國 立 交 通 大 學

電子工程學系 電子研究所

博 士 論 文

分離式閘極非揮發性記憶體技術及

新穎多晶矽電子抹除式唯讀記憶體之研究

Study on Split-Gate Non-Volatile Memory

Technology and A Novel Single Poly EEPROM

Memory Cell

研 究 生 ：宋 弘 政

指導教授 ：雷 添 福

中華民國 九十七 年 六 月

分離式閘極非揮發性記憶體技術及

新穎多晶矽電子抹除式唯讀記憶體之研究

Study on Split-Gate Non-Volatile Memory Technology and

A Novel Single Poly EEPROM Memory Cell

研究生：宋 弘 政　　Student：Hung-Cheng Sung

指導教授：雷添福 博士　　Advisor：Dr. Tan-Fu Lei

國立交通大學

電子工程學系 電子研究所

A Dissertation
Submitted to Department of Electronics Engineering
and Institute of Electronics
College of Electrical and Computer Engineering
National Chiao Tung University
In Partial Fulfillment of the Requirements
For the Degree of
Doctor of Philosophy
In
Electronics Engineering
June 2008
Hsinchu, Taiwan, Republic of China

中華民國九十七 年 六 月

分離式閘極非揮發性記憶體技術 及

新穎多晶矽電子抹除式唯讀記憶體之研究

學生：宋 弘 政　　　　指導教授：雷 添 福 博士

國立交通大學

電子工程學系 電子研究所博士班

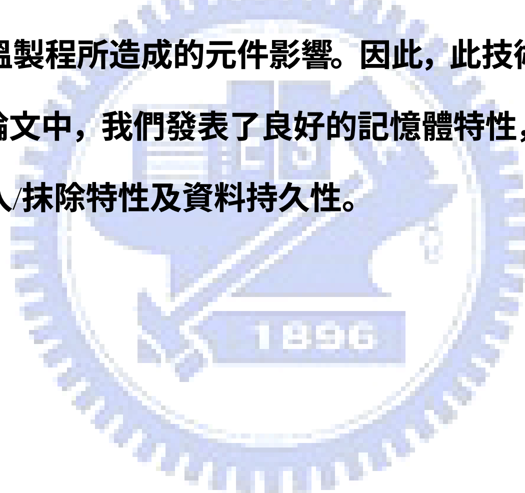
摘　　要

在此論文，首先，我們發展了一種新的方法來作分離式閘極快閃記憶體的寫入及干擾空間的量測。這個方法能幫助我們定量地了解操作空間的變化與電壓的關係，進而，這方法可以用來找到最佳化的寫入條件。由此方法找到的條件可以承受最大的電壓變化。我們成功地運用這方法在新一代的分離式閘極快閃記憶體的發展。

再者，一種新穎的三重自動對準分離式閘極快閃記憶元在此論文被揭露。此記憶元有 T 型的耦合結構，此新結構能大幅地增加源極和浮動閘極間的耦合電容而不需要記憶元面積的增加，此改善是藉由一個氧化層蝕刻的步驟來調變。此結構能用於寫入電壓的降低及記憶元的縮小。對於寫入電壓的降低，最高電壓可由

7.4V 降至 6.4V。而對於記憶元的縮小，我們成功地降低浮動閘極的長度由 0.18μm 到 0.14μm 而沒造成良率的下降或者可靠度的衰退。

最後，一種有著金屬控制閘極的新穎單多晶矽電子抹除式唯讀記憶體 EEPROM 在此論文中被發表。它的金屬閘極是由嵌刻(damascene)製程作成的鎢(W)線，它的閘極間的介電層是由原子磊晶長成的氧化鋁(Al_2O_3)。它的寫入/抹除的操作方式和傳統的堆疊閘極(stack-gate)記憶元是相同的，它用通道熱電子注入做寫入及用 FN 穿隧做抹除。因氧化鋁有著高介電常數的特性，所以我們可以用小於 6.5V 的電壓來執行寫入及抹除，而此電壓可以用 3.3V 的元件來操作，而不用使用到傳統的高壓元件。在製程相容方面，此記憶體只需比傳統 CMOS 製程多出二道光罩既可，此外，這氧化鋁是在後段製程中完成，所以此技術沒有汙染的顧慮以及額外高溫製程所造成的元件影響。因此，此技術非常能適用於嵌入式產品的應用。在此論文中，我們發表了良好的記憶體特性，如快速的寫入及抹除還有良好的重覆寫入/抹除特性及資料持久性。



Study on Split-Gate Non-Volatile Memory Technology and A Novel Single Poly EEPROM Memory Cell

Student: Hung-Cheng Sung   Advisor: Dr. Tan-Fu Lei

Department of Electronics Engineering &
Institute of Electronics
National Chiao Tung University

ABSTRACT

In this thesis, first, we developed a new methodology for program vs disturb window characterization on split gate flash. This method can help us to understand quantitatively how the window shifts vs bias conditions; furthermore, find the optimal program condition. The condition obtained by this method can withstand the largest program bias variations. This methodology was successfully implemented in the development for new generation of split-gate cell

Secondly, a new triple self-aligned (SA3) split-gate flash cell with a T-shaped source coupling structure is described in this paper. This novel structure can

significantly enhance coupling capacitance between the source and floating gate without increasing cell size. The enhancement can be simply modulated by an oxide-etching step. This new structure can be applied to program voltage reduction and cell size scaling. For program voltage reduction, the maximum program voltage of the new cell can be reduced from 7.4 to 6.4 V. For cell size scaling, we successfully reduce the floating length from 0.18μm to 0.14μm without showing the yield loss or reliability degradation.

Finally, a novel single poly EEPROM with metal control gate structure is presented in this paper. The control gate is tungsten (W) line made by a damascene process, and inter-gate dielectric is Al2O3 grown by Atomic Layer Deposition (ALD). The program and erase mechanism is the same as the one for traditional stacked-gate cell, which uses the channel hot electron injection for programming and Fowler-Nordheim (F-N) tunneling for channel erasing. With the high dielectric constant (K) property of Al2O3, we can perform the program and erase function with a voltage less than 6.5 V, which can be handled by 3.3 V devices instead of traditional high voltage devices. In the process compatibility aspect, this new cell needs only two extra masking steps over the standard CMOS process, and the high-K material is

deposited in the back-end metallization steps, so there is no cross-contamination issue caused by new material nor the device impact induced by the extra thermal cycle from conventional double poly process. Therefore, this new technology is suitable for embedded application. In this paper, the good cell performance is demonstrated; such as, fast programming/erasing, good endurance cycling and data retention.

誌謝

首先我要向我的指導教授雷添福博士致上最高的敬意。感謝他在學業研究與生涯規劃上給我的指導與鼓勵。此外，我要感謝台積電的王中樞資深處長和林詠濤處長的栽培和幫助、以及非揮發性記憶體部門同仁在技術方面的討論和建議，令我獲益良多。

同時也要感謝實驗室裡學弟的幫忙，你們的協助讓我的論文能順利完成。

最後，要感謝父母親多年的辛苦栽培以及我可愛的妻子和女兒的鼓勵。僅此論文獻給所有關心我的朋友。

Contents

Abstract (Chinese) I

Abstract (English)...III

Acknowledge VI

Contents VII

Figure Captions X

Chapter 1 Introduction 1

1.1 Embedded Flash Market and Application 1

1.2 Embedded Flash Technologies 3

1.2.1 1T Stack-gate Technologies 3

1.2.2 Split-gate Technology 4

1.2.3 2T Stack-gate Technology 5

1.2.4 Fully overlap Stack-gate Cell 6

1.3 Purpose of this work 6

Reference 20

Chapter 2 Split-Gate Flash Operation.. 22

2.1 Introduction 22

2.2 Cell cross-section and Layout..22

2.3 Cell Array Schematic 22

2.4 Programming 23

2.5 Erasing 26

2.6 Summary..27

Reference 34

Chapter 3 Novel Program vs Disturb Window Characterization for Split-Gate Flash Cell ………………………………………..……………………………... 36

3.1 Introduction ……………………….…………………………………...…..36

3.2 Experiment .....……………...………..........................……………………..37

3.3 Program vs. Disturb Window Characterization ……………………………..38

3.4 Application of the Window Characterization ………………………………..41

3.4.1 Finding optimal programming condition …………….………….41

3.4.2 Constant voltage vs. constant current programming ………………..41

3.5 Summary………………………………………………..…………………..…42

Reference ………………………………………………………..…………………..…57

Chapter 4 New Triple Self-aligned (SA3) Split-Gate Flash Cell with T-Shaped Source Coupling ……………………………………………………………………..…..58

4.1 Introduction…………………………………………………………………..58

4.2Device fabrication………………………………………………..……………..59

4.3 Array Bias condition ………………………………………...……………..60

4.4 The SCR Effect on Program and Erase ……………………………………60

4.4.1 Programming ………………………………….............................60

4.4.2 Erasing ……………………...…………………………………..60

4.5 Application to Voltage Reduction ..……………………..…..………………...61

4.6 Application to Cell Size Reduction ..………………….……...…………..…...62

4.7 Summary ..………………………………………………..………...……...63

Reference ..…………………………………………..………..…...……...73

Chapter 5 Novel Single Poly EEPROM with Metal Control Gate Structure..75

5.1 Introduction………………………………………………………………..75

5.2 Device Fabrication…..……………..………………………..……………..76

5.3 Result and Discussion………………………………………………….....77

5.4 Conclusion ……………………………………………………………......79

Reference ..……………………………………….……………………...88

Chapter 6 Conclusions and Further Recommendations ……………….……...90

6.1 Conclusions……………………………..…………………………….....90

6.2 Further recommendations………..……………………….………..…91

Reference …………………………………………………………..………....94

Vita ……………………………………………..…………………..………....95

Publication list ……………………………………………………….…..………...96

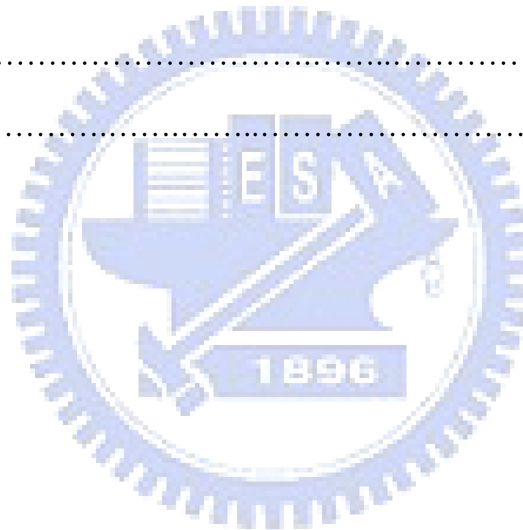


Figure Captions

Chapter 1

Fig. 1.1 (a)Estimate from 2004 to 2010 for unit shipments of flash MCU chips. (b) Flash MCU characteristics and memory density by application type. …….. 9

Fig. 1.2 Rapid increasing numbers of microcomputers are used in vehicles. The latest cars have microcomputers in from 30 to 100 different locations. ……….. 10

Fig. 1.3 (a) Schematic cross sectional view of stacked gate Flash memory cell along cell channel direction. The programming is through hot electron on drain side and erase is accomplished by FN tunneling in source side. (b) Top view and cross sectional view along with direction of stacked gate Flash memory cells fabricated in 0.35-mm technology. ………………………………………..… 11

Fig. 1.4 (a) Threshold voltage distribution of a 1-Mb Flash array after UV erasure, after CHE programming, and after FN erasure. (b) In NOR array architecture, the leaky bit can cause the read failure on its bit line [4]. ………………… 12

Fig. 1.5 VT shift during erase algorithm. A soft program after bulk erase can effectively tighten up the VT distribution. ……………………………..… 13

Fig. 1.6 Schematic illustration of self-aligned split-gate Flash cell (SuperFlash®) proposed by SST ……………………………………………………….. 14

Fig. 1.7 (a) Schematic of split gate nano-crystal bit cell for control gate-first and select gate-first integrations by Freescale, (b) Cell structure of MONOS split gate flash memory by Renesas, (c) Technology roadmap of Embedded Flash MCU by Renesas. ……………………………………………………… 15

Fig. 1.8 (a) 2-Tr stack gate NOR cell structure, (b) Top view of the cell [12] ………16

Fig. 1.9 Schematic diagram of the operation of 2 Tr stack gate cell. ………………. 17

Fig. 1.10 Top and cross-section drawing of the fully overlap cell ………………… 18

Fig. 1.11 (a) Circuit schematic drawing, (b) The operating conditions for word rewritable array ………………………………………………………… 19

Chapter 2

Fig. 2.1 (a) Cross-section of triple self-aligned split-gate Flash cell in bit-line direction, (b) Schematic diagram of cell top view and array configuration. ……….. 28

Fig. 2.2 (a) Equivalent circuit of split-gate Flash cell, (b) Cell array schematic… 29

Fig. 2.3 Bias condition and advantage of split-gate Flash cell. ……………………. 30

Fig. 2.4 Simulation results of the lateral field distribution along channel surface for (a) *region-A* (b) *region-B* (c) *region-C*. The enlarged schematic cross-section with electron flow lines during programming is shown in (b). In the inset of (b), point A (or A') marks the physical boundary between SG (or FG) and IPO. Point G (or G') marks the position of the maximum lateral field. The device parameters are: L_{SG} = 0.3 μm, L_{FG} = 0.183 μm, $\overline{L_{FG}}$ = 0.4 μm, T_{SGOX} = 180 Å, T_{FGOX} = 100 Å. ………………………………….. 31

Fig. 2.5 (a) Pinch-off point of FG and SG is A and B, respectively. Electron injection point is C, where is located in floating gate edge on the poly space. (b) Electric field and electron injection probability distribution. ……………. 32

Fig. 2.6 (a) Magnification of poly tip for tunneling, (b) energy band diagram during erasing, Emax is near the poly tip due to field-enhanced effect. ………… 33

Chapter 3

Fig. 3.1 (a)-(c) Cross section of he triple self-aligned split-gate Flash cell process sequence. ……………………………………………………………… 43

Fig. 3.2 (a) SEM picture of triple self-aligned split-gate cell, (b) TEM picture for a

sharp FG corner (indicated by the circle) created by poly etch. …………. 44

Fig. 3.3 Cell array and bias voltage for program, erase, read-out and three disturb conditions, which are: A. Column punchthrough disturb(PTC), B. Row punchthrough disturb(PTR), C. Reverse tunneling disturb(RT). Note that the cells outside the selected page are immune from disturb stress. ………… 46

Fig. 3.4 Program vs disturb window and the operation circle. The programming time is 10us and program current is 5μA. …………………………………… 47

Fig. 3.5 Program trend in PGM vs Disturb window. ……………………………… 48

Fig. 3.6 Bias condition for Column punch-through (PTC) and the disturb trend. … 49

Fig. 3.7 Bias condition for Row punch-through (PTR) and the disturb trend ………50

Fig. 3.8 Bias condition of Reverse tunneling (RT) and the disturb trend ………….. 51

Fig. 3.9 Trend of Vdp in operation window plot …………………………………… 52

Fig. 3.10 Operation circle in program and disturb window. The optimal program condition is at circle center, the condition is VSS=7.2V, Vg=1.75V, Idp=5μA. ……………………………………………………………………...53

Fig. 3.11 (a) Program vs disturb window varies with Idp from 1,5 to 9uA. Since the channel doping is well adjusted in this SA3 cell, no significant disturb boundary shift is observed under Idp variation (b) Operation circle comparison between Idp=1,5,9 μA. The optimal program condition is chosen at r2 center (Idp=5μA) because it has largest operation circle, the condition is VSS=7.2V, Vg=1.75V, Idp=5μA. ………………………………………… 54

Fig. 3.12 The circuit diagram for constant current and constant voltage programming ……………………………………………………………….. 55

Fig. 3.13 (a) The overlap window for constant current programming. The Idp varies from 1uA to 9uA, (b) The overlap window for constant voltage programming. Th eVdp varies from 0V to 0.6V. ………………………… 56

Chapter 4

Fig. 4.1 The cell cross-sectional view of tradition non-self-aligned cell vs. triple self-aligned(SA3) cell. 64

Fig. 4.2 Process flow of the T-shaped Triple Self-aligned (SA3) split-gate flash cell. 65

Fig. 4.3 (a) Traditional SA3 vs. T-shaped Vss coupling SA3 cell. (b) TEM picture of the new cell. The cell size is 0.38 μm2. 66

Fig. 4.4 Cell array and bias voltage for program, erase, read-out and three disturb conditions, which are: A. Column punchthrough disturb(PTC), B. Row punchthrough disturb(PTR), C. Reverse tunneling disturb(RT). Note that the cells outside the selected page are immune from disturb stress. 67

Fig. 4.5 (a) Pinch-off point of FG and SG is A and B, respectively. Electron injection point is C, where is located in floating gate edge on the poly space. (b) Electric field and electron injection probability distribution (c) Plot of factors effect on the electron injection probability. The factors include source voltage (Vss), source coupling ratio (SCR), substrate bias (Vsb) and bit-line voltage(Vbl). 68

Fig. 4.6 Source coupling ratio (SCR) vs. FG oxide spacer pull-back etching. 69

Fig. 4.7 (a) Program improvement vs FG oxide spacer pull-back. The program performance is characterizaed by Ir0/Ir1@Vs=6V, 10us, program current=3uA, where Ir0 and Ir1 is programmed and erased current, respectively. (b) Erasing improvement vs. FG oxide spacer etching. The erasing performance is characterized by Verase, which is the voltage to reach 50% Ir1 after 10ms erasing. 70

Fig. 4.8 (a) Program vs. disturb window of traditional SA3 cell. The source voltage for programming is 7.4V, (b) Program vs. disturb window of T-shape Vss coupling SA3 cell. Vss voltage can be reduced to 6.4V in new cell. ……. 71

Fig. 4.9 FG length reduction vs. yield. The yield of new SA3 cell remains stable when FG length is reduced, whereas the yield of old SA3 cell drops drastically with the shorter FG length. Note that the nominal FG length is around 180nm. . 72

Chapter 5

Fig. 5.1 (a) Logic OTP using oxide rupture mechanism [2]. The capacitor shown in the figure is a poly gate with junction overlap structure, it can be breakdown during programming, (b) schematic diagram of device and array structure for a single PMOS logic OTP memory [3], (c) The top and cross section view of eFuse. …………………………………………………… 80

Fig. 5.2 Schematic diagram and cross-sectional view of a logic MTP memory. …. 81

Fig. 5.3 (a) Cross-sectional view of a single cell, (b) Top view cell layout. The ideal 2T and 1T cell size is 26F2 and 18F2, respectively. …………………… 82

Fig.5.4 TEM picture of final cell. A proper oxidation treatment (~1 nm) before ALD and a post ALD anneal were done to ensure a good inter-gate dielectric quality. The physical thickness and the equivalent oxide thickness (EOT) of Al2O3 is about 20 nm and 9 nm, respectively. ………………………….. 83

Fig. 5.5 (a) Schematic diagram of memory array, (b) bias condition. ……………… 84

Fig. 5.6 (a) Program characteristics under different source (Vss) voltage, (b) Erasing characteristics with different control gate (CG) voltage. …………………. 85

Fig. 5.7 (a) Disturb cell location in the array. A,B,C &D cells are in the same page, which will have same high VSS during programming, (c) the cell current of

C, B,D cells after disturb. ………………………………………………… 86

Fig. 5.8 (a) Endurance Cycling characteristics with channel-hot-electron (CHE) programming and FN erasing. Only 10% current drop is observed after 100K cycling, (b) Data retention characteristics under various pre-cycling stress at 150C baking. ………………………………………………………… 87

Chapter 6

Fig. 6.1 Next generation of split-gate Flash cell. One extra CG gate is added to enhance the FG/VSS coupling. ………………………………………….. 93

Chapter 1

Introduction

1.1 Embedded Flash market and application

As the system getting more complicated, the request for both highs-speed logics and large memories arrays has increased in last decade. Masked ROMs have been intensively embedded in digital system to store data and program code for the micron-controller. However, the content in Masked ROM cannot be changed in the user end, it limits the flexibility of software upgrade, debug and impact the time to market of new product. As a result, more than of the half of micro-controller(MCU) are using non-volatile memory to store the code and data information.

The integration of non-volatile memory and logics can have following advantages. The most obvious one is the ***board area saving***, which is very important to many handheld consumer electronics. In addition, there are also several advantages in terms of performance and reliability over the two chips solutions with standard connections. The advantages are (1) ***Faster access time*** because of a reduced capacitive connection between microprocessor and memories; (2) ***Strongly reduced ground bouncing effects*** existing in stand-alone systems and caused by parasitic inductance when outputs are switching, (3) ***Increased number of memory input***, (4) ***Optimized bus, clock and control signals design***, (5) ***Reduction of the power consumption*** since the output buffers are removed, (6) ***Reduction of ElectroMagnetic Interferences(EMI) at board level*** [1].

Major application markets for Flash MCU include: automotive, household appliances, industrial and network controllers, consumer system, office automation, smart card controllers and USB controllers. Just for household appliance, USB controller, smart card and automotive segments, shipments of MCU chips with

embedded Flash memory are estimated to each nearly 7 billion units by 2010. The projected CAGR(compound annual growth rate) is 11.8% for revenue and 18.8% for unit shipment. Fig.1.1 (a)&(b) show an estimate from 2004 to 2010 for unit shipments of flash MCU chips. Segments considered in this estimate are flash MCU for: house appliances, smart cards, USB drive controllers and automotive [2]. The more detail discussion about the application for Embedded Flash is listed below.

- **Automotive Application**

Flash based MCU's are used in various subsystem modules throughout a vehicle including: power train, body and dash control, power steering control, and safety and navigation. They are being linked together in a networked vehicle using CAN and LIN controllers to handle control of the many MCU in the automobile. Flash based MCU's enable enhanced diagnostics capability within the vehicle. The field programmability of flash MCU's permits upgrading the vehicle and running diagnostics when at the dealer for routine maintenance and inspection.

- **Household Application**

Household appliances use digital controllers with non-volatile data storage for user parameter storage and for software updates. There is also potentially a flash MCU to link the appliance to a network or a motor control MCU.

- **Smart Card Application**

Smart cards are used in applications such as financial cards and citizen ID cards. There is a trend toward including more data in the card for new applications such as biometric identification. Another set of applications for SIM cards in mobile handsets is using embedded memory for storage of software for multimedia applications and tends to use more traditional non-volatile flash memory

- **Others Application**

Embedded Flash MCU is also widely used in sensor, battery and power supply

control, Zigbee protocol chips, USB controller chips and etc.

1.2 Embedded Flash Technologies

MCU with embedded memory already have more than 20 years history. At the beginning, only low density EEPROMs were integrated into the chip with microprocessor. As the system getting more complicated, higher density and higher performance embedded Flash technologies are developed to meet the application. Especially for automotive application, as shown in Fig.1.2, the latest cars have microcomputers in from 30 to 200 different locations, performing a wide range of powertrain control, vehicle ride, handling control, safety, comfort and convenience functions. There are many players in this market and each one has each own technology. In the following, four important technologies are chosen for brief discussion.

1.2.1 1T Stack-gate Technologies

The 1T stack-gate NOR array has dominated the high-end embedded Flash application for automotive application for years because of the small cell size and longtime good reliability record. Products with this structure in present market are mainly based on the ETOX concept (EPROM with Thin Oxide), which was proposed by Intel in 1984 [3]. Fig.1.3 (a) and (b) show the schematic cross sectional view along the channel direction and the top view of an industry standard stack gate cell.

Programming is performed by channel hot electron injection occurred near the drain side with control gate voltage(VCG) of 9-11 V, drain voltage (VD) of 4-6V, typically, in several μs. Erase is achieved by Folwer-Nordheim (FN) tunneling from FG to source or to both source and channel regions in a range of ms to several seconds. The threshold voltage(V_T) distribution of the Flash array after programming,

electrical erase and UV erasure is shown in Fig.1.4 (a) [4]. One of the major problems of this cell is the wide VT distribution after electrical erase. If the VT is too low to turn off the unselected cell completely, a high bit line current will occur and cause read failure as shown in Fig. 1.4 (b). Different models have been presented with the aim to explain the tail cells. For example, a distribution in the polycrystalline structure of the FG, with a barrier height variation at the grain boundaries, would give rise to a local enhancement of the tunnel barrier.[5]. Another model explains the tail cells as due to randomly distributed positive charge in the tunnel oxide [6]. To overcome the wide VT distribution after erase, a complicated erase and program algorithm is required. The algorithm includes the program/erase verification and a soft programming mechanism to tighten the VT distribution after erase. The graphical illustration of this operation is shown in Fig.1.5. This kind of algorithm will require a sophisticated state machine. For the high density embedded Flash application like engine control, the large overhead circuit might not be a concern because the penalty can be compensated by the small cell size [7]. However, for the commercial MCU application, the Flash density is typically smaller than 1-2Mb as shown in Fig.1.b, the large overhead circuit for program and erase control is not practical. In addition, the conventional channel hot electron programming is not efficient, it would require high current during data writing, which is not favored by the low power application. As a result, several technologies immune from over-erase issue are developed for the embedded Flash application; such as, split gate, 2T stack gate and fully overlap stack gate technology. We will briefly introduce these technologies in the following.

1.2.2 Split-gate Technology

Silicon Storage technology (SST) has developed one of the commercially most successful EmbFlash technology, a NOR type split gate cell with trademarked name

SuperFlash [8][9]. The advantage of this structure are low power consumption, fast access times and the immunity to over-erase issue. Generally speaking, split-gate structure uses source-side channel hot electron injection for programming, which has much higher program efficiency than drain side channel hot electron injection, and the erase is accomplished by field enhanced FN erase [10]. In Fig.1.6, due to the presence of the control gate (wordline), there is no concern for over-erase issue. Even though the unit cell size is larger than 1T stack gate, the chip area for small density array; like 1-2 Mb, is actually smaller than 1T stack gate cell because there is no extra circuit to control the V_T distribution.

The split gate structure can be used not only floating gate but also on local trap technology. As shown in the Fig.1.7(a) & (b) published by Freescale and Renesas, the advantage of split –gate technology is well recognized in the MCU application [11][12].

1.2.3 2T Stack-gate Technology

The cell consists of a select transistor(Tr) and a memory transistor(Tr) as shown in Fig. 1.8(a). Both select and memory transistor have the gate structure with the same with gate oxide, and the same symmetrical source/drain structure [12]. The memory array structure is shown in Fig.1.8(b).

The schematic operation diagrams are shown in Fig.1.9. In program mode, the VCG is applied with 12V and drain voltage is grounded on the selected cell , so the electron can be injected from channel to FG through FN tunneling. For unselected cell, the inhibit drain voltage is directly applied at the drain of memory Tr., and this inhibit drain voltage is cut-off by select Tr. In the erase mode, the erase voltage is divided into negative voltage and positive voltage, and Vcc is applied at a gate of select-Tr to reduce an electric field in the thin gate oxide of select-Tr. These approaches realize

channel FN program & erase and offer a high performance select-Tr with the thin gate oxide at the source side of memory cell. The advantages of this technology are low program and erase current, immune from over-erase issue, word re-writable and scalable to more advanced node. This cell is good for smart card application but the Flash macro size is still bigger than the split-gate Flash approach because of the 2T nature. Another drawback is slow program which is caused by FN tunneling.

1.2.4 Fully overlap Stack-gate Cell

Drawing of top view and cross-section views of the cell are shown in Fig.1.10. The active region, which has a heavy outline, has the form of the letter "Z" in the top view [13]. A first layer of polysilicon is used to form a floating gate memory transistor. A second layer of polysilicon, called the control gate, cover the floating gate and extends beyond the floating gat eon all sides. This second layer of poly forms a select transistor on each side of the floating gate. The coupling between FG and control gate is enhanced due to the fully overlap structure.

The programming is accomplished by applying a high positive voltage ~12-13V to the control gate of a cell in a grounded n-well causes electrons to tunnel from the accumulated well to floating gate. The cell is erased by biasing the n-well and bit line at a high positive voltage, typically 14-15V, and grounding the control gate. The operation condition is illustrated in Fig. 1.11.

Since this cell use FN tunneling for program and erase, and has the select transistor function because of the extended gate structure, so it has the advantage of low write current, immunity from over-erase and word re-writable function. However, the misalignment of the extended gate structure limits the scalability toward smaller geometry. Also, the thick oxide under the extended gate limits the cell current driving capability. The program speed is slow due to FN tunneling.

1.3 Purpose of this work

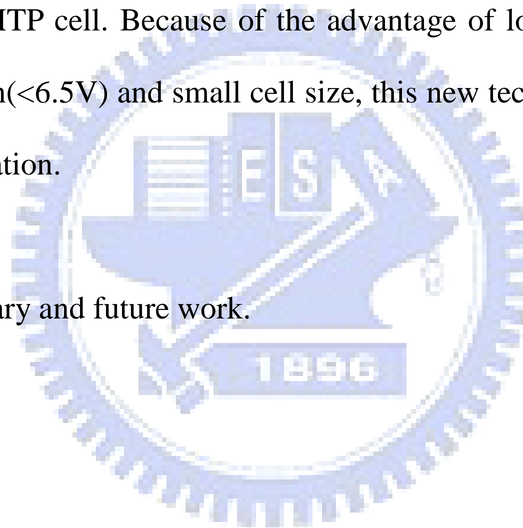
The traditional embedded Flash technology using ETOX cell requires more than 10 masking steps and complex state machine to control the programmed and erased V_T distribution. The split-gate Flash technology can perform the robust programming and erase function without state machine, however, it still needs the same complex process as stacked-gate floating technology. Moreover, it faces more serious hurdle to scale the cell size. The content of this thesis is divided into two parts; one is split-gate technology and another one is the novel metal gate EEPROM. For split-gate technology, we developed a new characterization to analyze the optimal program condition in terms of programming and disturb window, and we successfully developed a new high source coupling split-gate Flash cell which can operation at lower voltage and be more scalable. For novel metal gate EEPROM Flash, a new Flash EEPROM process with logic process compatibility is successfully demonstrated. The operation voltage can even be lower down to 6.5V and the extra masking steps over logic can be reduced to below 3. This new cell is very suitable to embedded application. The organization of this thesis is described in the following:

Chapter 2 describes the programming and erasure model for split-gate flash. It layouts the foundation for the programming optimization and new device development.

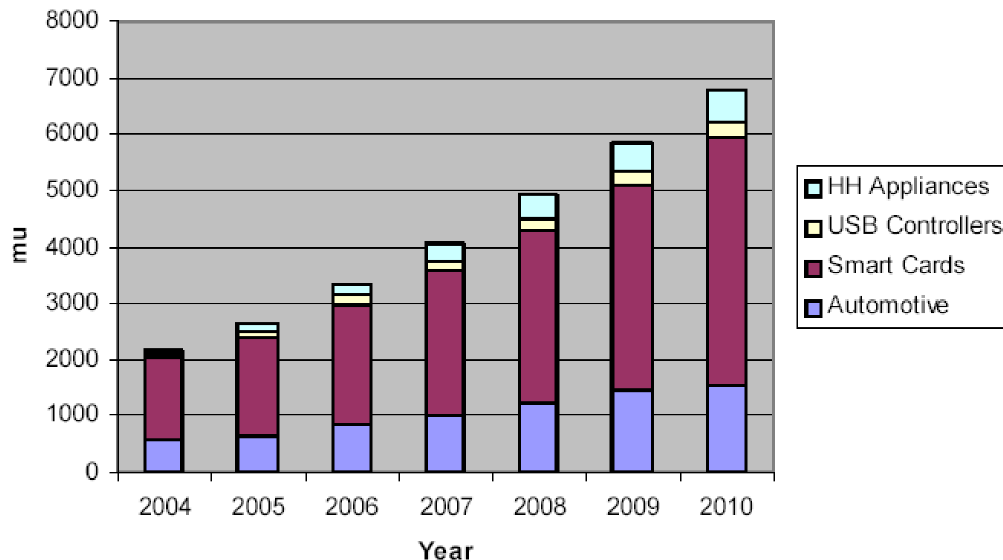
Chapter 3 introduces a new characterization technique to find the optimal programming condition, which can have the maximum programming and disturb window. This technique has been successfully utilized in the development for 0.18 split-gate Flash technology and beyond.

Chapter 4 disclose a new split-gate Flash cell with high source coupling ratio. With this new cell, we can have more room to reduce the floating length and lower down the operation voltage. We demonstrated the manufacturability of the cell by showing that the wafer sort yield (before and after retention bake) of 32Mb product is the same as the traditional process.

Chapter 5 reveals a novel metal gate EEPROM cell, which requires only 3 extra masking steps over standard logic process. The cell size is much smaller than conventional logic MTP cell. Because of the advantage of low extra masking steps, low voltage operation(<6.5V) and small cell size, this new technology is very suitable for embedded application.



Chapter is the summary and future work.



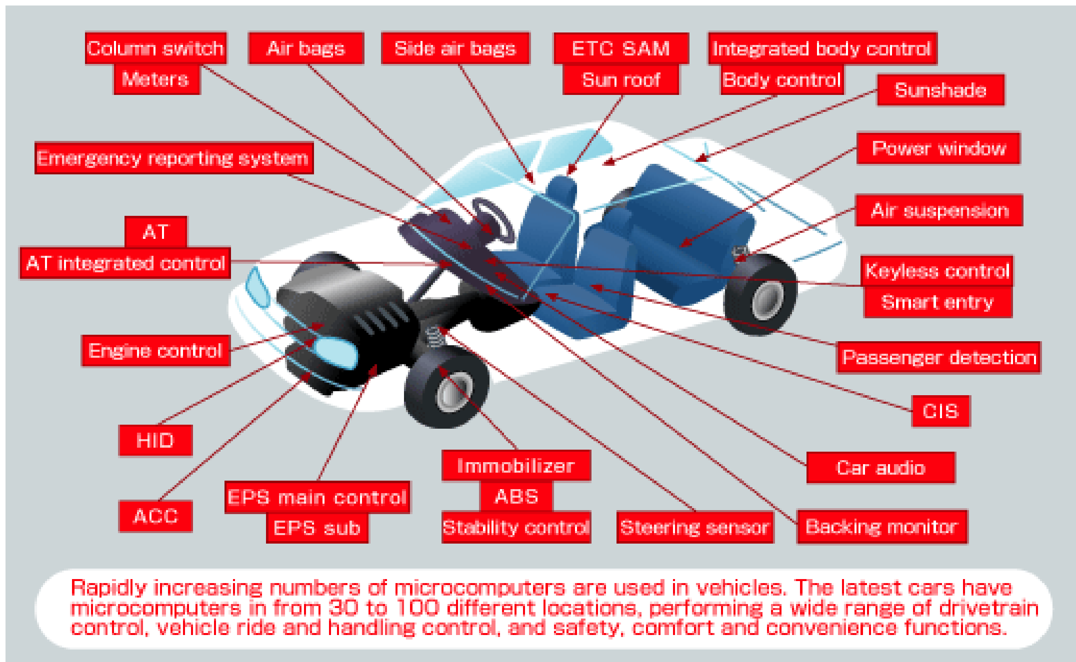
(a)

Flash MCU by Application

Market	Application	MCU Type	Flash
Automotive	Various	8-bit to 32-bit	16KB - 2MB
Industrial	Large Appliances	8-bit to 32-bit	60KB - 512KB
	Small Appliances	8-bit to 16-bit	128-KB
	Remote Control	8-bit to 16-bit	32-KB
	Motor Control	8-bit to 16-bit	16-KB
	Point of Sale	16-bit	128-KB
	Industrial Control	32-bit	512-KB
	Sensors	8-bit to16-bit	8-KB
	Electronic Locks	8-bit	8-KB
Consumer Digital	Baseband Chips	32-bit	224-KB
	Mobile Handsets	8-bit to 32-bit	60-KB
	Camera Phones	32-bit	64-KB
	Wireless Authentication	8-bit	3.5-KB
	LCD Driver	16-bit	256-KB
	Optical Disk Controller.	16-bit	768 KB
Networking	Wireless(ZigBee Protocol)	8-bit to 16-bit	128-KB
	Wireless Authentication	8-bit	3.5KB
	Indus. Network Control	32-bit	256-KB
Office Automation	USB Controllers	32-bit	128-KB - 1-MB
	Printers/Copiers	32-bit	384-KB
Smart Cards	Cell Phone/SIM	16-bit to 32-bit	36-KB to 384-KB
	ID Cards	8-bit to 16-bit	4KB - 72-KB

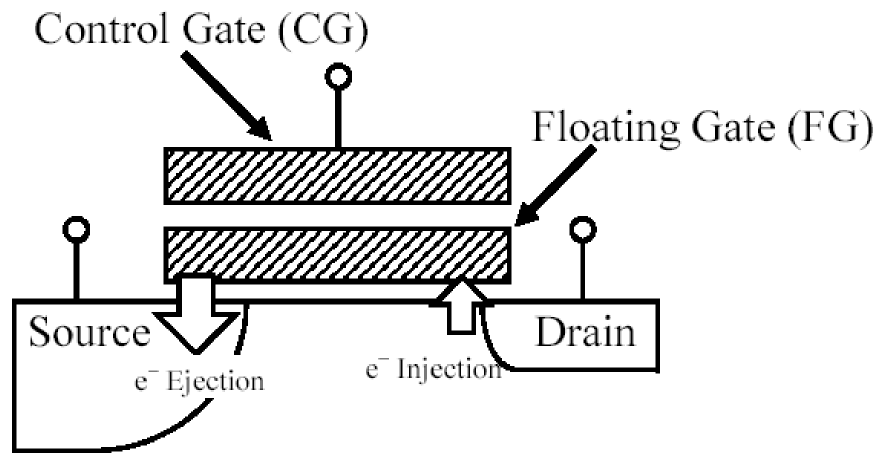
(b)

Fig. 1.1 (a)Estimate from 2004 to 2010 for unit shipments of flash MCU chips. (b) Flash MCU characteristics and memory density by application type [2].

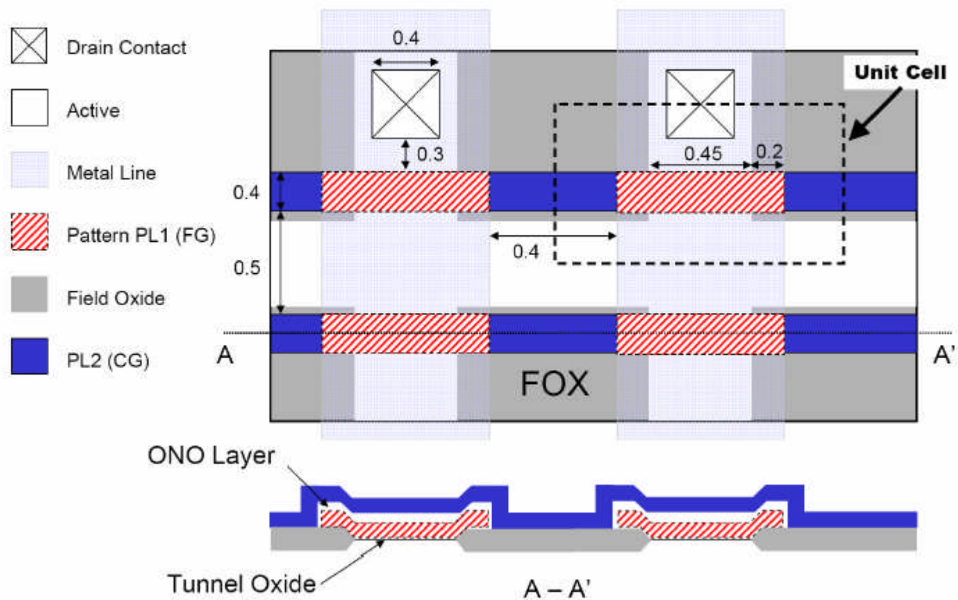


RENESAS
EDGE

Fig. 1.2 Rapid increasing numbers of microcomputers are used in vehicles. The latest cars have microcomputers in from 30 to 100 different locations.



(a)



(b)

Fig. 1.3 (a) Schematic cross sectional view of stacked gate Flash memory cell along cell channel direction. The programming is through hot electron on drain side and erase is accomplished by FN tunneling in source side. (b) Top view and cross sectional view along with direction of stacked gate Flash memory cells fabricated in 0.35-μm technology.

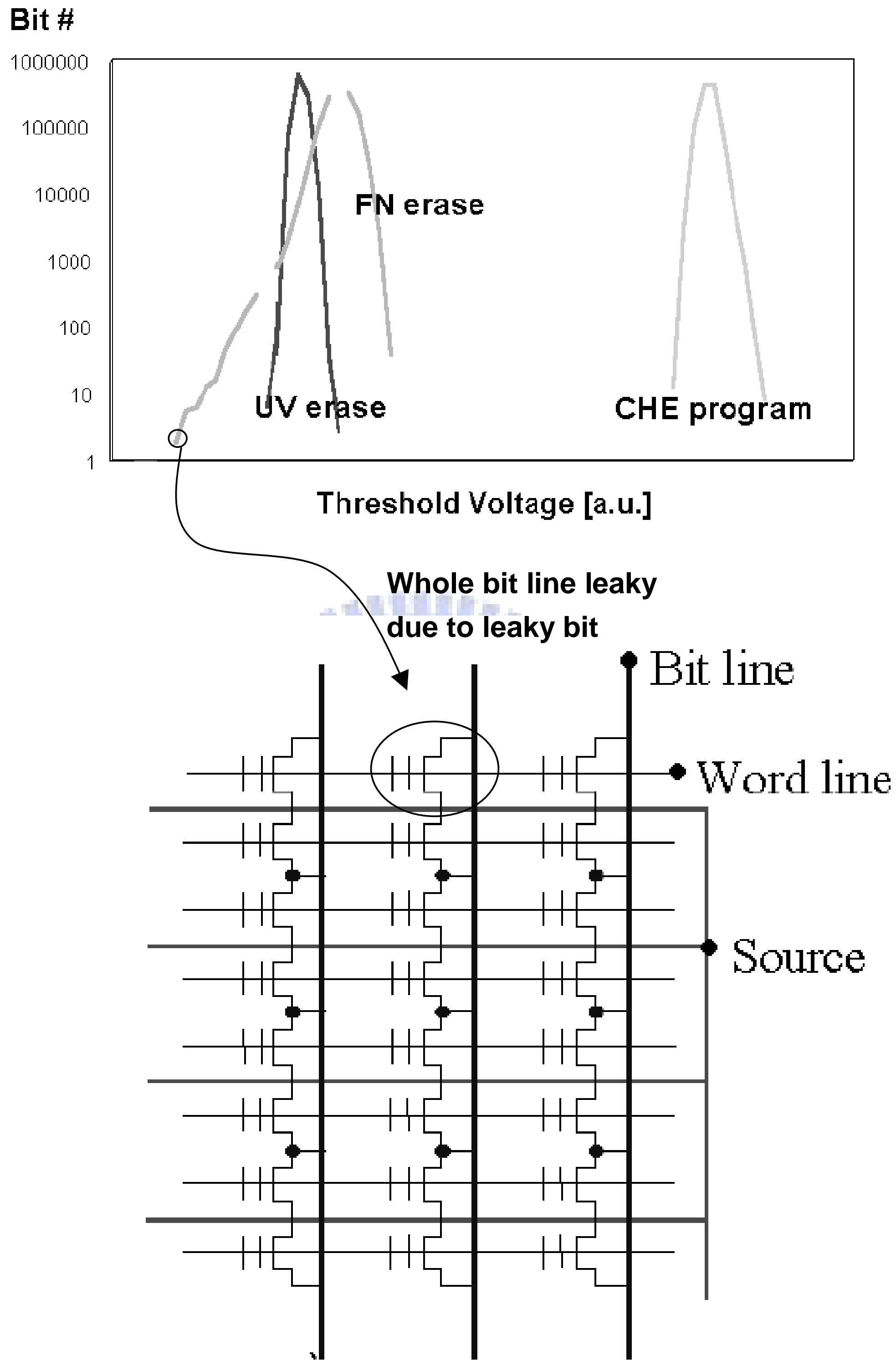


Fig. 1.4 (a) Threshold voltage distribution of a 1-Mb Flash array after UV erasure, after CHE programming, and after FN erasure. (b) In NOR array architecture, the leaky bit can cause the read failure on selected bit line [4].

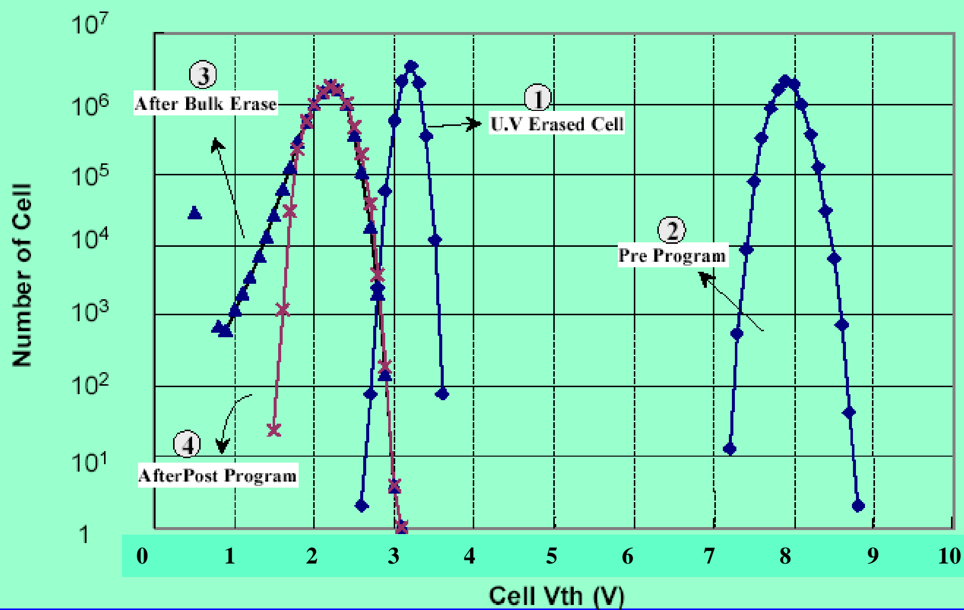


Fig. 1.5 V_T shift during erase algorithm. A soft program after bulk erase can effectively tighten up the V_T distribution.

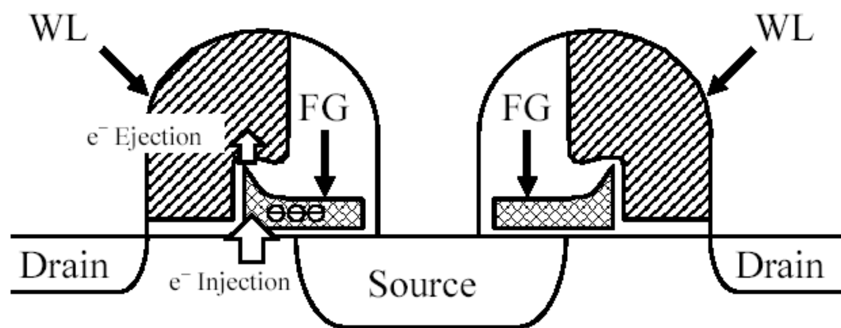


Fig. 1.6 Schematic illustration of triple self-aligned split-gate Flash cell (SuperFlash®) proposed by SST.

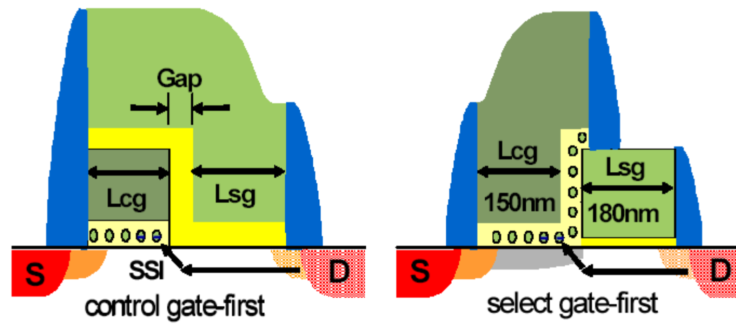
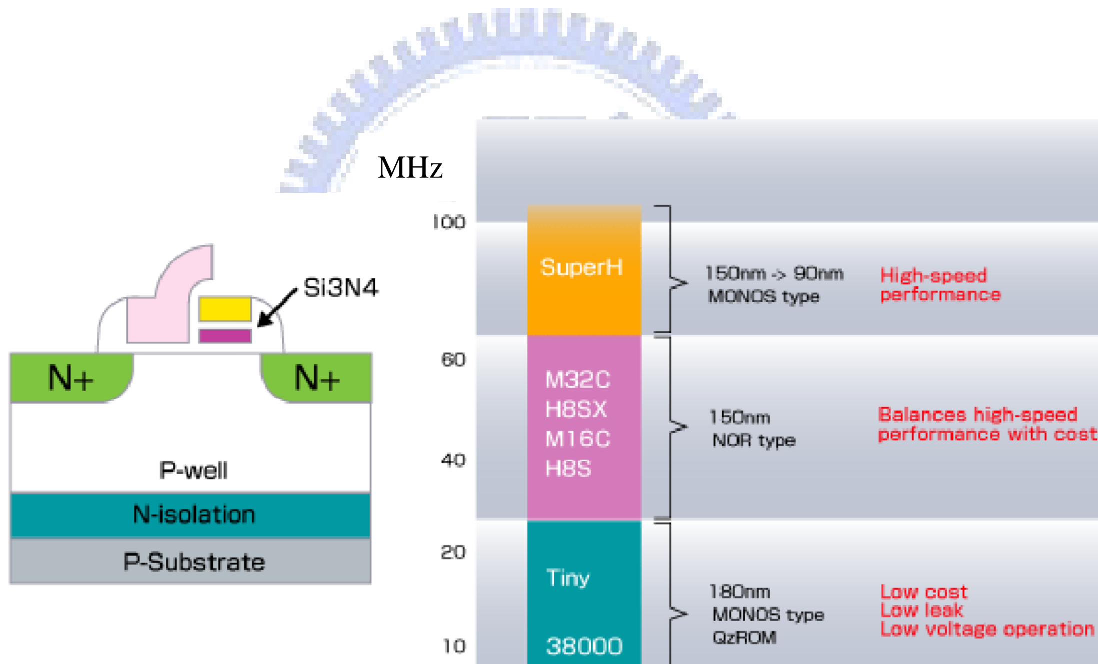


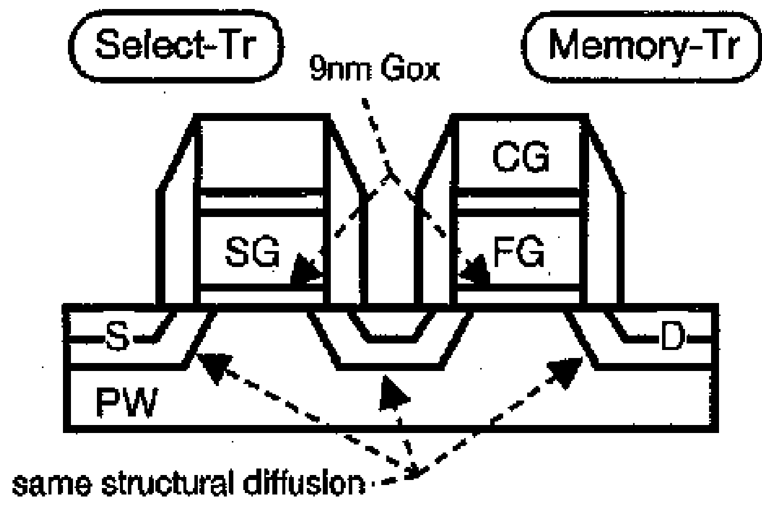
Figure 1. Schematic of split gate bitcell for control gate-first and select gate-first integrations. Control gate and select gate lengths are equivalent. Counterdoping under control gate is shown in grey.

(a)

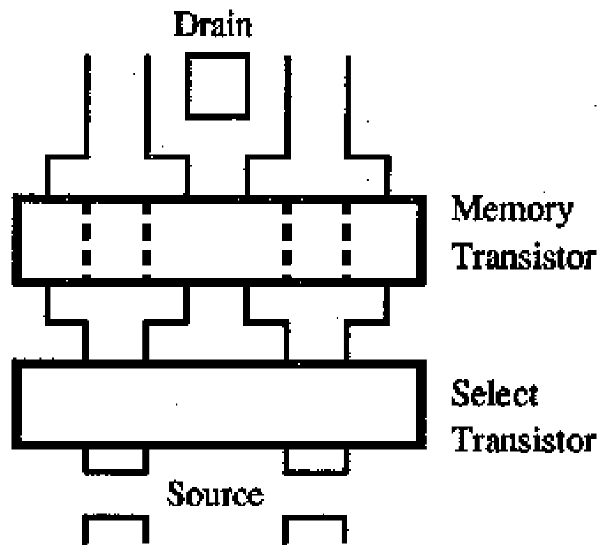


(b) (c)

Fig. 1.7 (a) Schematic of split gate nano-crystal bit cell for control gate-first and select gate-first integrations by Freescale. [11], (b) Cell structure of MONOS split gate flash memory by Renesas [12], (c) Technology roadmap of Embedded Flash MCU by Renesas [12].

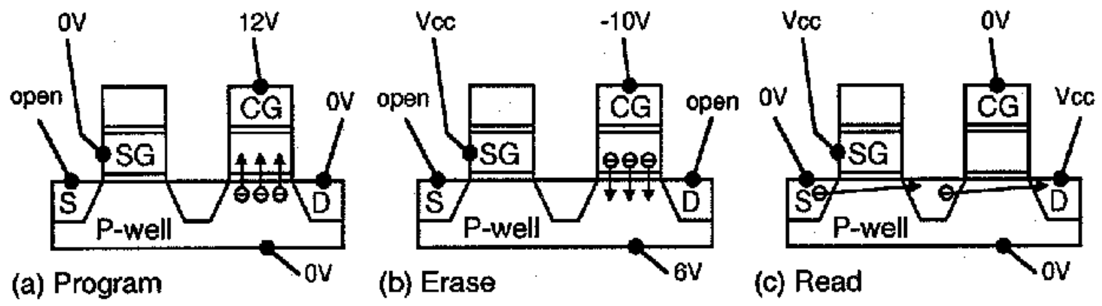


(a)



(b)

Fig. 1.8 (a) 2-Tr stack gate NOR cell structure, (b) Top view of the cell [12]



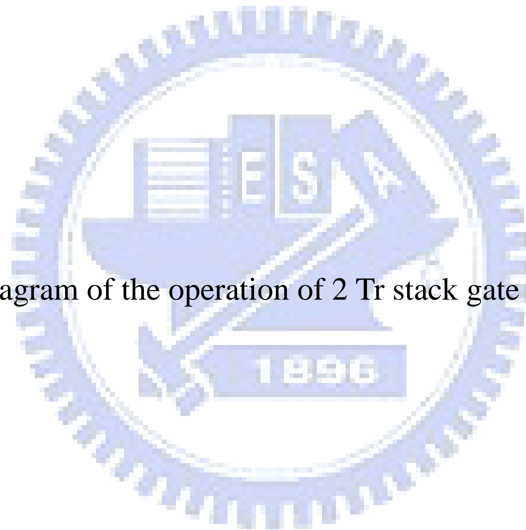


Fig. 1.9 Schematic diagram of the operation of 2 Tr stack gate cell [12].

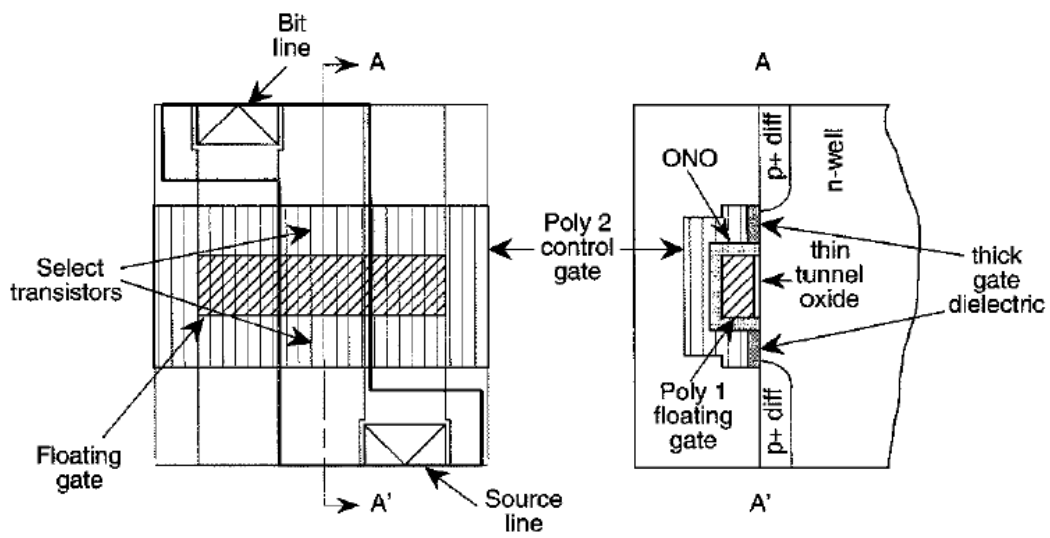
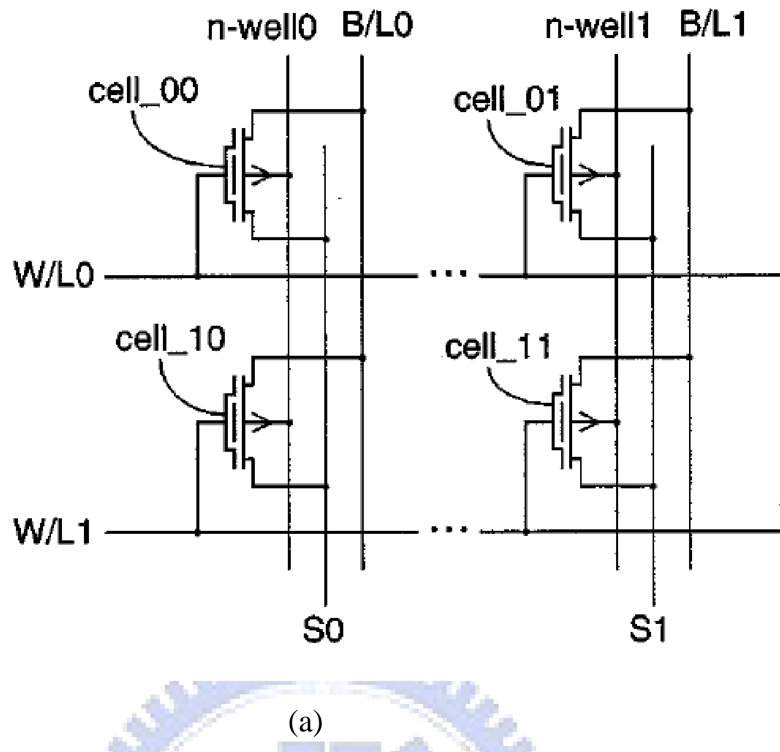


Fig. 1.10. Top and cross-section drawing of the fully overlap cell [13]



(a)

TABLE II
OPERATING CONDITIONS FOR WORD REWRITABLE ARRAY

		Clear (Program)	Write (Erase)	Read
Word line	Selected	V_{PP}	0	0
	Unselected	0	V_{PP}	V_{DD}
Bit line	Selected	0	V_{PP}	V_{DD} - 1.5
	Unselected	0	0	0
n-well	Selected	0	V_{PP}	V_{DD}
	Unselected	V_{PP}	V_{PP}	V_{DD}
Source	all	float	float	V_{DD}

(b)

Fig.1.11 (a) Circuit schematic drawing, (b) The operating conditions for word rewritable array

Reference

[1] P. Cappelletti, C. Golla, P. Olivo and E. Zanoni, *Flash memory*, Kluwer Academic Publishers, 2000, pp. 482

[2] " Non-Volatile Embedded Memories (Applications, Technologies, Users, IP Suppliers and Foundries)," *Memory Strategies International*, Sept. 2006, pp. 9-10.

[3] G. Verma and N. Mielke, "Reliability performance of ETOX based Flash memories," *Proc. IRPS*, 1988, p.158.

[4] R. Bez, E. Camerlenghi, A. Modelli, A. Visconti, "Introduction to Flash Memory," in *Proceedings of the IEEE*, Vol. 91, No.4, Apr. 2003, pp. 489-502.

[5] S. Maramatsu, T. Kubota, N. Nishio, H. Shirai, M. Matsuo, N. Kodama, M. Horikawa, S. Saito, K. Arai, and T. Okazawa, " The solution of over-erase problem controlling Poly-Si grain size – Modified scaling principles for Flash memory," in *IEDM Tech. Dig.*, 1994, pp.847-850

[6] C. Dunn, C. Kay, T. Lewis, T. Strauss, J. Screck, P. Hefley, M. Middendorf, and T. San, " Flash EEPROM disturb mechanism," in *Proc. Int. Rel. Phys. Symp.*, 1994, pp. 299-308

[7] C. Deml, M. Jankowski, C. Thalmaier, " A 0.13μm 2.125MB 23.5ns Embedded Flash with 2GB/s Read Throughput for Automotive Microcontrollers," *International Solid-State Circuits Conference*, 2007, pp. 478.

[8] S. Kianian, A. Levi, D. Lee, and Y.W. Hu, " A novel 3 volts-only, small sector erase, high density Flash E^2PROM," in *Symp. VLSI Technology Tech. Digest*, 1994, pp. 71-72.

[9] V. Markov, X.Liu, A. Kotov, A. Levi, T.N. Dang, and Y. Tkachev, " SuperFlash memory program/erase endurance," in *Proc. NVM Tech. Symp.*, 2003.

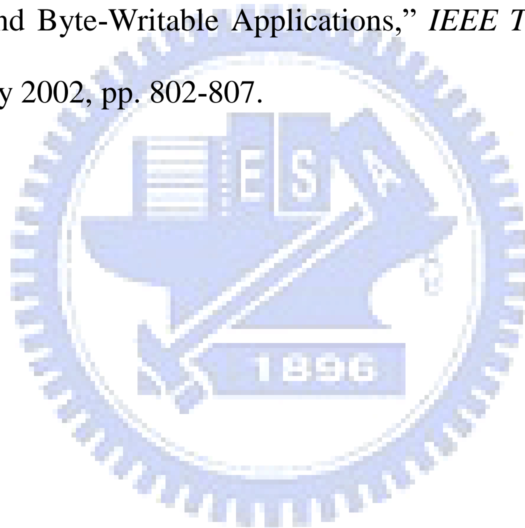
[10] J. Van Houdt, L. Haspeslagh, D. Wellekens, L. Deferm, G. Groeseneken, and H.E.

Maes, " HIMOS – A high effiencency Flash E^{2}PROM cell for embedded memory application," *IEEE Trans. Elect. Dev*., vol.40, pp.2255-2263, Dec, 1993.

[11] J.A. Yater, S.T. Kang, R. Steimle, C.M. Hing, B. Winstead, M. Herrick, G. Chindalore, "Optimization of 90nm Split Gate Nanocrystal Non-Volatile Memory," NVSMW 2007, pp. 77-78

[12] M. Hatanaka, T. Toya, "Applying Advanced Technology and Building a Stronger Product Rang to Make Our Flash&Flexible Microcomputer Superior Solutions", [online]. Available: http://resource.renesas.com/lib/eng/edge/11/special03.html

[13] J. Caywood, C.J. Huang, Y.J. Chang," A Nivel nonvolatile Memory Cell Suitable for Both Flash and Byte-Writable Applications," *IEEE Trans. Electron Devices,* Vol. 49, no.5, May 2002, pp. 802-807.



Chapter 2
Split-Gate Flash Operation

2.1 Introduction

The split-gate Flash cell in this thesis is built on 2nd generation of SuperFlash technology from Silicon Storage Technology, Inc.(SST)[1][2]. It use triple self-aligned technology to reduce cell size and improve coupling ration [3][4]. The superFlash technology and memory cell have a number of important advantages for designing and manufacturing flash EERPOM, or embedding the memory in logic devices, when compared with the thin oxide stacked gate or two transitor approaches. The split-gate memory cell is comparable in size to the single transistor stacked gate cell, yet provide the performance and reliability benefits of the traditional two transistor byte alterable EEPROM cell. By design, the split-gate memory cell eliminates the stacked gate issue if "overerase", by isolating each memory cell from the bit line. "Erase disturb" can not occur because all bytes are simultaneously erased in the same sector and each sector is completely isolated from every other sector during high voltage operation.

2.2 Cell cross-section and Layout

A top view and a cross –sectional view along bit line are presented in Fig 2.1. Polysilicon with silicide is used to connect control gate along the word line (row). Metal is used to connect the drain of each memory cell along the bit line (column). A common source is used for each sector, i.e., each pair sharing a common source along a row pair. A single word line is referred to as a row; the combination of pairs of rows form a sector, which is erased as a entity. Programming is done by word-by-word.

2.3 Cell Array Schematic

Fig.2.2 (a) is an equivalent memory cell, showing how the split-gate cell function as a select transistor and a memory transistor [5]. The memory array schematic is presented in Fig.2.2(b) , showing the logic organization of the memory array. This illustration represent a section of a typical cross-point memory array, arranged as 8 memory cells in 2 columns (bit lines), 2 sources lines, and 4 word lines. The cell operation condition is shown in Fig. 2.3. During the Read operation, reference voltage(2.5V) is applied to the control gate and the select gate via the word line. The reference voltage will "turn on" the select gate portion of the channel. If the floating gate is programmed (high threshold state), the memory transistor portion of the channel will not conduct. If the floating gate is erased (negative threshold state), this memory cell will conduct. The conducting state is output as a logic '1", the non-conducting state is a logic "0".

2.4 Programming

The cell programs using high efficient source-side channel hot electron injection. During programming, a voltage higher than the threshold (Vt) of the select transistor is placed on the control gate via the word line. This is sufficient to turn on the channel under the select portion of the control gate. The drain is biased at low voltage (~0.5) under a constant current circuitry if the cell is to be programmed. If the drain is at Vdd, programming is inhibited .The drain voltage is transferred across the select channel because of the voltage on the control gate. The source is bias at high voltage around 7~9V. Generally speaking, the source to drain voltage differential generate channel hot electrons. The source voltage is capacitively coupled to the floating gate. The field between the floating gate and the channel very efficiently sweeps the hot electrons across the Si-SiO2 barrier (~3.2eV) to the floating gate. To understand the programming mechanism in more detail, we can divide the programming operation into three regions [6]. The graphical illustration is shown in Fig. 2.4.

1) *Region-A*: $V_{FG} > V_{SS} + V_{TFG}$: (SG in saturation and FG in linear)

Where SG is the select gate transistor, FG is the floating gate transistor, V_{SS} is the potential at source node, V_{FG} is the FG potential, V_{TFG} is the threshold voltage of the FG.

The region happens when the cell is under deeply erased that the floating gate voltage is higher than the Vss voltage plus its threshold voltage. In region-A, the virtual drain extension concept [7] is valid and the channel surface potential at the gap is approximated as the drain potential. When V_{SG} is chosen much smaller than V_{SS}, a hyperbolic cosine-shaped lateral field [8] will be built up at the SG saturation region as shown in Fig. 2.4(a) and accelerates source-side injected electrons to jump into the FG.

2) *Region-B*: $V_{SG_SAT} + V_{TFG} \leq V_{FG} \leq V_{SS} + V_{TFG}$ (SG in saturation, FG in saturation)

Where V_{SG_SAT} is the saturation voltage of select transistor.

Most of the hot electron injection occurs in this region. FG transistor acts as a source follower, which can pass high voltage around V_{FG}-V_{TFG}. Since SG is slightly turn-on, this high voltage (V_{FG}-V_{TFG}) will make SG transistor operate in saturation region. This saturation region near SG/FG gap can generate high lateral electric field and trigger hot electron injection. The voltage drop between the pinch-off points of FG and SG is shown in Fig.2.5 (a), and the pinch-off potentials for FG at point A and for SG at B are expressed as:

$$V_A \cong V_{FG} - V_{TFG} \quad (1)$$

$$V_B \cong V_{SG} - V_{TSG}, \quad (2)$$

where V_A is the potential at point A, V_B is the potential at point B, V_{FG} is the FG potential, V_{TFG} is the threshold voltage of the FG, V_{SG} is the SG potential, and V_{TSG} is the threshold voltage of the SG.

In the split-gate flash program operation, a significant voltage drop in the longitudinal direction occurs between A and B; thus, the injection point is moved from the conventional high-voltage junction edge toward the poly gap between FG and SG. This is the so-called

"source-side" or "mid gate" hot carrier injection. From eqs. (1) and (2), the average longitudinal electric field (E_X) in the poly gap (A-B) region can be express as [9]-[11].

$$E_X = (V_A - V_B)/(X_A - X_B). \quad (3)$$

For first-order approximation, we can let V_{TFG} be equal to V_{TSG} and $(X_A - X_B)$ to $k \cdot L_G$, where L_G is the spacing between FG and SG, and k is the fitting parameter [9]. Thus, eq. (3) can be simplified to

$$E_X \cong (V_{FG} - V_{SG})/k \cdot L_G. \quad (4)$$

For a vertical oxide electric field at the injection point, the field is approximately

$$E_{OX} \cong (V_{FG} - V_C)/T_{OX}, \quad (5)$$

V_C will fall between V_A and V_B.

Assuming the storage charge in the floating gate is zero, the floating gate voltage (V_{FG}) is mainly coupled from FG and SG. It can be expressed in the following equation:

$$V_{FG} = Q_{FG}/C_T + \alpha_S V_{SS} + \alpha_G V_{SG} \cong \alpha_S V_{SS}, \quad (6)$$

where Q_{FG} is the floating gate charge, C_T is the total capacitance, α_S is the Vss coupling ratio on FG (SCR), and α_G is the select-gate coupling ratio on FG.

Insert eq. (6) into (4) and (5), we can further derive the E_X and E_{OX} 's relation with α_S and V_{SS}.

$$E_X \cong (\alpha V_{SS} - V_{SG})/k \cdot L_G \quad (7)$$

$$E_{OX} \cong (\alpha V_{SS} - V_C)/T_{OX} \quad (8)$$

With split –gate structure, we can achieve high E_X and a high E_{OX} favorable for electron injection at the same location. On the basis of the lucky-electron model (LEM), the injection current is determined by [12]-[13],

$$I_{FG} = K \times Is \left(\frac{\lambda n E_X}{\Phi_b} \right)^2 Exp\left(-\Phi_b/\lambda E_X\right), \quad (9)$$

where Is is the programming current, λ is the scattering mean free path, and Φ_b is the effective barrier height and commonly expressed as

$$\Phi_b = \Phi_{b0} - \beta E_{OX}^{1/2} - \theta E_{OX}^{2/3}. \quad (10)$$

Here, Φ_{b0} is the zero-oxide-field potential barrier (Φ_{b0} = 3.1 V), the second term is the Schottky barrier-lowering effect due to the image field, and the last term is the tunneling-related barrier-lowering coefficient [14].

From eqs. (9) to (10), we note that an increase in E_X and E_{OX} can significantly improve program efficiency because channel electrons can gain more kinetic energy along longitudinal direction and have lower barrier height for injection to floating. The simulation result is shown in Fig.2.5 (b). A high electron injection probability occurs when high E_X and E_{OX} are aligned in the FG/SG gap region, as a result the hot electron injection efficiency in split-gate cell is 1000 times more than the stack-gate counter part.

3) *Region-C*, $V_{FG} \le V_{SG_SAT} + V_{TFG}$ (SG in linear, FG is saturation)

In this region, V_{FG} is so low that the voltage passed through FG transistor cannot cause the SG transistor to be in saturation mode. The function of SSI programming will cease because source side injected electrons gain no more energy at the gap.

2.5 Field Enhanced Fowler-Nordheim Tunneling

The cell erases uses Fowler-Nordheim tunneling from the floating gate to the control gate. During erasing, the source and drain are grounded and the word line is raised to a high voltage. The low coupling ratio between the control gate and the floating gate provide a significant voltage drop across the inter-poly oxide. A local high electric field is generated primarily along the edge of the tunneling injector. Charge transfer is very rapid (~1ms) and is eventually limited by the accumulation of positive charge charges on the floating gate. This positive charge raises the floating gate voltage until there is insufficient voltage drop to

sustain Flower-Nordhein tunneling. The graphical illustration of field enhanced poly-to-poly FN tunneling is shown in Fig.2.6. A sharp tunneling injector is formed by special slop etching on FG in the triple self-aligned split gate flash technology. Unlike the traditional erasure by the Fowler-Nordheim tunneling though a triangular barrier, which requires a higher voltage across thin tunneling oxide. The field enhanced tunneling through sharp tip can be achieved at lower voltage and with thicker oxide to maintain good data retention performance. Using a cylindrical approximation, the electric field is highest (E_{MAX}) at the inner edge and lowest (E_{MIN}) at the outer edge [15];

$$E_{MAX} \cong V_{OX} / [a \times \ln(1 + T_{OX} / a)], \quad (11)$$

where a is the radius of curvature of the smaller cylinder, and

$$V_{OX} = V_{SG} - V_{FG} . \quad (12)$$

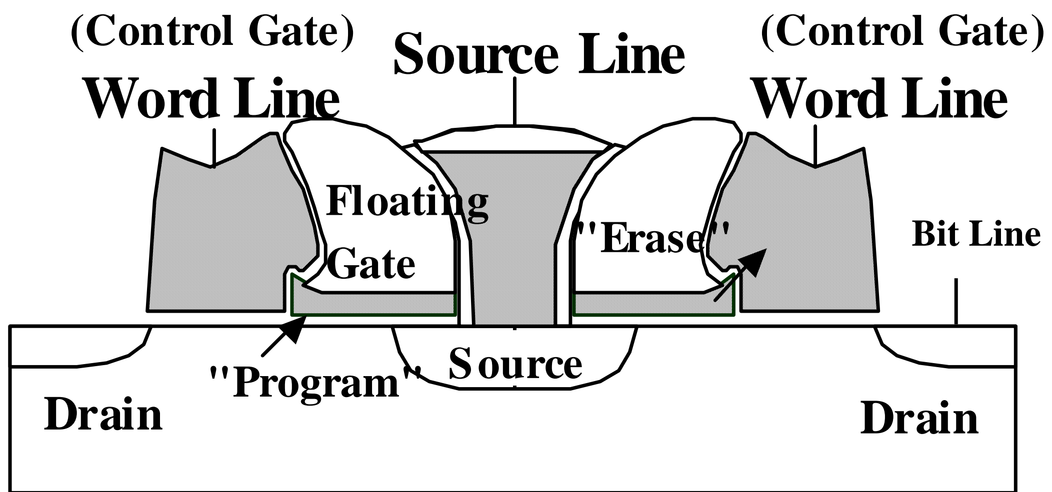
The tunneling current is described by the Fowler-Nordheim tunneling equation.

$$I_{fg}=AE_{OX}^{2}e^{-B/Eox} \quad (13)$$

Where A and B depends on the material property of the oxide and interface.

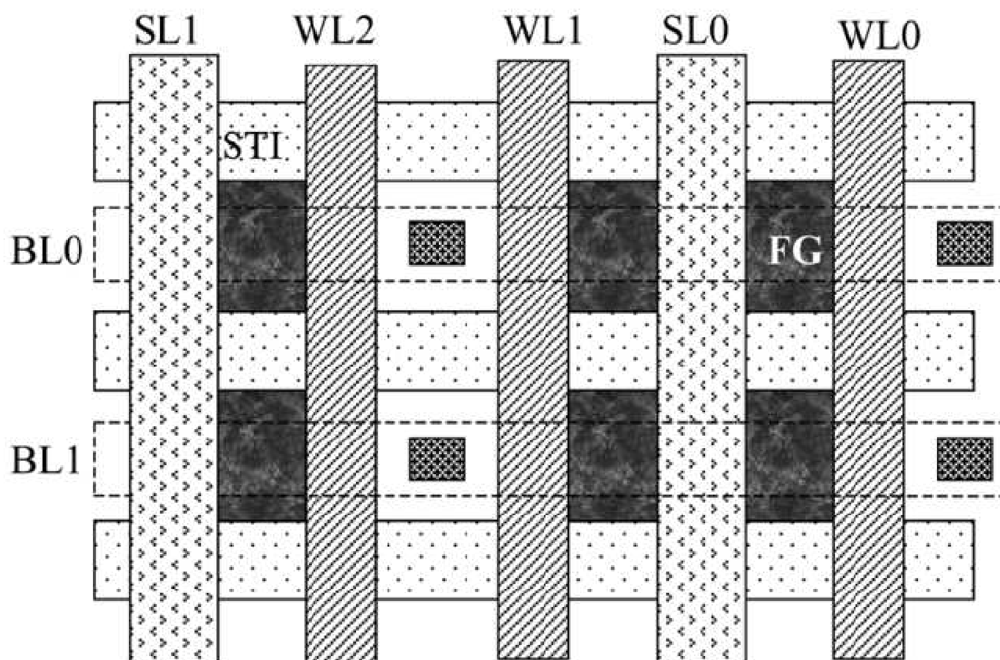
2.6 Summary

An analytical programming and erasing model is introduced in this chapter. It layouts the physical foundation for the following chapters on program/disturb window study and the new split-gate Flash cell design.



(a)

[watermark]



(b)

Fig.2.1 (a) Cross-section of triple self-aligned split-gate Flash cell in bit-line direction, (b) Schematic diagram of cell top view and array configuration

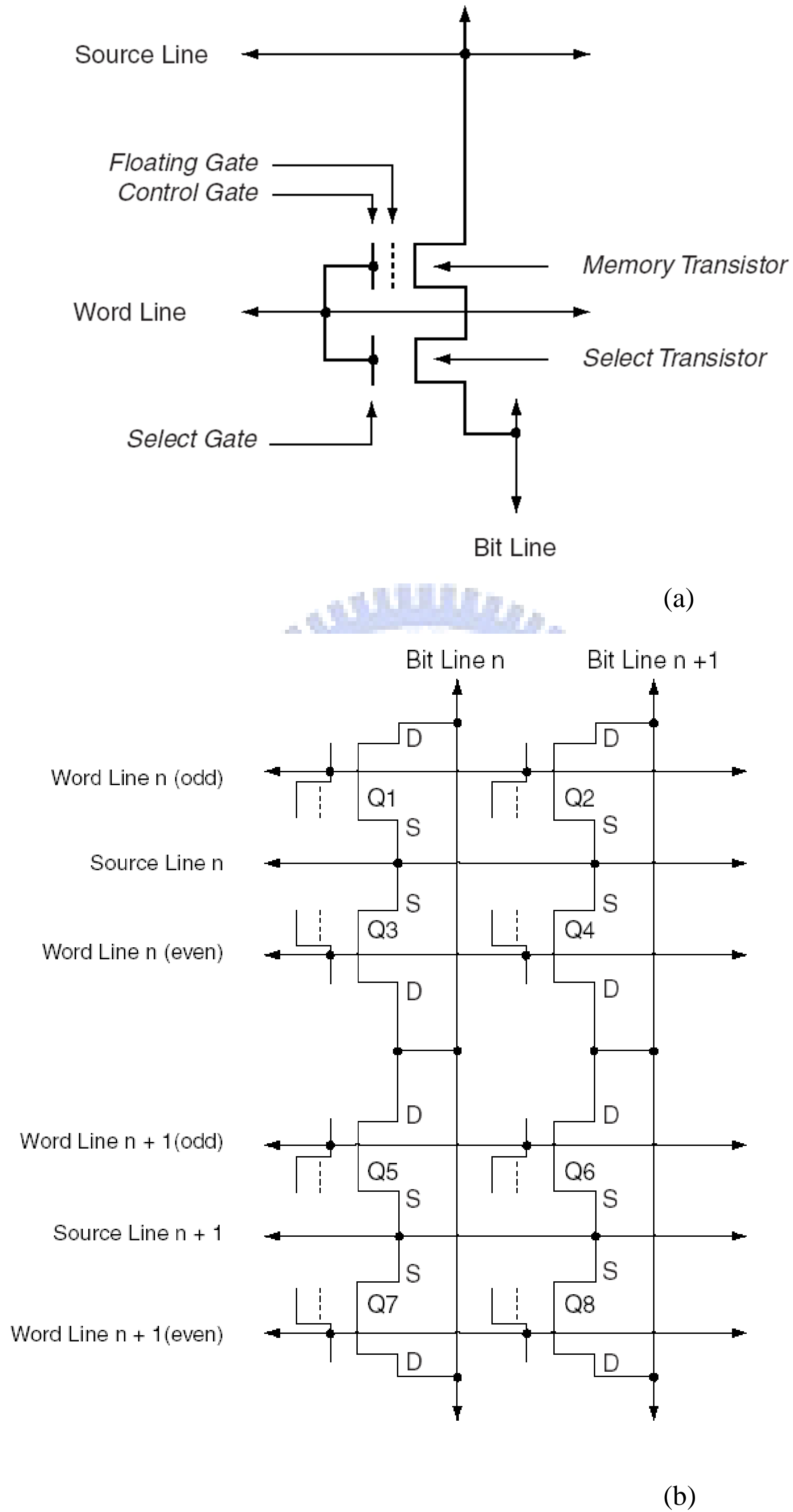
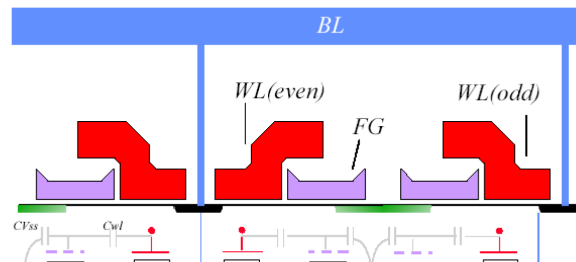


Fig.2.2 (a) Equivalent circuit of split-gate Flash cell, (b) Cell array schematic



	Vwl	Vbl	Vss	Vsb
PGM	1.8	Vdp	7~9V	0
ERA	12~13V	0	0	0
Read	~2.5	~1V	0	0

*Vdp = Vd @ Idp=2.5~5uA

1. **Faster programming (100X) efficiency by "source-side injection".**
2. **Better erase efficiency by "field-enhanced" FN tunneling.**
3. **No over-erase concerns because of select-gate structure.**
4. **One shot PGM and ERA, no verification CKT is needed.**
   1. **less over-head CKT → smaller die size.**
   2. **Easier design → good for embedded application.**

Fig.2.3 Bias condition and advantage of split-gate Flash cell.

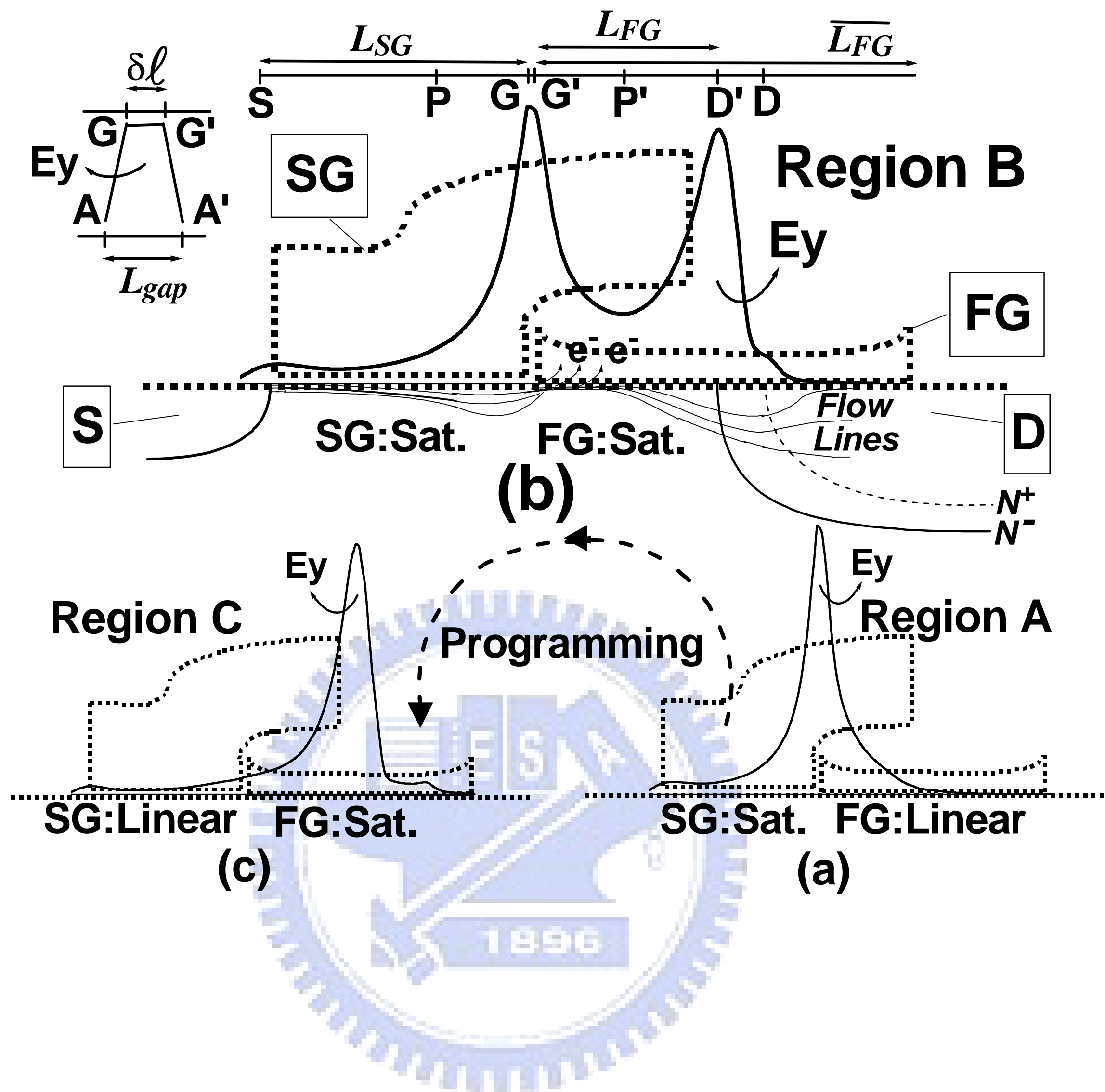


Fig.2.4 Simulation results of the lateral field distribution along channel surface for (a) *region-A* (b) *region-B* (c) *region-C*. The enlarged schematic cross-section with electron flow lines during programming is shown in (b). In the inset of (b), point A (or A') marks the physical boundary between SG (or FG) and IPO. Point G (or G') marks the position of the maximum lateral field. The device parameters are: $L_{SG} = 0.3$ μm, $L_{FG} = 0.183$ μm, $\overline{L_{FG}} = 0.4$ μm, $T_{SGOX} = 180$ Å, $T_{FGOX} = 100$ Å [6]

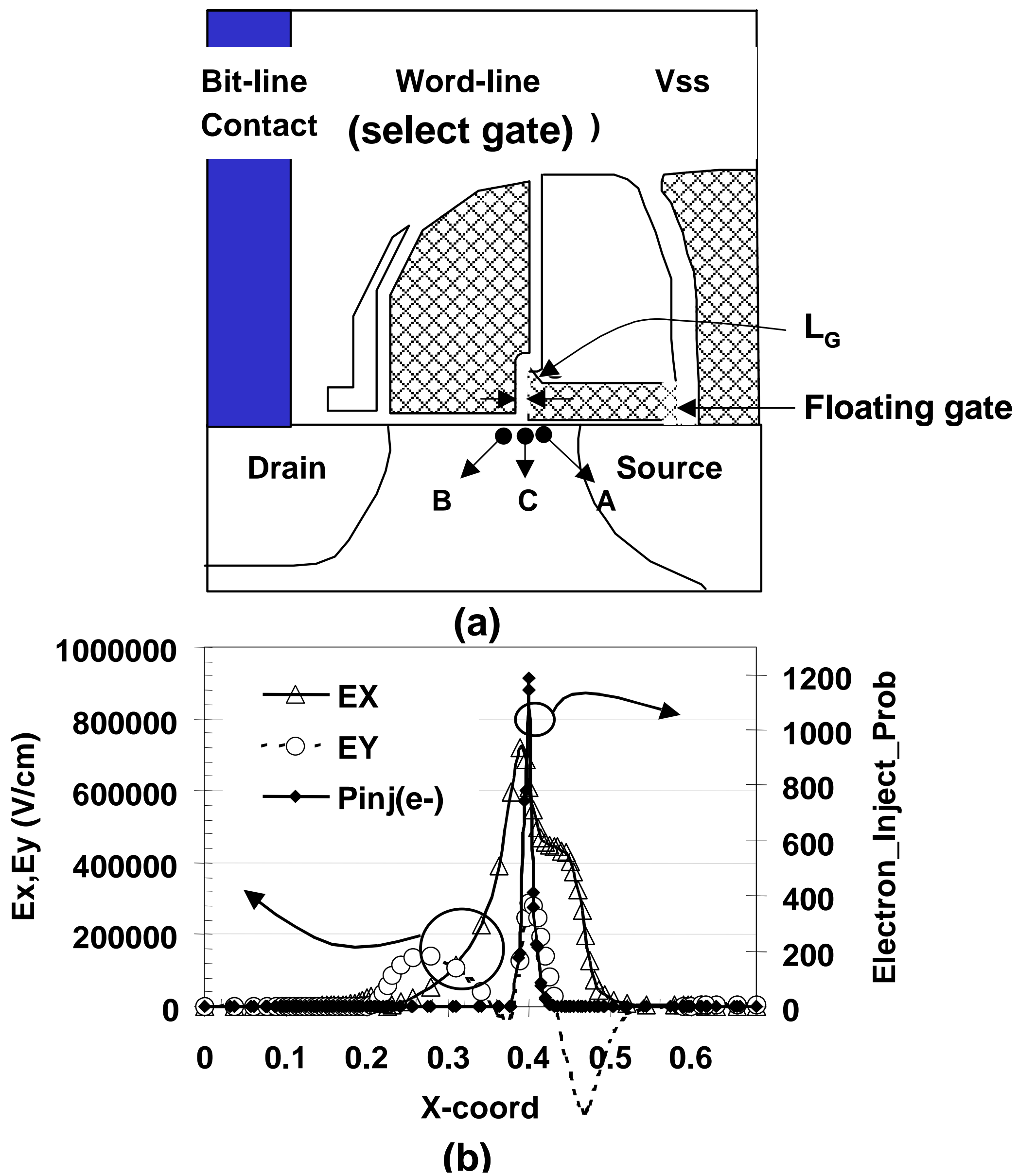
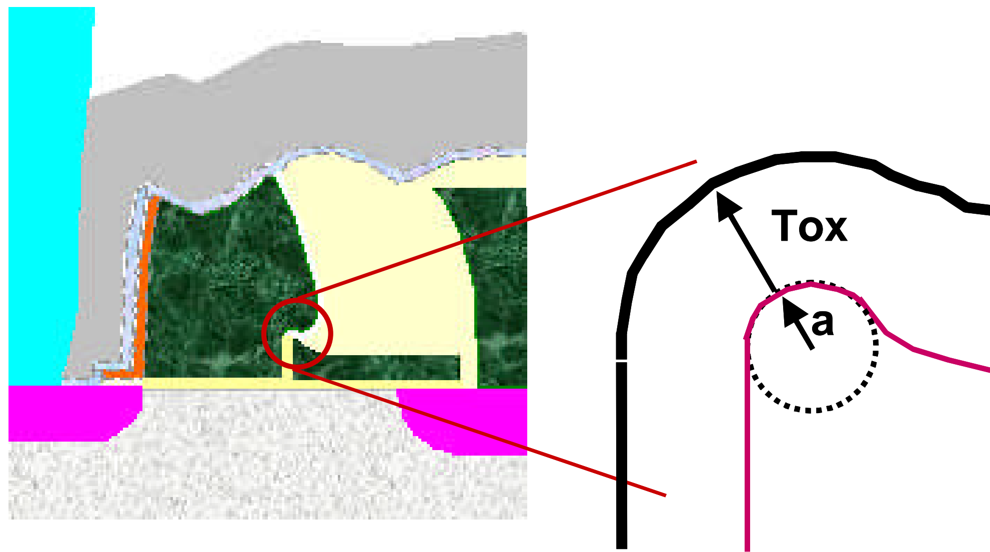
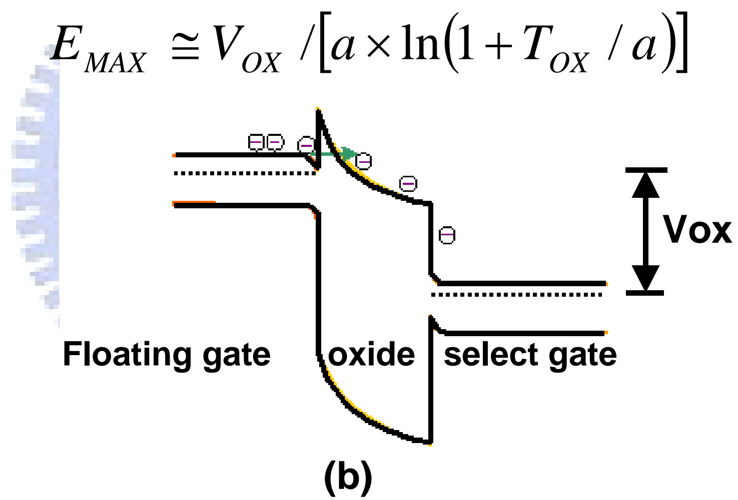


Fig.2.5 (a) Pinch-off point of FG and SG is A and B, respectively. Electron injection point is C, where is located in floating gate edge on the poly space. (b) Electric field and electron injection probability distribution.



(a)

$$E_{MAX} \cong V_{OX} / \left[a \times \ln\left(1 + T_{OX} / a\right) \right]$$



(b)

Fig.2.6 (a) Magnification of poly tip for tunneling, (b) energy band diagram during erasing, Emax is near the poly tip due to field-enhanced effect.

.

Reference

[1] S. Kianian, A. Levi, D. Lee, and Y.-W. Hu, “A novel 3 volts-only, small sector erase, high density Flash E^2PROM,” in *Symp. VLSI Technol. Dig.,* 1994, p. 71-72.

[2] B. Yeh, “Single transistor non-volatile electrically alterable semiconductor memory device,” *United States Patent* 5,029,130, 1991.

[3] W.T. Chu, H.H. Lin, C.T.Hsieh, H.C. Sung, Y.H. Wang, Y.T. Lin, C.S. Wang, “Shrinkable Triple Self-Aligned Field-Enhanced Split-Gate Flash Memory,” *IEEE Trans. Electron Devices*, vol. 51 no. 10, pp. 1667-1671, 2004

[4] R. Mih, J.Harrington, K. Houlihan, H.K. Lee, K. Chan, J. Johnson, B. Chen, J. Yan, A. Schmidt, C. Gruensfelder, K. Kim, D. Shum, C. Lo, D. Lee, A. Levi, and C. Lam, “0.18um Modular Triple Self-Aligned Embedded Split-Gate Flash Memory,” in *Symp. VLSI Technol. Dig., 2000*, pp. 120-121.

[5] “SuperFlash EEPROM Technology,” *SST Technical Paper*, November, 2001

[6] Y.H.Wang, M.C. Wu, C.J. Lin, W.T.Chu, Y.T. Lin, C.S. Wang, K.Y. Cheng, “An Analytical Programming Model for the Drain-Coupling Source-side Injection Split Gate Flash EEPROM,” *IEEE Trans. Electron Devices*, vol. 52, no. 3, Mar 2005, pp. 385-391

[7] H.S.Wang, “Gate current injection in MOSFET’s with a split-gate (virtual drain) structure “ IEEE ELECTRON Devices Lett., vol. EDL-14, pp.262-264, May 1993

[8] P.K.Ko, “ Hot-electron effect in MOSFET” Ph.D. dissertation, Univ. California, Berkeley, 1982

[9] H. Guan, D. Lee and G. P. Li, “An analytical model for optimization of programming efficiency and uniformity of split gate source-side injection superflash memory” *IEEE Trans. Electron Devices*, vol. 50, pp. 809–815, Mar. 2003.

[10]] J. V. Houdt, G. Groeseneken and H. E. Maes, “An analytical model for the optimization of source-side injection flash EEPROM devices” *IEEE Trans. Electron Devices*, vol. 42, pp. 1314–1320, July 1995

[11] J. V. Houdt, P. Heremans, L. Deferm, G. Groeseneken and H. E. Maes, “Analysis of the Enhanced Hot-Electron Injection in Split-Gate Transistors Useful for EEPROM Applications” *IEEE Trans. Electron Devices*, vol. 39, pp. 1150–1156, May 1992

[12] C. Hu, S. Tam, F.C. Hsu, P.K. Ko, T.Y. Chan and K.W. Terill, “ Hot-electron-induced MOSFET degradation – model, monitor and improvement,” IEEE Trans. Electron Devices, vol. ED-32, Feb. 1985, pp.375-384

[13] S. Tam, P. K. Ko and C. Hu, “Lucky-electron model of channel hot-electron injection in MOSFET’s” *IEEE Trans. Electron Devices*, vol. 31, pp. 1116–1125, Sep. 1984.

[14]] T. H. Ning, C. M. Osburn, and H. N. Yu, “Emission probability of hot electrons from silicon into silicon dioxides” *J. Appl. Phys.*, vol. 48, pp. 286–293, Jan. 1977.

[15] A. Kotov, A. Levi, Yu. Tkachev, and V. Markov, “Tunneling phenomenon in superflash cell” *Proc. NVM Tech. Symp.,* 2002. pp. 110-115

Chapter 3
Novel Program vs Disturb Window Characterization for Split-Gate Flash Cell

3.1 Introduction

In this chapter, a new methodology for program vs. disturb window characterization on split gate flash cell is presented. The window can be graphically illustrated in V_{WL} (word-line)-V_{SS} (source) domain under a given program current. This method can let us understand quantitatively how the window shifts vs. bias conditions and help us to find the optimal program condition. The condition obtained by this method can have the largest operation window in programming and disturb. This methodology has been successfully implemented in the development for 0.18um triple self-aligned (SA3) split-gate technology and beyond.

The split-gate flash memory proposed by Silicon Storage Technology Inc. (SST) is commonly used in stand-alone and embedded non-volatile memory because of the advantages of fast erase speed, high programming efficiency, and most important, no verification after program and erase. The erase is achieved by field-enhanced F-N tunneling through sharp poly tip, and the program is accomplished by source-side hot carrier injection (SSI) [1][2]. Although the above mechanisms provide good physics foundation for fast erase and program, a robust characterization methodology for determining operation condition is still crucial to guarantee reliable one shot program and erase. For erase condition, the window characterization is more straightforward because word-line voltage is the only parameter. Generally, higher word-line voltage yields faster erase speed and better cycling performance [3]. The upper limit of erase voltage is constrained by the reliability of word lines and high

voltage transistors. The optimization of erase voltage is typically determined by the trade-off between HV circuit reliability and endurance cycling. However, the program window characterization is much more difficult due to its complex bias conditions for both selected and un-selected cells. A good program condition requires not only fast program efficiency on selected cells but also very limited disturb effect on un-selected cells. The variables involved in the programming are programming current, word-line voltage and source-line voltage, and the figure of merit is the programming speed and three types of disturb behaviors. It is not a simple task to find an optimal condition among those variables. In this paper, we successfully develop a new methodology for program vs. disturb window characterization. This method can convert the constraints on programming and disturb into a clear graphical illustration. We can quantify the programming window from the 2D graph and use the data to find the optimal program condition from single cell measurement.

3.2 Experiment

The flash memory cells used in this study were fabricated by 0.18 μm triple self-aligned (SA3) Split-gate Flash technology [4]. The graphical illustration of process flow is shown in in Fig.3.1 [4][5]. Firstly, floating gate oxide is grown prior to floating gate poly deposition. Next, shallow-trench isolation (STI) is formed to become the 1st self-alignment to floating gate. After memory well implantation and thick nitride deposition, the region for floating gate and source line is defined and opened. Then, thick TEOS spacers and source poly lines are formed to become the 2nd self-alignment, which is source-line to floating gate. After thick nitride is removed, floating gate is defined by TEOS hard mask, and HTO is deposited to act as tunneling oxide. Afterwards, word-line poly was deposited and etched to become spacer word-line, which is the 3rd self-alignment to floating gate. After this step, the Flash cell is finished. The standard logic process flow for source & drain implant and metallization will be

followed. The cross-sectional view of SEM and TEM pictures of the cell are shown in Fig. 3.2 [4].

3.3 Program vs. Disturb Window Characterization

As shown in Fig.3.3, the memory cells are arranged in a cross point array, using a word line and bit line for address location selection; thus, unselected cells within same page will suffer the stress from high V_{SS} voltage, and this is where the program disturb comes from. There are three types of possible program disturbs, and they will be described in the following paragraph. The cells in different page will not have program disturb because each sector is individually isolated. Each cell is only exposed to higher voltage within the selected page along row or source line; there is no high voltage on the bit line.

In triple self-aligned split-gate Flash, programming is operated at following conditions: Source node (V_{SS}) biased at high voltage, word-line (V_{WL}) slightly turned-on and bit-line connected to a constant current source (Idp). This program condition can cause three disturb stress modes; they are: Column Punch-through Disturb (PTC), Row Punch-through Disturb (PTR) and Reverse Tunneling disturb (RT) [6][7]. The disturbed bits' location and stress conditions are shown in Fig.3.3. The program time and disturb duration are determined by product spec and array architecture. In this paper, we use 32Mbit Flash as a target vehicle. The program time is 10us, and the disturb time for PTC, PTR and RT is 1ms, 40ms and 260ms, respectively. For the criteria in this characterization, the program spec is defined as "programmed state current (Ir0) smaller than 5% of erased state current (Ir1)", and the disturb spec is defined as "cell current drop ratio (ΔIr1/Ir1) smaller than 10% after program disturb".

For a given program current (Idp) and program time, the program vs. disturb window can be presented in the V_{WL} (word-line)-V_{SS} (source) domain. The plot is shown in Fig.3.4. A valid program condition must be enclosed by the following five boundary conditions to guarantee fast programming and a very limited disturb behavior. The detail discussion is

shown below. Note that the V_{SS} voltage is constrained at 9V to prevent device form damage.

- **Curve 1: Programming.**

The curve meets Ir0 = 5% of Ir1 after 10us programming. As shown in Fig.3.5, the lower right side of curve 1 (higher V_{SS} and lower V_{WL}) is the region satisfying program specifications. Based on the eqs. (7) & (8) shown in Chap.2, we can find that a higher lateral and vertical field can be induced by higher V_{SS} and lower $V_{SG}(=V_{WL})$, as a result, the programming is better in this region.

$$E_X \cong (\alpha V_{SS} - V_{SG}) / k \cdot L_G$$

$$E_{OX} \cong (\alpha V_{SS} - V_C) / T_{OX}$$

- **Curve 2: Column punch-through (PTC) disturb.**

The curve meets ΔIr1/Ir1 = 10% after a 1 ms disturbance. The upper left side of the curve 2 (lower V_{SS} and higher V_{WL}) is the region satisfying the PTC disturbance specifications. The bias condition and the disturb trend is shown in Fig.3.6. The source and bit lines of PTC are the same as those of programming-selected bits, while the word lines of PTC are grounded to turn current off. In a real situation, a small amount of leakage can still exist and cause undesired injection under a high source-to-drain voltage drop. The disturb will get worse when the source-to-drain voltage drop get higher. The higher V_{SS} can cause larger V_{DS} is obvious. However, the effect of V_{WL} on PTC is through the modulation of constant current programming circuitry. To maintain constant current programming, Vdp will be lower when the V_{WL} in selected cell becomes lower, and thus the source-to-drain voltage drop in PTC cell get larger. Therefore, the region with higher V_{SS} and lower V_{WL} will have larger source-to-drain voltage drop and will be easier to get PTC disturb. As shown in the illustration of the PTC disturb trend on the V_{SS}-V_{WL} domain, the upper left side of the curve 2 (lower V_{SS} and higher V_{WL}) is the region satisfying the PTC disturbance specifications.

- **Curve 3: Row punch-through (PTR) disturb**

The curve meets ΔIr1/Ir1 = 10% after a 40 ms disturbance. PTR occurs on the same word

line of program-selected bits, so the word line is turned on and the source line is biased at a high voltage. The bias condition and the PTR trend is is shown at Fig.3.7. To prevent undesired programming on the erased cell, an inhibited voltage (~2 V at the worst case) is applied to unselected bit lines to stop the leakage flowing from the source to the drain. However, when the back bias is not strong enough to shut off the leakage, the undesired programming will occur. Thus, PTR tends to occur under higher V_{WL} and V_{SS} voltages. As a presentation of this trend on the V_{SS}-V_{WL} domain, the bottom left side of curve 3 (lower V_{WL} and lower V_{SS}) is the region satisfying the PTR disturbance specifications.

- **Curve 4: Reverse tunneling (RT) disturb**

The curve meets ΔIr1/Ir1 = 10% after a 260ms disturbance. RT occurs on the bits at the condition that the source line is connected to a high voltage, word line is grounded and bit line is biased under an inhibited voltage. This is the most minor disturb mode in split-gate Flash. The bias condition and RT disturb trend is shown in Fig.3.8. The disturb mechanism is solely caused by reverse tunneling from the word line to the floating gate and is only dependent on V_{SS} voltage. Higher the V_{SS} voltage, worse the reverse tunneling disturb. Therefore, the left side of the curve is the region satisfying the RT disturbance specifications. This disturb mode can be totally eliminated if a good process is chosen to suppress FG undercut during floating gate etching. In this characterization, the V_{SS} is clamped at 9V to prevent device from damage.

- **Curve 5: Drain voltage during programming (Vdp)=0V**

The lower boundary of the program vs disturb window is enclosed by Vdp=0 V. Beyond this boundary, Vdp will have a negative voltage, which is not allowed in standard split-gate Flash design.

3.4 Application of the Window Characterization

3.4.1 Finding optimal program condition

To find an optimal program condition from the program/disturb window characterization, a maximum circle, named as operation circle, is drawn within the enclosed window, as shown in Fig. 3.10. The circle center can be chosen as the best program condition for a given Idp. The reason is that the circle center has largest voltage variation allowance for V_{WL} and V_{SS}. Note that the scale at X and Y axis is kept the same.

Next step is to find the best Idp setting. As shown in Fig.3.11 (a), we can see that the operation window changes with Idp. When Idp is lowered from 5μA to 1μA, the window shifts to the higher V_{SS} because higher V_{SS} is needed to compensate the programmability loss caused by lower Idp. On the other hand, when Idp is increased from 5μA to 9μA, the window shifts toward lower V_{SS} and high V_{WL} (weaker programming region) because the programmability is enhanced by the higher Idp. Comparing the maximum circle size between Idp=1, 5, and 9μA in Fig.3.9 (b), we can find that Idp=5μA has the largest circle size, which means 5μA is the best program current setting. Therefore, we choose the center of operation circle at Idp=5μA as the best program condition. The bias condition is: V_{SS} =7.2V, V_{WL} =1.8V at Idp=5μA. This methodology was successfully implemented in 0.18um SA3 single cell characterization to determine program setting for 32Mbit-product.

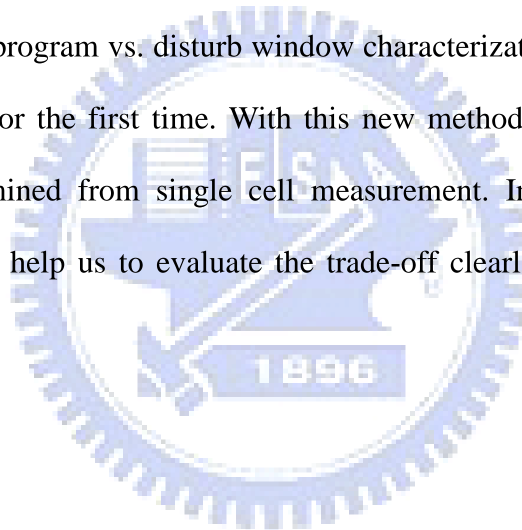
3.4.2 Constant voltage vs. constant current programming

For ETOX stack gate Flash cell, constant voltage programming is commonly used [8], while for super Flash Flash technology the constant current method is chosen to improved the operation window. Using the characterization method developed in this thesis, we can compare the operation window difference between these two program schemes. The circuit diagram is shown in Fig.3.12. We check the overlap window for constant current programming with Idp varying from 1-9uA, and the one for constant voltage programming

with Vdp varying from 0 to 0.6V. As shown in Fig. 3.13, we can find the constant current programming has much larger window than the constant voltage method. In constant voltage programming, the PTC disturb is very serious when Vdp ~0V, and the programmability is poor in low VWL when Vdp is ~0.6V, so the overlap window is very small. Under constant current programming, Vdp will vary to supply the stable programming current and the back bias effect can suppress the punchthrough from happening.

2.5 Summary

A new methodology for program vs. disturb window characterization in split gate flash cell is presented in this paper for the first time. With this new methodology, the optimal program condition can be determined from single cell measurement. In addition, the quantitative window information can help us to evaluate the trade-off clearly between various program settings.



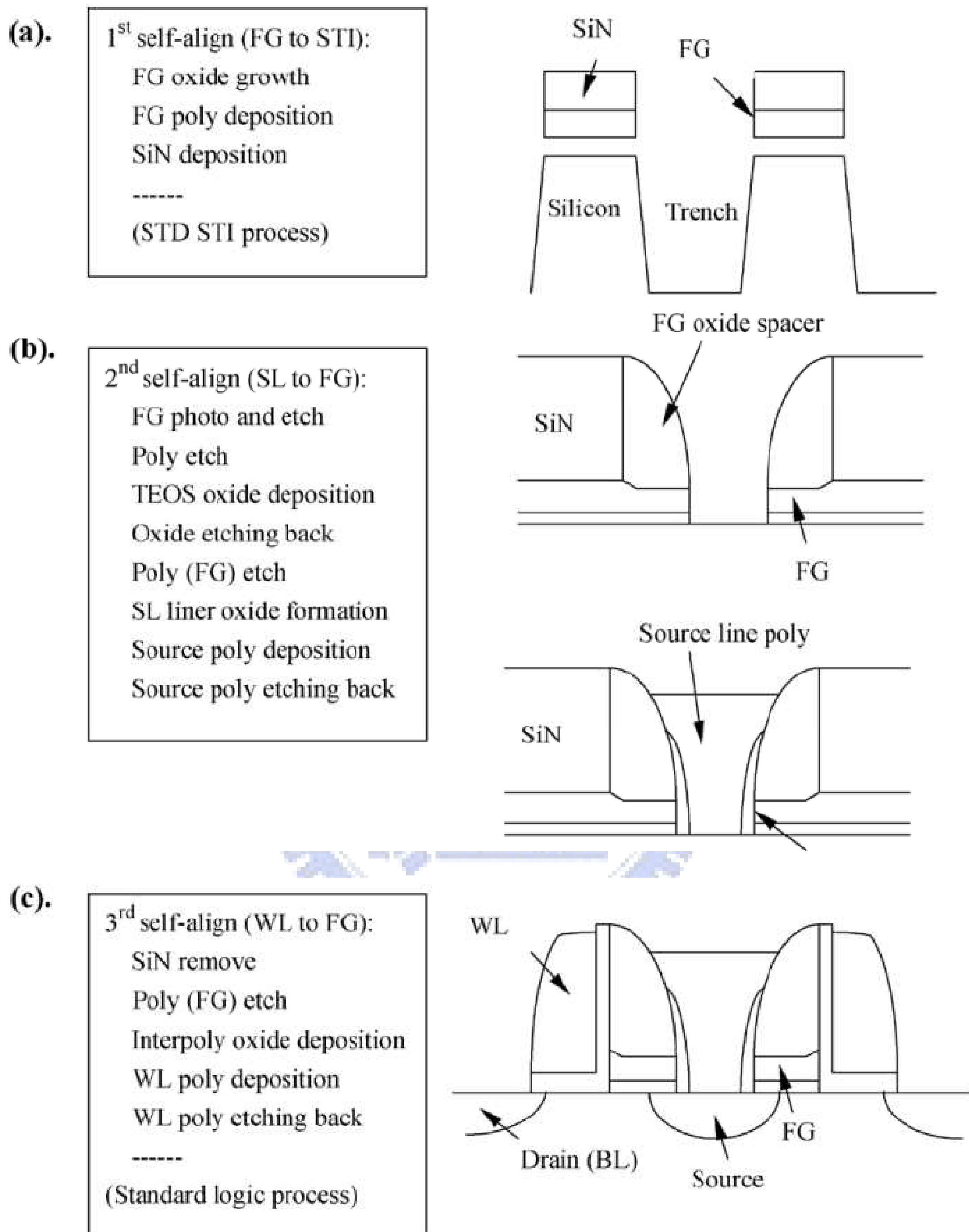
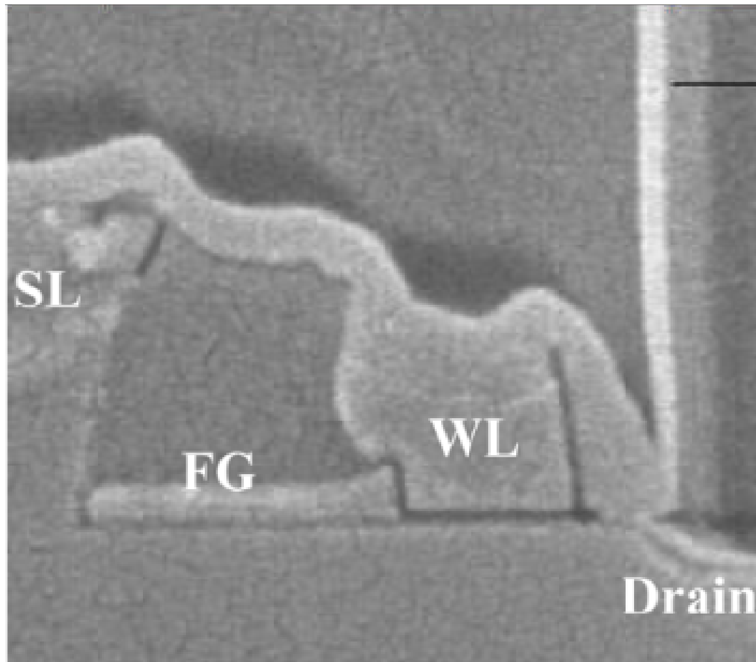
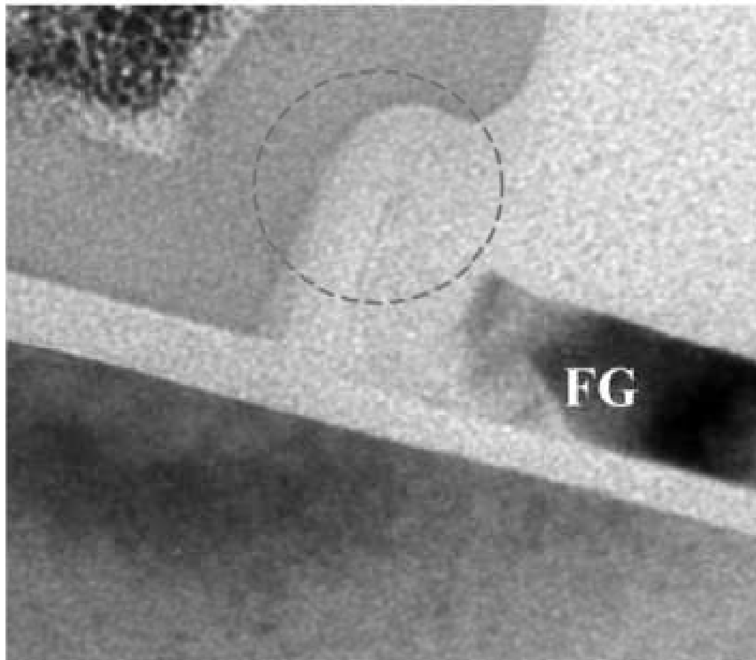


Fig.3.1 (a)-(c) Cross section of he triple self-aligned split-gate Flash cell process sequence.



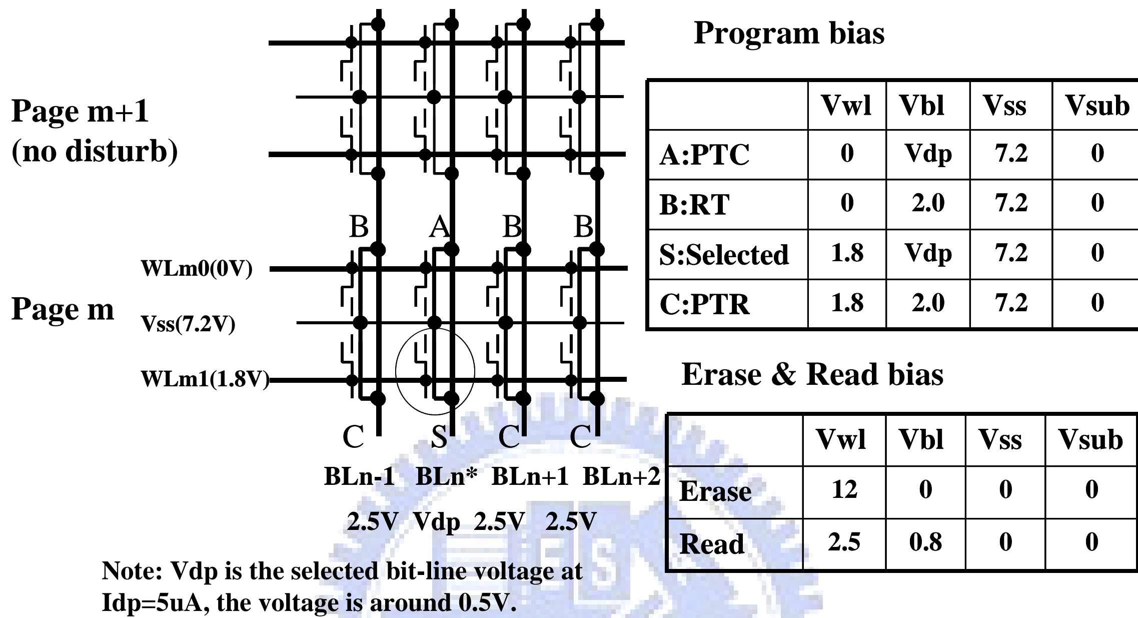
(a)



(b)

Fig.3.2. (a) SEM picture of triple self-aligned split-gate cell, (b) TEM picture for a sharp FG corner (indicated by the circle) created by poly etch. [4]

Keep blank



Program bias

	Vwl	Vbl	Vss	Vsub
A:PTC	0	Vdp	7.2	0
B:RT	0	2.0	7.2	0
S:Selected	1.8	Vdp	7.2	0
C:PTR	1.8	2.0	7.2	0

Erase & Read bias

	Vwl	Vbl	Vss	Vsub
Erase	12	0	0	0
Read	2.5	0.8	0	0

Fig.3.3 Cell array and bias voltage for program, erase, read-out and three disturb conditions, which are: A. Column punchthrough disturb(PTC), B. Row punchthrough disturb(PTR), C. Reverse tunneling disturb(RT). Note that the cells outside the selected page are immune from disturb stress.

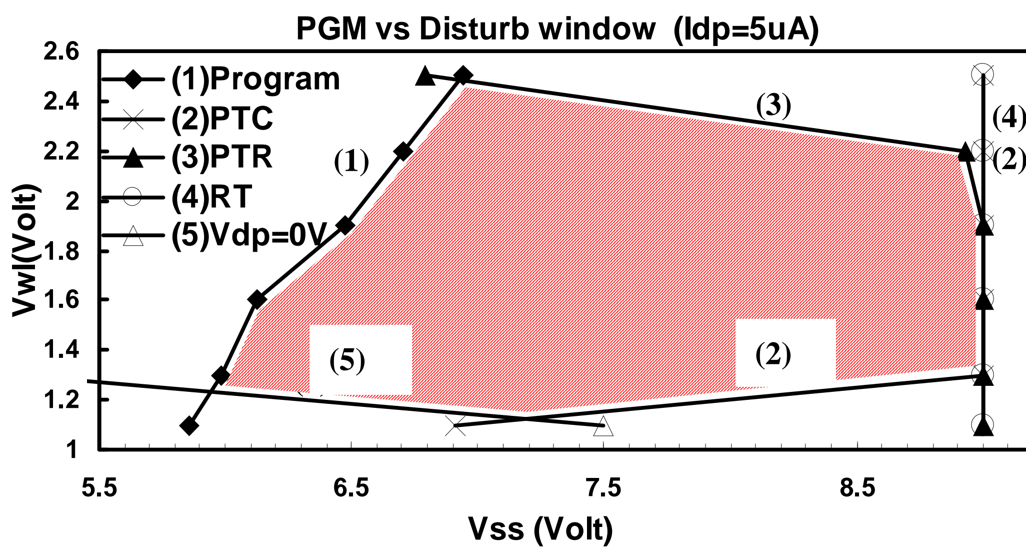


Fig.3.4 Program vs disturb window and the operation window. The programming time is 10us and program current is 5μA

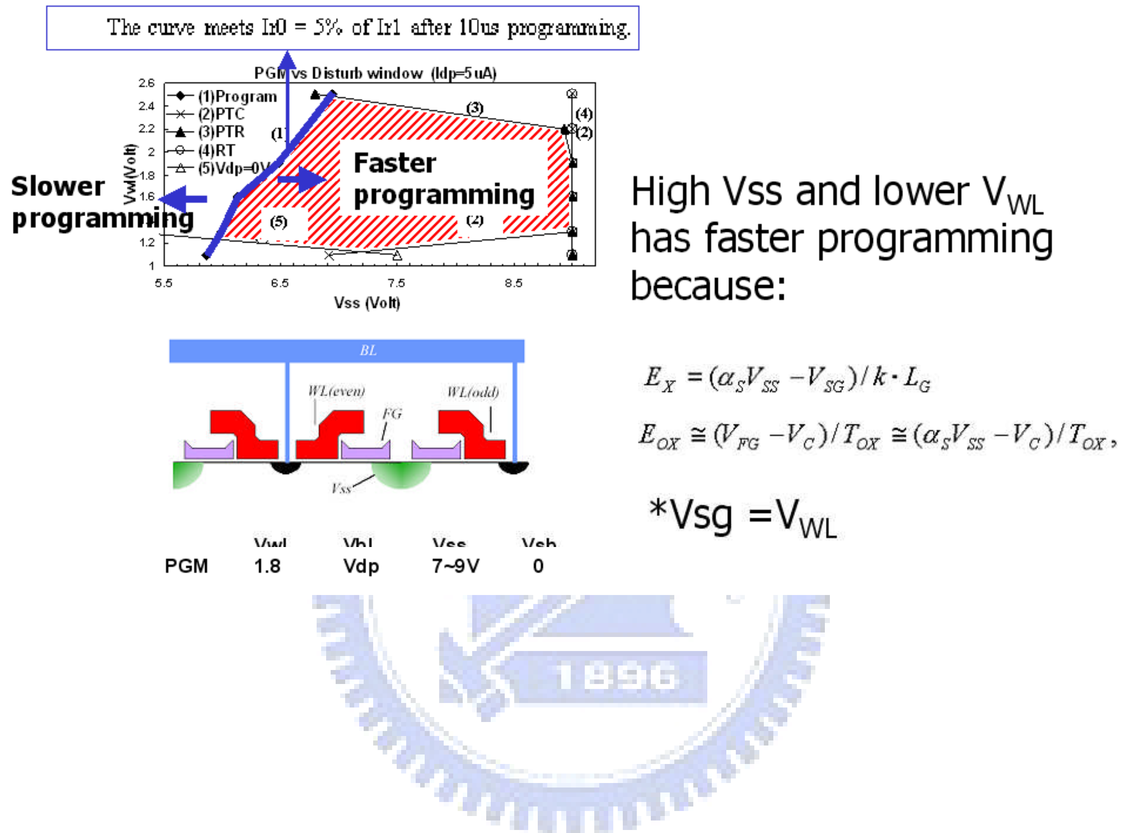


Fig. 3.5 Program trend in PGM vs Disturb window

Curve-2: PTC (Column) Disturb Trend

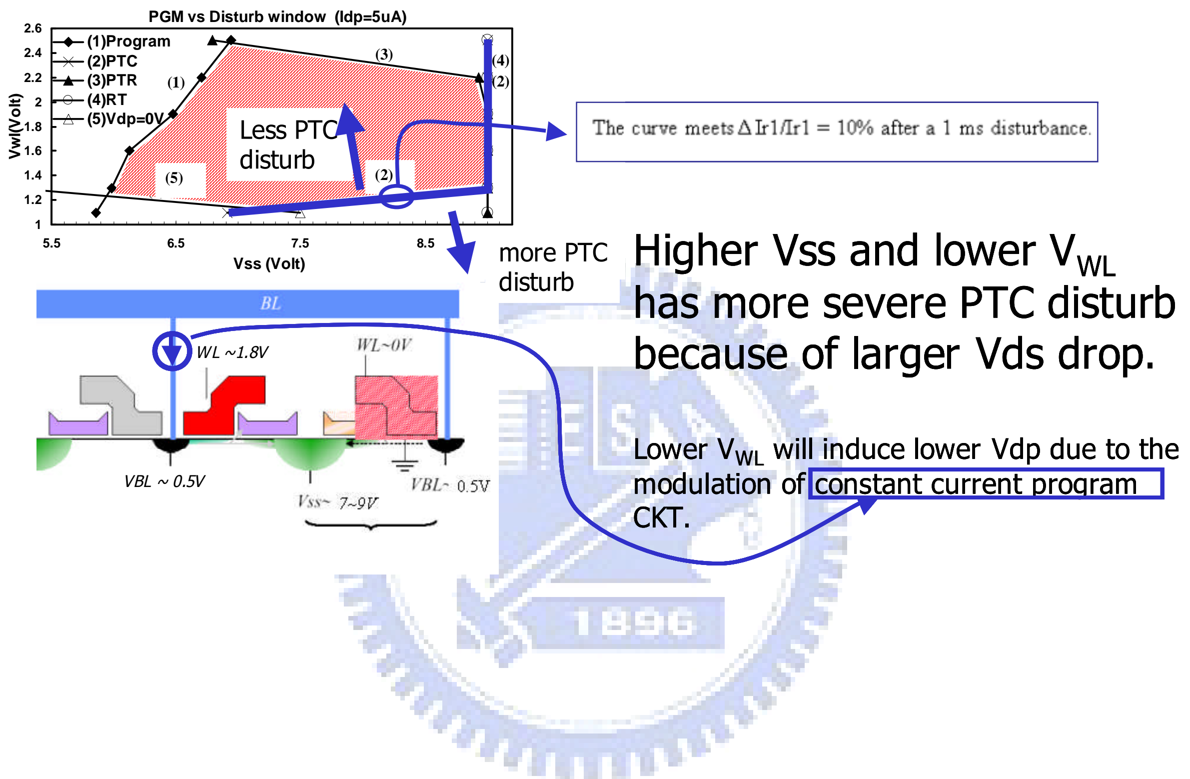
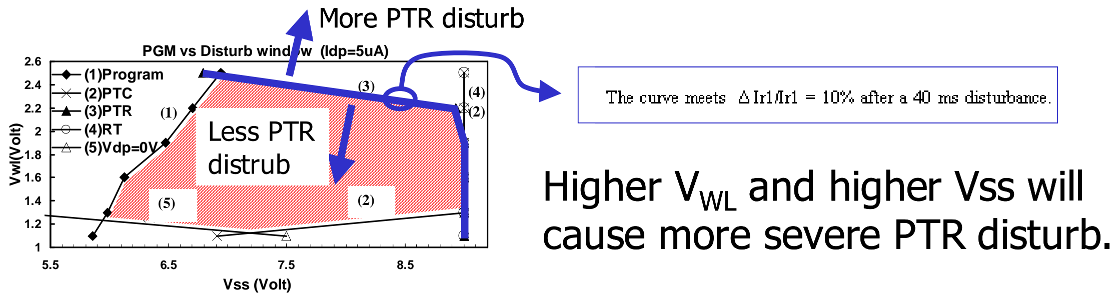


Fig. 3.6 Bias condition for Column punch-through (PTC) and the disturb trend.

Curve-3: PTR (Row) Disturb Trend



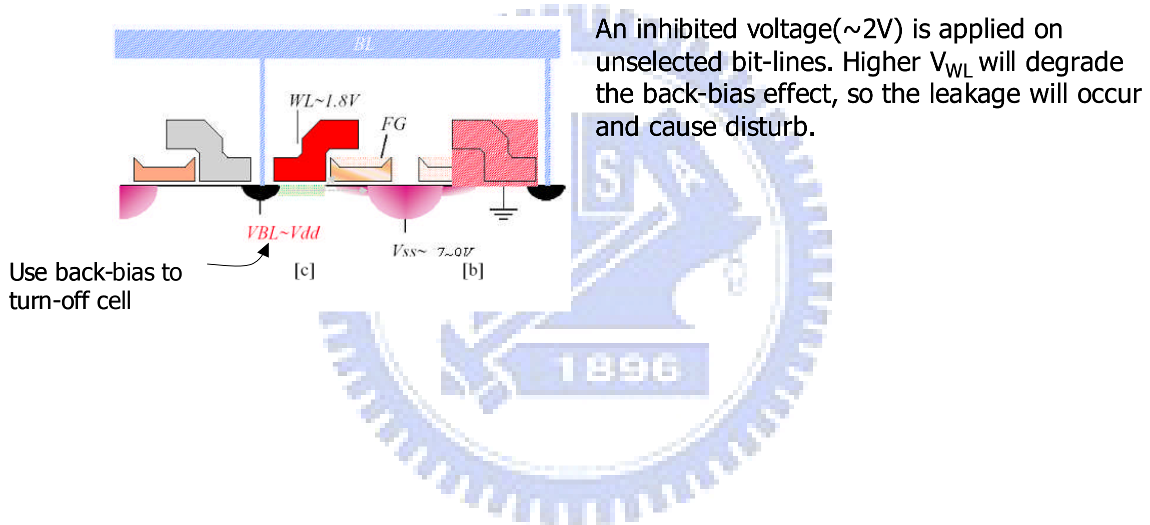
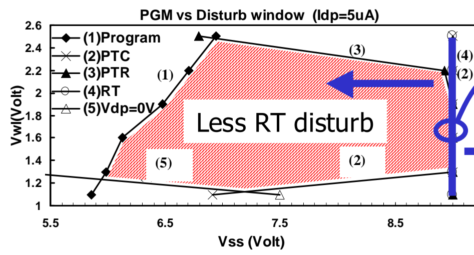


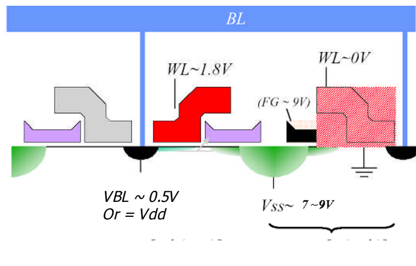
Fig.3.7 Bias condition for Row punch-through (PTR) and the disturb trend

Curve-4: RT (Reverse Tunneling) Disturb Trend



The spec is ΔIr1/Ir1 < 10% after a 260ms disturbance. In this characterization, it is limited at 9V.

More RT disturb

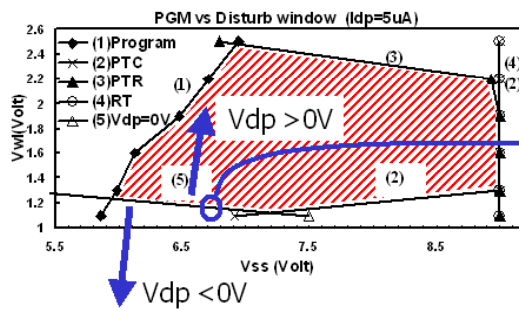


Higher Vss will cause more severe RT disturb.

The disturb mechanism is solely caused by reverse tunneling from the word line to the floating gate and is only dependent on VSS voltage.

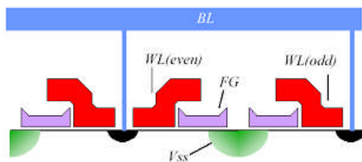
Fig.3.8 Bias condition of Reverse tunneling (RT) and the disturb trend

Curve 5: Drain voltage during programming (Vdp)=0V



The curve meets Vdp=0V

The lower boundary of the program vs disturb window is enclosed by Vdp=0 V. Beyond this boundary, Vdp will have a negative voltage, which is not allowed in the split-gate Flash design.



	Vwl	Vbl	Vss	Vsb
PGM	1.8	Vdp	7~9V	0

Fig. 3.9 Trend of Vdp in operation window plot

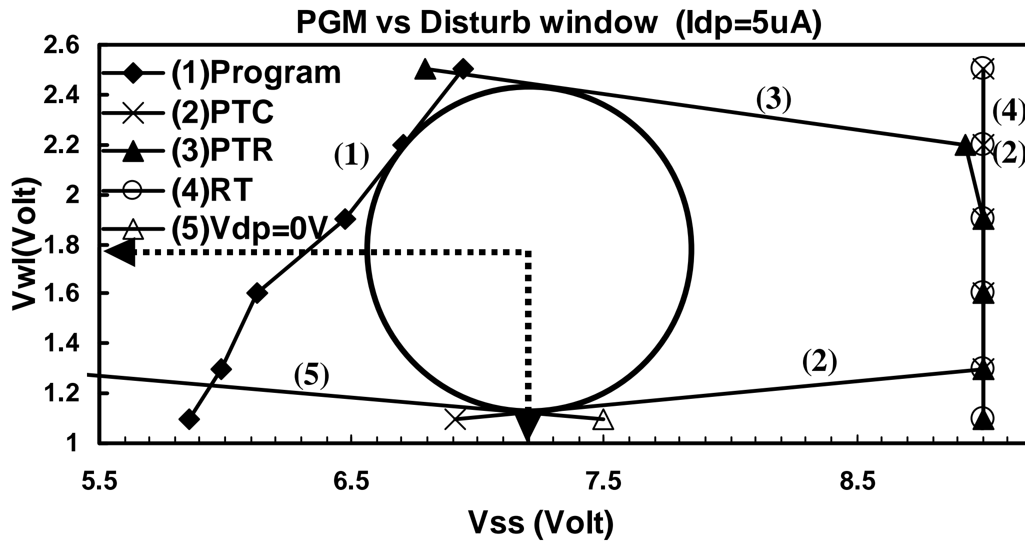
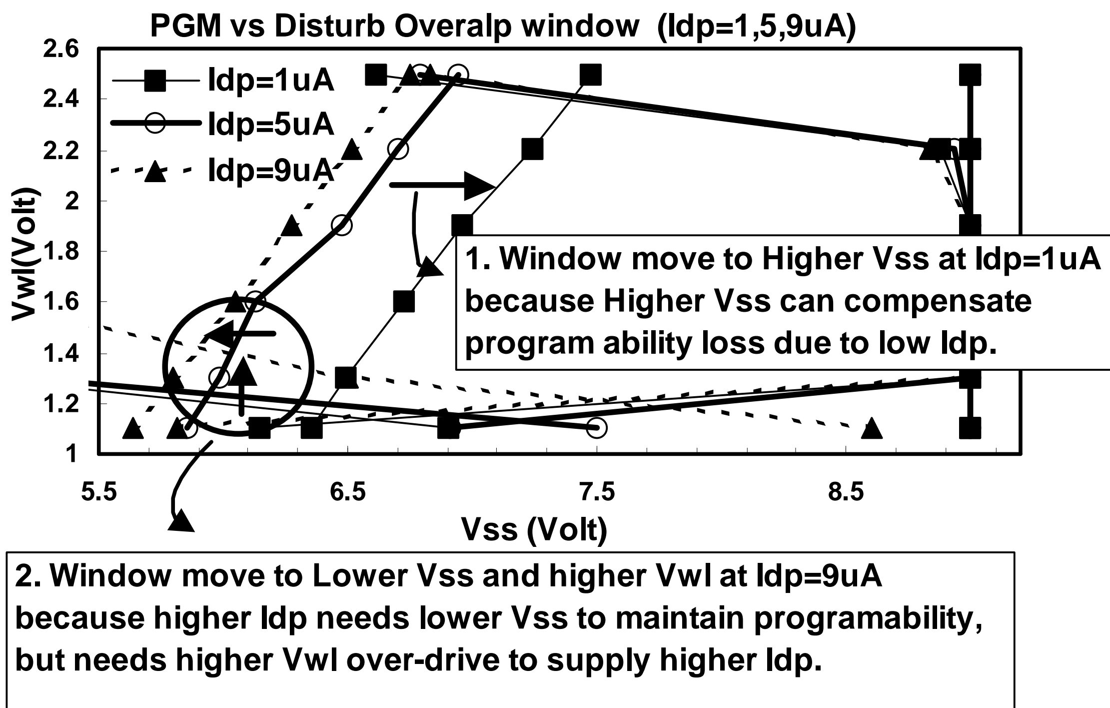
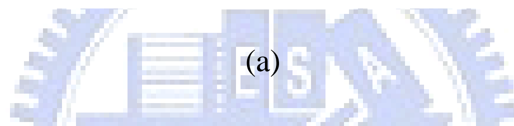
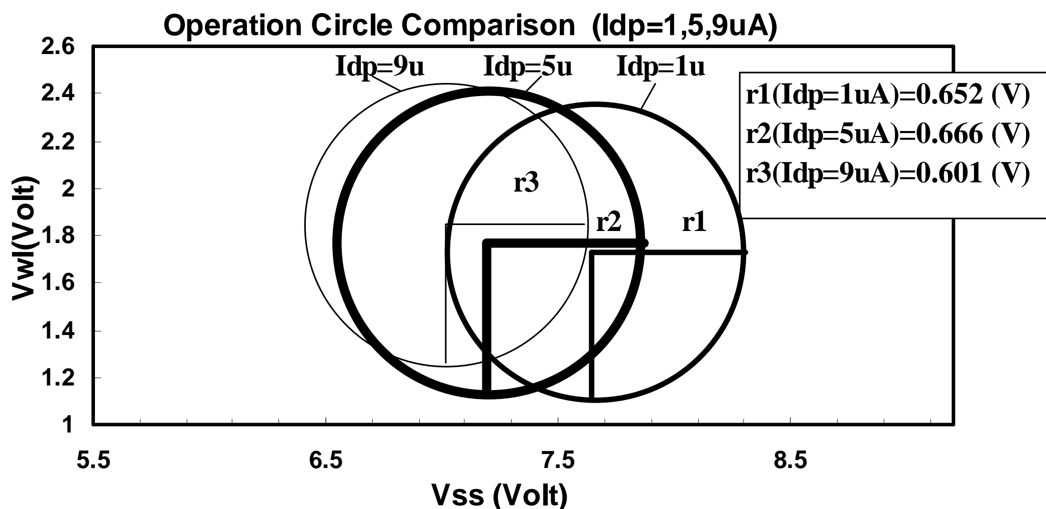


Fig.3.10 Operation circle in program and disturb window. The optimal program condition is at circle center, the condition is VSS=7.2V, Vg=1.75V, Idp=5μA.





(a)



(b)

Fig.3.11 (a) Program vs disturb window varies with Idp from 1, 5 to 9uA. Since the channel doping is well adjusted in this SA3 cell, no significant disturb boundary shift is observed under Idp variation (b) Operation circle comparison between Idp=1,5,9 μA. The optimal program condition is chosen at r2 center (Idp=5μA) because it has largest operation circle, the condition is V_{SS}=7.2V, V_{WL}=1.75V, Idp=5μA.

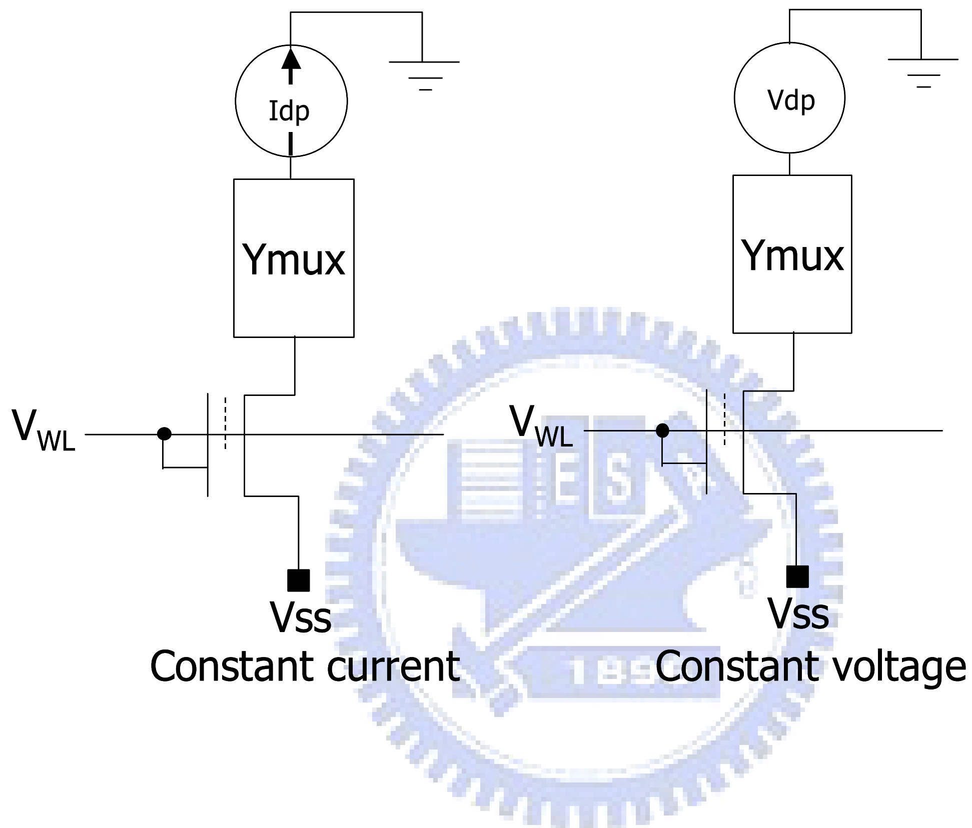
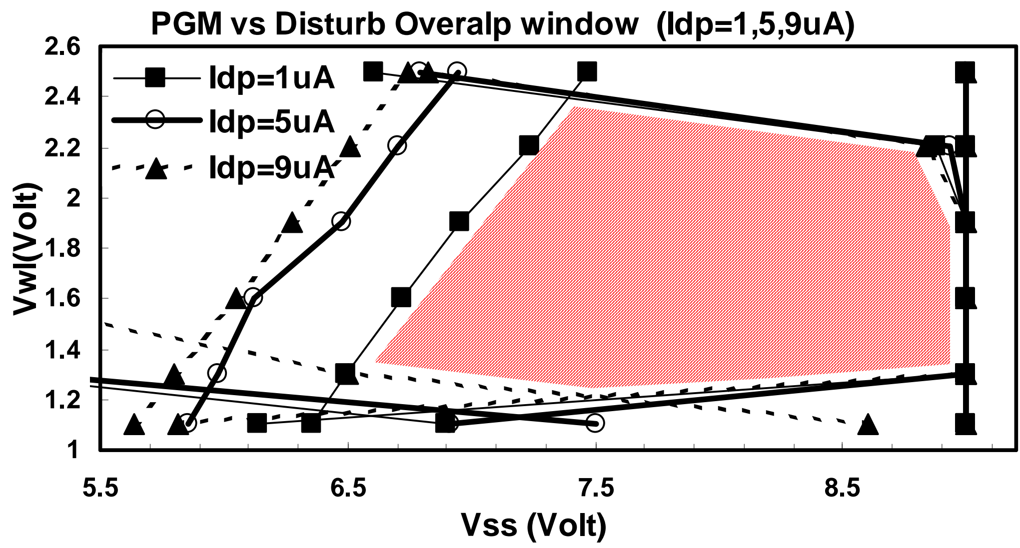
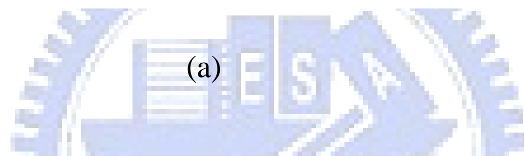
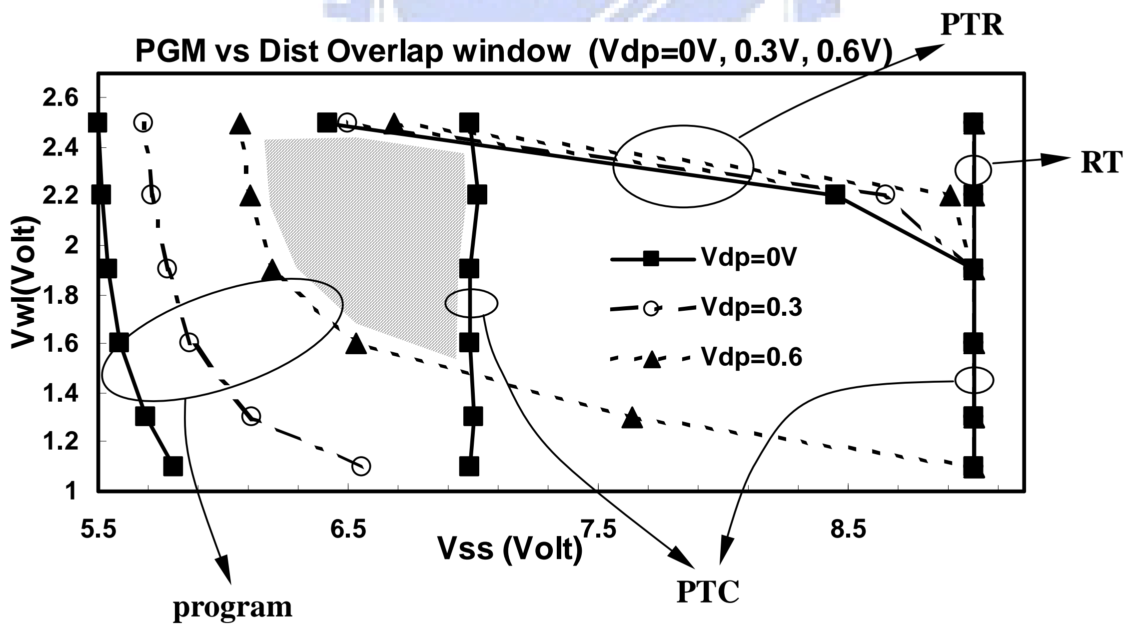


Fig.3.12 The circuit diagram for constant current and constant voltage programming.



(a)





(b)

Fig. 3.13. (a) The overlap window for constant current programming. The Idp varies from 1uA to 9uA, (b) The overlap window for constant voltage programming. The Vdp varies from 0V to 0.6V.

Reference

[1] S. Kianian, A. Levi, D. Lee, and Y.-W. Hu, "A novel 3 volts-only, small sector erase, high density Flash E^2PROM," in *Symp. VLSI Technol. Dig.,* 1994, p. 71-72.

[2] B. Yeh, "Single transistor non-volatile electrically alterable semiconductor memory device," *United States Patent* 5,029,130, 1991.

[3] K.C. Huang, Y.K. Fang, D.N. Yaung, C.W. Chen, H.C. Sung, D.-S. Kuo, C.S. Wang, "The impacts of control gate voltage on the cycling endurance of split gate flash memory," *IEEE Electron Device Lett.,* vol. 21, p. 359-361, July 2000.

[4] W.T. Chu, H.H. Lin, C.T.Hsieh, H.C. Sung, Y.H. Wang, Y.T. Lin, C.S. Wang, "Shrinkable Triple Self-Aligned Field-Enhanced Split-Gate Flash Memory," *IEEE Trans. Electron Devices*, vol. 51 no. 10, pp. 1667-1671, 2004

[5] R. Mih, J.Harrington, K. Houlihan, H.K. Lee, K. Chan, J. Johnson, B. Chen, J. Yan, A. Schmidt, C. Gruensfelder, K. Kim, D. Shum, C. Lo, D. Lee, A. Levi, and C. Lam, "0.18um Modular Triple Self-Aligned Embedded Split-Gate Flash Memory," in *Symp. VLSI Technol. Dig., 2000*, pp. 120-121.

[6] M.G. Mohammad, K.K. Saluja, "Electrical model for program disturb faults in non-volatile memories," *VLSI Design,* 2003. Proceedings. 16th International Conference, pp. 217 – 222, 2003

[7] D. Wellekens, J. Van Houdt, L. Haspeslagh, J. Tsouhlarakis, P. Hendrickx, L. Deferm, H.E. Maes, "Embedded HIMOS(R) flash memory in 0.35 um and 0.25 um CMOS technologies," *IEEE Transactions on Electron Devices*, vol. 47, no. 11, pp. 2153-2160, 2000

[8] P. Cappelletti, C. Golla, P. Olivo, E. Zanoni. (1999) *Flash Memory*. pp. 317-319.

Chap.4

New Triple Self-aligned (SA3) Split-Gate Flash Cell with T-Shaped Source Coupling

4.1 Introduction

The split-gate flash technology proposed by Silicon Storage Technology Inc. (SST) is commonly used in stand-alone and embedded nonvolatile memory because of the advantages of its high efficient program/erase performance and one-shot program/erase algorithm [1][2]. The erasing is performed by "field-enhanced Flower-Nordheim (F-N) tunneling" through a sharp poly tip, which allows erasing at a lower voltage and a higher speed than its stack-gate counterpart [3]. For programming, "source-side hot carrier injection" (SSI) is used, which can have an electron injection efficiency two orders of magnitude larger than that of conventional channel hot electron injection [4]. In addition, its 1.5 transistor (T) cell structure is immune to over-erasing, so no program/erase verification circuitry is needed [5][6]. As a result, better performance can be achieved by a smaller and simpler overhead circuitry because the program/erase voltage is lower and no complex state machine for verification is required. Therefore, the split-gate technology is very attractive for high-performance portable systems and also suitable for embedded silicon-on-chip (SOC) applications [7][8].

However, the major issues of split-gate flash technology in deep-submicron generation are that the cell size is difficult to scale down due to its 1.5T structure and that the operation voltage is high for the program/erase function. In 1998, a triple self-aligned (SA3) structure was proposed, the cell size is reduced by eliminating the redundant overlap and misalignment margin, and the operation voltage is also lowered due to the reduction of WL to FG coupling [9][10]. The comparison between traditional non-self-aligned cell vs SA3 cell is shown in Fig.4.2. Despites the advantage of SA3 cell, the split-gate cell scaling is still limited by a large source (Vss)-to-floating gate (FG) overlap, which is used to provide a sufficient source coupling to the floating gate for electron injection. With such a large overlap, a large floating-gate length is necessary to prevent punchthrough induced by the high source

voltage during programming; thus, scaling down the cell is still difficult [11].

In general, there are two ways to overcome the scaling obstacles. One is to use a thinner oxide or a high-K material for the coupling dielectric to increase coupling [12], the other is to rearrange the FG/Vss overlap to increase coupling without increasing cell size. For the first approach, it is limited by retention concerns, after many years' effort, the industry still stay with the thermal grown tunnel oxide. Therefore, the 2nd approach with FG/Vss rearrangement is a more practical choice.

In this chapter, we proposed a T-shaped Vss poly structure to gain additional FG-Vss coupling on the side and the top of floating gate. In the following sections, we will describe the device fabrication and discuss how the program/erase performance is improved by using the new structure. Finally, a 16Mbit test vehicle is used to demonstrate the sort-1 and sort-2 yield performance of this new approach.

4.2 Device Fabrication

The device fabrication sequence is shown in Fig. 4.2. The process begins with coupling oxide growth and floating gate poly deposition. Next, shallow-trench isolation (STI) is formed to become the first self-alignment; i.e., STI to floating gate. After memory well implantation, a thick nitride was deposited and then patterned by FG photolithography and etching, which opens the region for the floating gate and source line. Then, a thick oxide spacer is formed by chemical vapor deposition (CVD) and etching. The exposed FG region is later removed by dry etching. Afterward, three new steps are implemented to form the additional Vss overlap on FG: oxide spacer pull-back etching, source liner oxide deposition and Vss poly spacer deposition/etching. The additional coupling between Vss and FG is controlled by the amount of oxide spacer pull-back etching and source liner oxide thickness. Afterward, a thick Vss poly is deposited and polished to form a source line. After this step, the Vss poly line is self-aligned to FG. This is the second self-alignment. Then, the thick nitride is removed using hot phosphoric acid and the exposed poly region is removed by dry etching. Next, the tunneling oxide is deposited, and the word-line poly is deposited and etched to form a poly spacer, which becomes the third self-alignment to the FG. The cross-sectional profile comparison between a conventional cell and the new SA3 cell is shown in Fig. 4.3(a), and a TEM image of the new cell is shown in Fig. 4.3(b). The cell size is close to 0.35 to 0.4 μm^2 and the technology node is 0.18

μm.

4.3 Array Bias Condition

The array bias condition is shown in Fig. 4.4. During programming, V_{SS} is biased at about 7.2 V and the bit line is round 0.5V, which connected to a contact current circuitry, and the select-gate (SG) is slightly turned on. For erasing, a 12V high voltage is applied on word-line to perform FN erasing. During read, the word-line voltage is connect to ~2.5V and bit-line is biased around 0.8V.

4.4 The SCR Effect on Program and Erase

The program and erase mechanisms for the SST's split-gate flash cell has been described in Chap.2. In the following sessions, special focus will be given on why source coupling ratio (SCR) has a significant effect on both mechanisms.

4.4.1. Programming

From the discussion in Chap.2, we derive the E_X and E_{OX} as shown in the following equations. (eqs.7&8 in Chap.2)

$$E_X \cong (\alpha V_{SS} - V_{SG}) / k \cdot L_G$$

$$E_{OX} \cong (\alpha V_{SS} - V_C) / T_{OX}$$

From above equations, we can find that the the SCR (α_S) has a linear effect on the a large longitudinal electric field (Ex) and vertical oxide electric field (E_{OX}).

On the basis of the lucky-electron model (LEM), the injection current equation is shown below. The linear effect of SCR on E_X and E_{OX} will turn into exponential effect on injection current.

$$I_{FG} = K \times Is \left(\frac{\lambda n Ex}{\Phi_b} \right)^2 Exp\left(-\Phi_b / \lambda Ex \right),$$

The simulation result shown in Fig.4.5, it indicates that SCR is the most important factor influencing electron injection probability; 10% increase on SCR can improve electron injection probability by 300%.

4.4.2 Field- enhanced F-N tunneling

As shown in eqs (9)-(11) in Chap2, using a cylindrical approximation, the electric field is highest (E_{MAX}) at the inner edge and lowest (E_{MIN}) at the outer edge.

$$E_{MAX} \cong V_{OX} / [a \times \ln(1 + T_{OX} / a)], \quad (7)$$

where a is the radius of curvature of the smaller cylinder, and

$$V_{OX} = V_{SG} - V_{FG} = V_{SG} - (\alpha_G V_{SG} + \alpha_S V_{SS}). \quad (8)$$

Since $V_{SS} = 0$ and $\alpha_G + \alpha_S \approx 1$ during erasing, eq. (11) can be simplified to

$$V_{OX} = \alpha_S V_{SG}. \quad (9)$$

Simply stated, a higher α_S (SCR) can result in a higher voltage drop across the inter-poly oxide. Based on the Fowler-Nordheim tunneling equation, the α_S (SCR) will also have exponential effect on the injection efficiency.

From the above discussion, we can understand the improvement on α_S (SCR) will have significant advantage on the programming and erasing performance for split-gate Flash. The improvement is very important for lower voltage operation and cell size scaling.

4.5 Application to Voltage Reduction

As was revealed in section 4.2, we can modulate the SCR by varying the etching time for FG oxide spacer pull-back. The result in Fig. 4.6 shows that more pull-back etching can induce a higher source coupling. Next, by measuring the cells at different etching times, we can obtain the relationship between SCR and program/erase performance [16]-[19]. As shown in Fig. 4.7(a) and (b), a higher SCR can result in a more efficient program/erase performance as we predicted in the previous section. The criteria for evaluating the program/erase performance is described in the following: (a) The programmability is characterized as Ir0/Ir1@Vs=6V, 10us with program current=3uA, where Ir0 and Ir1 is programmed and erased current, respectively. (b) The erasing performance is characterized by Verase, which is the voltage to reach 50% Ir1 after 10ms erasing.

Since programming and erasing can be improved by a higher SCR, we can perform the functions with a lower voltage, which is important for low-Vcc operation. To determine the program condition for the T-shaped SA3 cell, we used the new

program-disturb window characterization methodology described in Chap.3 [20]. Comparing the program-disturb windows between conventional SA3 and new T-shaped coupling cell, shown in Figs. 4.8(a) and (b), we can find that the Vss voltage for programming can be reduced from 7.4 to 6.4 V in new cell. It is important to let the voltage lower than 6.5V because it can be handled by the 3.3V IO device, which is a requirred device in many application. The junction of IO devices typically can sustain up to 7V, so it can handle the 6.5V operation with proper cascade design. Using IO device for high voltage circuit design has several benefits: (1) Masking steps saving. About 4 masking step for well and LDD implant for traditional HV devices can be eliminated, (2) Area saving. The 3.3V design rule can be much tighter than the rule for >10V. Despite the low Vs voltage advantage, there is a minor side effect caused by the reverse-tunneling-disturbance (RT). The boundary drops from 9 V to 8 V in the new cell. The RT disturbance is caused by the undesired electron tunneling from WL to FG when FG is coupled to a high voltage during programming. The RT disturbance tends to occur under the conditions of a high Vss bias and a high SCR. In general, the degradation can be improved by FG profile optimization.

For the erase performance improvement by the new SA3 cell, we found that the Verase can be reduced by 0.5V as shown in Fig. 4.7(b). Basically, the erase voltage reduction is not as critical as the reduction in program voltage because of the two following reasons. First, the erase operation consumes much less power than the program operation; thus, the high-voltage design for erase is much easier. Second, the erase voltage can be greatly reduced by splitting the voltage into two polarities [8].

4.6 Application to Cell Size Reduction

For the traditional SA3 cell, a deep Vss junction is necessary to induce a sufficient coupling capacitance between Vss and FG. Thus, a large FG length is required to accommodate the large source junction and to prevent the punchthrough triggered by the high Vss voltage. With the help of T-shaped Vss coupling, the same SCR can be achieved with a shallower Vss junction. As a result, we can reduce FG length while keeping the same cell performance. As we can observe in Fig. 4.9, the yield of a 16 Mbit test vehicle with a shallower Vss junction and a T-shaped coupling structure remains stable when FG length is reduced, whereas the yield of the conventional SA3 cell drops markedly due to punchthrough-induced program failure.

Note that the nominal FG length is 0.18 μm. In addition, the data retention performance of the new cell remains as good as that in the case using the traditional approach. The data retention failure rate after 72 hrs baking at 250°C is less than 1% out of 3000 good die.

4.7 Summary

A triple self-aligned (SA3) split-gate flash cell with a T-shaped source coupling approach is described in this paper. This novel structure can significantly enhance the coupling capacitance between the source and the floating gate (SCR) with no complex process steps. This new structure can be applied to program voltage reduction and cell size scaling. For the program voltage reduction, the maximum program voltage of the new cell can be reduced from 7.4 to 6.4 V, which is characterized by a newly developed methodology for program-disturb window characterization. For erasing, the voltage reduction is about 0.5 V. Regarding cell size scaling, comparable sort-1 and sort-2 yields are demonstrated using the new cell with a smaller floating gate length and a shallower source junction. In summary, the new T-shaped source coupling approach is a novel, well controlled, and inexpensive approach, which does not need additional masking step. Therefore, it is a very attractive solution for the next-generation SA3 split-gate cell.

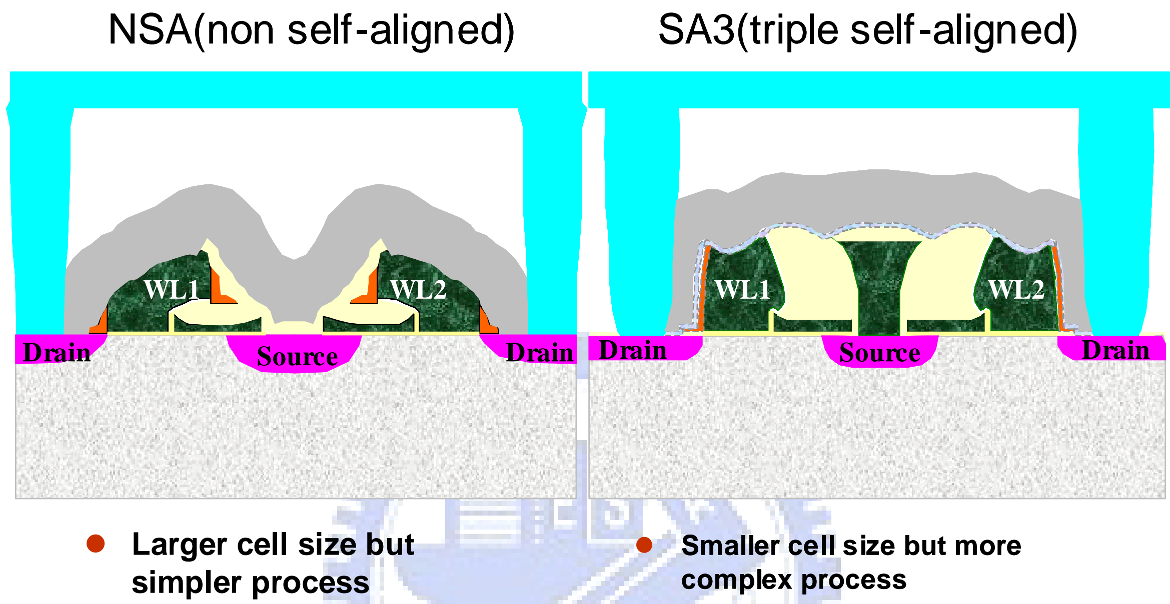
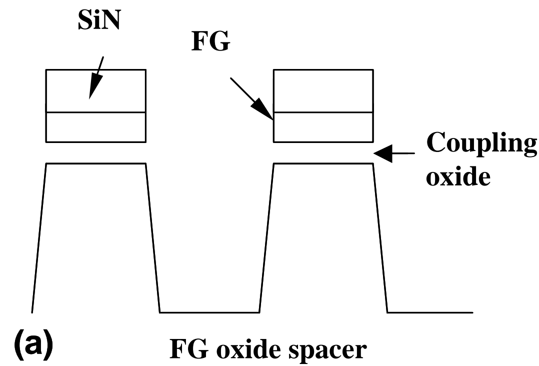


Fig.4.1 The cell cross-sectional view of tradition non-self-aligned cell vs. triple self-aligned(SA3) cell.

<u>I. 1st self-align (FG to STI):</u>
1. Coupling oxide oxide growth
2. FG poly deposition
3. SiN deposition
4. Standard STI process

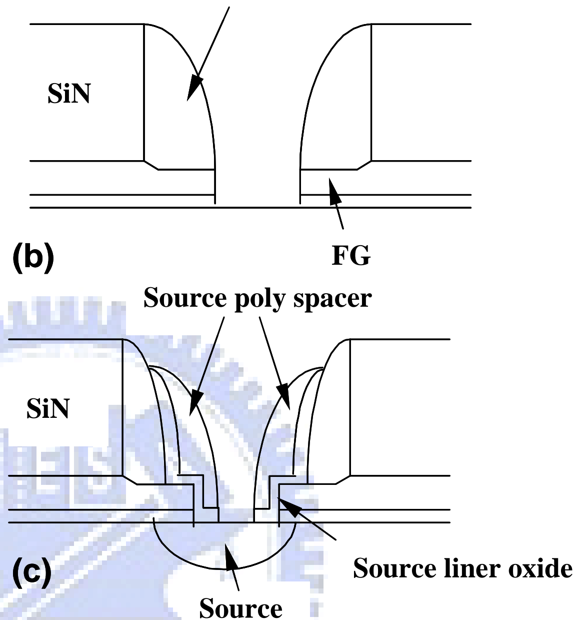


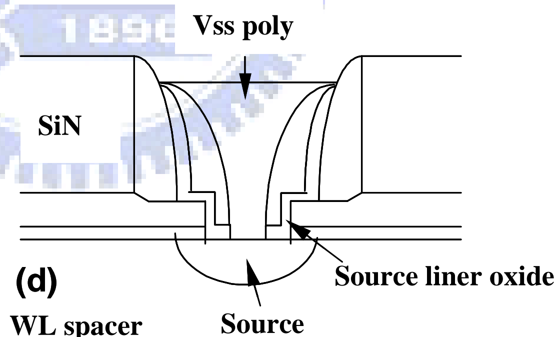
<u>II. 2nd self-align (SL to FG):</u>
1. FG photo and etch
2. Poly slope etch
3. FG oxide spacer formation
    -- oxide deposition/etching
4. Poly (FG) etch

5*. Oxide spacer pull-back etching
6*. Source liner oxide deposition
7*. Source poly spacer formation
    -- poly deposition/etch

8. Source implantation
9. Vss poly deposition and
    etching back

Steps 5-7 are the new process other than conventional SA3 process [10].





<u>III. 3rd self-align (WL to FG):</u>
1. SiN removal
2. FG poly etching
3. Inter-poly(tunnel oxide)
    deposition
4. WL poly spacer deposition/etch

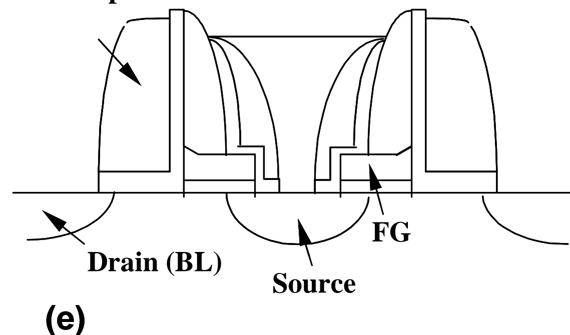
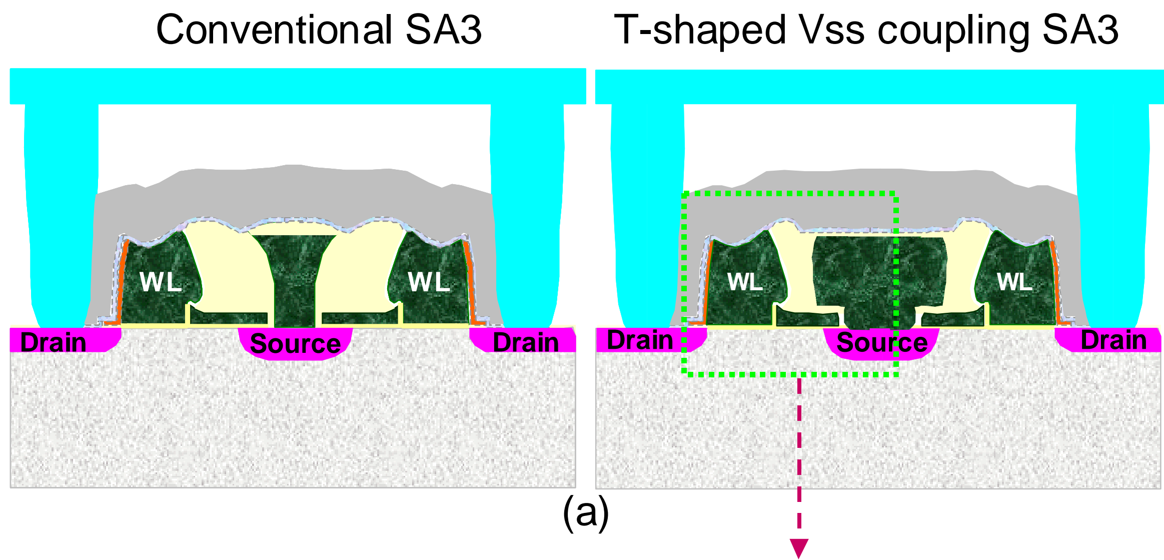
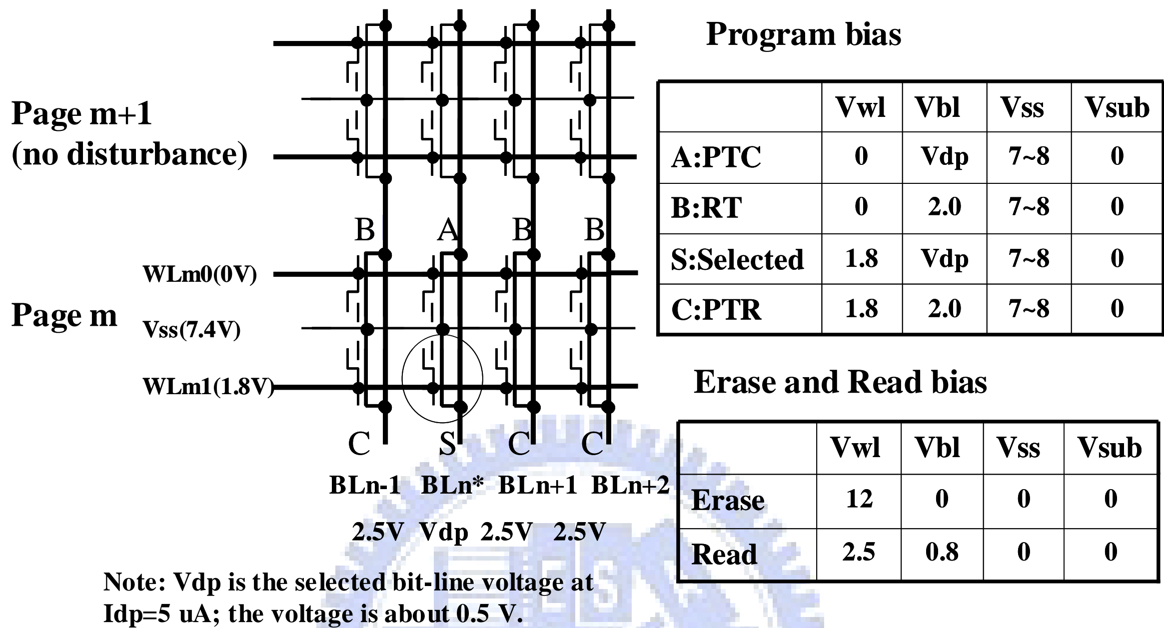


Fig.4.2. Process flow of the T-shaped Triple Self-aligned (SA3) split-gate flash cell.



(b)

Fig. 4.3 (a) Traditional SA3 vs. T-shaped Vss coupling SA3 cell. (b) TEM picture of the new cell. The cell size is 0.38 μm^2.



Program bias

	Vwl	Vbl	Vss	Vsub
A:PTC	0	Vdp	7~8	0
B:RT	0	2.0	7~8	0
S:Selected	1.8	Vdp	7~8	0
C:PTR	1.8	2.0	7~8	0

Erase and Read bias

	Vwl	Vbl	Vss	Vsub
Erase	12	0	0	0
Read	2.5	0.8	0	0

Fig. 4.4 Cell array and bias voltage for program, erase, read-out and three disturb conditions, which are: A. Column punchthrough disturb(PTC), B. Row punchthrough disturb(PTR), C. Reverse tunneling disturb(RT). Note that the cells outside the selected page are immune from disturb stress.

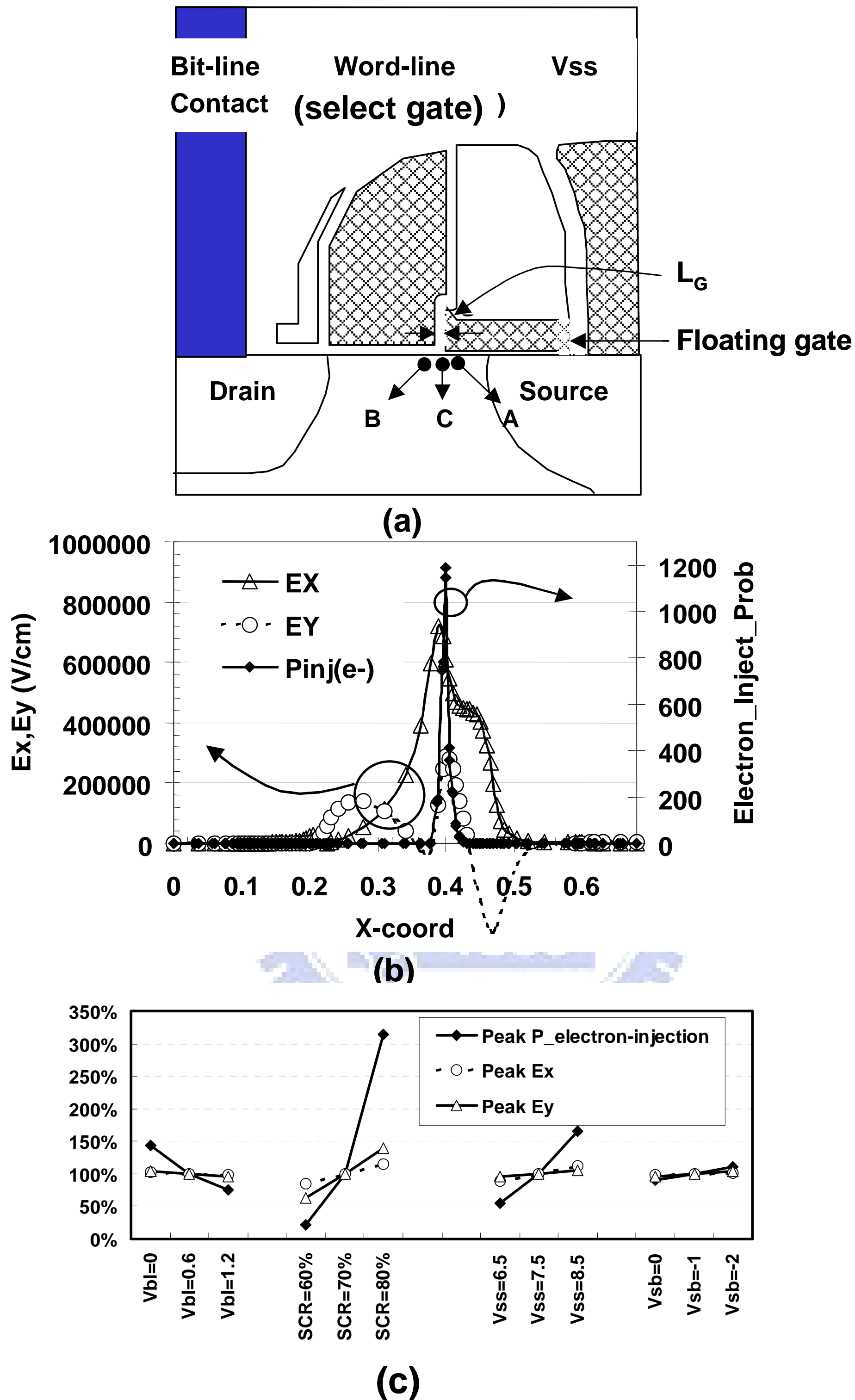


Fig.4.5 (a) Pinch-off point of FG and SG is A and B, respectively. Electron injection point is C, where is located in floating gate edge on the poly space. (b) Electric field and electron injection probability distribution (c) Plot of factors effect on the electron injection probability. The factors include source voltage (Vss), source coupling ratio (SCR), substrate bias (Vsb) and bit-line voltage(Vbl)

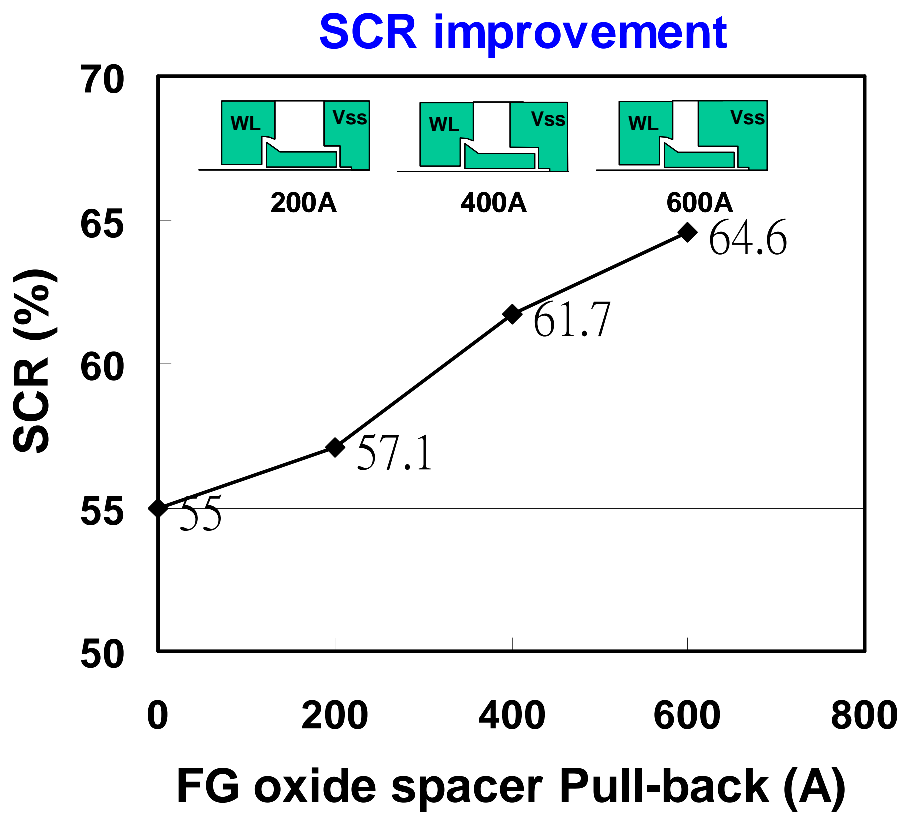
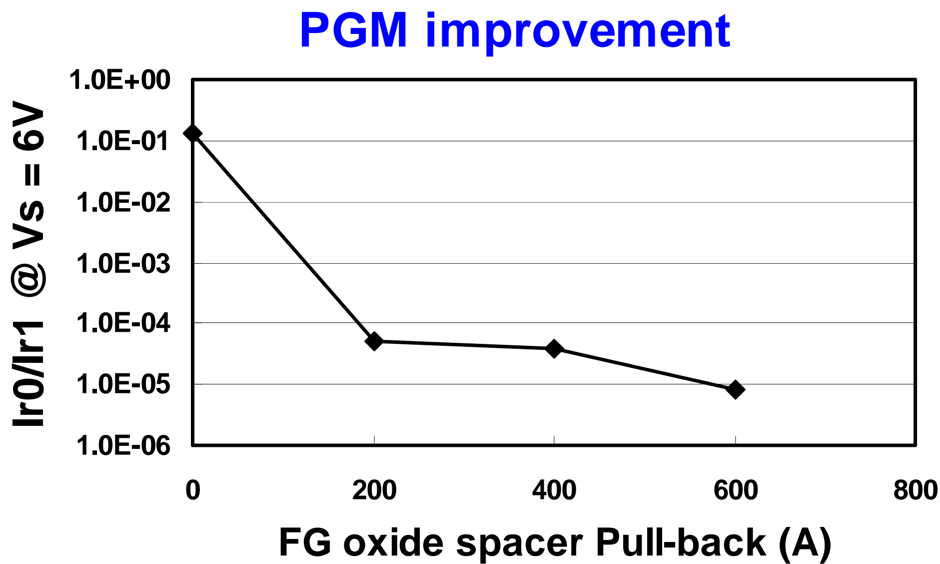
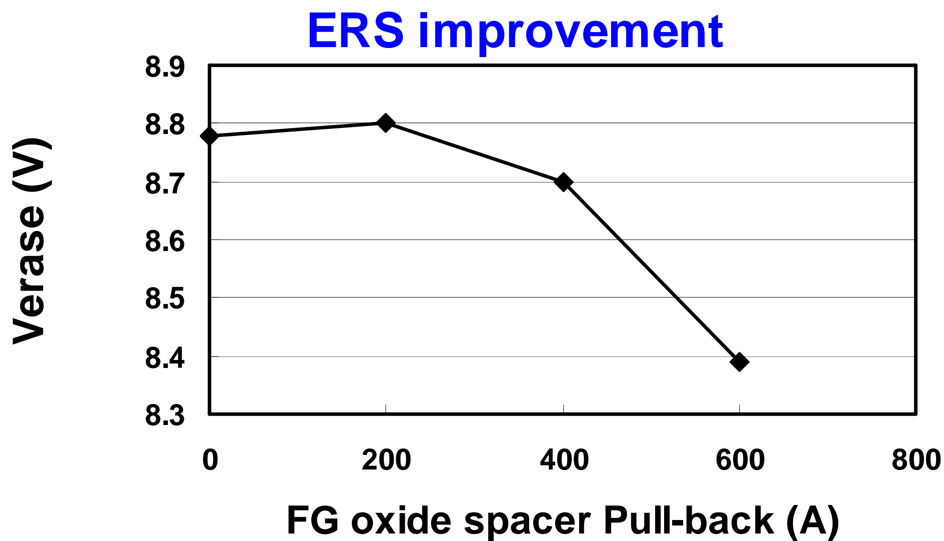


Fig.4.6 Source coupling ratio (SCR) vs. FG oxide spacer pull-back etching.

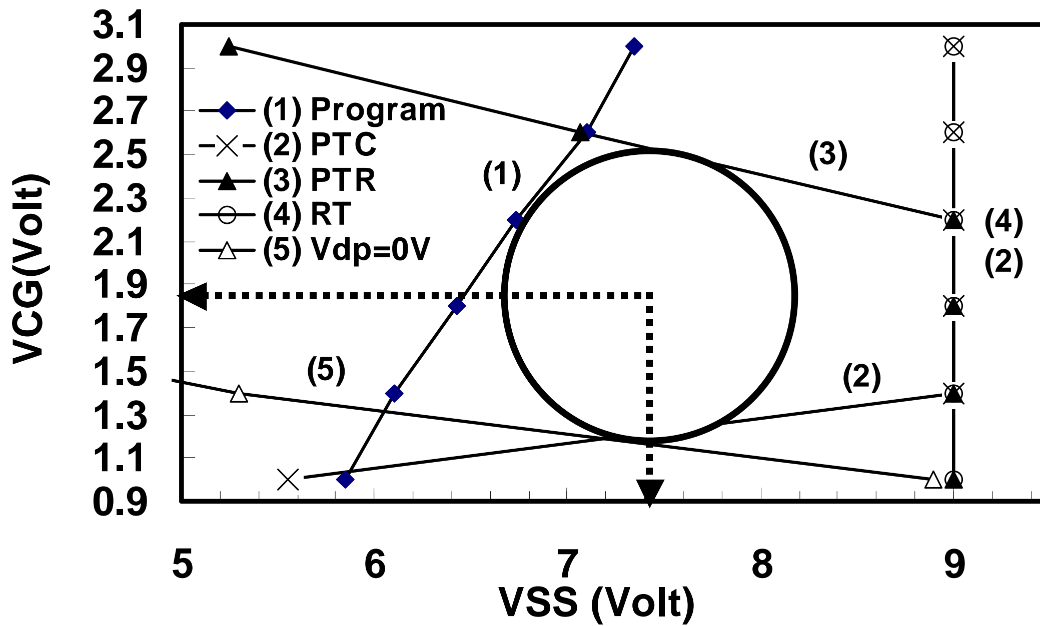


(a)



(b)

Fig.4.7 (a) Program improvement vs FG oxide spacer pull-back. The program performance is characterizaed by Ir0/Ir1@Vs=6V, 10us, program current=3uA, where Ir0 and Ir1 is programmed and erased current, respectively. (b) Erasing improvement vs. FG oxide spacer etching. The erasing performance is characterized by Verase, which is the voltage to reach 50% Ir1 after 10ms erasing



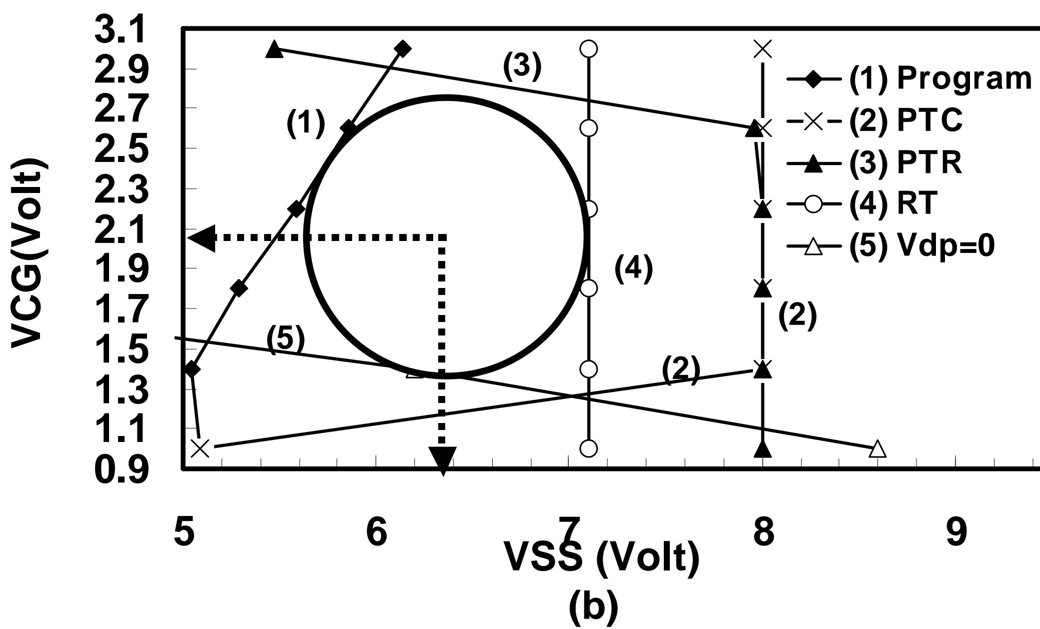


Fig.4.8 (a) Program vs. disturb window of traditional SA3 cell. The source voltage for programming is 7.4V, (b) Program vs. disturb window of T-shape Vss coupling SA3 cell. Vss voltage can be reduced to 6.4V in new cell.

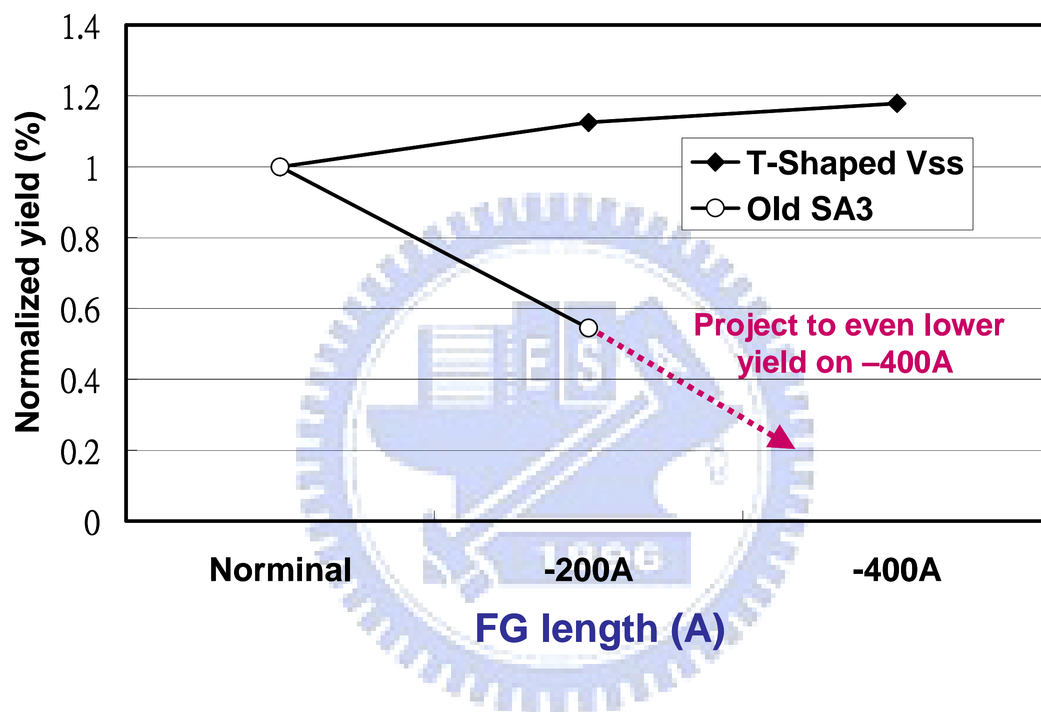


Fig. 4.9 FG length reduction vs. yield. The yield of new SA3 cell remains stable when FG length is reduced, whereas the yield of old SA3 cell drops drastically with the shorter FG length. Note that the nominal FG length is around 180nm.

Reference

[1] S. Kianian, A. Levi, D. Lee, and Y.-W. Hu, "A novel 3 volts-only, small sector erase, high density Flash E^2PROM," in *Symp. VLSI Technol. Dig.,* 1994, p. 71-72.

[2] B. Yeh, "Single transistor non-volatile electrically alterable semiconductor memory device," *United States Patent* 5,029,130, 1991.

[3] *SST Data Book Flash Memory*, 2004.

[4] A. T. Wu, T. Y. Chan, P.K. Ko and C. Hu, "A novel high-speed, 5-volt programming EPROM structure with source-side injection" in *IEDM Tech. Dig.,* 1986. pp. 584–587.

[5] A. Chimenton, P. Pellati, P. Olivo, " Overerase Phenomena: An Insight Into Flash Memory Reliability," in *Proceedings of the IEEE*, vol. 91, no. 4, April 2003, pp. 617-626.

[6] Muramatsu S., Kubota T., Nishio N., Shirai H., Matsuo M., Kodama N., Horokawa M., Saito S., Arai K. and Okazawa T, "The solutions of over-erase problem controlling poly-Si grain size-modified scaling principles for flash memory" *IEDM Tech. Dig.,* 1994. pp. 847.

[7] D.S. Kuo, C.S. Wang, et al., "A Foundry Operation with Flash Technology," in *Proc. 17th IEEE Nonvolatile Semiconductor Memory Workshop*, 2000, pp14-21

[8] J. V. Houdt et al., " The HIMOS flash technology: the alternative solution for low-cost embedded memory," in *Proceedings of the IEEE*, vol. 91, no. 4, April 2003, pp. 627-635.

[9] R. Mih, J.Harrington, K. Houlihan, H.K. Lee, K. Chan, J. Johnson, B. Chen, J. Yan, A. Schmidt, C. Gruensfelder, K. Kim, D. Shum, C. Lo, D. Lee, A. Levi, and C. Lam, "0.18um Modular Triple Self-Aligned Embedded Split-Gate Flash Memory," in *Symp. VLSI Technol. Dig., 2000*, pp. 120-121.

[10] W.T. Chu, H.H. Lin, C.T. Hsieh, H.C. Sung, Y.H. Wang, Y.T. Lin, C.S. Wang, "Shrinkable Triple-Aligned Field-Enhanced Split-Gate Flash Memory," *IEEE Transactions on Electron Devices*, vol. 51, no. 10, pp. 1667-1671, 2004

[11] S. M. Sze, *Physics of Semiconductor Devices*, 2nd ed. New York: Wiley. 1981.

[12] Melik-Martirosian, A.; Ma, T.P.; Wang, X.W.; Guo, X.; Widdershoven, F.P.; Wolters, D.R.; van der Wal, V.J.D.; van Duuren, M.J., **“**Demonstration of a flash memory cell with 55 Å EOT silicon nitride tunnel dielectric **,”** *VLSI Technology, Systems, and Applications, 2001.* Proceedings of Technical Papers. 2001 Int. Symp., 18-20 April 2001, pp. 138 -141

[13] C. Hu et al., “Hot-electron induced MOSFET degradation – Model, monitor and improvement,” *IEEE Trans. Electron Devices*, vol. ED-32, p.375, 1985.

[14] S. Tam, P. K. Ko and C. Hu, “Lucky-electron model of channel hot-electron injection in MOSFET’s” *IEEE Trans. Electron Devices*, vol. 31, pp. 1116–1125, Sep. 1984.

[15] T. H. Ning, C. M. Osburn, and H. N. Yu, “Emission probability of hot electrons from silicon into silicon dioxides” *J. Appl. Phys.*, vol. 48, pp. 286–293, Jan. 1977.

[16] D. Lee et al, “Vertical Floating-Gate 4.5F^2 Split-Gate NOR Flash Memory at 110nm Node,” in *Symp. VLSI Technol. Dig.,* 2004, p. 72-73.

[17] C. M. Liu, J. Brennan, Jr., K. Chan, P. Guo, A. V. Kordesch and K. Y. Su, “On the capacitance coupling ratios of a source-side injection flash memory cell” *Jap. J. Appl. Phys.*, vol. 40, pp. 2958-2962, Apr. 2001

[18] R. Bez, E. Camerlenghi, D. Cantarelli, L. Ravazzi and G. Crisenza “A novel method for the experimental determination of the coupling ratios in submicron EPROM and flash EEPROM cells” in *IEDM Tech. Dig.,* 1990. pp. 99-102.

[19] H. Fujiwara, M. Arimoto, T. Hkaida, S. Sudo, K. Kurooka, H. nagassawa, T. Hiroshima and K. Mamero, “A new method for measuring the coupling coefficient of a split-gate flash EEPROM” *Proc. IEEE Int. Conf. On Microelectronic Test Structures*, vol. 14, pp. 43-46, Mar. 2001.

[20] H.C. Sung, T.F. Lei, T.H. Hsu, Y.C. Kao, Y.T. Lin, C.S. Wang, “Novel Program vs Disturb Window Characterization for Split-Gate Flash Cell,” *IEEE Electron Device Lett.,* vol.26, no.3, Mar 2005 pp. 194-196

Chap.5
Novel Single Poly EEPROM with Metal Control Gate Structure

5.1 Introduction

The consumer electronics is the most important market sector in IC industry, the household spending on consumer electronics doubles since 1994 [1]. The characteristics of consumer electronics are the short lifetime cycle and fast cost erosion. To be successful in this business, the time to market and the total cost is the main key factor. Recently, the Logic NVM (Non-Volatile Memory) technology has gained high attention because it can offer design flexibility to make chip meet the spec without going through design revision and can use pure logic process without extra masking for the non-volatile memory.

Currently, there are two categories in Logic NVM. One is Logic OTP and another is Logic MTP. The Logic OTP is the One Time Programming memory. There are several ways to perform the one time programming function using pure logic process. For example: (1) oxide rupture mechanism to cause gate to junction short[2], (2) electro-migration method to change the poly resistance [3], and (3) use single floating PMOS structure, which relies on the weak coupling between floating gate and junction to get electrons injected to the floating gate [4]. Since OTP memory can only be programmed once, the application is limited to code storage, calibration/trimming, feature selection, memory repair and ROM replacement. The summary of Logic OTP technology is shown in Fig. 5.1. The 2nd type of LogicNVM is the logic MTP. It can perform Multi-Time Programming up to 10,000 times like the typical Flash or EEPROM . It uses N-well as the coupling gate instead of traditional double poly, so it can be built on pure logic process [5,6]. As shown in Fig. 5.2, the cell size is very big because it needs to use a N-well region as the coupling gate. Due to the big cell size, the biggest density

usually can only be several K bits. As a result, the application is limited in the precision analog trimming code, digital right management (DRM), RFID and data storage.

In this chapter, a novel single poly EEPROM with small cell size and fewer extra masking steps is presented. The uniqueness of this cell is that the control gate structure is a tungsten line formed by damascene process and the inter-gate dielectric is the high-K material grown by Atomic Layer Deposition (ALD) [7]. Because of a stronger coupling between Control gate (CG) and Floating gate (FG) by using the high K material like Al_2O_3, the program and erase voltage can be lowered from ~10 V to 6.5 V. Therefore, the program/erase circuitry could be handled by 3.3 V IO devices instead of conventional high voltage devices [8]. Furthermore, the cell size can be very compact because the control gate is on top of the floating gate but not from the huge well diffusion. For CMOS process compatibility, this new approach can be built on pure logic process but adding two extra masking steps for Deep Nwell (DNW) and Control gate (CG). In addition, the high-K material is deposited during back-end metallization steps, so there is no high-K material contamination concerns because the material is deposited during back-end metallization steps. Therefore, the new cell presented in this paper is very suitable for mid-density embedded Multi-Time-Program (MTP) applications.

5.2 Device Fabrication

The cross-sectional and top view of a single cell is shown in Fig. 5.3 (a)&(b). The ideal 2T and 1T cell size for this new cell is 26 and 18 F^2, respectively [10]. For feasibility study, a relaxed 2T cell, 1.5 μm^2, is used in this experiment. The 2T structure is chosen because it is immune from over-erase concern [11]. The channel width is 0.32 μm, and the channel length for floating gate and select gate is 0.4 and 0.18 μm, respectively. The floating gate overlap with STI is 0.32 μm per-side. The starting material is the P-type wafer with 8-12 ohm-cm

resistance. The device fabrication begins with a deep N-well photo and implantation, then uses the standard 0.18 μm CMOS process from Shallow-Trench-Isolation (STI) to contact plug formation. The gate oxide (~7 nm) and gate poly (150 ~ 200nm) for periphery devices are acted as the tunneling oxide and floating poly for this EEPROM device. After contact plug formation, the control gate is formed by a tungsten (W) damascene process, which includes Control gate photo & etching, Al_2O_3 deposition, barrier metal (TiN) deposition, W fill and CMP. A proper oxidation treatment (~1 nm) before ALD and a post ALD anneal were done to ensure a good inter-gate dielectric quality. The physical thickness and the equivalent oxide thickness (EOT) of Al_2O_3 is about 20 nm and 9 nm, respectively. After the control gate damascene process, the typical back-end metal process is followed. The TEM picture of final cell is shown in Fig. 5.4. Note that the high-K film, Al_2O_3, is deposited in the back-end metallization steps but not in the front-end process shown in the previous works [8], [12], so there is no cross-contamination issue caused by new material nor the device impact induced by the extra thermal cycle from conventional double poly process. To prevent data retention problems induced by salicidation, the cell area is blocked with protective oxide during the salicidation process [13].

5.3 Result and Discussion

- **Programming & Erasing Performance**

The program and erase mechanism is the similar to the one for traditional stacked-gate cell, which uses channel-hot-electron injection for programming and Fowler-Nordheim (F-N) tunneling for channel erasing. The array schematic and bias condition is shown in Fig. 5.5 (a) & (b). As shown in Fig. 5.6 (a) & (b), the programming and erasing can be accomplished in 500 μs and 100 ms, respectively. The maximum voltage is 6.5 V and 5 V for programming and erasing, respectively. The voltage is less than the typical stacked-gate flash memory

because of the stronger FG-CG coupling contributed by the high-K film, Al_2O_3, whose dielectric constant is 9, which is 2.3 times and 28% higher than oxide and nitride, respectively. The coupling ratio is around 70%, which is calculated from the cell layout and the equivalent oxide thickness. We use 0.18μm logic well and junction for this memory cell, so we can save the masking step for memory fabrication. However, since the drain and well is engineered for logic device, the programming performance is not as good as typical commodity flash cell.

With maximum 6.5V operation, the voltage could be handled by 3.3V IO device, which typically has breakdown voltage higher than ~7V. Using 3.3V IO device for the programming/erasing circuit can save the extra masking & process for high voltage device fabrication, also the chip area can be greatly reduced because of the tighter design rule for 3.3V devices.

- **Disturb Characteristics**

No disturb behavior is found on the non-selected cells in bit-line and word-line directions during programming. The data is shown in the Fig. 5.7 (b). Similar to the split-gate Flash memory array reported in Chap.3, there are three types of disturb in the same page, which will have the high Vss disturb during program. The disturbed bits are labeled as B,C,D in the Fig. 5.7 (a)

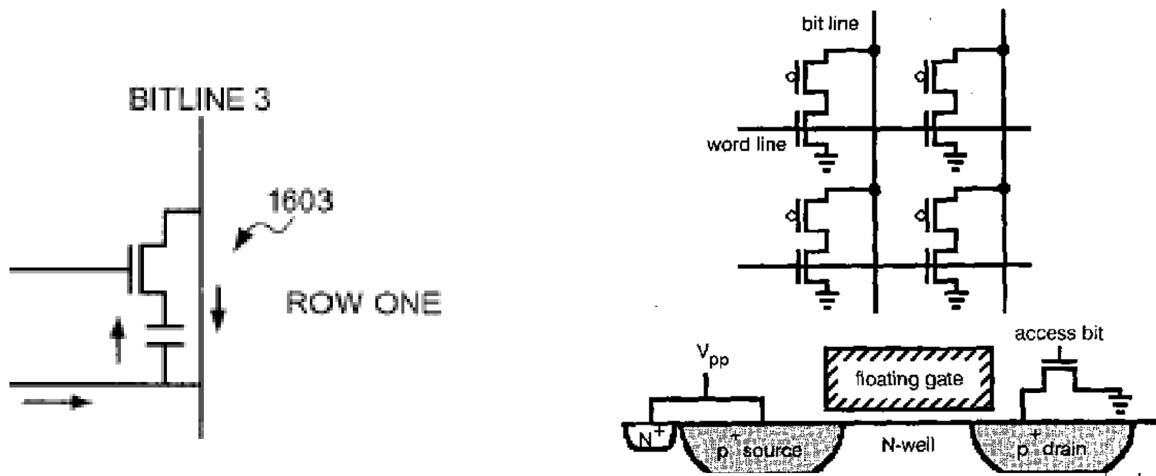
- **Reliability Performance**

Fig.5.8 (a) and (b) show the single cell reliability on cycling and retention. As shown in Fig. 5.8 (a), the cell current only degrades 10% after 100K cycling. In Fig.5.8 (b), data retention is projected to last more than 10 years at 150°C for the cells with cycling up to 100 ~ 1K. However, the current of the cell with 10K cycling drops seriously within 24hrs but reaches a saturation level on the further extended bake. The 1st possible reason for this behavior is the trapping characteristic of Al_2O_3. Because we have not optimized the top and bottom barrier for the Al_2O_3 in this study, the existing barrier could be too thin to prevent charge trapped in the Al_2O_3 film. It is well known that the Al_2O_3 film is a very good charge

trapping material [12]. More cycling times could result in more trapped charges, and these charges could drift during the subsequent high temperature retention bake and then cause cell current shift afterward [14]. The second possible cause could be the high stress directly on the Flash cell during W CMP, it could physically damage the dielectric (tunneling oxide or Al_2O_3) and degrade the cell reliability performance. To clarify the root cause and to further improve the retention performance, the experiments on the Inter-Poly-Dielectric (IPD) film and cell process integration are on going.

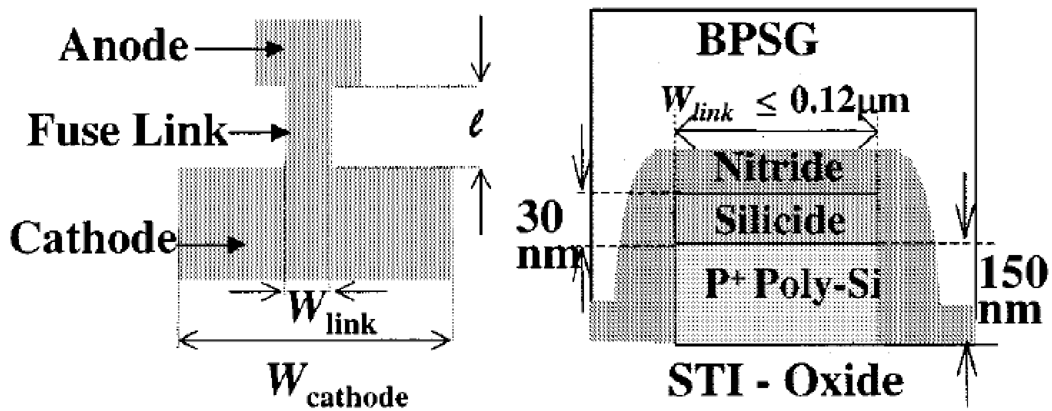
5.4 Conclusion

The EEPROM cell with W damascene control gate is presented for the first time. The device fabrication is very compatible to standard CMOS process because the extra masking step can be only 2 (DNW and CG) over the CMOS process. Actually, DNW is a common layer for mixed-signal design, if it is counted as a default mask layer, the extra masking requirement for this new cell could be only one layer. Moreover, the extra process step (high K film deposition and W CMP for CG) for memory cell is done at the back-end metallization step, so there is no concerns of cross contamination nor the device impact by the extra thermal cycle from the conventional double poly process. Since the CG is formed directly on FG like the way on ETOX cell, we can layout the very competitive cell size for 2T and 1T cell with 26 F^2 and 18 F^2 respectively, which is much smaller than the other single poly EEPROM technologies with N-well coupling. In addition, the good single cell performance with 6.5V program/erase is demonstrated in this paper, it could allow 3.3V IO device to replace typical HV device. Chip area could be saved in addition to the masking layer saving mentioned earlier. Owing to the significant advantages like Logic compatible process, compact cell size and low voltage program/erase operation, this new cell is suitable for mid-density embedded Multi-Time-Program (MTP) application.



(Ref. U.S. Patent #'s 6,667,902 & 6,671,040)

(a) (b)



(c)

Fig. 5.1 (a) Logic OTP using oxide rupture mechanism [2]. The capacitor shown in the figure is a poly gate with junction overlap structure, it can be breakdown during programming, (b) schematic diagram of device and array structure for a single PMOS logic OTP memory [3], (c) The top and cross section view of eFuse [4].

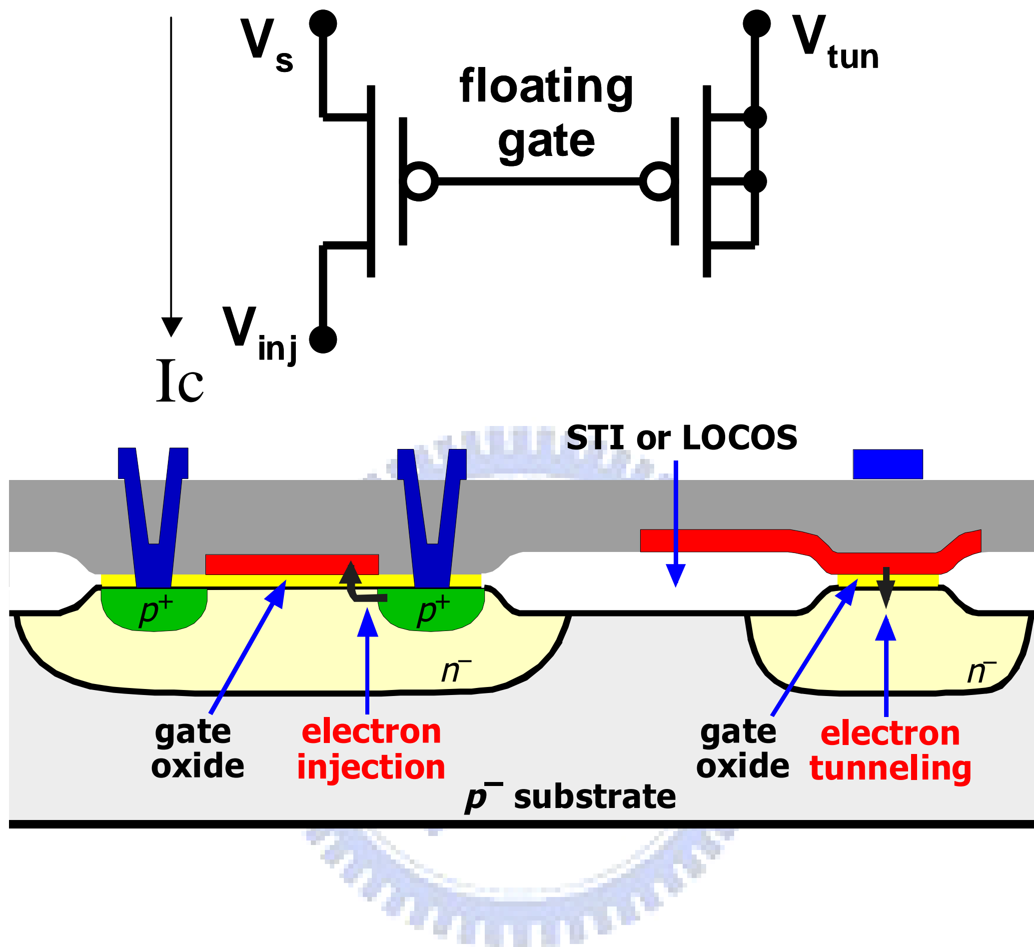
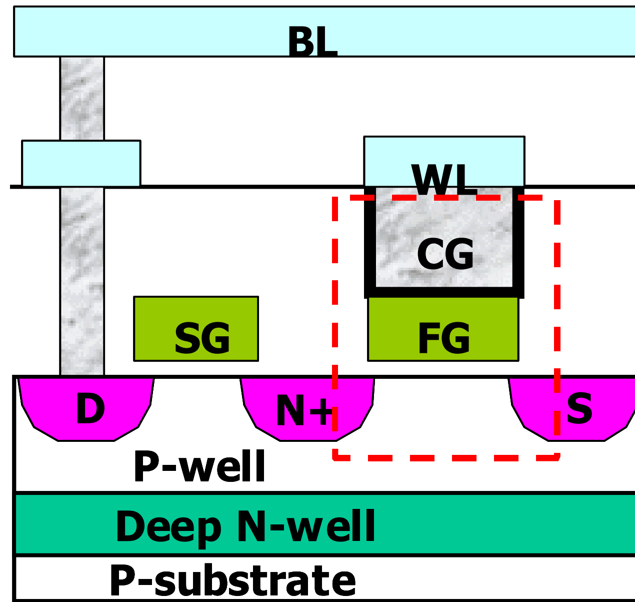
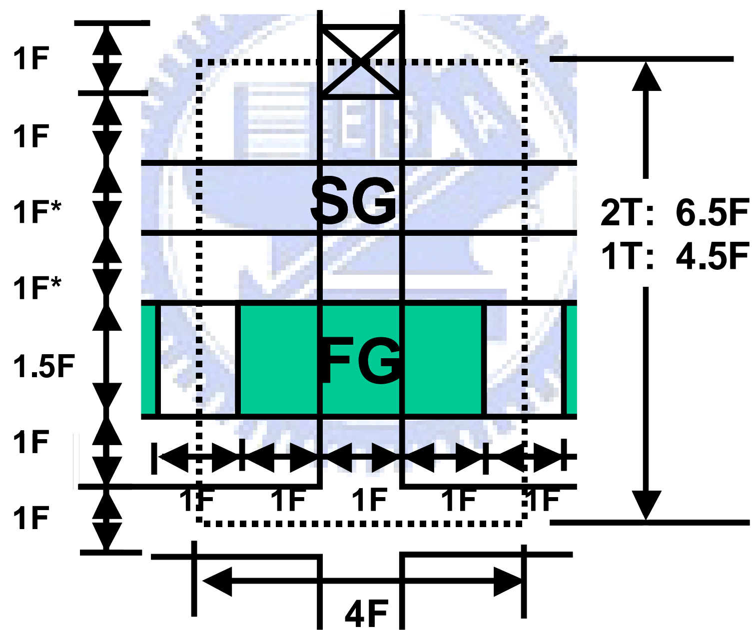


Fig. 5.2 Schematic diagram and cross-sectional view of a logic MTP memory cell [6]



(a)



(b)

Fig.5.3 (a) Cross-sectional view of a single cell, (b) Top view cell layout. The ideal 2T and 1T cell size is $26F^2$ and $18F^2$, respectively.

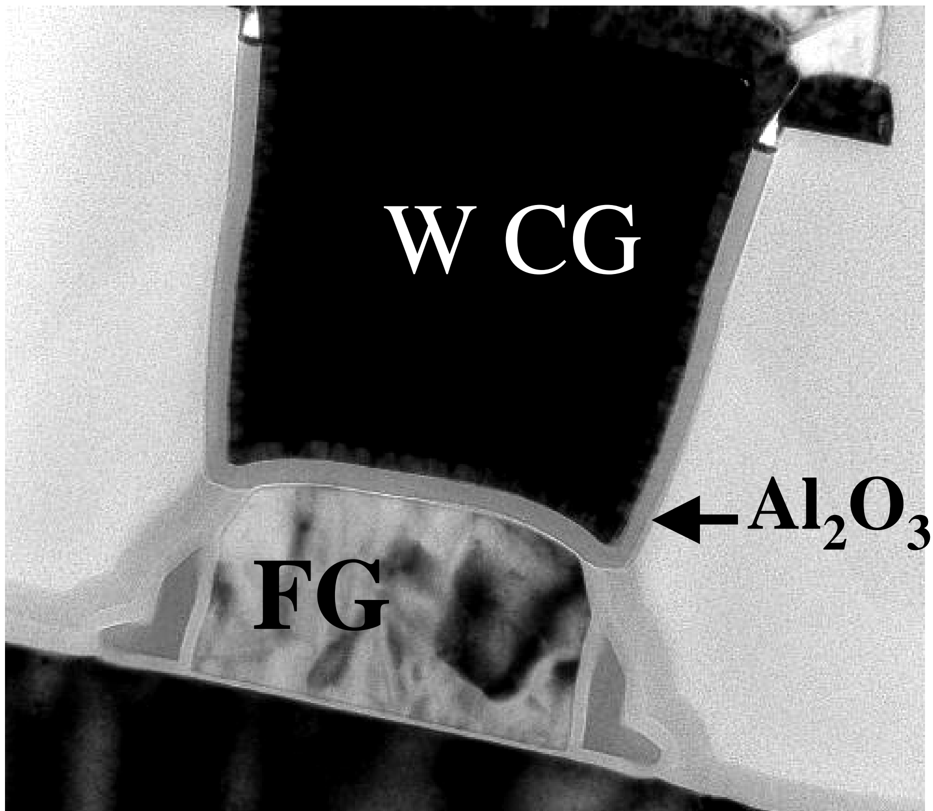
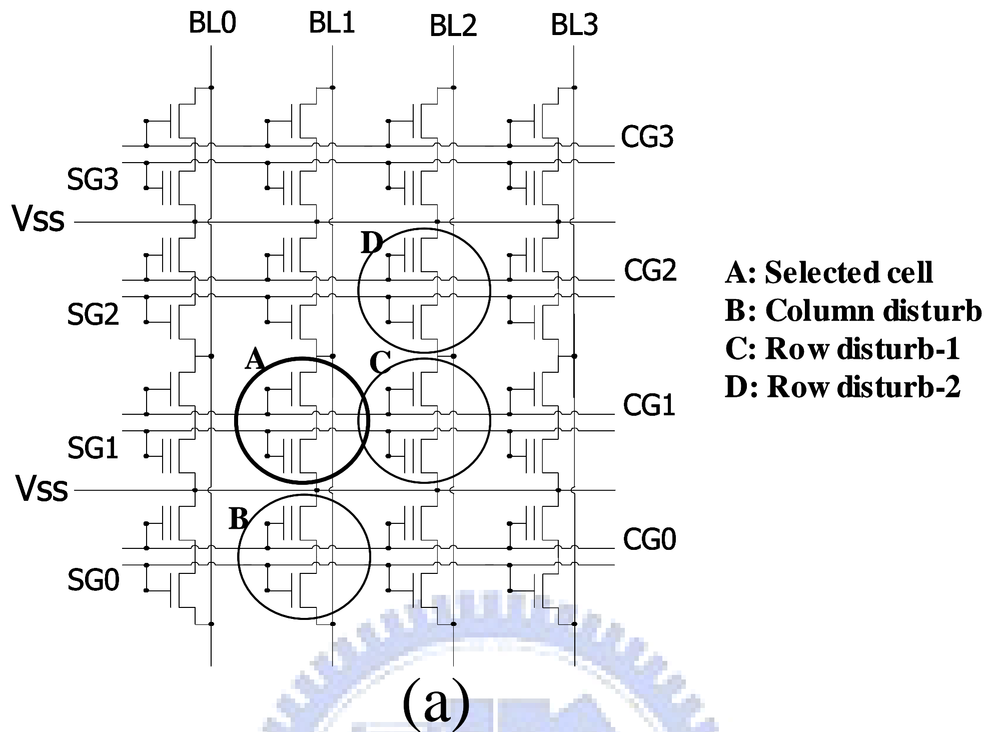


Fig.5.4 TEM picture of final cell. A proper oxidation treatment (~1 nm) before ALD and a post ALD anneal were done to ensure a good inter-gate dielectric quality. The physical thickness and the equivalent oxide thickness (EOT) of Al_2O_3 is about 20 nm and 9 nm, respectively.



(a)

		SG	WL	BL	Source	PW
Read	Selected	2.5	3	1	0	0
	Unselected	0	0	0	0	0
Program	Selected	2.5	6.5	0	6.2	0
	Unselected	0	0	3	0	0
Erase	Selected	0	-4	0	0	5
	Unselected	0	0	0	0	0

(b)

Fig.5.5 (a) Schematic diagram of memory array, (b) bias condition

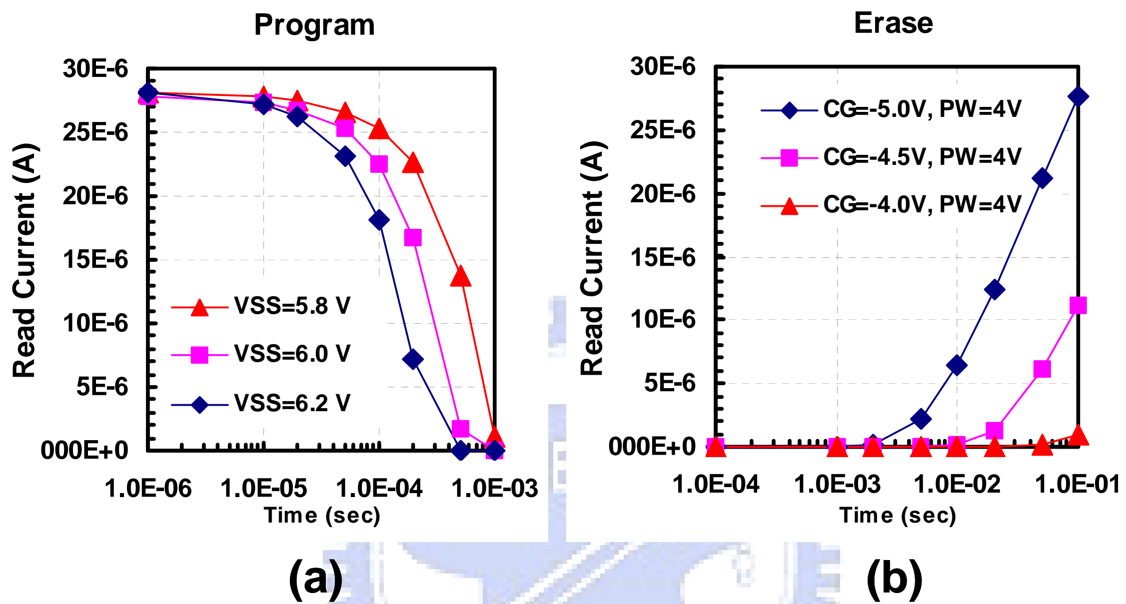
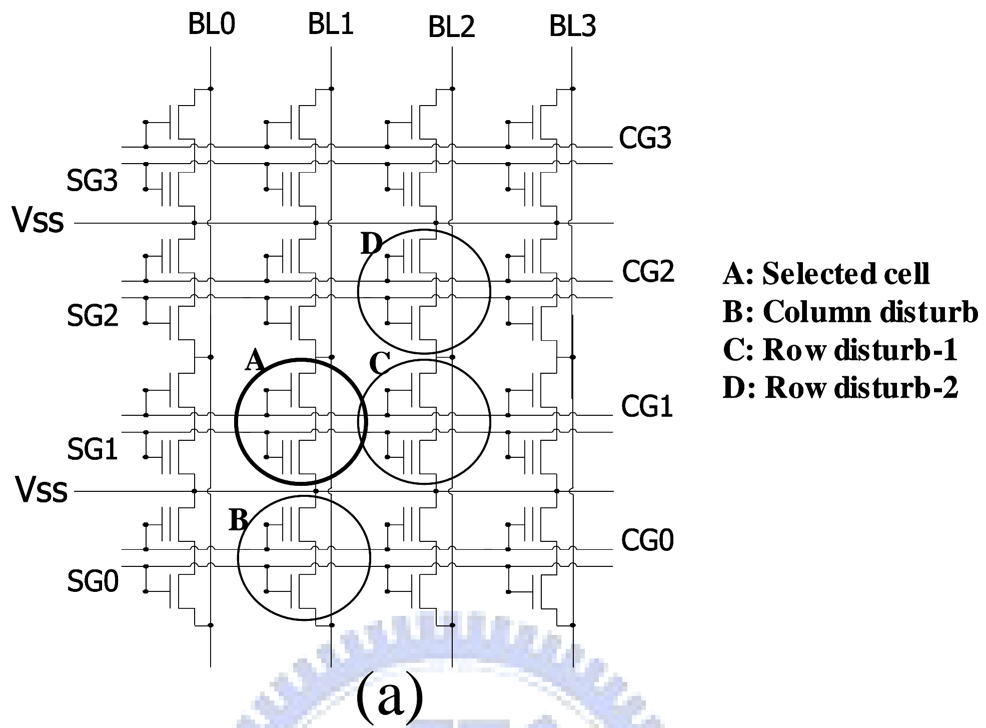
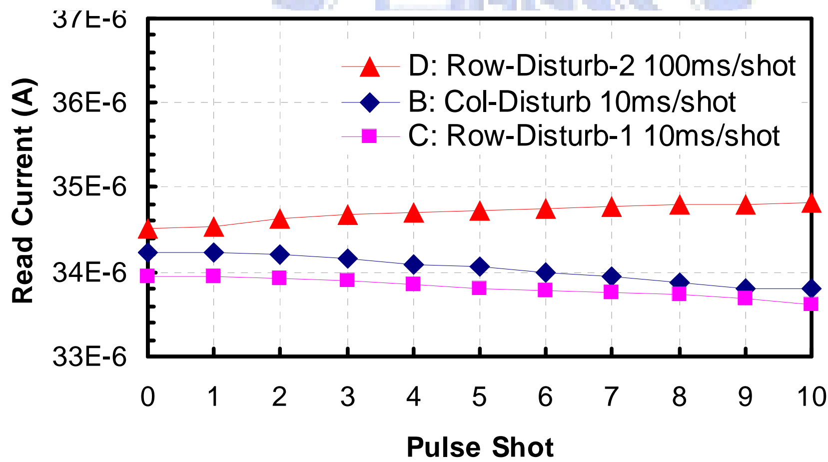


Fig.5.6 (a) Program characteristics under different source (Vss) voltage, (b) Erasing characteristics with different control gate (CG) voltage.

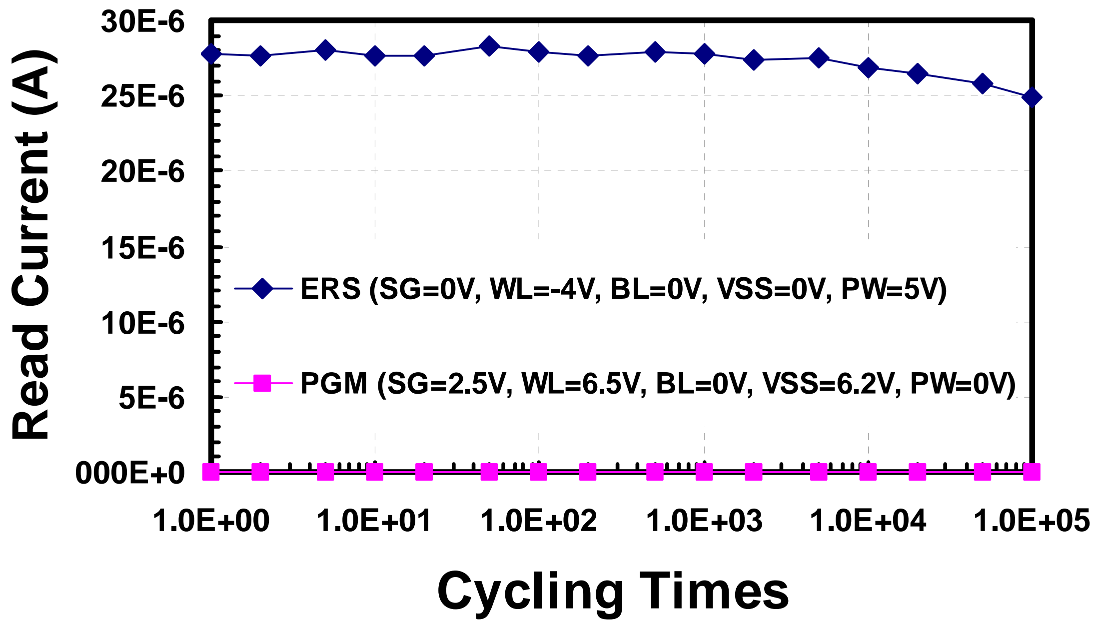


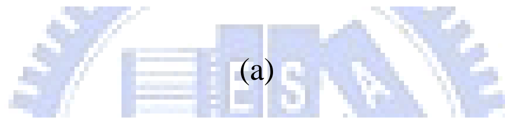
(a)



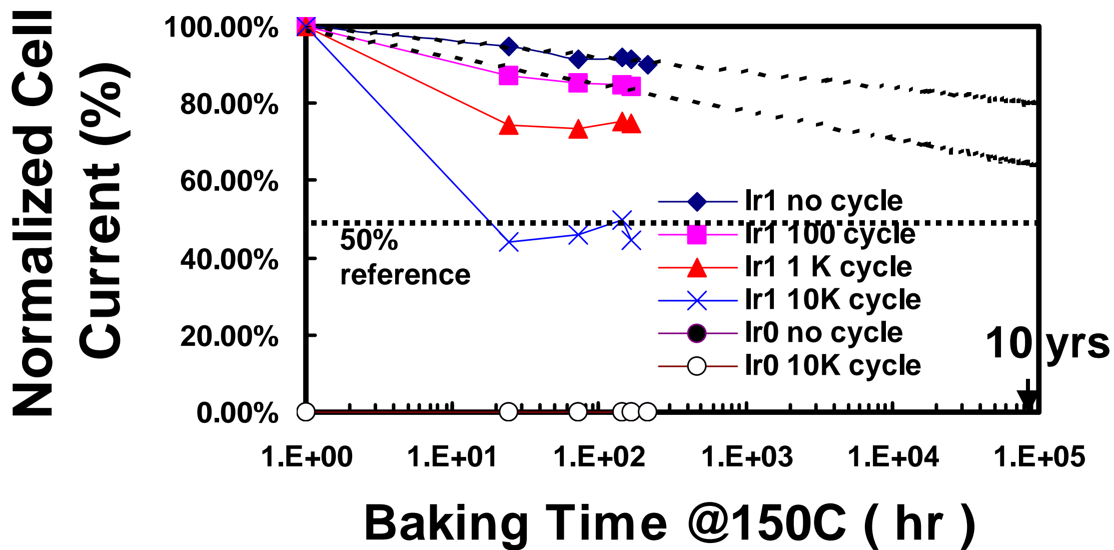
(b)

Fig. 5.7 (a) Disturb cell location in the array. A,B,C &D cells are in the same page, which will have same high V_{SS} during programming, (c) the cell current of C, B,D cells after disturb





(a)



(b)

Fig.5.8 (a) Endurance Cycling characteristics with channel-hot-electron (CHE) programming and FN erasing. Only 10% current drop is observed after 100K cycling, (b) Data retention characteristics under various pre-cycling stress at 150C baking.

Reference

[1] J. Harding, " Smaller, fater, Cheaper, Better: The Relentless Pressure on Consumer Electronics," Logic NVM Symp, 2007

[2] US patent 6,667,902 & 6,671,040

[3] C. Kothandaraman, S.K. Iyer and S. S. Iyer, " Electrically Programmable Fuse (eFUSE) Using Electromigration in Silicides," *IEEE Electron Device Lett.*, vol. 23, no. 9, Sep. 2002, pp. 523-525

[4] L. Chang, C. Kuo, Chenming. Hu, A. Kalnitsky, A. Bergemont and P. Francis, " Non-volatile memory device with true CMOS compatibility," *Electronics Letter*, vol.35, no.17, 19th Aug. 1999, pp. 1443-1444

[5] A. Pesavento, T. Gilliland, C. lindhorst, S. Srinivas, F. Bernard, S. Salazar, C. Diorio, S. King, C. Bockorick, B. Wang, Y. Ma, C.H. Wang, T. Humes, J. Caywood, "Embedded nonvolatile Memory in Logic CMOS," in *IEEE NVSMW*, 2004, pp. 49-50.

[6] K. Ohsaki, N. Asamoto, A. Takagaki, "A single poly EEPROM Cell Structure for Use in Standard CMOS Processes," *IEEE J. Solid-State Circuits*, vol. 29, no. 3, Mar. 1994, pp. 311-316.

[7] Y.L. Tu, H.L. Lin, L.L. Chao, Danny Wu, C.S. Tsai, C. Wang, C.F. Huang, C.H. Lin, Jack Sun, "Characterization and Comparison of High-k Metal-Insulator-Metal (MiM) Capacitors in 0.13 μm Cu BEOL for Mixed-Mode and RF Applications," in *Symp. VLSI Technol. Dig., 2003*, pp. 79-80.

[8] B. Govoreanu, P. Blomme, M. Rosemeulen, J. Van Houdt, K. DeMeyer, "VARIOT: A Novel Multilayer Tunnel Barrier Concept for Low-Voltage Nonvolatile Memory Devices," *IEEE Electron Device lett.*, vol. 24, no. 2, Feb. 2003, pp. 99-101.

[9] J. C. Mitros, C.Y. Tsai, H. Shichijo, K. Kunz, A. Morton, D. Goodpaster, D. Mosher, T.R. Efland, "High-Voltage Drain Extended MOS Transistors for 0.18-um Logic CMOS

Process" *IEEE Trans. Electron Devices*, vol. 48, no.8, August 2001, pp. 1751–1755.

[10] Y.H. Song, J.I. Han, J.W. Kim, J.H. Park, S.Y. Kim, D.W. kwon, Y.M. Park, J.S. Lee, W.K. Lee, D.Y. Lee, J.W. Kim, M.S. Kang, J. Kim, and K.D. Sub, "A High Density and Low-cost Self-aligned Shallow Trench Isolation NOR Flash Technology with 0.14 μm^2 cell size" in *IEDM Tech. Dig.,* 2001. pp. 2.4.1-2.4.4.

[11] A. Chimenton, P. Pellati, P. Olivo, " Overerase Phenomena: An Insight Into Flash Memory Reliability," in *Proceedings of the IEEE*, vol. 91, no. 4, April 2003, pp. 617-626.

[12] T. Sugizaki, M. Kobayashi, M. Ishidao, H. Minakata, M. Yamaguchi, Y. Tamura, Y. Sugiyama, T. Nakahishi, H. Tanaka, "Novel Multi-bit SONOS Type Flash Memory Using a High-K Charge Trapping Layer" in *Symp. VLSI Technol. Dig., 2004*, pp. 27-28.

[13] J.W. Liou, C.J. Huang, H.H. Chen, G. Hong, "Characterization of Process-Induced Mobile Ions on the Data Retention in Flash Memory" *IEEE Trans. Electron Devices*, vol. 50, no.4, April 2003, pp. 995–1000.

[14] M. Janai, B. Eitan, A. Shappir, E. Lusky, I. Bloom, G. Cohen, "Data retention Reliability Model of NROM Nonvolatile Memory Products" *IEEE Trans. On Device and Materials Reliability*, vol. 4, no.3, September 2004, pp. 404–415.

Chapter 6

Conclusions and Further Recommendations

6.1 Conclusions

In this thesis, a new methodology for program vs. disturb window characterization on split gate flash cell is presented in Chapter 3. The window can be graphically illustrated in V_{WL}(word-line)-V_{SS}(source) domain under a given program current. This method can help us to understand quantitatively how the window shifts vs bias conditions, and lead us to find the optimal program condition. The condition obtained by this method can have the largest operation window. This methodology was successfully implemented in 0.18um triple self-aligned (SA3) split-gate cell development.

Then, in the chapter 4, a new triple self-aligned (SA3) split-gate flash cell with a T-shaped source coupling approach is described. This novel structure can significantly enhance coupling capacitance between the source and floating gate without increasing cell size. The enhancement can be simply modulated by an oxide-etching step. This new structure can be applied to program voltage reduction and cell size scaling. For program voltage reduction, the maximum program voltage of the new cell can be reduced from 7.4 to 6.4 V, which is characterized by the newly developed characterization methodology presented in Chapter 3. For cell size scaling, comparable wafer sort yield was demonstrated using the new cell with a shorter floating length and a shallower source junction. To understand the relationship between source coupling ratio (SCR) and the program/erase mechanism, an insightful

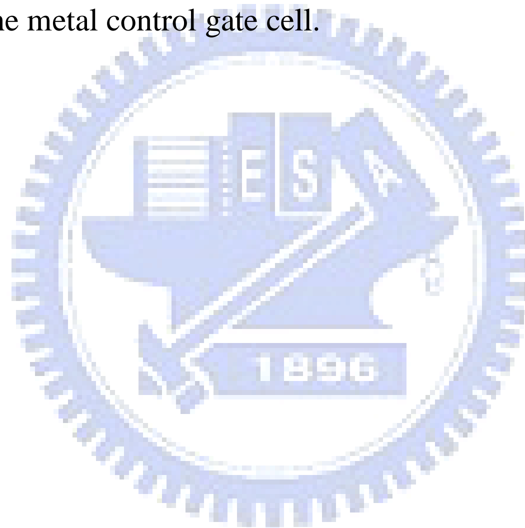
discussion on the program and erase mechanisms for our split–gate flash cell is described in Chapter 2.

In this chapter 5, a novel single poly EEPROM using metal control gate is presented in this paper. The control gate is tungsten (W) line made by a damascene process, and inter-gate dielectric is Al_2O_3 grown by Atomic Layer Deposition (ALD). The program and erase mechanism is the same as the one for traditional stacked-gate cell, which uses the channel hot electron injection for programming and Fowler-Nordheim (F-N) tunneling for channel erasing. With the high dielectric constant (K) property of Al_2O_3, we can perform the program and erase function with a voltage less than 6.5 V, which could be handled by 3.3 V devices instead of traditional high voltage devices. In the process compatibility aspect, this new cell needs only two extra masking steps over the standard CMOS process, and the high-K material is deposited in the back-end metallization steps without the concern for cross-contamination nor the device impact from the extra thermal cycle. Therefore, this new technology is suitable for embedded application. In this paper, the good cell performance is demonstrated; such as, fast programming/erasing, good endurance and data retention.

6.2 Further Recommendations

There are some interesting topics for further study. For window characterization, the choosing of program and disturb spec needs further investigation because the operation window would have strong dependence on how we define the spec. For split gate technology evolution, the direction to scale down the cell is to continuously improve the CG/FG coupling. The cell described in the reference [1] is a very good

candidate for the next generation of split –gate Flash. It uses an additional control gate on top of floating to fully utilize the FG area for coupling. By doing so, the high voltage in source voltage can be moved to additional CG, so the source junction can be shallower and the cell can be much smaller. Also, the addition of control gate can provide more flexibility to control the program disturb mechanism, so the operation window can be enlarged. In addition, the erase gate can be moved to the poly on top of source junction, so the select transistor can use low voltage oxide to increase driving capability. For the single poly EEPROM with metal control gate, the future work we would focus on the trapping mechanism study and the experiment on new high K material for the metal control gate cell.



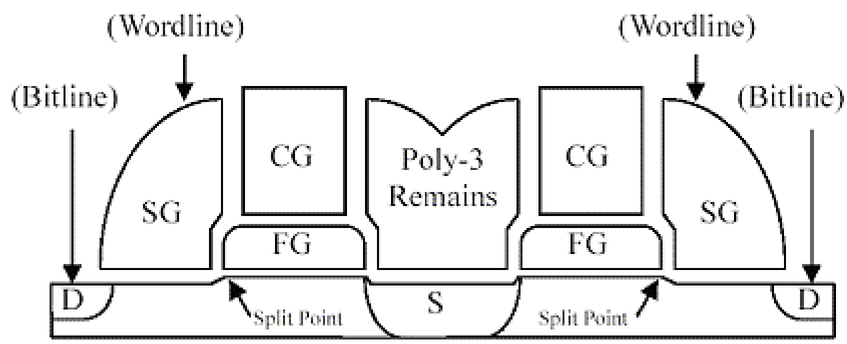


Fig.6.1. Next generation of split-gate Flash cell. One extra CG gate is added to enhance the FG/V_{SS} coupling [1].

Reference

[1] Y.S. Cho, M.J. Chen, " A novel Highly Reliable Flash Memory – characteristics, Reliability Evaluation, and application," National Chiao-Tung University PHD Thesis, 2004.

學經歷

姓名: 宋弘政

性別: 男

出生: 民國 54 年 10 月 25 日

籍貫: 高雄市

住址: 高雄市楠梓區宏毅二路南五巷四十一號

學歷: 國立交通大學電物系 [72 年 9 月- 76 年 6 月]

國立交通大學光電研究所碩士班 [76 年 9 月- 78 年 7 月]

國立交通大學電子研究所博士班 [88 年 8 月- 97 年 6 月]

經歷: 臺灣積體電路記憶體 工程師~專案經理 [80 年 7 月- 94 年 8 月]

臺灣積體電路北美分公司 專案經理 [94 年 12 月至今]

博士論文題目:

分離式閘極非揮發性記憶體技術及新穎多晶矽電子抹除式唯讀記憶體之研究

Study on Split-Gate Non-Volatile Memory Technology and A Novel Single Poly EEPROM Memory Cell

Publication List

1. International Journal:

[1] Hung-Cheng Sung, Tan Fu Lei, Te-Hsun Hsu, Chen-Ming Huang, Ya-Chen Kao, Yung-Tao Lin, C.S. Wang; "New Triple Self-aligned(SA3) Split-gate Flash Cell with T-shaped Source Coupling," ***JJAP***, Vol. 44, no. 10, 2005, pp. 7377-7383

[2] Wen-Ting Chu, Hao-Hsiung Lin, Chia-Ta Hsieh, **Hung-Cheng Sung,** Yu-Hsiung Wang, Yung-Tao Lin, Wang, C.S,; "Shrinkable triple self-aligned field-enhanced split-gate flash memory," ***IEEE Trans. Electron Devices***, Vol. 51, Issue 10, Oct. 2004 pp. 1667 - 1671

2. International Letter:

[1] Hung-Cheng Sung, Tan Fu Lei, Te-Hsun Hsu, Ya-Chen Kao, Yung-Tao Lin, C.S. Wang "Novel program versus disturb window characterization for split-gate flash cell," ***IEEE Electron Device Lett.,*** Vol. 26, Issue 3, March 2005 pp. 194 - 196

[2] Hung-Cheng Sung, Tan Fu Lei, Te-Hsun Hsu, S.W. Wang, Ya-Chen Kao, Yung-Tao Lin, Wang, C.S, "Novel Single Poly EEPROM with Damascene Control Gate Structure," ***IEEE Electron Device Lett***., vol. 26, no.10, 2005, pp. 770-772

[3] Wen-Ting Chu, Hao-Hsiung Lin, Yeur-Luen Tu, Yu-Hsiung Wang, Chia-Ta Hsieh, **Hung-Cheng Sung,** Yung-Tao Lin, Chia-Shiung Tsai, C.S. Wang "Using an ammonia treatment to improve the floating-gate spacing in split-gate flash memory," **IEEE Electron Device Letters**, Vol. 25, Issue 9, Sept. 2004 pp. 616 - 618

[4] Wen-Ting Chu, Hao-Hsiung Lin, Yu-Hsiung Wang, Chia-Ta Hsieh, **Hung-Cheng Sung**, Yung-Tao Lin, C.S. Wang, "High SCR design for one-transistor split-gate full-featured EEPROM," ***IEEE Electron Devices Lett***., Vol. 25, Issue 7, July 2004 pp.498 - 500

[5] Kuo-Ching Huang, Yean-Kuen Fang, Dun-Nian Yang, Chii-Wen Chen, **Hung-Cheng Sung,** Di-Son Kuo, C.S.Wang, Mong-Song Liang , "The impacts of control gate voltage on the cycling endurance of split gate flash memory," *IEEE Electron Devices lett.* Vol. 21, Issue 7, July 2000 pp. 359 - 361

[6] Kuo-Ching Huang, Yean-Kuen Fang, Dun-Nian Yaung, Chii-Wen Chen, **Hung-Cheng Sung**; Di-Son Kuo, C.S. Wang, Mong-Song Liang, "Effect of substrate bias on the performance and reliability of the split-gate source-side injected flash memory," IEEE Electron Device Lett., Vol. 20, Issue 8, Aug. 1999 pp. 412 - 414