and the thickness of the HfO_2 trapping layer is about 6 nm. The blocking layer Al_2O_3 is about 20nm thick. It is found that the Al_2O_3 blocking layer and the HfO_2 trapping layer are crystallized. This crystallized Al_2O_3 blocking layer may result in the electron back injection from the gate electrode. From Fig. 4-3, it can be found that the HfO_2 charge trapping layer extends to the underneath of the spacer (surround by the dash circle), and its thickness is about 2~3 nm. This structure provides larger space to the memory device to store charges.

4-3 Multi-bit Operation

The S/D of sample B was activated by 900℃ for 60 seconds, and I_d-V_g curve and I_s-V_g curve of multi-bit operation are shown in Fig. 4-4. It is found that this memory device can be operated by the V_t modulation and the GIDL current modulation. It should be noted that the V_t modulation on sample B is the movement of the whole subthreshold curve. We infer that there is no barrier between the S/D and the channel while applying +16V to the gate so that a lot of electrons pass through the tunneling oxide into the trapping layer above the channel center. The S/D of sample A was activated at 900℃ for 20 seconds. The I_d-V_g curve and I_s-V_g curve of the multi-bit operation are shown in Fig. 4-5. It should be noted that the V_t modulation is the movement of the upper half subthreshold curve. According to the simulated results of chapter 3, we can infer that the stored charges in the trapping layer are positioned above the channel and near the S/D junctions. We suspect that there is an electron barrier between the S/D and the inverted channel while applying +14V to the gate so that only few electrons can pass through this barrier and be injected to the trapping layer above the channel center.

Therefore, we can infer that the S/D junctions do not overlap with the gate electrode on sample A because the S/D annealing time is only 20 seconds so that the lateral diffusion of the S/D dopants is not long enough. On the other hand, the S/D junctions overlap with the gate electrode on sample B because the S/D annealing time is 60s so that the S/D dopants diffuse to the underneath of the gate electrode.

The modulation of the whole subthreshold curve, the modulation of the upper half subthreshold curve and the modulation of the GIDL current are presented by combining the I-V curves of sample A with sample B. Therefore, we can demonstrate the effects of the stored-charge distribution on the device characteristics as shown in chapter 3. In addition, we demonstrate that this multi-bit operation is practicable.

4-3-1 Sample B and C: Sufficient gate to source/drain overlap

At the beginning, the trapping layer on sample B was divided into 3 regions laterally as shown in Fig. 4-6. They are Region 1 "above drain junction", Region 2 "above channel center", and Region 3 "above source junction".

The I_d-V_g curve and I_s-V_g curve of the multi-bit operation on sample B have been shown in Fig. 4-4. How to operate this 3-bits memory device is explained as follows. First, we let the electrons from the gate tunnel into the trapping layer by applying -16V to the gate and 0V to the S/D, and then we call it the 1st state. The electrons are stored at Region 1 and 3, so the GIDL current increase largely on the forward read operation or the reverse read operation. However, only few electrons were injected into Region 2 which resulted in small V_t shift. Then, the voltage (+11V) was applied to the drain and the voltage (0V) was applied to the gate and the source so that the 2nd state can be obtained. During the 2nd operation, the electrons stored at Region 1 tunneled through the tunneling oxide into the drain due to the high drain voltage (+11V) so the GIDL current on the forward read could be reduced. Next, the 3rd state was achieved by applying +11V to the source and 0V to the gate and the drain, and then the stored charges at Region 3 could tunnel through the tunneling oxide to the source which results in the reduction of the GIDL current on the reverse read. Afterward, in order to restore charges at Region 1 and 3, the gate was biased at -16V, and the S/D was biased at 0V. After that, the voltage (+11V) biased to the source to reduce charges stored at Region 3, and then we obtained the 4th state which the GIDL current was large on the forward read and was small on the reverse read. Then, the drain was biased at +11V to reduce the charges stored at Region 1. Next, the gate was biased at +16V and the S/D were biased at 0V. The electrons were injected from the S/D into Region 1, 2 and 3, and then the 5th state was achieved. Since the S/D overlap with the gate, there is no barrier between the S/D and the channel while applying +16V to the gate so that a lot of electrons passed through the tunneling oxide into the center of trapping layer, Region 2. It is obvious that whole subthreshold curve was moved to the right because of the charges stored at Region 2. Then, we apply voltage (+11V) to the drain or the source to decrease the charges stored at Region 1 or 3 respectively, and then it resulted in the 6th, 7th and 8th state.

In addition, the charges stored at Region 2 can be reduced by applying high negative voltage to the gate and 0V to the S/D. Table 4-1 summarizes the 8 states of I_d-V_g and I_s-V_g characteristic of the memory device on sample B.

The S/D of sample C activated by 900℃ for 180 seconds, and I_d-V_g curve of multi-bit operation is shown in Fig. 4-7. Since the S/D annealing time for sample C is 180s, the S/D junctions overlap with the gate. Therefore, the I_d-V_g characteristics of sample C are similar to sample B so that we demonstrated the basic memory performances only for sample B.

4-3-2 Sample A: Insufficient gate to source/drain overlap

At the beginning, we divided the trapping layer on sample A into five regions laterally such as shown in Fig. 4-8. They are Region 1 "above drain junction", Region 2 "above channel and near drain junction", Region 3 "above channel center", Region 4 "above channel and near source junction", and Region 5 "above source junction".

The I_d-V_g curve and I_s-V_g curve of the multi-bit operation on sample A have been shown in Fig. 4-5. The operation of the 3-bits memory device is explained as follows. First, due to the crystallized Al_2O_3 blocking layer, we let electrons from the gate tunnel into the trapping layer by applying -16V to the gate and 0V to the S/D, and then we call it the 1st state. Because the electric field between the gate and the S/D is larger than the electric field between the gate and the channel, the electrons were injected from the gate into Region 1 and 5. Since the electrons are stored at Region 1 and 5, the GIDL current increase largely on both forward read operation and reverse read operation. However, only few electrons were injected into Region 1, 2, and 3 and then resulted in small V_t shift. Then, the voltage (+10V) was applied to the drain and the voltage (0V) was applied to the gate and the source so the 2nd state can be obtained. During the 2nd operation, the electrons stored at Region 1 tunneled through the tunneling oxide into the drain due to the high drain voltage (+10V) so the GIDL current of the forward read could be reduced. Next, the 3rd state was achieved by applying +10V to the source and 0V to the gate and the drain, and then the stored charges at Region 5 could tunnel through the tunneling oxide into the source which resulted in the reduction of the GIDL current on the reverse read. Afterward, in order to let Region 1 and 5 restore charges, the gate was biased at -16V, and the S/D was biased at 0V. After that, the voltage (+10V) biased to the source to reduce charges stored at Region 5, and then we obtained the 4th state which the GIDL current was large on the forward read and was small on the reverse read. Then, the drain is biased to +10V to reduce the charges stored at Region 1. Next, the gate was biased at +14V and the S/D were biased at 0V. The electrons were injected from the S/D into Region 1, 2, 4 and 5, and then the 5th state was achieved. Since the S/D does not overlap with the gate, there is the electron barrier between the S/D and the channel while applying +14V to the gate. Therefore, only few electrons passed through this barrier into the center of trapping layer, Region 3. From the results of the simulations in chapter 3, it is obvious that the upper half subthreshold curve is moved to the right because of the charges stored at Region 2 and 4. Then, we applied the voltage (+10V) to the drain or the source to decrease the charges stored at Region 1 or 5 respectively, and it resulted in the 6th, 7th and 8th state.

In addition, the charges stored at Region 2 and 5 can be reduced by applying high negative voltage to the gate and 0V to S/D. Table 4-2 summarizes the 8 states of I_d-V_g and I_s-V_g characteristic of the memory device on sample A.

4-4 P/E Speed and Memory Window

4-4-1 Sample B: Sufficient gate to source/drain overlap

The I_d-V_g characteristics after applying different gate voltages with pulse width 1ms are shown in Fig. 4-9. The V_t shifts after program and erase operations at different gate voltages with different pulse widths are shown in Fig. 4-10 and Fig. 4-11, respectively. It can be found that the program and erase speed increase apparently when the gate voltage increases. The maximum memory window is about 7V. The programming time could be as short as 1μs at $V_g = +18V$, and the memory window is about 1.7V. However, the erase speed is slower than the program speed, so the memory device requires pulse width 100μs at $V_g = -18V$ to erase the 1.7V memory window. The electrons are injected into the trapping layer from the channel under the program operation. The holes are injected into the trapping layer from the channel under the erase operation. Because the hole barrier between the channel and tunneling oxide is larger than the electron barrier, the erase speed is slower than program speed. When the pulse width increases, the saturation phenomenon of erase threshold voltage occurs due to some electrons injected from the gate into the trapping layer. On the other hand, the saturation phenomenon of program threshold voltage occurs due to some holes injected from the gate into the trapping layer when the pulse width increases.

The GIDL current is measured at constant gate voltages of -7V as the drain voltage is fixed at +1V. The increase of the GIDL current after programming at different gate voltages with different pulse widths is shown in Fig. 4-12. Then, as shown in Fig. 4-13, the erase speed diagram is evaluated by applying different drain voltages with different pulse widths to reduce the GIDL current on the forward read. It is observed that the program and erase speed increase apparently when the operation voltage increases. The maximum memory window is about two orders of magnitude and can be obtained by applying -18V to gate for 10ms. However, at gate voltage -14V, -16V and -18V, the memory window after programming with pulse width 1s is narrower than those with pulse width 0.1s. The possible reason of this phenomenon is that the holes generated from the substrate are injected into the trapping layer to eliminate some electrons stored in the trapping layer. This memory device can be programmed with pulse width 1μs at V_g = -18V, and the increase of the GIDL current is about one order. However, this memory device requires pulse width 100μs to erase one order memory window at V_d = +11V. Generally, in order to erase the maximum memory window of the GIDL current, V_d requires +11V, and pulse width requires 10ms. When the pulse width is longer than 10ms at V_d = +11V, the level of the GIDL current would not change anymore because all stored charges are already detraped. It should be noted that the maximum V_d is limited by the drain junction breakdown voltage. Therefore, the allowable erase voltage would be lower than the program voltage typically. The erase speed at $V_d = +10V$ is much faster than $V_d = +9V$. The possible reason is that stored electrons are detraped from the trapping layer into the drain by direct tunneling when drain voltage is lower than +9V. On the contrary, some stored electrons can be detraped by FN tunneling while drain voltage is higher than +10V.

4-4-2 Sample A: Insufficient gate to source/drain overlap

The V_t shift after the program and erase operation at different gate voltages with different pulse widths is shown in Fig. 4-14 and Fig. 4-15, respectively. It is observed that the program and erase speed increase dramatically when the gate voltage increases. The maximum memory window is about 5.5V. The programming time could be as short as 1μs at $V_g = +18V$, and the memory window is about 1.4V. However, the erase speed is slower than program speed so the memory device requires a pulse width of 100μs at $V_g = -18V$ to erase the 1.7V memory window. The saturation phenomenon of program and erase threshold voltage occurs when the pulse width increases, and it is due to same reason explained on sample B.

The GIDL current is measured at constant gate voltages of -5V as the drain voltage is fixed at +1V. The increase of the GIDL current after programming at different gate voltages with different pulse widths is shown in Fig. 4-16. Then, as shown in Fig. 4-17, the erase speed is evaluated by applying different drain voltage with different pulse widths to reduce the GIDL current on forward read. It is observed that the program and erase speed increase dramatically when the operation voltage increases. The maximum memory window is about two orders of magnitude and can be obtained by applying -18V to V_g for 10ms. However, at gate voltage -14V, -16V and -18V, the memory window after programming with pulse width 1s is always narrower than those with pulse width 0.1s. It is similar to sample B. This memory device can be programmed with pulse width of 1μs at V_g = -18V, and the increase of the GIDL current is about 6 times. However, this memory device requires a pulse width of 100μs at V_d = +10V to erase the program state. Generally, in order to erase the maximum memory window of the GIDL current, V_d requires +10V, and the pulse width requires 10ms. When the pulse width is longer than 10ms at V_d = +10V, the level of the GIDL current would not change anymore because all stored charges are already detraped. As on sample B, the maximum V_d on sample A is also limited by the drain junction breakdown voltage. When V_d is +10V, the erase speed increases largely. It confirms our inference discussed on sample B.

4-5 Retention Performance

The initial memory window of the V_t modulation is divided into the small memory window and the large memory window. The large memory window means that the crowded charges are stored in the trapping nodes. The crowded charges easily encounter strong Coulomb repulsive force between the charges. The strong Coulomb repulsive force will discharge the charges stored in the trapping nodes quickly and increase the charge loss rate. Therefore, when we evaluate the retention performance by extrapolating to 10 years, the large memory window may have worse performance. For this reason, we also evaluate retention performance of the small memory window.

4-5-1 Sample B: Sufficient gate to source/drain overlap

Fig. 4-18 shows the retention characteristics of the V_t on sample B at 25℃ and 85℃. The small memory window was achieved by the low program voltage, and the retention performances were evaluated at both 25℃ and 85℃. After 10^5 seconds, the small memory window still maintained the same level at 25℃, and the stored charges were almost reserved while extrapolating to 10 years. The retention characteristic of the small memory window at 85℃ showed poorer retention performance than at 25℃, but it is still good. Only 27% charges were lost at 85℃ after 10^5 seconds. Extrapolating to 10 years, more than 55% charges are reserved at 85℃.

The large memory window is about 4.5V and is achieved by the high program voltage. After 10^5 seconds, the large memory window narrowed 14% at 25℃. After 10 years retention time at 25℃, the remained memory window is about 3V. On the other hand, 35% charges were lost at 85℃ after 10^5 seconds, and the remained memory window is about 1.5V while extrapolating to 10 years.

The V_t shift on sample B is as a result of the charges stored in trapping layer above the channel center, so the lateral charge migration effect [24] is small. Therefore, the retention performance of the V_t on sample B is very good.

The retention characteristics of the GIDL current on sample B are evaluated at both 25℃ and 85℃ as shown in Fig. 4-19. The GIDL current is measured at constant gate voltages of -7V as the drain voltage is fixed at +1V. Then, we can have the initial memory window about two orders of magnitude. At the beginning, the current ratio between the program state and the erase state was about 90. After 10^5 seconds, the current ratio reduced to 20 at 25℃. Extrapolating to 10 years, the current ratio is only 2. In addition, the retention performance at 85 ℃was much poorer than at 25℃. In summary, the retention performance of the GIDL current on sample B is poor. The possible reasons are discussed as follows. As we have inferred, the S/D junctions overlap with the gate on sample B. However, the overlap region is quite small because the S/D annealing time is only 60 seconds. In order to modulate the GIDL current, the positions of the stored charges must be above S/D junctions so they may be near the corner of the spacer. At the corner of the spacer, there are a lot of defects generated from the device fabrication process. These defects will enhance the charge loss rate. The other reason is due to “lateral charge migration” [24]. When the charges are stored in the HfO_2 trapping layer at the source side or the drain side, the charges will move toward the trapping site above the channel center, and then the retention performance will degrade [24].

<u>4-5-2 Sample A: Insufficient gate to source/drain overlap</u>

Fig. 4-20 shows the retention characteristics of the V_t on sample A at 25℃ and 85℃. The small memory window is about 3.2V and is achieved by the low program voltage. The retention characteristics at 25℃ and 85℃ are shown in the figure. After 10^5 seconds, the small memory window narrowed 31% at 25℃. Extrapolating to 10 years, the small memory window vanishes at 25℃. The retention characteristics of the small memory window at 85℃ showed much poorer performance than at 25℃. The large memory window was about 4.6V and achieved by the high program voltage. After about 10^5 seconds, the memory window narrowed 40% at 25℃. The large memory window also vanishes after 10 years retention time at 85℃.

The retention characteristics of the GIDL current on sample A at 25℃ is shown in Fig.4-21. The GIDL current is measured at constant gate voltage of -5V as the drain voltage is fixed at +1V. Then, we can have the initial memory window of about two orders of magnitude which is as same as sample B. At the beginning, the current ratio between the program state and the erase state is 150. After 10^5 seconds, the current ratio reduces to 40. Extrapolating to 10 years, the memory window vanishes.

Both the V_t and the GIDL current showed poor retention characteristics. There are two possible reasons. The one is the escape of the charges which are stored in the trapping layer. As we have mentioned, the S/D junctions do not overlap with the gate on sample A. On sample A, the positions of the stored charges are near S/D junctions or above S/D junctions so they must be under the spacer or near the corner of the spacer. There is no Al_2O_3 blocking oxide under the spacer or near the corner of the spacer so the charges can easily escape from the trapping layer. Moreover, above the charge trapping positions on sample A, there are a lot of defects generated from the device fabrication process. These defects will enhance the charge loss rate. The other reason is due to “lateral charge migration” [24]. When the charges are stored in the HfO_2 trapping layer at the source side or the drain side, the charges will move toward the trapping site above the channel center, and then the retention performance will degrade [24].

4-6 Endurance Performance

4-6-1 V_t modulation

The endurance performance of the V_t on sample A and B are shown in Fig. 4-22 and Fig. 4-23, respectively. If the stored electrons are not completely eliminated after every erase operation, then every P/E cycle leaves a few electrons in trapping layer resulting in a slight increasing on the V_t. On the other hand, if the tunneling layer is damaged during the P/E operations, the device can not have the memory function anymore. In this work, the memory windows of the V_t did not degrade after 10^4 P/E cycles on sample A and 10^5 P/E cycles on sample B. However, the V_t of both the program state and the erase state slightly increased after every P/E cycle. For this reason, it seems more correctly to define the endurance performance as the difference between the 1st program state and the final erase state. Due to this reasonable definition of the endurance performance, the endurance performance degraded 10% on sample A after 10^4 P/E cycles and degraded 22% on sample B after 10^5 P/E cycles. Therefore, sample B can be operated more than 10^5 times and then still have the allowable memory window on the V_t.

4-6-2 GIDL modulation

The endurance performance of the GIDL current on sample A and B are shown in Fig. 4-24 and Fig. 4-25, respectively. The endurance performance degraded noticeably after every P/E cycle on both sample A and B. After 200 P/E cycles, the memory windows are quite small on sample A and B. One possible reason is the change of the stored charges distribution after every P/E cycle.

In order to evaluate the changes of vertical electric field after every P/E cycle, the simulated structure (channel length = 0.4μm) defined in chapter 2 is used. Only one difference is that the stored-charge position is defined as SiO_2 replacing silicon. As shown in Fig. 4-26, the lateral charge position is set above the drain junction. The vertical electric field at program operation is drawn along upper line, and the vertical electric field at erase operation is drawn along lower line. Fig. 4-27 shows the vertical electric field along X-axis at program operation (gate voltage = -5V, source and drain voltage = 0V). It is found that the maximum electric field is slightly reduced and the electric field at the left side is increased. Fig. 4-28 shows the vertical electric field along X-axis at erase operation (drain voltage = +4V, gate and source voltage = 0V). It is found that the maximum electric field is moved toward the right side.

The illustration of stored charges distribution at the drain side after every P/E cycle is shown in Fig. 4-29. After the 1st program operation, the electrons are stored above the drain junction and then result in the high GIDL current. However, a few electrons are left in the trapping layer after the 1st erase operation. Therefore, at the 2nd program operation, the vertical electric field above the drain junction is slightly reduced by those few remaining electrons, so the electrons injected into the trapping layer above the drain junction decrease during the 2nd program operation. In addition, some injected electrons are stored close to the center of trapping layer as a result of the increase of the vertical electrical field which is at the left side of the drain junction. During the 2nd erase operation, the maximum vertical electric field is moved toward the right side so that more stored electrons can't be erased. After the 2nd erase operation, more stored electrons remain resulting in the increase of GIDL current. Consequently, after every P/E cycle, the GIDL current of the program state decreases, and the GIDL current of the erase state increases. Finally, the memory window will vanish.

4-7 Disturbance Performance

As a result of sharing the same word line and bit line in a memory array, the disturbance problems occur. Generally speaking, the main voltage disturbance on flash memory can be divided into "read disturbance", "gate disturbance", and "drain disturbance". According to the arrangements of word lines and bit lines, the memory array is divided into the NAND array and the NOR array. Therefore, we should consider which disturbance characteristics according to which memory array is adopted. In order to operate these 3-bits memory devices, the NOR array must be adopted as shown in Fig. 4-30. If we operate the device A, the memory state of the other memory devices which share the same word line or bit line with the device A will be disturbed. For example, as shown in Fig. 4-30 (a), the device B shares the same bit line with the device A, and the device C shares the same word line with the device A.

Fig. 4-30 (b) shows the voltage of the word lines and the bit lines under different operations. When the device A operates at read procedure, the read disturbance is defined as the change of the V_t or the GIDL current on the device A itself. When the device A operates at high gate voltage bias, the device C sharing the same word line may exhibit the variation of the V_t and the GIDL current. This phenomenon is called the gate disturbance. When the device A is programmed by the high positive gate voltage (+14V), the device C shares the same word line with the device A and has the same gate voltage (+14V). In order to reduce the vertical electrical field between the gate and the S/D region, both the S/D of the device C is biased at +7V. In contrast, when the device A is erased by the high negative gate voltage (-14V), both the S/D of the device C are necessary biased at -7V. When the device A operates at the high drain voltage bias (+11V) to erase the forward GIDL current, the device B sharing the same bit line may exhibit the variation of the V_t and the GIDL current. This phenomenon is called the drain disturbance. For reducing the vertical electrical field between the gate and the drain, the gate voltage (+3V) is applied to the drain of the device B.

Consequently, we have to discuss all voltage disturbance effects including “read disturbance”, “gate disturbance”, and “drain disturbance”. All disturbance characteristics discussed in this chapter are based on sample B. The V_t modulation of sample B is the movement of the whole subthreshold curve so sample B have more potential to be adopted into industry than sample A. In addition, the retention performance of sample B is better than sample A. Therefore, we discuss the disturbance characteristics only on sample B. The disturbance time is set to 1000 seconds, which is equivalent 10^6 read, program, or erase operations as each of the operation time is 1ms.

4-7-1 Read Disturbance

The positive word-line voltage and the bit-line voltage may cause the electrons injection or the channel hot electrons injection during read operation. Then, the unwanted V_t shift or the unwanted change of the GIDL current occurs. Fig. 4-31 and Fig. 4-32 show the read disturbance characteristics of the V_t and the GIDL current respectively when the gate voltage is biased at 4V. It is observed that the read disturbance in these SAHOS non-volatile memory is weak and negligible at V_g = 4V. This is a reasonable result because 4V is well lower than the P/E voltages for both V_t modulation and GIDL modulation. On the other hand, when the gate voltage is biased at -8V, Fig. 4-33 and Fig. 4-34 show the read disturbance characteristics of the V_t and the GIDL current respectively. The reason we choose the gate voltage as -8V is that we want to make sure the memory windows are one order of magnitude at least on every device. Therefore, the gate voltage (-8V) is the worst case for these memory devices during the read operation. After 1000 seconds, the V_t of the program state is slightly reduced as shown in Fig. 4-33 because the gate voltage (-8V) is high enough to detrap some stored electrons. Because the V_t shift is only -0.6V, it is not a serious issue for the device reliability. Fig. 4-34 reveals that the GIDL current is also affected by the high negative gate voltage (-8V). Some stored electrons at the program state were detrapped and then resulted in the reduction of the GIDL current. Besides, at the erase state, some electrons were injected from the gate into the trapping layer and then resulted in the increase of the GIDL current. However, after 1000 seconds, the changes of the GIDL current for both the program state and the erase state are small. The current ratio reduced from 20 to 9 and was still distinguishable, so it is not big deal for the device reliability. In summary, the read disturbance of this SAHOS memory device is under the acceptable condition.

4-7-2 Gate Disturbance

Under the program operation, Fig. 4-35 and Fig. 4-36 show the gate disturbance characteristics of the V_t and the GIDL current respectively. After 1000 seconds, it can be found that the gate disturbance is not an issue under the program operation. Under the erase operation, Fig. 4-37 and Fig. 4-38 show the gate disturbance characteristics of the V_t and the GIDL current respectively. The V_t shift is small and negligible. However, the GIDL current of the program state is reduced at high negative gate voltage (-14V). One possible reason is that some stored electrons were detrapped by the high negative gate voltage. The other possible reason is the lateral charge migration as a result of the lateral electrical field in the trapping layer. Because we apply -7V to the source and the drain terminal, and the substrate terminal is always fixed at 0V. Therefore, the lateral electrical field may exist in the trapping layer. Even so, the gate disturbance is also under the acceptable situation. Because current ratio is still 4 after 1000 seconds stress and the P/E state can be distinguishable.

4-7-3 Drain Disturbance

The drain disturbance characteristics of the V_t are shown in Fig. 4-39. After 1000 seconds, it reveals that the drain disturbance of the V_t is insignificant at this specific stress condition. On the other hand, Fig. 4-40 shows the drain disturbance characteristics of the GIDL current. It can be found that the GIDL current of the program state decreased monotonically. The difference in voltage between the drain and the gate is +8V and is high enough to cause that the stored electrons tunnel into the drain. One solution is to raise the gate voltage to reduce the voltage difference between the gate and the drain. However, the higher gate voltage may induce the channel hot electron injections because the drain voltage (+11V) is high. Therefore, the drain disturbance of the GIDL current is a serious issue for this device reliability.

4-8 Summary

The 3-bit operations of this n-channel SAHOS memory device are demonstrated by combining the V_t modulation, the GIDL current modulation on forward read, and the GIDL current modulation on reverse read. The memory performances of the P/E speed, the retention, the endurance and the disturbance are evaluated separately in this chapter. Compared with sample A, we can find that sample B shows better memory performances. For example, sample B have faster P/E speed on the V_t modulation, better retention performance, and the movement of whole subthreshold curve which is more feasible in the production.

The memory performances of the V_t modulation on sample B are summarized as follows. The threshold voltage (V_t) can shift with large memory window of 7V, and the memory device shows good retention of nearly no charge loss after 10^5 seconds. Moreover, sample B presents high endurance of 22% degradation after 10^5 P/E cycles. In addition, the disturbance effects on the V_t modulation are negligible when this memory device is implemented by the NOR array architecture.

The memory performances of the GIDL current modulation on sample B are summarized as follows. The GIDL current can be modulated about two orders of magnitude, but it shows poor retention and poor endurance. The read disturbance and the gate disturbance are under the acceptable condition when this memory device is applied to the NOR array architecture. However, the drain disturbance is an issue for the device reliability.

Table 4-1: The 8 states of the I_d-V_g and I_s-V_g characteristics on sample B.

	Forward read (Vd=1V, Vs=0V) GIDL current increase	Reverse Read (Vd=0V, Vs=1V) GIDL current increase	V_t shift (Whole subthresold)
1st state	Yes	Yes	No
2nd state	No	Yes	No
3rd state	No	No	No
4th state	Yes	No	No
5th state	Yes	Yes	Yes
6th state	Yes	No	Yes
7th state	No	No	Yes
8th state	No	Yes	Yes

Table 4-2: The 8 states of the I_d-V_g and I_s-V_g characteristics on sample A.

	GIDL current increase Forward read (V_d=1V, V_s=0V)	GIDL current increase Reverse Read (V_d=0V, V_s=1V)	V_t shift (upper half subthresold)
1st state	Yes	Yes	No
2nd state	No	Yes	No
3rd state	No	No	No
4th state	Yes	No	No
5th state	Yes	Yes	Yes
6th state	No	Yes	Yes
7th state	No	No	Yes
8th state	Yes	No	Yes

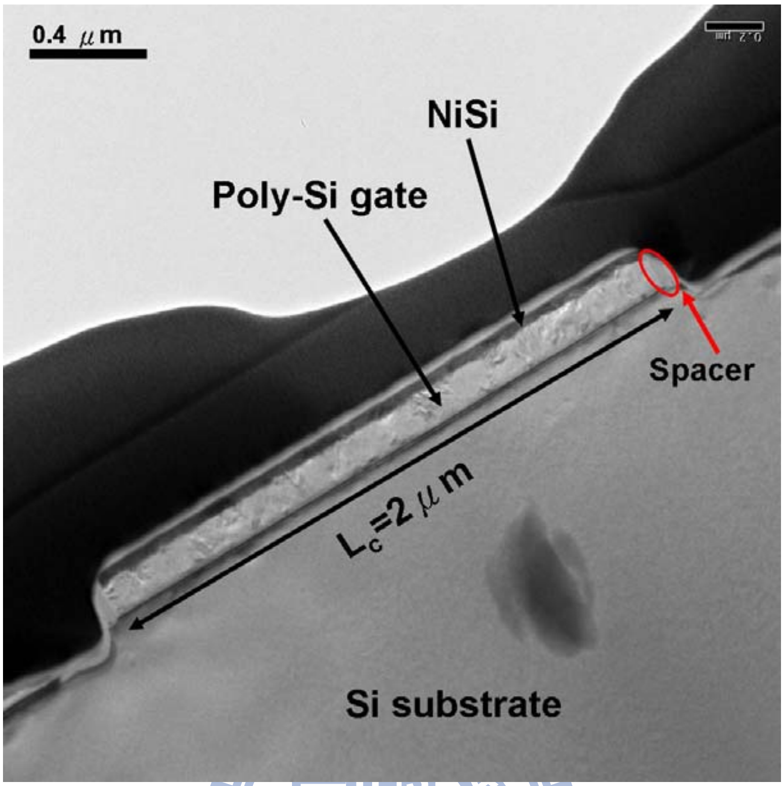


Fig. 4-1: The TEM image of the cross-section view on sample C.

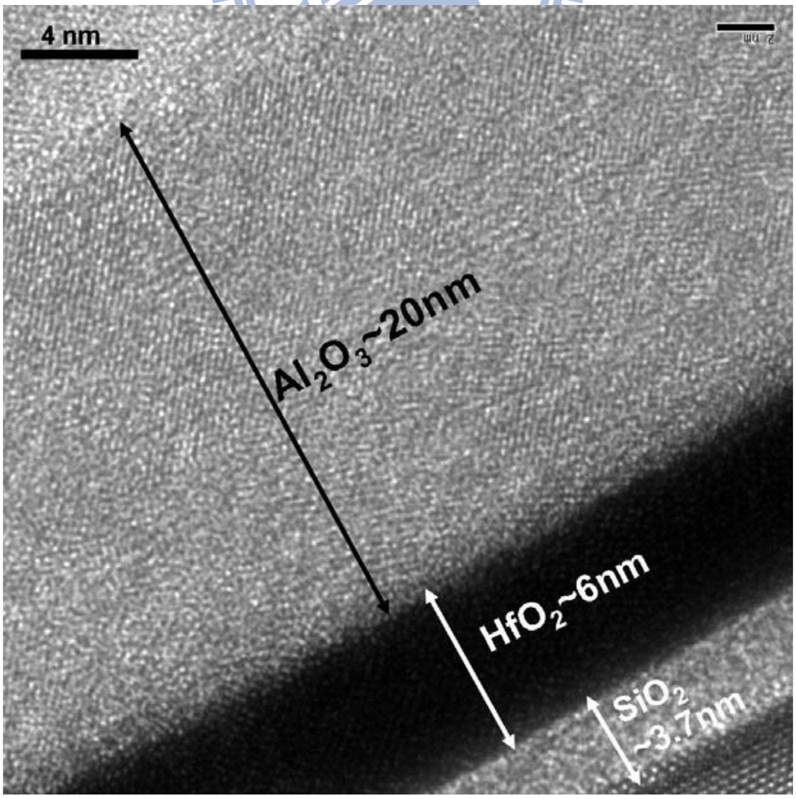


Fig. 4-2: The high resolution TEM image of the gate stack dielectric on sample C.

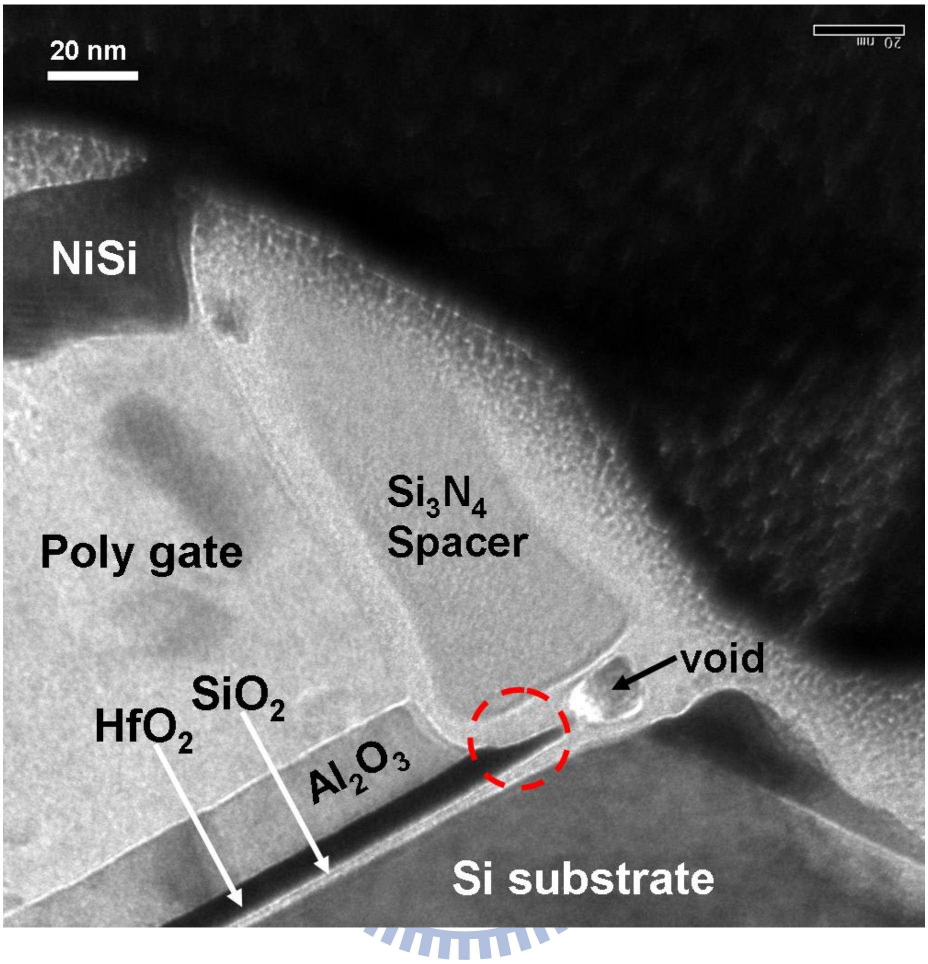
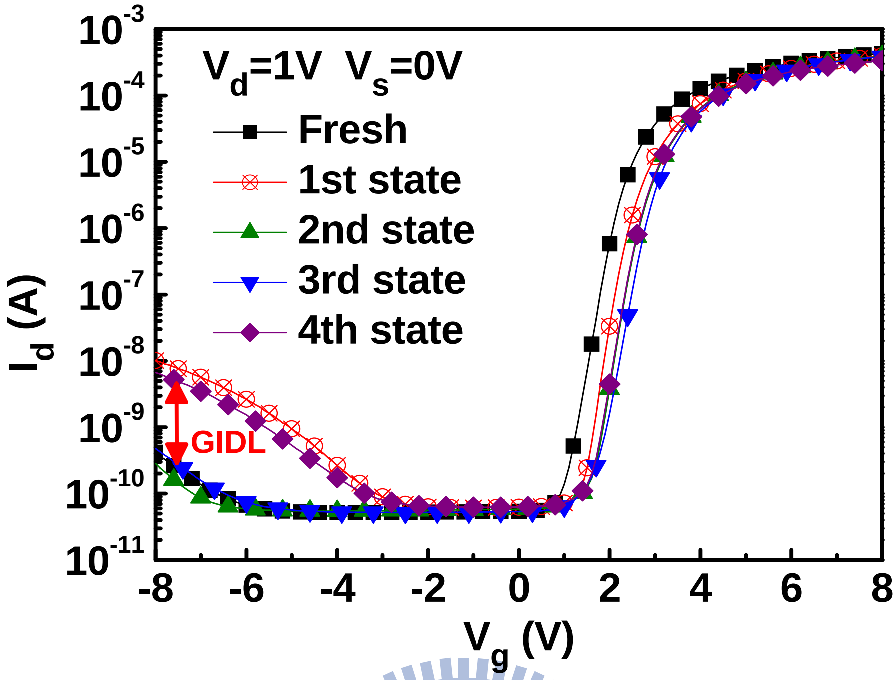
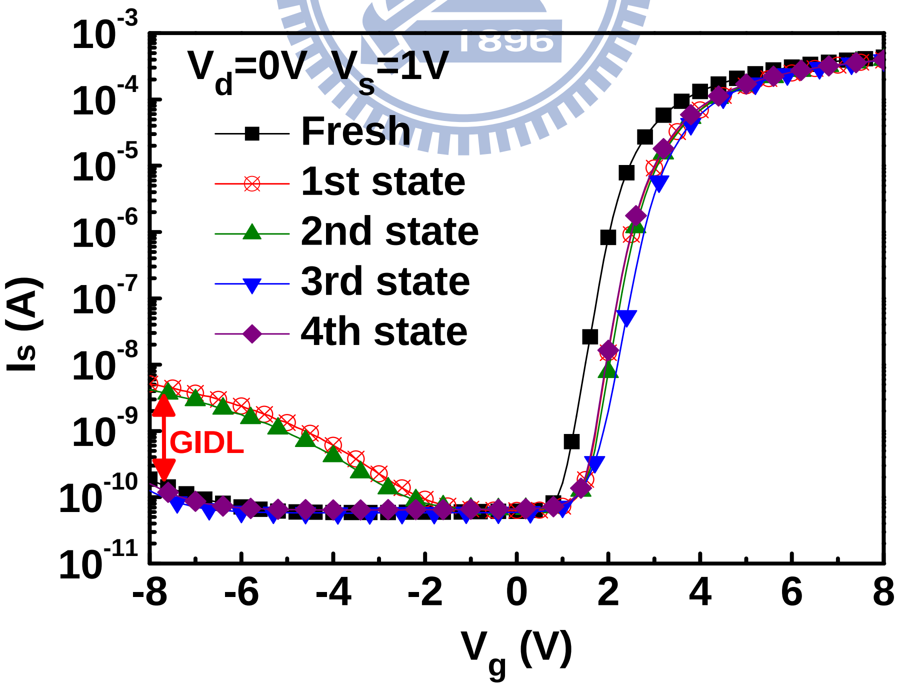


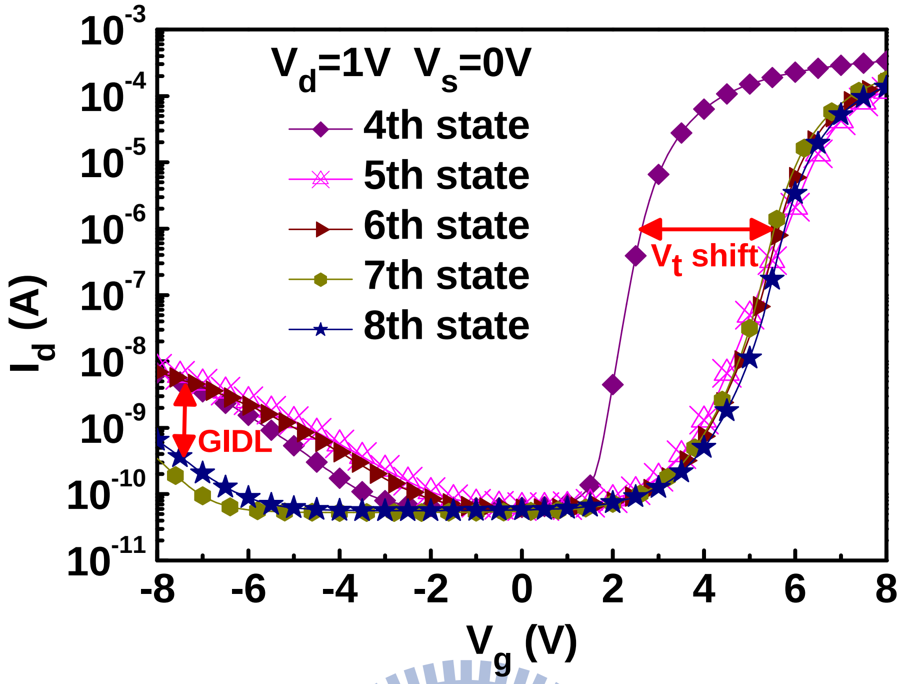
Fig. 4-3: The high resolution TEM image of the spacer region on sample C.


V_d=1V V_s=0V
Fresh
1st state
2nd state
3rd state
4th state
GIDL
I_d (A)
V_g (V)
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
10^{-9}
10^{-10}
10^{-11}
-8
-6
-4
-2
0
2
4
6
8

(a)

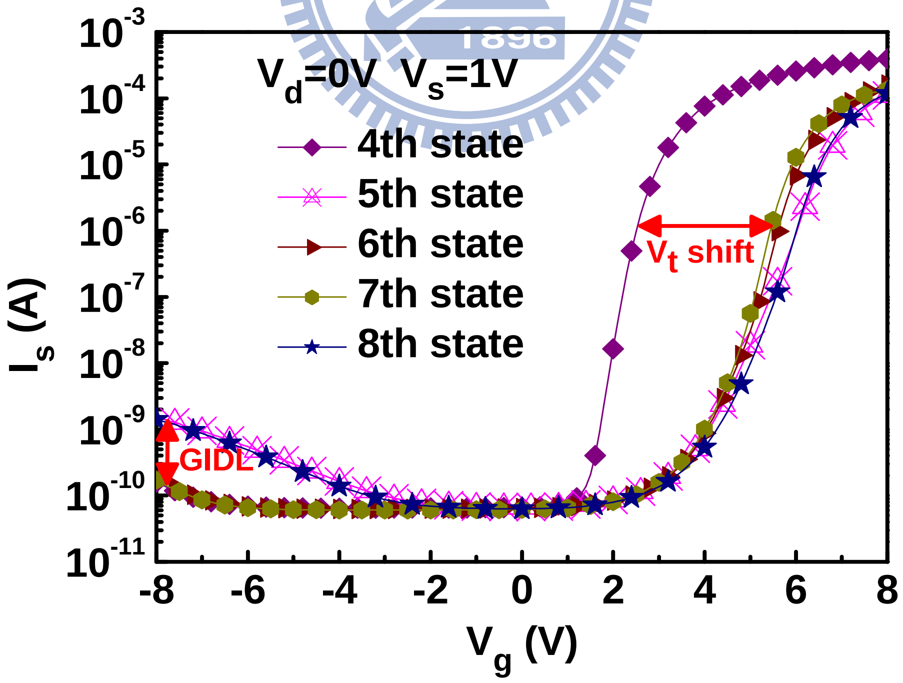

V_d=0V V_s=1V
Fresh
1st state
2nd state
3rd state
4th state
GIDL
Is (A)
V_g (V)
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
10^{-9}
10^{-10}
10^{-11}
-8
-6
-4
-2
0
2
4
6
8

(b)



V_d=1V V_s=0V
4th state
5th state
6th state
7th state
8th state
V_t shift
GIDL
I_d (A)
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
10^{-9}
10^{-10}
10^{-11}
-8
-6
-4
-2
0
2
4
6
8
V_g (V)

(c)



V_d=0V V_s=1V
4th state
5th state
6th state
7th state
8th state
V_t shift
GIDL
I_s (A)
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
10^{-9}
10^{-10}
10^{-11}
-8
-6
-4
-2
0
2
4
6
8
V_g (V)

(d)

Fig. 4-4: I_d-V_g and I_s-V_g curves on sample B during the operations.

1st state: (V_g = -16V, V_d = 0V, V_s = 0V, for 0.1sec)

2nd state: 1st state + (V_g = 0V, V_d = +11V, V_s = 0V, for 0.1sec)

3rd state: 2nd state + (V_g = 0V, V_d = 0V, V_s = +11V, for 0.1sec)

4th state: 3rd state + (V_g = -16V, V_d = 0V, V_s = 0V, for 0.1sec)

+ (V_g = 0V, V_d = 0V, V_s = +11V, for 0.1sec)

5th state: 4th state + (V_g = 0V, V_d = +11V, V_s = 0V, for 0.1sec)

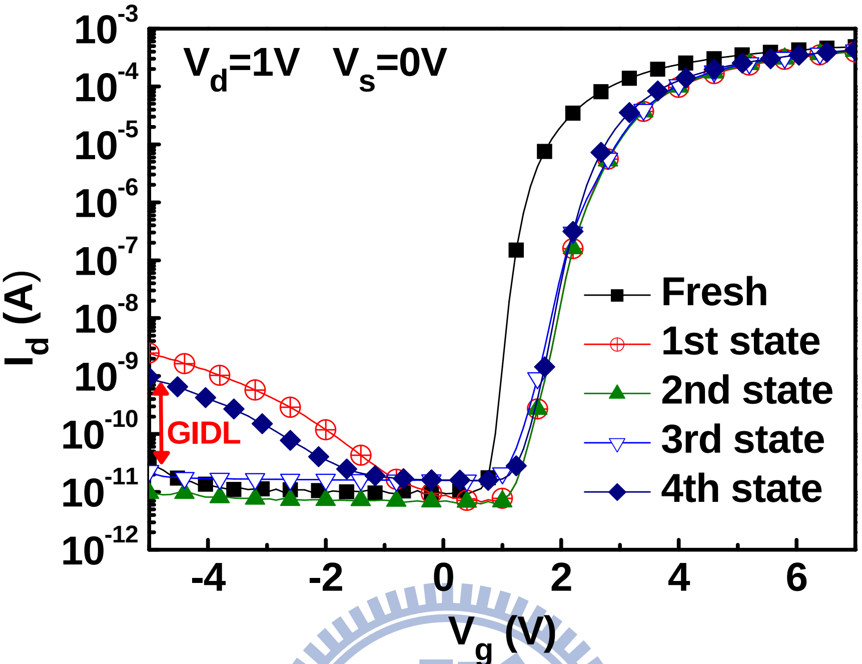
+ (V_g = +16V, V_d = 0V, V_s = 0V, for 0.1sec)

6th state: 5th state + (V_g = 0V, V_d = 0V, V_s = +11V, for 0.1sec)

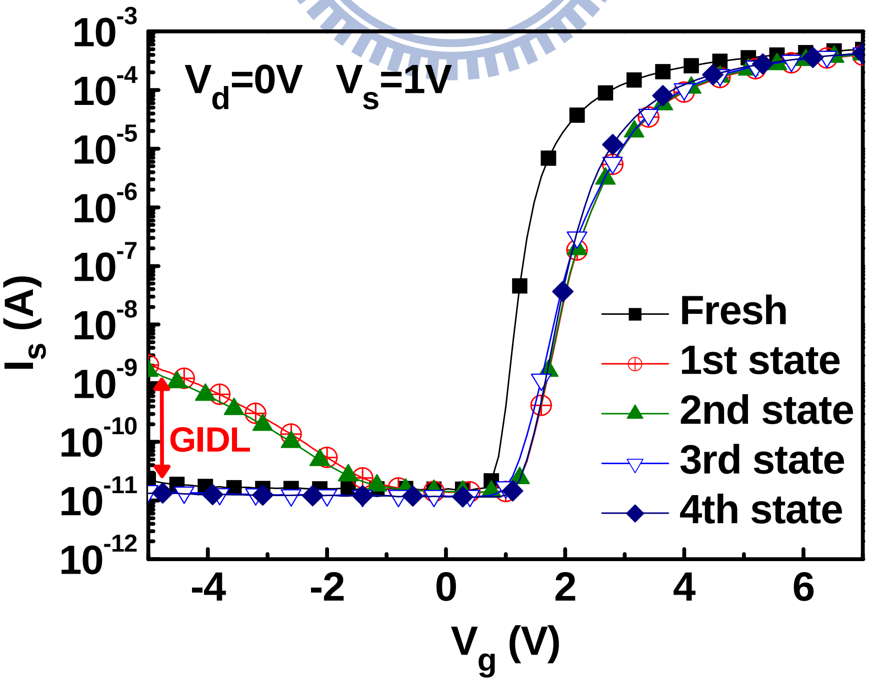
7th state: 6th state + (V_g = 0V, V_d = +11V, V_s = 0V, for 0.1sec)

8th state: 7th state + (V_g = +16V, V_d = 0V, V_s = 0V, for 0.1sec)

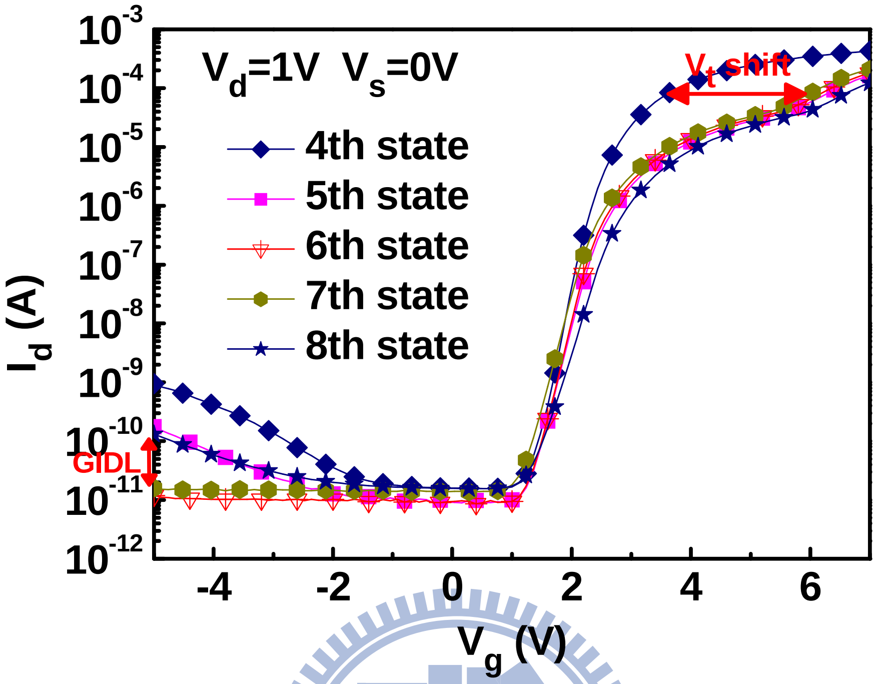
+ (V_g = 0V, V_d = +11V, V_s = 0V, for 0.1sec)


V_d=1V V_s=0V
I_d (A)
V_g (V)
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
10^{-9}
10^{-10}
10^{-11}
10^{-12}
-4
-2
0
2
4
6
GIDL
Fresh
1st state
2nd state
3rd state
4th state

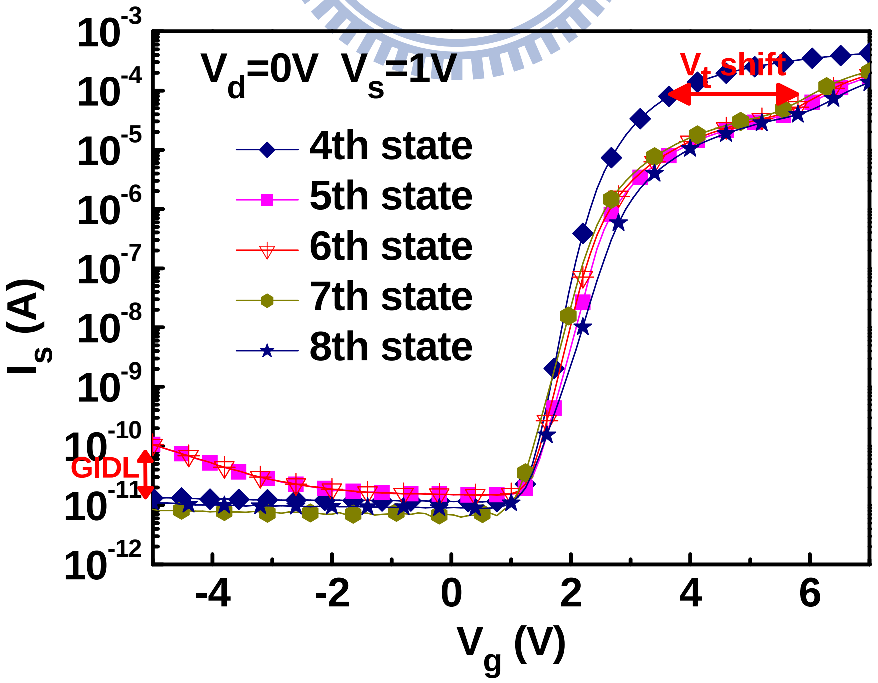
(a)


V_d=0V V_s=1V
I_s (A)
V_g (V)
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
10^{-9}
10^{-10}
10^{-11}
10^{-12}
-4
-2
0
2
4
6
GIDL
Fresh
1st state
2nd state
3rd state
4th state

(b)


V_d=1V V_s=0V
4th state
5th state
6th state
7th state
8th state
V_t shift
GIDL
I_d (A)
V_g (V)
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
10^{-9}
10^{-10}
10^{-11}
10^{-12}
-4
-2
0
2
4
6

(c)


V_d=0V V_s=1V
4th state
5th state
6th state
7th state
8th state
V_t shift
GIDL
I_s (A)
V_g (V)
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
10^{-9}
10^{-10}
10^{-11}
10^{-12}
-4
-2
0
2
4
6

(d)

Fig. 4-5: I_d-V_g and I_s-V_g curves on sample A during the operations.

1st state: (V_g = -16V, V_d = 0V, V_s = 0V, for 0.1sec)

2nd state: 1st state + (V_g = 0V, V_d = +10V, V_s = 0V, for 0.1sec)

3rd state: 2nd state + (V_g = 0V, V_d = 0V, V_s = +10V, for 0.1sec)

4th state: 3rd state + (V_g = -16V, V_d = 0V, V_s = 0V, for 0.1sec)
+ (V_g = 0V, V_d = 0V, V_s = +10V, for 0.1sec)

5th state: 4th state + (V_g = 0V, V_d = +10V, V_s = 0V, for 0.1sec)
+ (V_g = +14V, V_d = 0V, V_s = 0V, for 0.1sec)

6th state: 5th state + (V_g = 0V, V_d = +10V, V_s = 0V, for 0.1sec)

7th state: 6th state + (V_g = 0V, V_d = 0V, V_s = +10V, for 0.1sec)

8th state: 7th state + (V_g = +14V, V_d = 0V, V_s = 0V, for 0.1sec)
+ (V_g = 0V, V_d = 0V, V_s = +10V, for 0.1sec)

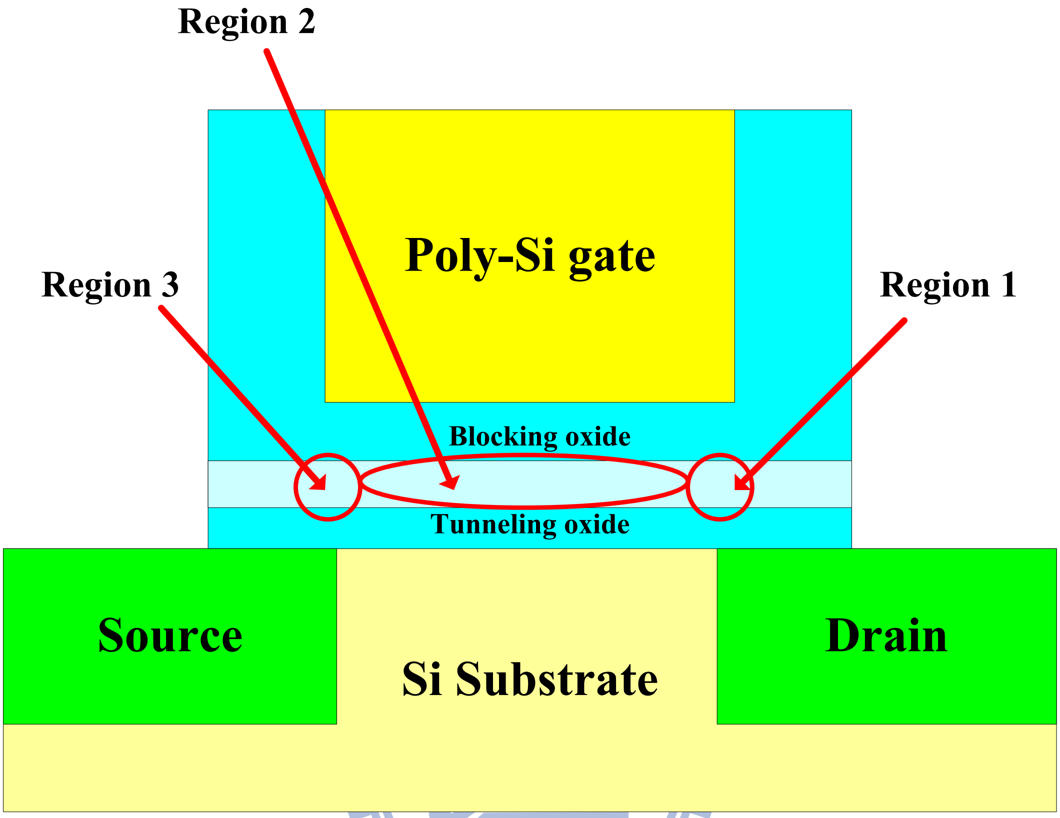


Fig. 4-6: There are three regions in the trapping layer on sample B, such as Region 1 “above drain junction”, Region 2 “above channel center”, and Region 3 “above source junction”.

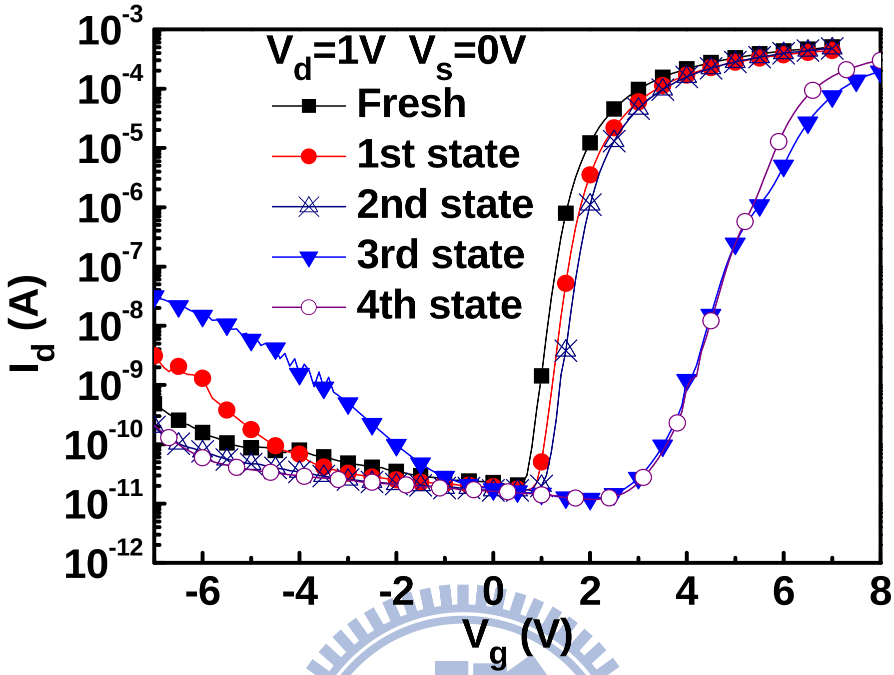


Fig. 4-7: I_d-V_g curves of sample C during the operations.

1st state: (V_g = -14V, V_d = 0V, V_s = 0V, for 0.1sec)

2nd state: 1st state + (V_g = 0V, V_d = +10V, V_s = 0V, for 0.1sec)

3rd state: 2nd state + (V_g = +14V, V_d = 0V, V_s = 0V, for 0.1sec)

4th state: 3rd state + (V_g = 0V, V_d = +10V, V_s = 0V, for 0.1sec)

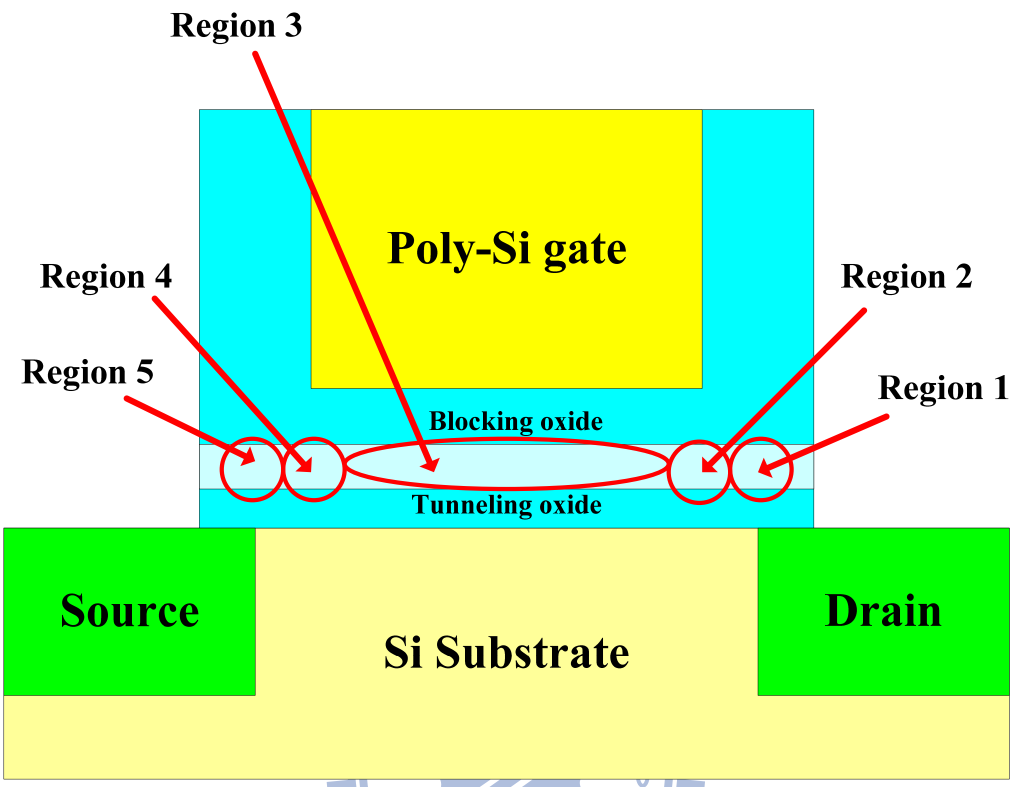


Fig. 4-8: There are five regions in the trapping layer on sample A, such as Region 1 "above drain junction", Region 2 "above channel and near drain junction", Region 3 "above channel center", Region 4 "above channel and near source junction", and Region 5 "above source junction".

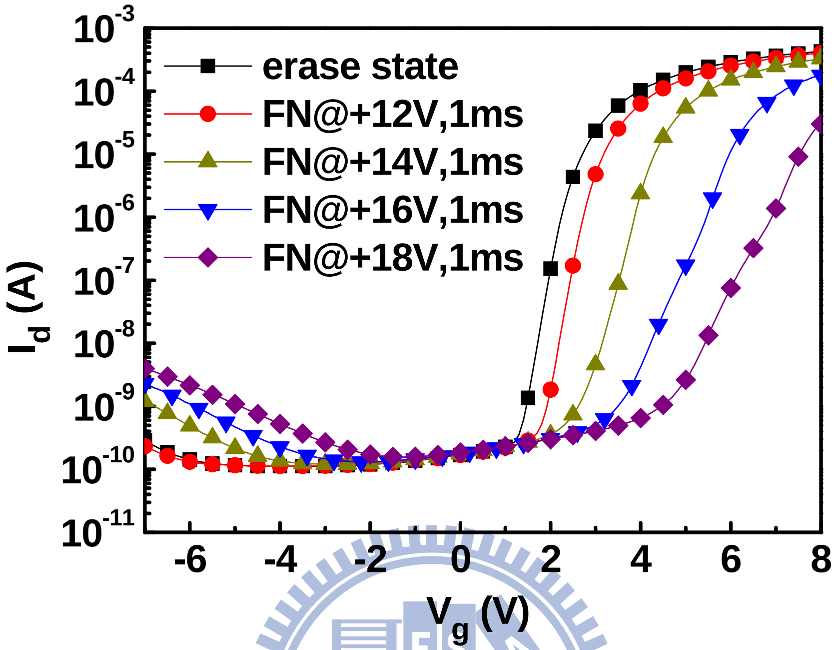


Fig. 4-9: The I_d-V_g characteristics on sample B after applying different voltages with same pulse width 1ms.

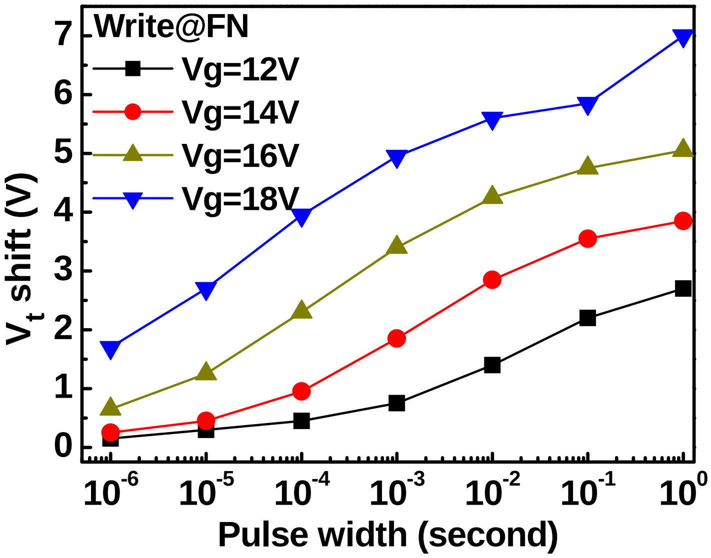


Fig. 4-10: The V_t shift after the program operations at different gate voltages with different pulse widths on sample B.

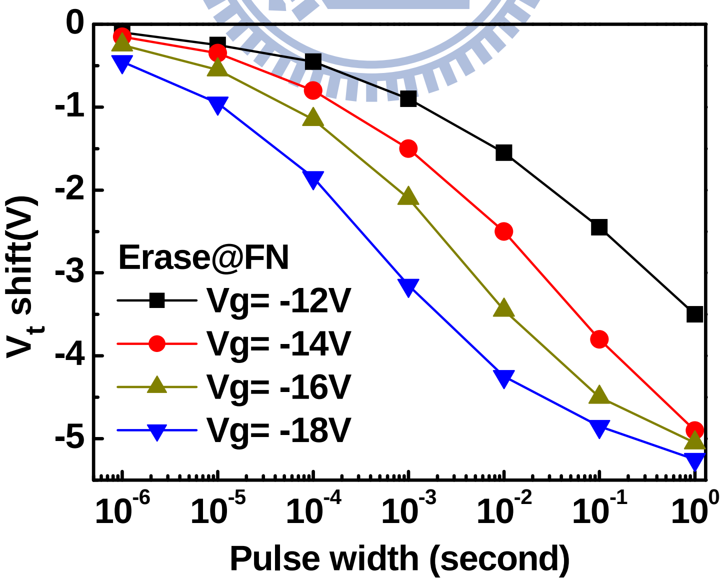


Fig. 4-11: The V_t shift after the erase operations at different gate voltages with different pulse widths on sample B.

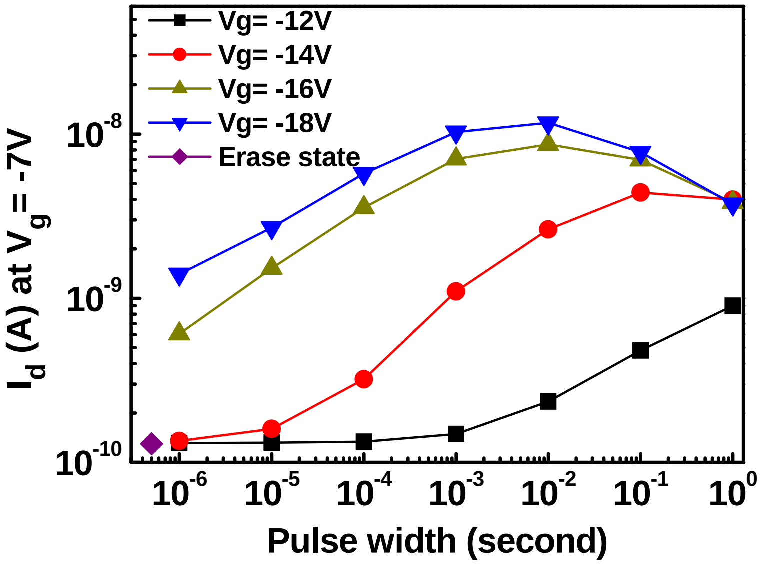


Fig. 4-12: The increase speed of the GIDL current after the program operations at different gate voltages with different pulse widths on sample B.

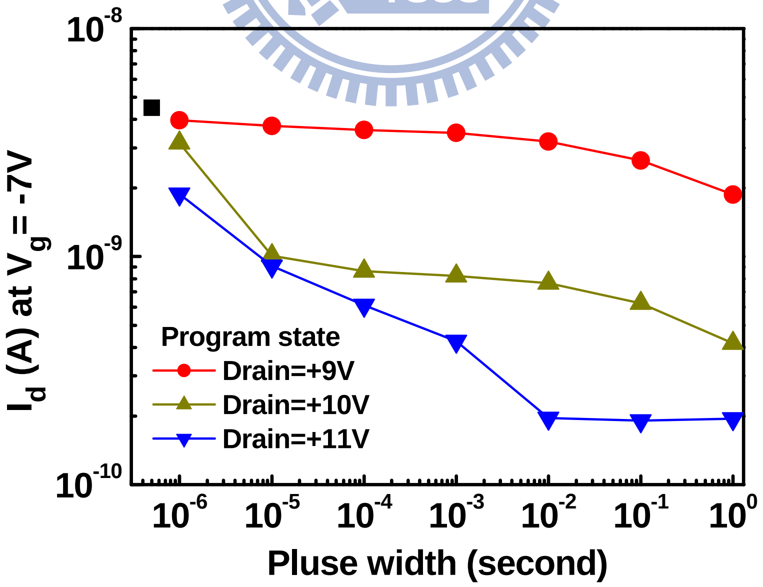


Fig. 4-13: The decrease speed of the GIDL current after the erase operations at different gate voltages with different pulse widths on sample B.

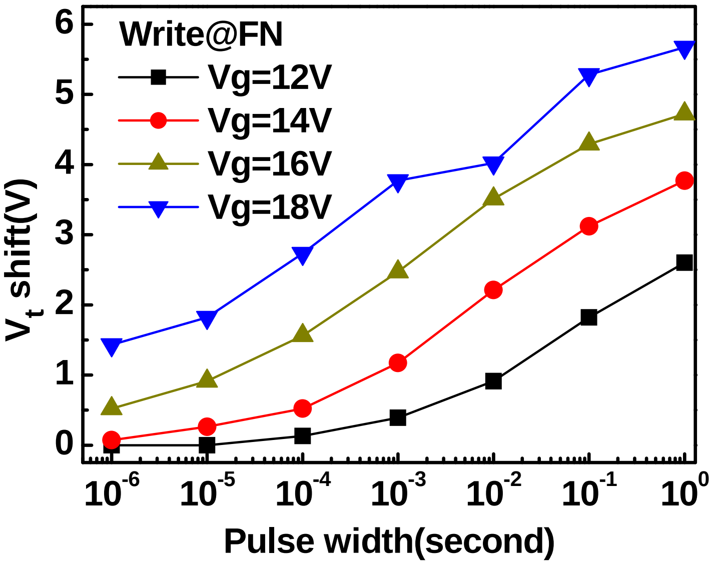


Fig. 4-14: The V_t shift after the program operations at different gate voltages with different pulse widths on sample A.

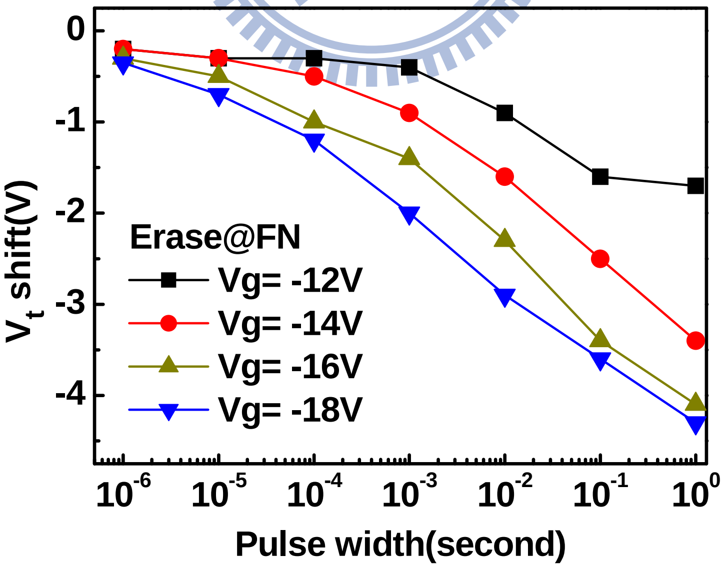


Fig. 4-15: The V_t shift after the erase operations at different gate voltages with different pulse widths on sample A.

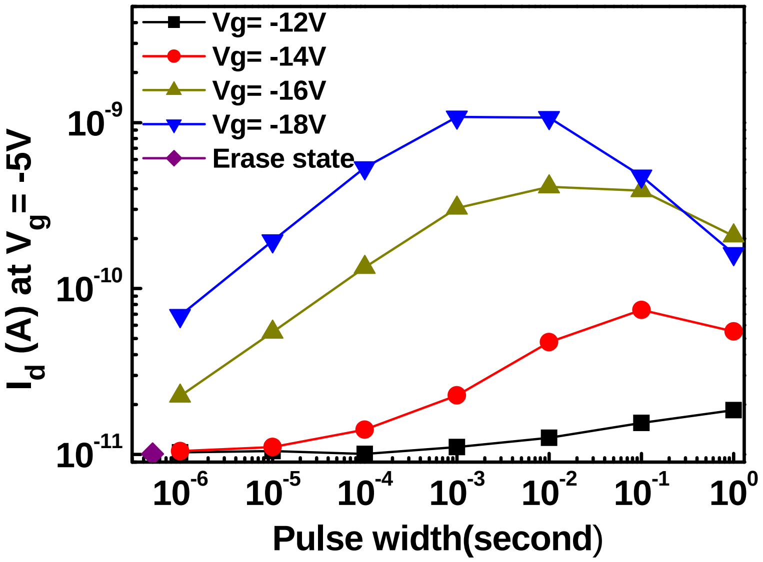


Fig. 4-16: The increase speed of the GIDL current after the program operations at different gate voltages with different pulse widths on sample A.

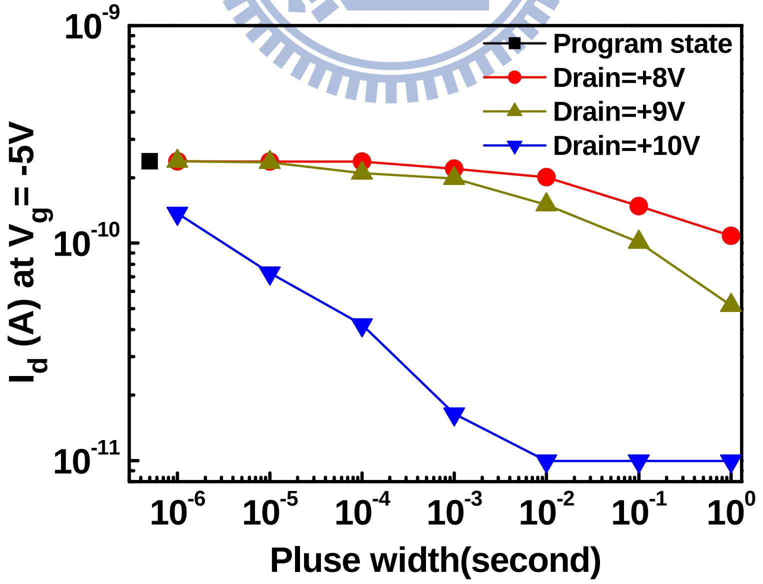


Fig. 4-17: The decrease speed of GIDL current after the erase operations at different gate voltages with different pulse widths on sample A.

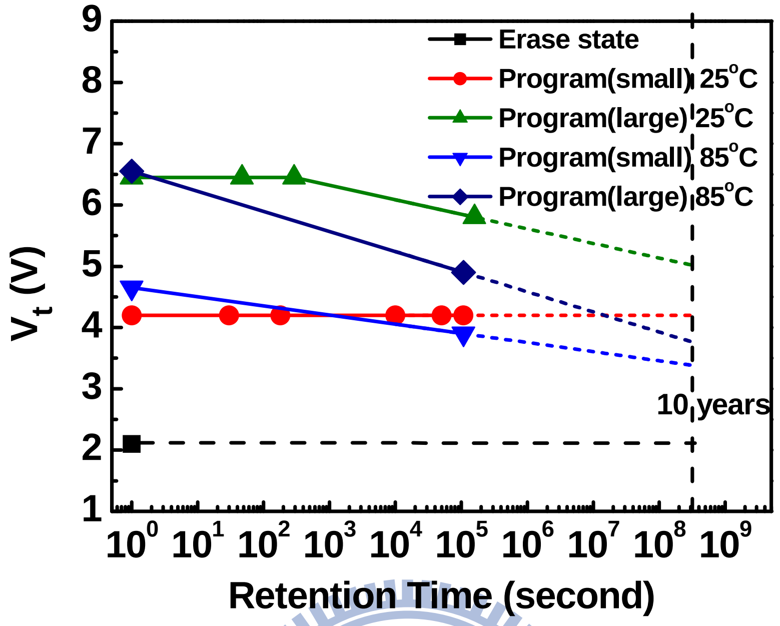


Fig. 4-18: The retention characteristics of the V_t on sample B at 25℃ and 85℃.

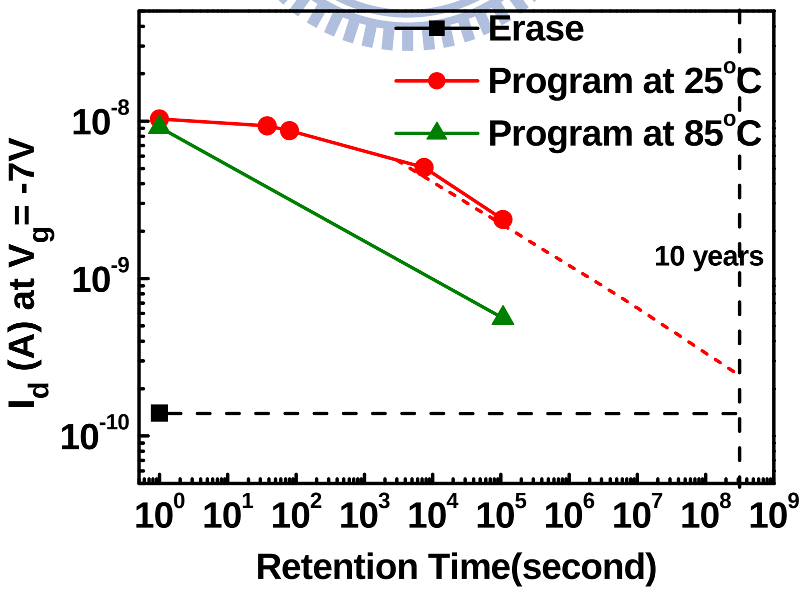


Fig. 4-19: The retention characteristics of the GIDL current on sample B at 25℃ and 85℃.

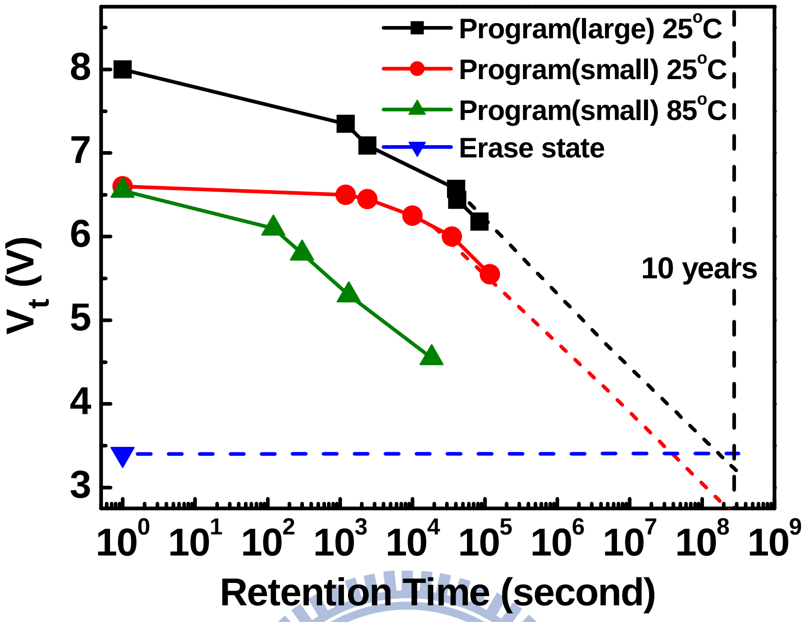


Fig. 4-20: The retention characteristics of the V_t on sample A at 25℃ and 85℃.

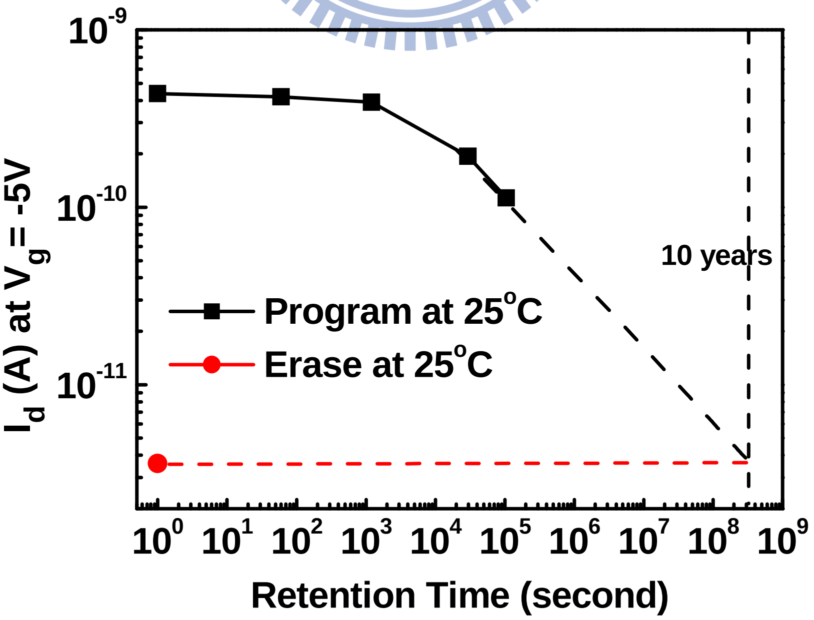


Fig. 4-21: The retention characteristics of the GIDL current on sample A at 25℃.

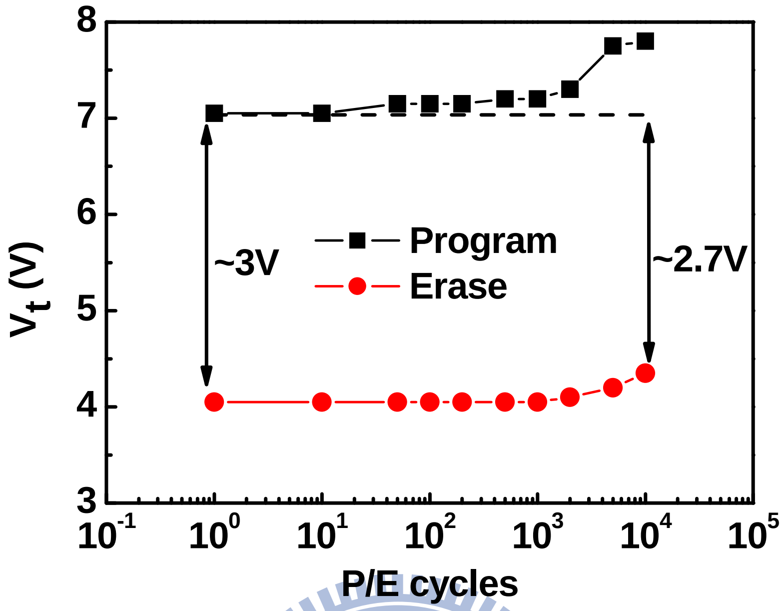


Fig. 4-22: The endurance performance of the V_t on sample A.

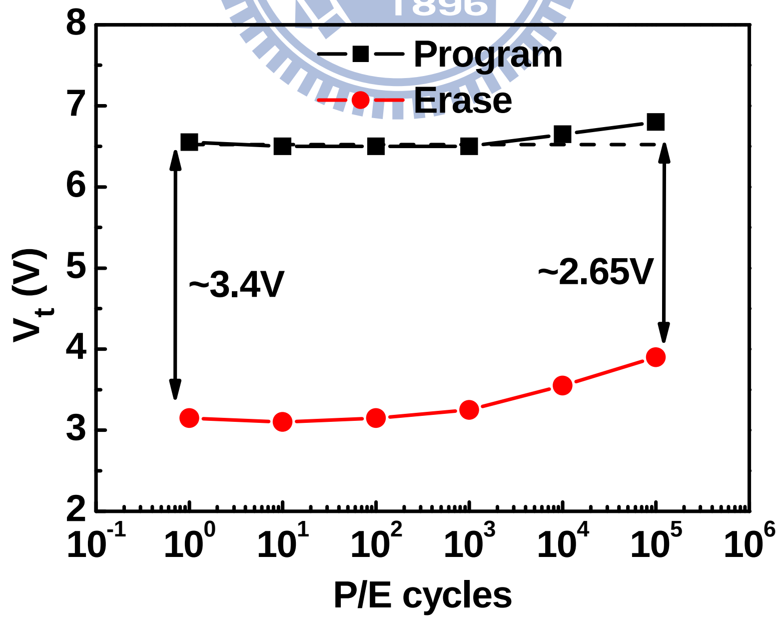


Fig. 4-23: The endurance performance of the V_t on sample B.

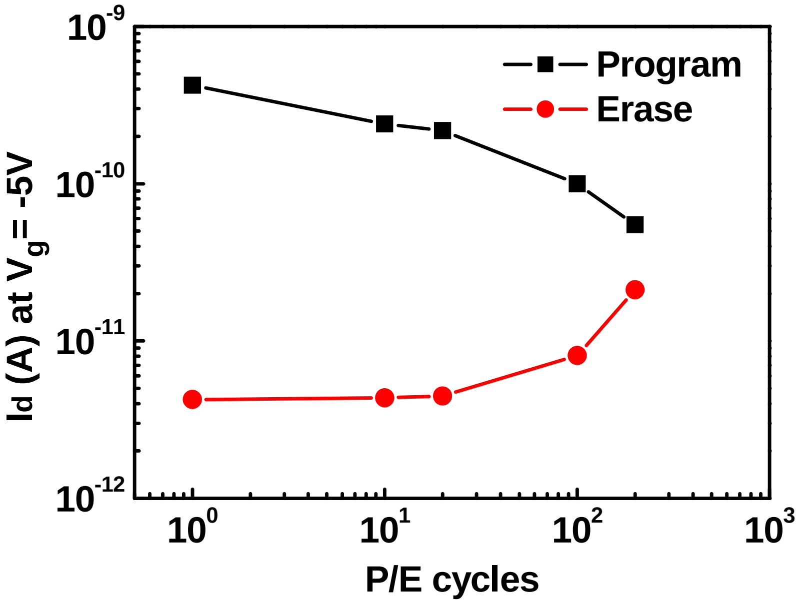


Fig. 4-24: The endurance performance of the GIDL current on sample A.

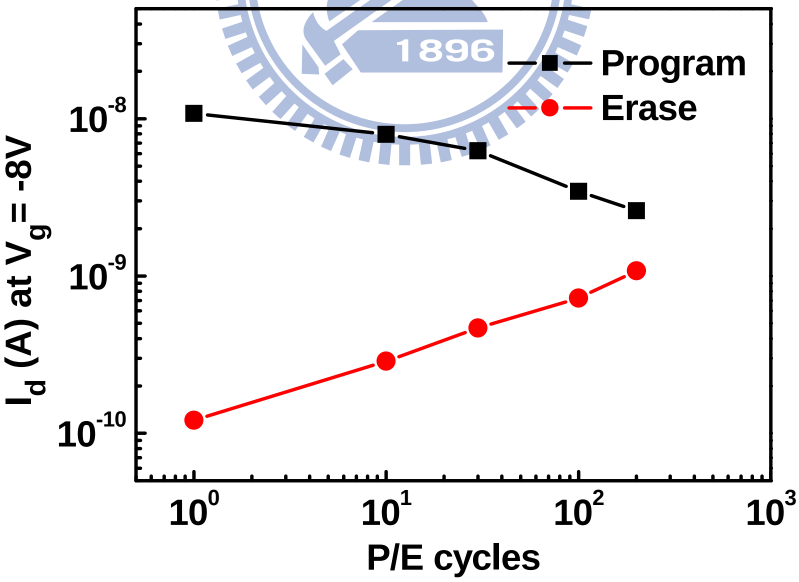


Fig. 4-25: The endurance performance of the GIDL current on sample B.

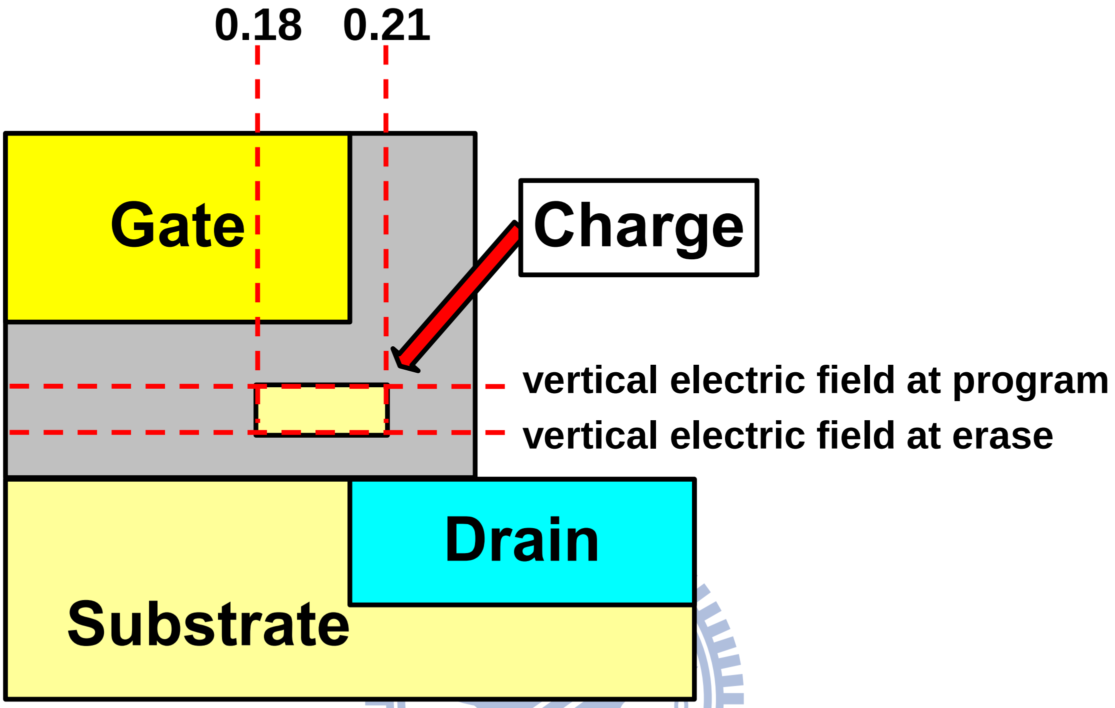


Fig. 4-26: The simulated structure: the lateral charge position is between 0.18 and 0.21μm; the vertical electric field at program operation is drawn along upper line, and the vertical electric field at erase operation is drawn along lower line.

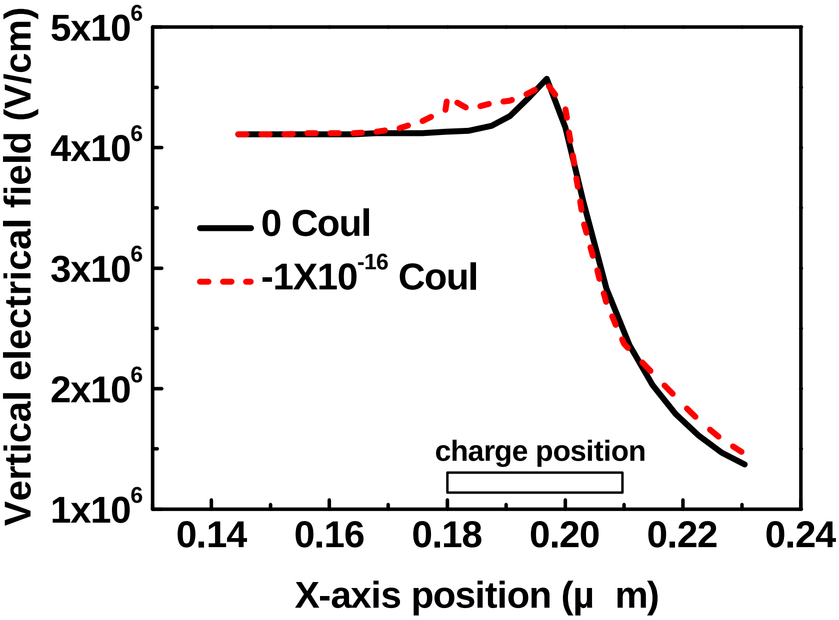


Fig. 4-27: The vertical electrical field along X-axis at program operation (gate voltage = -5V, source and drain voltage = 0V).

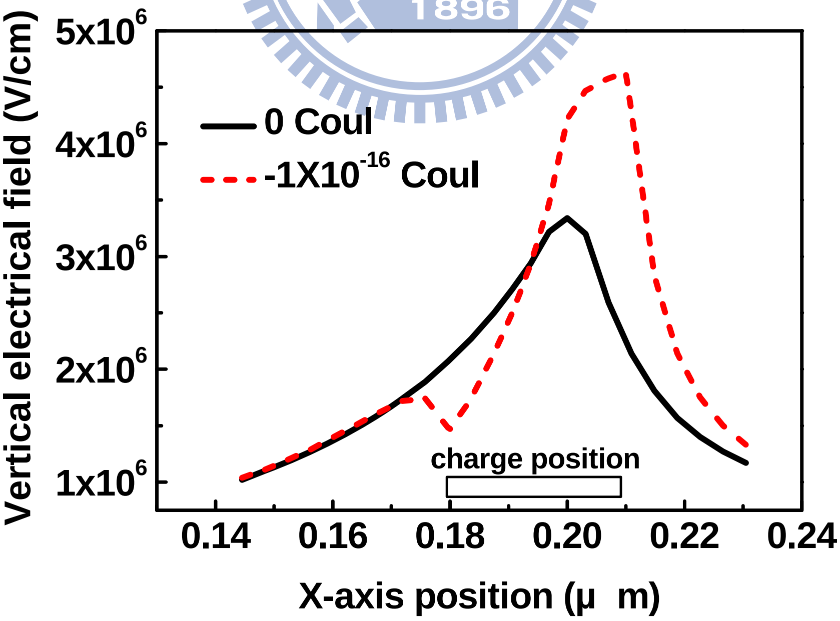
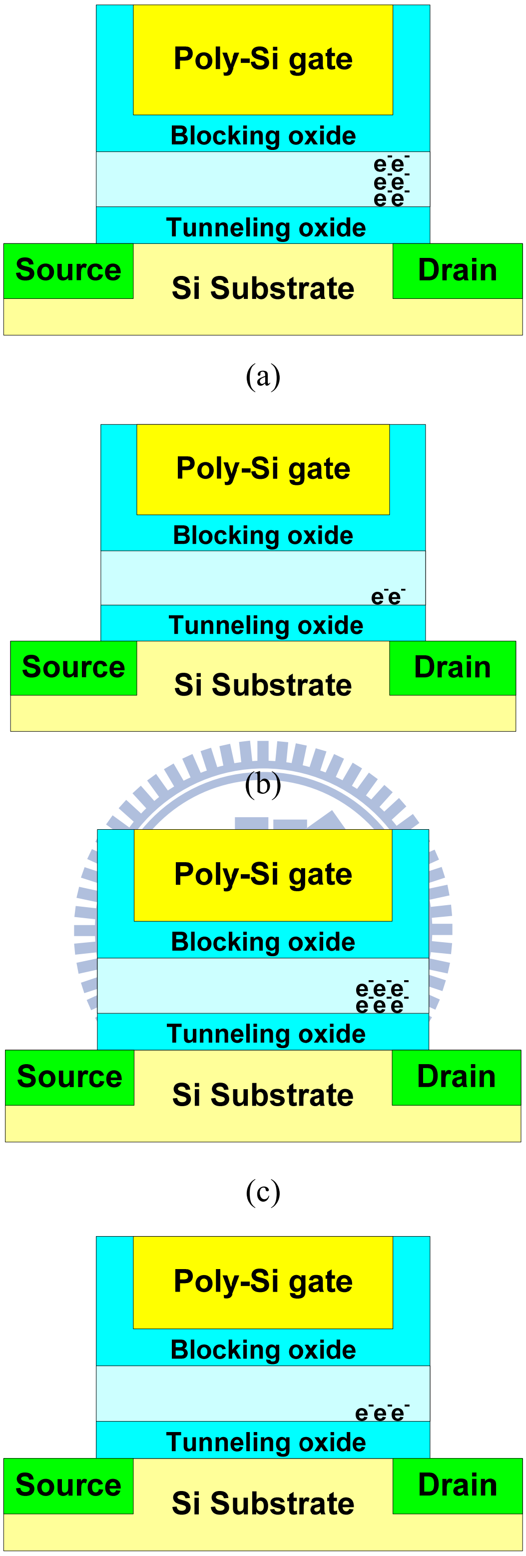


Fig. 4-28: The vertical electrical field along X-axis at erase operation (drain voltage = +4V, gate and source voltage = 0V).



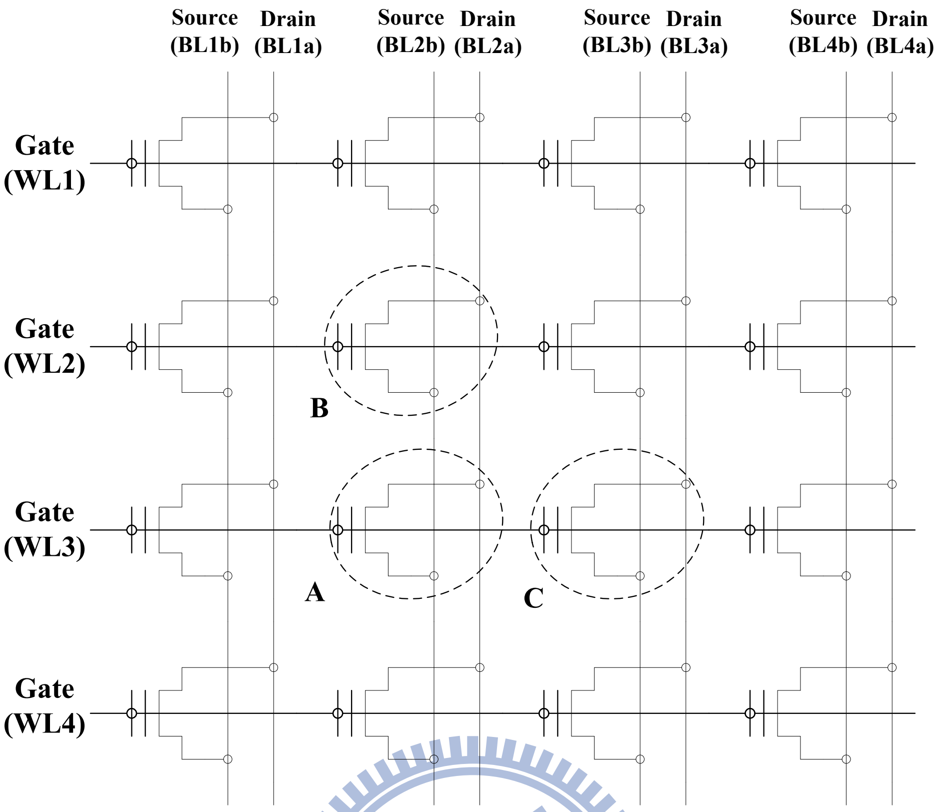
(a)

(b)

(c)

(d)

Fig. 4-29: The illustration of the stored-charges distribution at the drain side after every P/E cycle (a) the 1st program state (b) the 1st erase state (c) the 2nd program state (d) the 2nd erase state.



(a)

Operate device A (WL3, BL2a, BL2b) **Other WL and BL** (WL2, BL3a, BL3b)

	WL3	BL2a	BL2b	WL2	BL3a	BL3b
Read (V_t)	4V	1V	0V	0V	0V	0V
Read (Forward GIDL)	-8V	1V	0V	0V	0V	0V
Program (both V_t and GIDL)	14V	0V	0V	7V	7V	7V
Erase (V_t)	-14V	0V	0V	-7V	-7V	-7V
Erase (Forward GIDL)	0V	11V	0V	3V	0V	0V

Disturb for device B (BL2a, BL2b, WL2) **Disturb for device C** (BL3a, BL3b, WL3)

(b)

Fig. 4-30: Disturbance effects of the SAHOS memory for 3-bits operations (a) NOR array structure (b) voltage of bit line and word line during the operations.

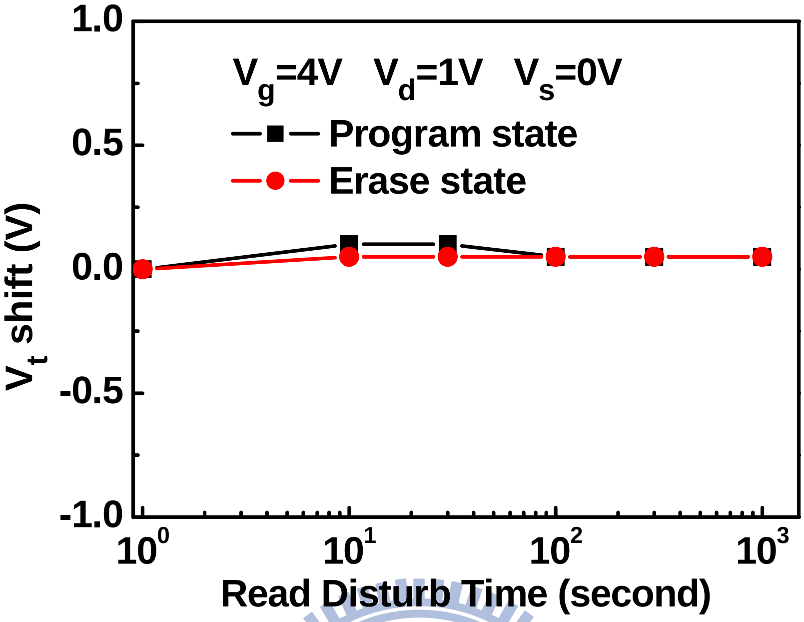


Fig. 4-31: Read disturbance characteristics of the V_t shift at V_g = 4V.

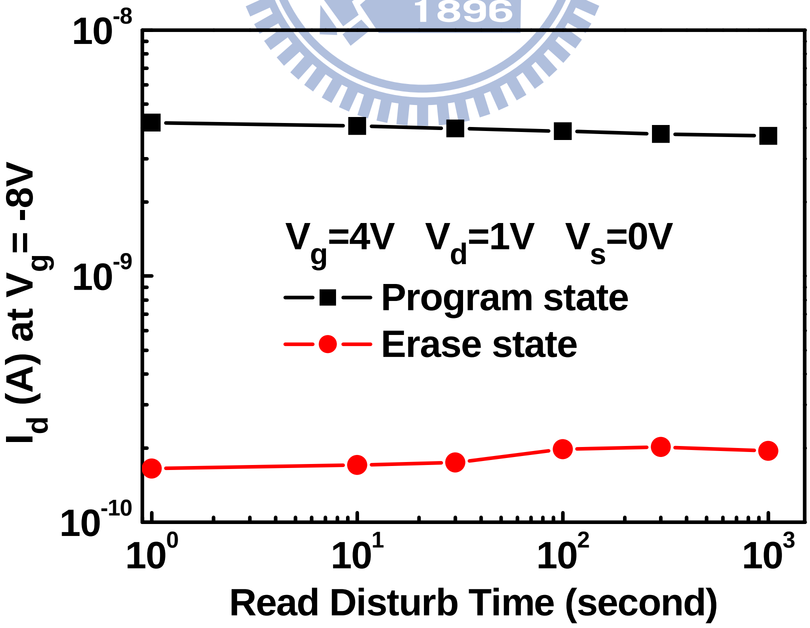


Fig. 4-32: Read disturbance characteristics of the GIDL current at V_g = 4V.

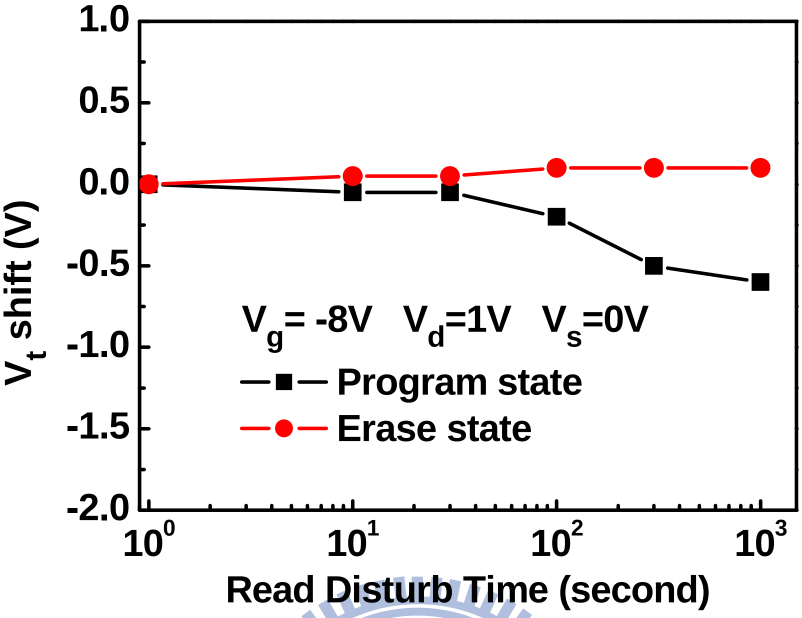


Fig. 4-33: Read disturbance characteristics of the V_t shift at V_g = -8V.

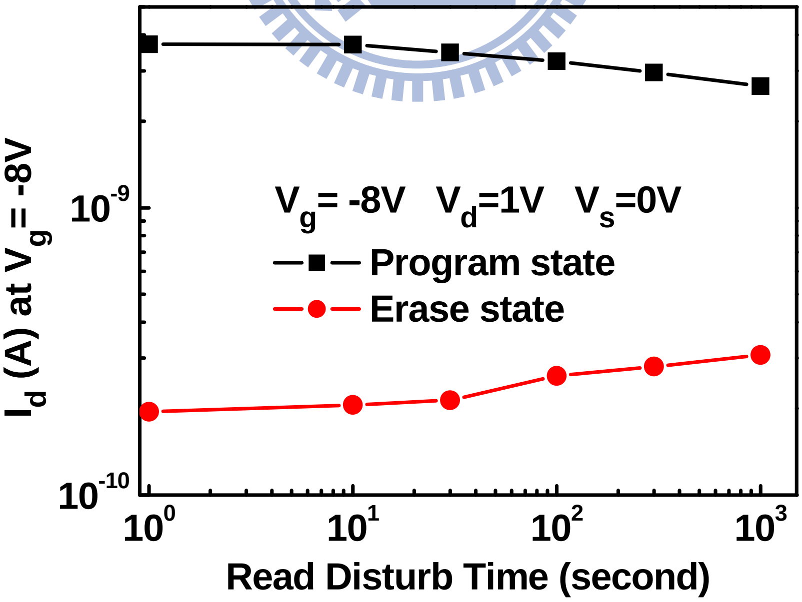


Fig. 4-34: Read disturbance characteristics of the GIDL current at V_g = -8V.

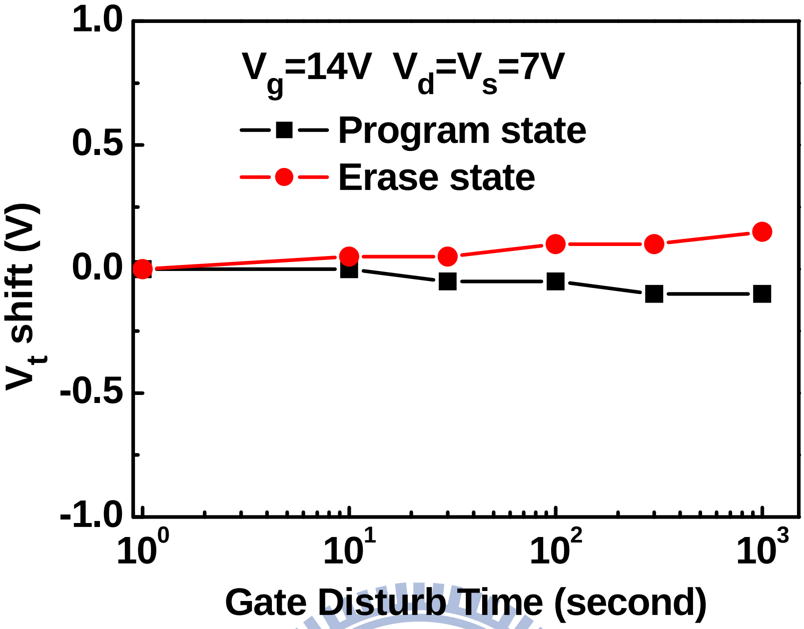


Fig. 4-35: Gate disturbance characteristics of the V_t shift under the program operation.

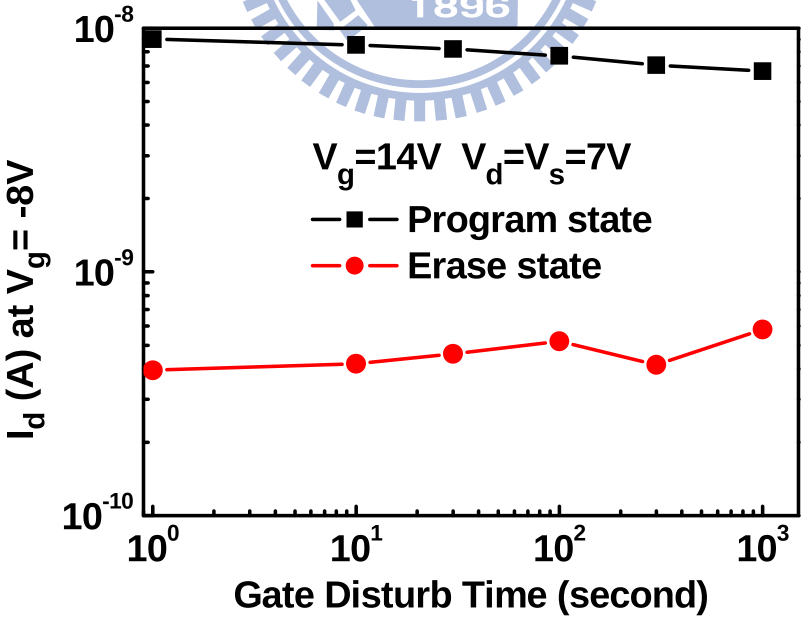


Fig. 4-36: Gate disturbance characteristics of the GIDL current under the program operation.

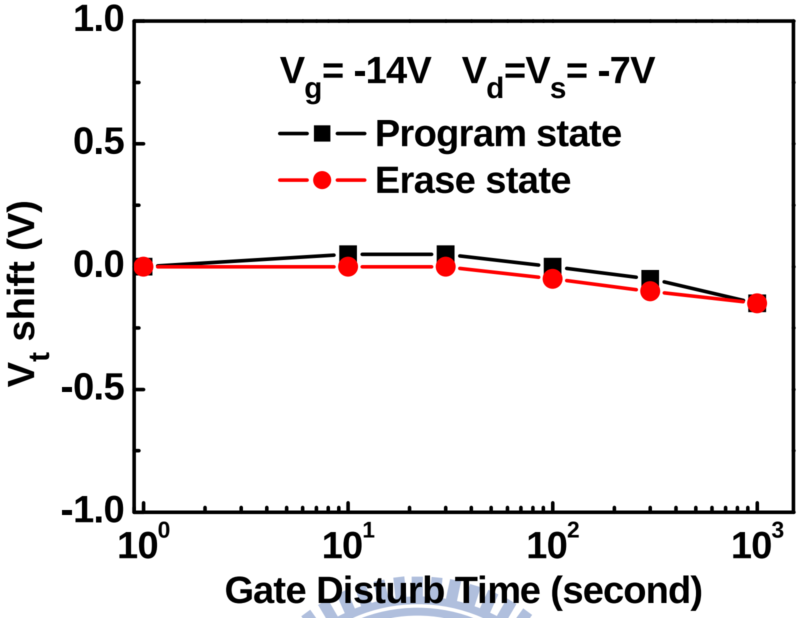


Fig. 4-37: Gate disturbance characteristics of the V_t shift under the erase operation.

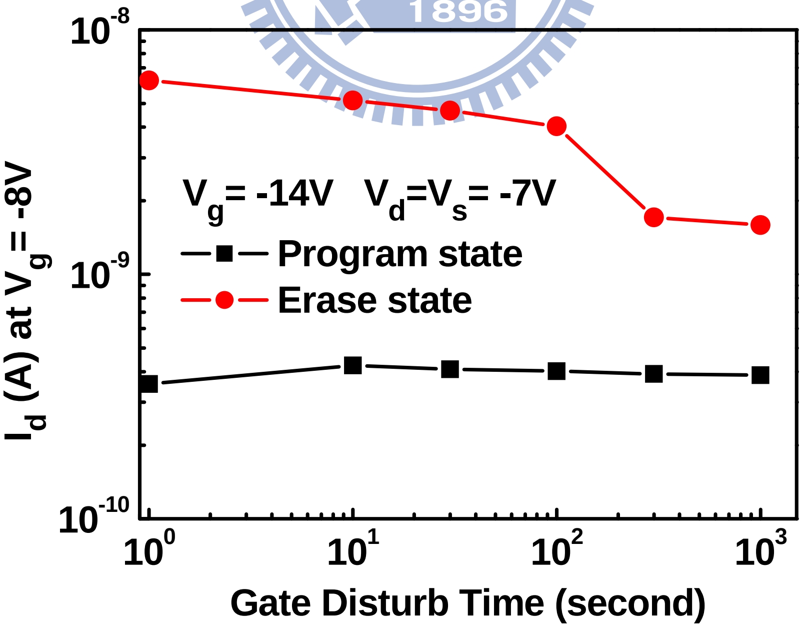


Fig. 4-38: Gate disturbance characteristics of the GIDL current under the erase operation.

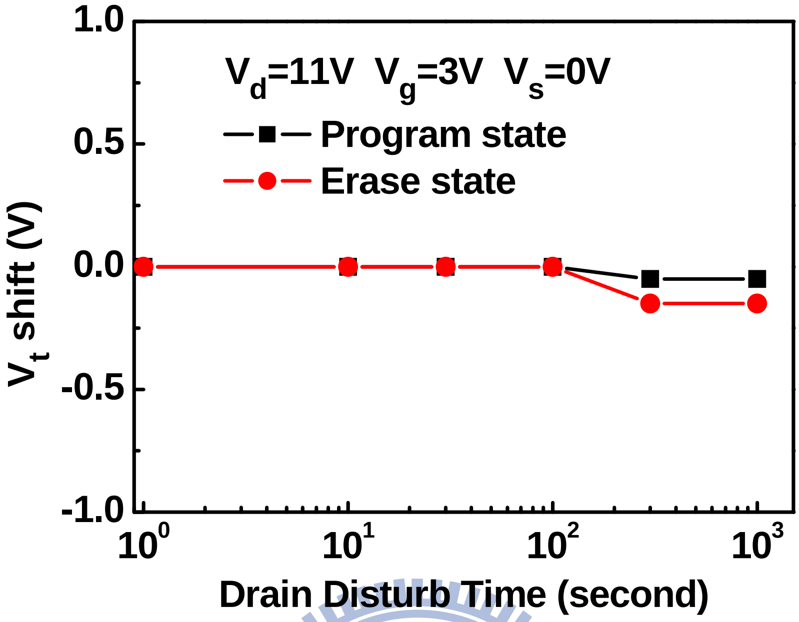


Fig. 4-39: Drain disturbance characteristics of the V_t shift.

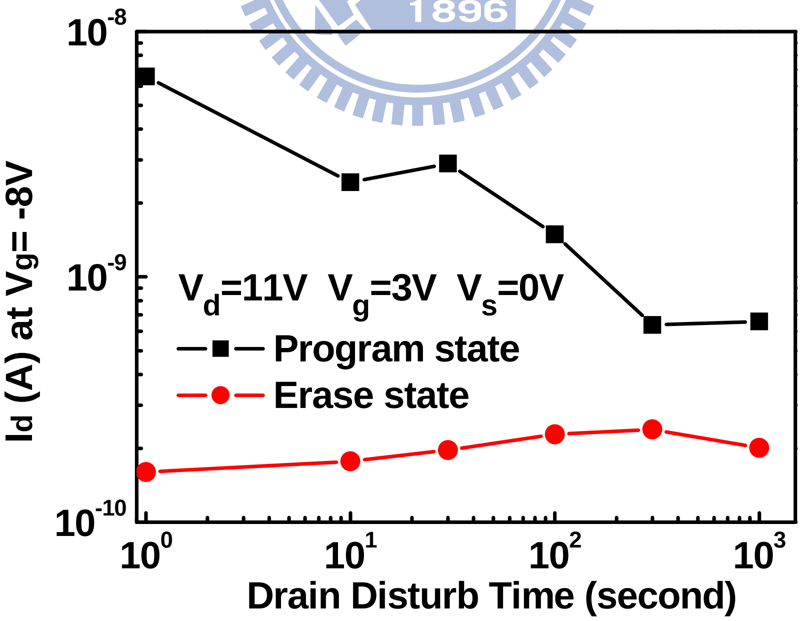


Fig. 4-40: Drain disturbance characteristics of the GIDL current.

Chapter 5
Conclusions and Future Works

5-1 Conclusions

In this study, it can be found that the modulation of the V_t and the modulation of the GIDL are related to the stored-charge positions according to the simulated results. Then, the 3-bit per cell memory characteristics are demonstrated on the fabricated SAHOS memory device.

The simulated results show that both the long channel device and the short channel device have the similar effects of stored-charge distribution on device characteristics. When the charges are stored at the position which is above the channel center, the whole subthresthold curve is moved to the right because of the negative charges stored at the center of the trapping layer resulting in a raise of the electron barrier of the channel region. When the charges are stored at the position which is above the channel and near the drain junction, only the upper half subthresthold curve is moved to the right. Because the electron barrier of the channel region is slightly dragged down by the drain voltage, the lower half subthreshold curve keeps. However, the electron barrier is not low enough to allow complete electron conduction, so the upper half subthresthold curve is moved to the right. Next, when the charges are stored just above the drain junction, the GIDL current increase largely. Because the vertical electric field between the gate and the drain is enhanced and then results in larger GIDL current. In summary, the simulated results show that placing charges at the above-mentioned charge storage positions can result in the movement of the whole subthreshold curve, the movement of the upper half subthreshold curve, or the increase of the GIDL current.

In order to realize 3-bit per cell memory by utilizing the modulation of the V_t and the modulation of the GIDL, the n-channel SAHOS memory devices with the charge trapping layer HfO_2 extending to the underneath of the spacer are fabricated. The crystallized Al_2O_3 blocking layer which is found in TEM images should be responsible for the electron back injection from the gate. The 3-bit operations of these n-channel SAHOS memory devices are demonstrated by combining the V_t modulation, the GIDL current modulation on forward read, and the GIDL current modulation on reverse read.

Moreover, the devices with S/D annealing time of 60 seconds show the modulation of the whole subthreshold curve, but the devices with S/D annealing time of 20 seconds show the modulation of the upper half subthreshold curve only. The inference is that the S/D junctions do not overlap with gate on the devices of 20 seconds annealing time so there is an electron barrier between the S/D and the channel while programming. On the contrary, the devices with S/D annealing time of 60 seconds are the S/D-to-gate overlap structure.

The devices with the S/D-to-gate overlap structure show better memory performances than those with S/D-to-gate non-overlap structure. The stored charges on non-overlap structure are under the spacer, and there is no Al_2O_3 blocking layer and exists a lot of defects. Therefore, the charge lose rate is fast.

For the devices with the S/D-to-gate overlap structure, the threshold voltage (V_t) can shift with large memory window of 7V, and shows good retention of nearly no charge loss after 10^5 seconds. Moreover, high endurance of 22% degradation after 10^5 P/E cycles is presented on the V_t modulation. In addition, the disturbance effects of the V_t are negligible when this memory device is implemented by the NOR array architecture.

For the devices with the S/D-to-gate overlap structure, the GIDL current can be modulated by about two orders of magnitude, but it shows poor retention and poor endurance. The stored charges are near the corner of the spacer and then result in poor retention. At the corner of the spacer, there are a lot of defects generated from the device fabrication process. These defects will enhance the charge loss rate. The other reason of poor retention is due to “lateral charge migration”. When the charges are stored in the HfO_2 trapping layer at the source side or the drain side, the charges will move toward the trapping site above the channel center. The poor endurance is possible as a result of the change of the stored charges distribution after every P/E cycle. The read disturbance and the gate disturbance are under the acceptable condition when this memory device is applied to the NOR array architecture. However, the drain disturbance is an issue for the device reliability.

In conclusion, the V_t modulation shows good memory performance, but the device reliability of the GIDL current modulation requires improvement.

5-2 Future Works

Although the feasibility of 3-bit per cell memory is proved, the device dimension with the gate length of 2μm is too large in this study. Nowadays, the memory device is already scaled down to 45nm in the production. Therefore, the device dimension of this 3-bit per cell memory requires scaling down.

To improve the retention performance of the GIDL current modulation, there are three possible solutions. First one is extending the Al_2O_3 blocking layer to the underneath of the spacer. It can prevent the vertical escape of the stored charges. Second one is replacing the HfO_2 trapping layer by HfAlO because HfAlO shows better retention performance than HfO_2 [42]. The last one is changing the trapping layer into nano-crystal structure.

The poor endurance performance also requires improvement. If the degradation of the endurance performance is due to the change of the stored charges distribution after every P/E cycle, one possible solution is to increase the S/D junction breakdown voltage. Therefore, the higher S/D voltage can be adopted during the erase operation, and then the stored charges can be totally removed. The other method is by varying the thickness of the gate dielectric and modulating the position of S/D junctions to optimize the endurance performance.

Finally, in order to reduce the drain disturbance effect, the gate voltage needs to increase. Therefore, how to reduce channel hot electron injection is an issue needed to be solved.

References

[1] Y. S. Shin, “Non-volatile memory technologies for beyond 2010,” in *Proc. Symp. VLSI Circuits Tech. Dig.*, 2005, pp. 156-159.

[2] J. M. Slaughter, R. W. Dave, M. Durlam, G. Kerszykowski, K. Smith, K. Nagel, B. Feil, J. Calder, M. DeHerrera, B. Garni, and S. Tehrani, “High speed toggle MRAM with MgO-based tunnel junctions,” in *IEDM Tech. Dig.*, 2005, pp. 873-876.

[3] C. C. Hung, Y. S. Chen, D. Y. Wang, Y. J. Lee, W. C. Chen, Y. H. Wang, C. T. Yen, S. Y. Yang, K. H. Shen, C. P. Chang, C. S. Lin, K. L. Su, H. C. Cheng, Y. J. Wang, D. D. L. Tang, M. J. Tsai, and M. J. Kao, “Adjacent-reference and self-reference sensing scheme with novel orthogonal wiggle MRAM cell,” in *IEDM Tech. Dig.*, 2006, pp. 334-337.

[4] T. Nakamura, Y. Fujimori, N. Izumi, and A. Kamisawa, “Fabrication technology of ferroelectric memories,” *Jpn. J. Appl. Phys.*, vol. 37, no. 3B, pp. 1325-1327, Mar. 1998.

[5] R. Moazzami, “Ferroelectric thin film technology for semiconductor memory,” *Semicond. Sci. Technol.*, vol. 10, no. 4, pp. 375-390, Apr. 1995.

[6] N. Yamada, E. Ohno, K. Nishiuchi, N. Akahira, and M. Takao, “Rapid-phase transitions of GeTe-Sb_2Te_3 pseudobinary amorphous thin films for an optical disk memory,” *J. Appl. Phys.*, vol. 69, no. 5, pp. 2849-2856, Mar. 1991.

[7] S. Lai and T. Lowrey, “OUM – a 180 nm nonvolatile memory cell element technology for stand alone and embedded applications,” in *IEDM Tech. Dig.*, 2001, pp. 803-806.

[8] I. G. Baek, M. S. Lee, S. Seo, M. J. Lee, D. H. Seo, D. S. Suh, J. C. Park, S. O. Park, H. S. Kim, I. K. Yoo, U. I. Chung, and J. T. Moon, “Highly scalable

non-volatile resistive memory using simple binary oxide driven by asymmetric unipolar voltage pulses," in *IEDM Tech. Dig.*, 2004, pp. 587-590.

[9] S. Seo, M. J. Lee, D. H. Seo, E. J. Jeoung, D. S. Suh, Y. S. Joung, I. K. Yoo, I. R. Hwang, S. H. Kim, I. S. Byun, J. S. Kim, J. S. Choi, and B. H. Park, "Reproducible resistance switching in polycrystalline NiO films," *Appl. Phys. Lett.,* vol. 85, no. 23, pp. 5655-5657, Dec. 2004.

[10] S. Seo, M. J. Lee, D. H. Seo, S. K. Choi, D. S. Suh, Y. S. Joung, I. K. Yoo, I. S. Byun, I. R. Hwang, S. H. Kim, and B. H. Park, "Conductivity switching characteristics and reset currents in NiO films," *Appl. Phys. Lett.*, vol. 86, no. 9, p. 093509, Feb. 2005.

[11] M. Fujimoto, H. Koyama, M. Koyama, M. Konagai, Y. Hosoi, K. Ishihara, S. Ohnishi, and N. Awaya, "TiO_2 anatase nanolayer on TiN thin film exhibiting high-speed bipolar resistive switching," *Appl. Phys. Lett.*, vol. 89, no. 22, p. 223509, Nov. 2006.

[12] D. Lee, H. Choi, H. Sim, D. Choi, H. Hwang, M. J. Lee, S. A. Seo, and I. K. Yoo, "Resistance switching of the nonstoichiometric zirconium oxide for nonvolatile memory application," *IEEE Electron Device Lett.*, vol. 26, no. 10, pp. 719-721, Oct. 2005.

[13] C. Y. Lu, T. C. Lu, and R. Liu, "Non-volatile memory technology-today and tomorrow," in *Proc. IEEE-IPFA*, 2006, pp. 18-23.

[14] W. N. Papian, "The MIT magnetic-core memory," in *Proc. Eastern Joint Comp. Conf.*, Dec. 1953, pp. 37-42.

[15] D. Kahng and S. M. Sze, "A floating gate and its application to memory devices," *Bell Syst. Tech. J.*, vol. 46, no. 4, pp. 1288-1295, 1967.

[16] J. D. Blauwe, "Nanocrystal nonvolatile memory devices," *IEEE Trans. Nanotechnol.*, vol. 1, no. 1, pp. 72-77, Mar. 2002.

[17] P. Pavan, R. Bez, P. Olivo, and E. Zanomi, “Flash memory cells—an overview,” *Proc. IEEE,* vol. 85, no. 8, pp. 1248-1271, Aug. 1997.

[18] M. H. White, D. A. Adams, and J. Bu, “On the go with SONOS,” *IEEE Circuits Devices Mag.*, vol. 16, no. 4, pp. 22-31, Jul. 2000.

[19] Process Integration, Devices and Structures in International Technology Roadmap for Semiconductors, pp. 35-36, 2007.

[20] D. Frohman-Bentchkowsky, and M. Lenzlinger, “Charge transport and storage in metal-nitride-oxide-silicon (MNOS) structures,” *J. Appl. Phys.*, vol. 40, no. 8, pp. 3307-3319, Jul. 1969.

[21] D. Frohman-Bentchkowsky, “The metal-nitride-oxide-silicon (MNOS) transistor —characterisctics and applications,” *Proc. IEEE*, vol. 58, no. 8, pp. 1207-1219, Aug. 1970.

[22] J. Bu, and M. H. White, “Effects of two-step high temperature deuterium anneals on SONOS nonvolatile memory devices,” *IEEE Electron Device Lett.*, vol. 22, no. 1, pp.17-19, Jan. 2001.

[23] K. T. Chang, W. M. Chen, C. Swift, J. M. Higman, W. M. Paulson, and K. M. Chang, “A new SONOS memory using source-side injection for programming,” *IEEE Electron Device Lett.*, vol. 19, no. 7, pp. 253-255, Jul. 1998.

[24] T. Sugizaki, M. Kobayashi, M. Ishidao, H. Minakata, M. Yamaguchi, Y. Tamura, Y. Sugiyama, T. Nakanishi, and H. Tanaka, “Novel multi-bit SONOS type flash memory using a high-k charge trapping layer,” in *VLSI Symp. Tech. Dig.*, 2003, pp. 27-28.

[25] P. Xuan, M. She, B. Harteneck, A. Liddle, J. Bokor, and T. J. King, “FinFET SONOS flash memory for embedded applications,” in *IEDM Tech. Dig.*, 2003, pp. 609-613.

[26] C. H. Lee, J. Choi, C. Kang, Y. Shin, J. Sel, J. Sim, S. Jeon, B. I. Choe, D. Bae, K.

Park, and K. Kim, “Multi-level NAND flash memory with 63 nm-node TANOS (Si-Oxide-SiN-Al_2O_2-TaN) cell structure,” in *VLSI Symp. Tech. Dig.*, 2006, pp. 21-22.

[27] H. T. Luo, S. Y. Wang, E. K. Lai, Y. H. Shih, S. C. Lai, L. W. Yang, K. C. Chen, J. Ku, K. Y. Hsieh, R. Liu, and C. Y. Lu, “BE-SONOS: A bandgap engineered SONOS with excellent performance and reliability,” in *IEDM Tech. Dig.*, 2005, pp. 547-550.

[28] H. T. Luo, 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),” in *IEDM Tech. Dig.*, 2006, pp. 495-498.

[29] C. H. Lee, K. I. Choi, M. K. Cho, Y. H. Song, K. C. Park, and K. Kim, “A novel SONOS structure of SiO_2/SiN/$A1_2O_3$ with TaN metal gate for multi-giga bit flash memories,” in *IEDM Tech. Dig.*, 2003, pp. 613-616.

[30] Z. Liu, C. Lee, V. Narayanan, G. Pei, and E. C. Kan, “Metal nanocrystal memories- Part I: device design and fabrication,” *IEEE Trans. Electron Devices*, vol. 49, no. 9, pp. 1606-1613, Sep. 2002.

[31] T. H. Hou, C. Lee, V. Narayanan, U. Ganguly, and E. C. Kan, “Design optimization of metal nanocrystal memory- Part I: nanocrystal array engineering,” *IEEE Trans. Electron Devices*, vol. 53, no. 12, pp. 3095-3102, Dec. 2006.

[32] J. J. Lee and D. L. Kwong, “Metal nanocrystal memory with high-κ tunneling barrier for improved data retention,” *IEEE Trans. Electron Devices*, vol. 52, no. 4, pp. 507-511, Apr. 2005.

[33] S. Choi, S. S. Kim, M. Chang, H. Hwang, S. Jeon, and C. Kim, “Highly thermally stable TiN nanocrystals as charge trapping sites for nonvolatile memory device applications,” *Appl. Phys. Lett.*, vol. 86, no. 12, p. 123110, Mar.

2005.

[34] M. She and T. J. King, “Impact of crystal size and tunnel dielectric on semiconductor nanocrystal memory performance,” *IEEE Trans. Electron Devices*, vol. 50, no. 9, pp. 1934-1940, Sep. 2003.

[35] S. T. Kang, J. Yater, C. Hong, J. Shen, N. Ellis, M. Herrick, H. Gasquet, W. Malloch, and G. Chindalore, “Si Nanocrystal Split Gate Technology Optimization for high performance and reliable embedded microcontroller applications,” in *Proc. Non-Volatile Semiconductor Memory Workshop*, 2008, pp. 59-60.

[36] K. Kim, J. H. Choi, J. Choi, and H. S. Jeong, “The future prospect of nonvolatile memory,” in *Proc. Int. Symp. VLSI-TSA*, 2005, pp. 88-94.

[37] R. Waser, “Resistive non-volatile memory devices (Invited Paper),” *Microelectronic Engineering*, vol. 86, no. 7-9, pp. 1925-1928, 2009.

[38] A. Padilla, S. Lee, D. Carlton, and T. J. K. Liu, “Enhanced endurance of dual-bit SONOS NVM cells using the GIDL read method,” in *VLSI Symp. Tech. Dig.*, 2008, pp. 142-143.

[39] E. Lusky, Y. Shacham-Diamand, G. Mitenberg, A. Shappir, I. Bloom, and B. Eitan, “Investigation of channel hot electron injection by localized charge-trapping nonvolatile memory devices,” *IEEE Trans. Electron Devices*, vol. 51, no. 3, pp. 444-451, Mar. 2004.

[40] A. Padilla, K. Shin, T. J. K. Liu, J. W. Hyun, I. Yoo, and Y. Park, “Dual-bit gate-sidewall storage FinFET NVM and new method of charge detection,” *IEEE Electron Device Lett.*, vol. 28, no. 6, pp. 502-505, Jun. 2007.

[41] A. Padilla and T. J. K. Liu, “Dual-bit SONOS FinFET non-volatile memory cell and new method of charge detection,” in *Proc. Int. Symp. VLSI-TSA*, 2007, pp. 24-25.

[42] Y. N. Tan, W. K. Chim, W. K. Choi, M. S. Joo, T. H. Ng, and B. J. Cho, "High-κ HfAlO charge trapping layer in SONOS-type nonvolatile memory device for high speed operation," in *IEDM Tech. Dig.*, 2004, pp. 889-892.

[43] C. H. Lai, A. Chin, H. L. Kao, K. M. Chen, M. Hong, J. Kwo, and C. C. Chi, "Very low voltage SiO_2/HfON/HfAlO/TaN memory with fast speed and good retention" in *VLSI Symp. Tech. Dig.*, 2006, pp. 44-45.

[44] C. H. Lee, K. C. Park, and K. Kim, "Charge-trapping memory cell of SiO_2/SiN/high-k dielectric Al_2O_3 with TaN metal gate for suppressing backward-tunneling effect," *Appl. Phys. Lett.*, vol. 87, no. 7, p. 073510, Aug. 2005.

[45] S. Maikap, P. J. Tzeng, L. S. Lee, H. Y. Lee C. C. Wang, P. H. Tsai, K. S. Chang-Liao, W. J. Chen, K. C. Liu, P. R. Jeng, and M. J. Tsai, "High-κ Hf-based charge trapping layer with Al_2O_3 blocking oxide for high-density flash memory," in *Proc. Int. Symp. VLSI-TSA*, 2006, pp. 36-37.

[46] S. Maikap, P. J. Tzeng, S. S. Tseng, T. Y. Wang, C. H. Lin, H. Y. Lee, C. C. Wang, T. C. Tien, L. S. Lee, P. W. Li, J. R. Yang, and M. J. Tsai, "High-κ HfO_2/TiO_2/HfO_2 multilayer quantum well flash memory devices," in *Proc. Int. Symp. VLSI-TSA*, 2007, pp. 18-19.

[47] S. Maikap, H. Y. Lee, T. Y. Wang, P. J. Tzeng, C. C. Wang, L. S. Lee, K. C. Liu, J. R. Yang, and M. J. Tsai, "Charge trapping characteristics of atomic-layer-deposited HfO_2 films with Al_2O_3 as a blocking oxide for high-density non-volatile memory device applications," *Semicond. Sci. Technol.*, vol. 22, no. 8, pp. 884-889, Aug. 2007.

[48] Sentaurus-TCAD Mannuls

[49] B. Y. Tsui and C. P. Lin, "A novel 25-nm modified Schottky-barrier FinFET with high performance," *IEEE Electron Device Lett.*, vol. 25, no. 6, pp. 430-433, Jun.

2004.

[50] J. Chen, T. Y. Chan, I. C. Chen, P. K. Ko, and C. Hu, “Subbreakdown drain leakage current in MOSFET,” *IEEE Electron Device Lett.*, vol. 8, no. 11, pp. 515-517, Nov. 1987.

個人簡歷（Vita）

姓　　名：余昆武

性　　別：男

出生年月日：民國七十四年八月十三日

出 生 地：台灣 桃園縣

學　　歷：

建國國小	（1991.9 ~ 1997.6）
桃園國中	（1997.9 ~ 2000.6）
國立武陵高級中學	（2000.9 ~ 2003.6）
國立交通大學電子工程學系學士	（2003.9 ~ 2007.6）
國立交通大學電子研究所碩士	（2007.9 ~ 2009.10）

碩 士 論 文：

非揮發性記憶體的儲存電荷的空間分佈對元件特性的影響

Effects of Spatial Distribution of the Stored-Charge on Device Characteristics of Non-volatile Memory