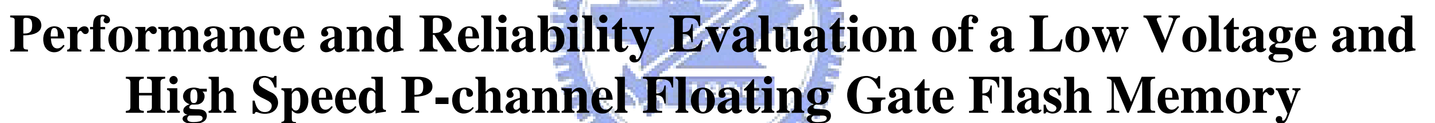
國立交通大學

電機學院微電子奈米科技產業研發碩士班

碩 士 論 文

低電壓且高速操作的Ｐ通道快閃式記憶體元件
性能及可靠性研究



Performance and Reliability Evaluation of a Low Voltage and High Speed P-channel Floating Gate Flash Memory

研 究 生：黃燿賢

指導教授：莊紹勳　教授

中 華 民 國 九 十 六 年 六 月

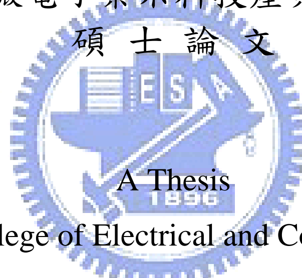
低電壓且高速操作的P通道快閃式記憶體元件性能及可靠性研究

Performance and Reliability Evaluation of a Low Voltage and High Speed P-channel Floating Gate Flash Memory

研 究 生：黃耀賢　　Student：Yao-Hsien Huang

指導教授：莊紹勳　　Advisor：Steve. S. Shung

國 立 交 通 大 學
電機學院微電子奈米科技產業研發碩士班
碩 士 論 文



A Thesis
Submitted to College of Electrical and Computer Engineering
National Chiao Tung University
in partial Fulfillment of the Requirements
for the Degree of
Master
in

Industrial Technology R & D Master Program on
Microelectronics and Nano Sciences

June 2007

Hsinchu, Taiwan, Republic of China

中華民國九十六年六月

低電壓且高速操作的P通道快閃式記憶體元件
性能及可靠性研究

學生：黃耀賢 指導教授：莊紹勳

國立交通大學電機學院產業研發碩士班

摘　　要

快閃式記憶體於近年來已為非揮發性記憶體產品之主流。就一個先進的快閃式記憶體元件來說，高效能(High Performance)與高可靠度(Reliability)是兩個主要的設計考量重點。目前的研究方向主要是朝低電壓操作及高效能的目標前進，為了達到低電壓操作及高效能的需求，我們可以利用改善元件結構或是改變操作方式來達成。本論文即是利用改變操作方式來達到低電壓操作及高效能的目的。

本論文探討P通道快閃式記憶體，我們提出的新式操作方法稱之為順向偏壓輔助汲極熱電子注入 (Forward Bias Assisted Drain Hot Electron Injection, FBADHE)。我們利用一個基極-汲極二極體的順向偏壓(Bias Forward)加速載子，再改變偏壓狀態成為逆向偏壓(Reverse Bias)，這會在二極體的深空乏區中產生更大的碰撞游離 (Impact Ionization)，碰撞產生的大量載子再經由垂直電場的作用而注入至懸浮閘極中，產生臨限電壓的變化，藉以達到改變邏輯儲存的目的。我們將此新的操作方式和汲極崩潰熱電子注入操作 (Drain Avalanche Hot Electron Injection, DAHE)以及一般p通道浮動閘極記憶體常用的價帶-導帶間穿隧模式(Band-to-Band Induced Hot

Electron Injection, BBHE)進行操作速度及耐久度(Endurance)、資料保存特性(Data Retention)之比較，並研究各自的優缺點。最後，為了改善P通道快閃式記憶體的最大可靠度問題—汲極擾動(Drain Disturb)，我們提出一種方式：改變陣列結構，利用加上一個微小的負基極(Substrate)電壓在未選擇的元件(Unselected Cell)上，即可大幅度改善汲極擾動問題。

Performance and Reliability Evaluation of a Low Voltage and High Speed P-channel Floating Gate Flash Memory

Student：Yao-Hsien Huang Advisors：Dr. Steve S. Chung

Industrial Technology R & D Master Program of
Electrical and Computer Engineering College
National Chiao Tung University

ABSTRACT

Recently, the flash memory has become the main stream of nonvolatile semiconductor memory products. High performance and reliability are two major issues for the design and manufacturing. The goal of research and development of flash memory cells is to lower the operational voltage and to enhance the performance and reliability. Two approaches are widely used to reach the goal: one is to improve the cell structure, and the other is to change the operation scheme. This thesis is to develop an operation scheme to achieve low voltage, low power consumption, and high reliability.

P-channel flash memories are studies in this work. We propose a new programming scheme to inject electrons into the floating gate. It is called Forward Bias Assisted Drain Hot Electron Injection (FBADHE). First, we apply a small positive drain bias on the Drain-Substrate junction. Then, we apply an appropriate negative bias and switch the junction to reverse-bias. The change of the mode of p-n junction causes more impact ionization in the

deep depletion region and more electron-hole pairs are generated. Carriers are then injected into the floating gate via the assistance of vertical electric field due to positive gate voltage. We compare the performance and reliability of this new operation scheme with other traditional ones used on p-channel flash memories: Band-To-Band Induced Hot Electron Injection (BBHE) and Drain Avalanche Hot Electron Injection (DAHE). Finally, in order to improve the major reliability problem of p-channel flash memories, the drain disturb, we propose an alternative way to solve the problem by applying a moderate negative substrate bias on unselected cell, but a new structure is needed.

誌謝

光陰荏苒，兩年的碩士生涯，就在汗水和歡笑的交織中渡過了。在這段時間當中，我們不斷的努力，從修課到研究，再由研究到實驗，經過了理論與實務上的磨合，最後將其大成集結成我們的畢業論文。

在這段研究生涯中，首先要感謝我的老師莊紹勳教授給我的諄諄教誨，使我從一個非本科系生，完全沒有基礎的情況下，進步到能夠研讀期刊會議論文而進一步具有獨自進行研究的能力。除了學業上的收穫外，也讓我學習到研究應有的正確態度及精神，尤其是面對問題的處理能力，讓我受用無窮。

實驗室生活相當的充實，首先要感謝研究群的所有夥伴：亞峻、元亨和大正，除了在研究中互相討論彼此切磋以及心得成果分享外，還在閒暇的休閒生活中充分的放鬆自己，以迎接日後的挑戰。此外要感謝同實驗室汪老師研究群的博士班學長：志昌、小馬和俊榮，在我還不熟悉實驗儀器操作以及在研究遇到阻礙時，給予我適時的幫助及啟發，讓我的研究可以順利進行下去。

另外，我要感謝以前工作時的長官及同事，在你們身上，我學到了做事情該有的負責態度，以及面對挫折時要如何解決，還有如何在繁忙的工作中調配時間來加強自己的競爭力。有了這樣的歷練，我才能順利的轉換領域，朝自己的目標前進。我還要感謝茂德科技，有你們提供的樣品，才讓我的研究可以順利進行。

最後，我要感謝我的家人，你們是我精神上最大的支柱，有你們的支持，我才能順利完成學業。

謹將這份榮耀獻給關心我的大家。

Contents

Chinese Abstract i

English Abstract iii

Acknowledgement v

Contents vi

Table Captions viii

Figure Captions ix

Chapter 1 Introduction 1

1.1 The Motivation of This Work 1

1.2 Organization of the Thesis 2

Chapter 2 Device Fabrication and Experimental Setup 3

2.1 Device Fabrication 3

2.2 Experimental Setup 5

Chapter 3 Basic I-V Characteristics and Capacitance Coupling Model of P-channel Flash Memories 10

3.1 Threshold Voltage Determination 10

3.2 Capacitance Model and Extraction of Coupling Ratio 11

3.3 GIDL Measurement 12

Chapter 4 Mechanism and Characteristics of Forward Bias Assisted Drain Hot Electron Injection 18

4.1 Introduction 18

4.2 Mechanism of BBHE and DAHE in P-channel Flash cells 18

4.3 Mechanism of Forward Bias Assisted Drain Hot Electron Injection in N-Channel Flash Cell 27

4.4 Mechanism of Forward Bias Assisted Drain Hot Electron Injection in P-channel Flash Cell 32

4.5 Study of FBADHE at Sub-breakdown Region in P- Channel Flash Cell 44

4.6 Erase Operation in P- Channel Flash Cell 49

4.7 Comparison of Program/Erase Performance with Different Doped Poly Gate 53

Chapter 5 Reliability of FBADHE in P-channel Cells 56

5.1 Introduction 56

5.2 Program/Erase Cycling Endurance 56

5.3 Data Retention Characteristics 61

5.4 Disturb Characteristics 61

Chapter 6 Summary and Conclusions 73

References 74

Table Captions

Table 2.1 The spilt table of stacked-gate flash memories used in this study. 7

Table 4.1 Comparison of BBHE and DAHE in p-channel flash cells............. 26

Table 4.2 Operational table of P/E schemes in this study. 42

Figure Captions

Fig. 2.1 Sketch and cross-sectional view of stacked-gate flash memory. (a) Layout of experimental stacked-gate flash memory. (b) The cross-sectional view of the flash cell in the length direction. (c) The cross-sectional view of the flash cell in the width direction................ 4

Fig. 2.2 The experimental setup of the current-voltage and the transient characteristics measurement in flash cells. Automatic controlled characterizations system is setup based on the PC controlled instrument environment. 8

Fig. 2.3 The timing diagram of the trigged pattern mode method during (a) program and (b) erase operation. 9

Fig. 3.1 The schematic cross section of a flash cell showing a four-capacitance model... 14

Fig. 3.2 Energy band diagram of a flash cell at the onset of inversion (with no charge in the floating gate). 15

Fig. 3.3 The subthreshold characteristic of p-channel flash cell and dummy cell with different drain voltages. 16

Fig. 3.4 GIDL current of (a) dummy cell and (b) flash cell in p-channel cells. For flash cells, the more negative drain voltage causes electrons inject into floating gate by Band-to-Band tunneling so the threshold voltage shifts. 17

Fig. 4.1 Schematic illustration and Band Diagram of BBHE on the p-channel flash cells. Electrons in the valence band edge can move to conduction band, and holes are left on the valence band so that electron-hole pairs generate. 21

Fig. 4.2 Schematic illustration and band diagram of DAHE on the p-channel flash cells. Electrons with high energy cause more and more impact

ionization when moving from p (Drain) region to n (Substrate) region and generate a large amount of electron-hole pairs. 22
Fig. 4.3 The relationship of gate voltage and gate current under the condition of drain junction breakdown on p-channel flash cells. 23
Fig. 4.4 The relationship of drain current, gate current and drain voltage under different gate voltage ((a) V_G= 0V and (b) 9V respectively). 24
Fig. 4.5 Injection efficiency (I_G/I_D) with different gate voltages. The injection efficiency of BBHE (~10^{-3}) is better than DAHE (~10^{-7}).................. 25
Fig. 4.6 The programming setup and pulse pattern of PASHEI in n-channel flash cells... 28
Fig. 4.7 The injection mechanism of PASHEI in n-channel flash cells. Substrate electron current inject to floating gate by the effect of vertical electric field. 29
Fig. 4.8 The programming setup and pulse pattern of FBASEI in n-channel flash cells. The substrate (p)-drain (n) diode in emitting period is in forward bias and is in reverse bias in collecting time......................... 30
Fig. 4.9 The characteristics of transient program of CHE, PASHEI and FBASEI in n-channel flash cells. 31
Fig. 4.10 The (a) program setup, pattern mode and (b) injection mechanism of PASHEI in p-channel flash cells. 36
Fig. 4.11 The combination of gate voltage and drain voltage of different programming schemes. 37
Fig. 4.12 I-V characteristics of drain (p)-substrate (n) diode. 38
Fig. 4.13 The I_G-V_D characteristics with different gate voltage in p-channel dummy cells. 39
Fig. 4.14 The gate current components by fixing the potential difference between gate and drain terminal. 40
Fig. 4.15 Transient program characteristics under different operational

conditions. 41
Fig. 4.16 Transient program characteristics with the assistance of drain forward bias under different operating conditions. 43
Fig. 4.17 The relationship of drain (a) forward bias ($V_{D\text{-}High}$) and (b) reverse bias ($V_{D\text{-}Low}$) on FBADHE in p-channel flash cells. 45
Fig. 4.18 The relationship of drain emitting time on FBADHE in p-channel flash cells. 46
Fig. 4.19 The relationship of pulse count and threshold voltage shift in p-channel flash cells.. 47
Fig. 4.20 The relationship of pulse count and threshold voltage shift in p-channel flash cells.. 48
Fig. 4.21 Experimental setup and band diagram of Channel F-N erase in p-channel flash cells.. 51
Fig. 4.22 Transient characteristics of Channel F-N erase in p-channel flash cells 52
Fig. 4.23 Transient program characteristics of FBADHE with different doped poly gate. 54
Fig. 4.24 Transient erase characteristics of FBADHE with different doped poly gate. 55
Fig. 5.1 Cycling endurance of BBHE and FBADHE in p-channel flash cells. Experimental setup: BBHE: V_G= 11V, V_D=-5V, programming time is 20 μs; FBADHE is: V_G= 6V, V_D=1 to -5.9V, programming time is 1 μs; Erase setup is: V_G= -10V, V_B= 6V and erase time is 10ms...... 58
Fig. 5.2 Cycling endurance of FBADHE with different doped gate in p-channel flash cells. Experimental setup: FBADHE is: V_G= 6V, V_D=1 to -5.9V, 1 μs; Erase setup is: V_G= -10V, V_B=5V, 10ms................................. 59
Fig. 5.3 Subthreshold characteristics comparison of fresh and cycled cells with different programming scheme. The subthreshold slope doesn't show

significant difference. 60
Fig. 5.4 Data retention characteristics of FBADHE and BBHE at 80°C in p+ poly gate flash cells. By the benefit of thick interpoly ONO and tunnel oxide, the data retention ability is excellent even in room temperature or high temperature (80°C).. 65
Fig. 5.5 Types of disturb in conventional NOR array. D.D. represents the drain disturb, which is along the Bit Line (BL1). G. D. represents the gate disturb, which is along the Word Line (WL1).. 66
Fig. 5.6 Gate disturb characteristics with different gate voltage in p-channel flash cells.. 67
Fig. 5.7 Read disturb characteristics with different doped floating gate in p-channel flash cells after 10k cycling. Read disturb is very small due to the thick dielectric layer.. 68
Fig. 5.8 Drain disturb characteristics with different gate voltage in fresh p-channel flash cells.. 69
Fig. 5.9 Drain disturb characteristics with different drain voltage before and after 10k cycling in p-channel flash cells. 70
Fig. 5.10 Improved NOR array with triple well technology for applying the substrate bias. 71
Fig. 5.11 Substrate disturb characteristics with different doped floating gate after 10k cycling in p-channel flash cells.. 72

Chapter 1
Introduction

1.1 The Motivation of This Work

With the introduction of the flash memory by MASUOKA in 1984 [1], the flash memory has become a significantly solid-state memory technology in the past two decades. Nowadays, there are two major applications of flash memories: embedded system and mass storage system.

In the past, many programming and erase schemes have been studied for the performance and the reliability issues. In n-channel, we adopt Channel Hot Electron Injection (CHEI) [6] for programming and Fowler-Nordheim tunneling (F-N tunneling) for erasing on NOR array. In NAND array, we use Fowler-Nordheim tunneling (F-N tunneling) for both programming and erasing schemes. Except the above programming/erase schemes, there are some other programming schemes which have been used in to n-channel flash memories, such as: Substrate Hot Electron Injection (SHEI) [2], Band-to-Band tunneling Induced Substrate Hot Electron (BBISHE) [3], and Channel Hot-electron Initiated Secondary Electron (CHISEL) [4]. For p-channel flash memories, Band-to-Band tunneling Induced Hot Electron Injection (BBHE) [5], Drain Avalanche Hot Electron (DAHE) [7], and Channel Hot Hole Initiated Hot Electron Injection (CHHIHE) [8] are usually proposed in reported papers.

Comparing with n-channel flash memories, only a few papers have been focused p-channel ones. Two kinds of flash memories have their own advantages. For p-flashes, faster programming speed, lower power consumption, and higher injection efficiency can be achieved. For n-flashes, better reliability is a huge advantage on widely using on data storage. In this paper, a comprehensive discussion of performance and reliability of a new

programming scheme, Forward Bias Assisted Drain Hot Electron Injection (FBADHE), used in p-channel flash memories will be investigated.

1.2 Organization of the Thesis

This thesis is divided into six chapters. In chapter 1, we introduce the motivation of this study. In chapter 2, we introduce the device fabrication, the setup of experimental instruments, and the basic principle of flash memories. The fundamental principle and basic I-V characteristics are shown in Chapter 3. The mechanism and characteristic of this new program scheme (FBADHE) is described in Chapter 4. The reliability of FBADHE on p-channel flash memories is discussed in Chapter 5. Finally, the summary and conclusion will be given in Chapter 6.

Chapter 2
Device Fabrication and Experimental Setup

This chapter is divided into two sections. First, the semiconductor devices we used in this study will be described. Second, the instrument setup and the experimental techniques to accurately control these instruments are illustrated.

2.1 Device Fabrication

Figure Fig. 2.1 (a) shows the layout of the stacked-gate flash memory cell. The cross-sectional views of the cell in length and width directions are illustrated in Fig. 2.1 (b) and Fig. 2.1 (c), respectively. As shown in Fig. 2.1 (a), the floating gate and the control gate are patterned in the width direction and the length direction, respectively. Moreover, in order to increase the gate coupling coefficient, the floating gate is extended into the Local Oxidation of Silicon (LOCOS) region for enhancing the fringing capacitance between control-gate and floating-gate as given in Fig. 2.1 (c). In addition, the dummy cell with shorted floating gate and control gate is also used for gate current measurement.

In this work, the flash memory cells are prepared based on standard 0.35μm CMOS process and with conventional stacked-gate flash memory cell. The cell sizes which we used in this study are W/L = 0.7/0.65 μm and 0.7/0.45 μm respectively. The floating gate width W_{FG} = 1.2 μm, tunnel oxide thickness T_{OX} = 90 Å, and effective ONO inter-poly dielectric thickness T_{ONO} = 165 Å (The individual thickness of interpoly ONO is $T_{Top\text{-}Oxide}$/ T_{SiN}/ $T_{Bottom\text{-}Oxide}$=65Å/90Å/55 Å). Interpoly dielectric thickness heavily influences program/erase speed and the magnitude of read current for an industry-standard flash cell [8]. MDD

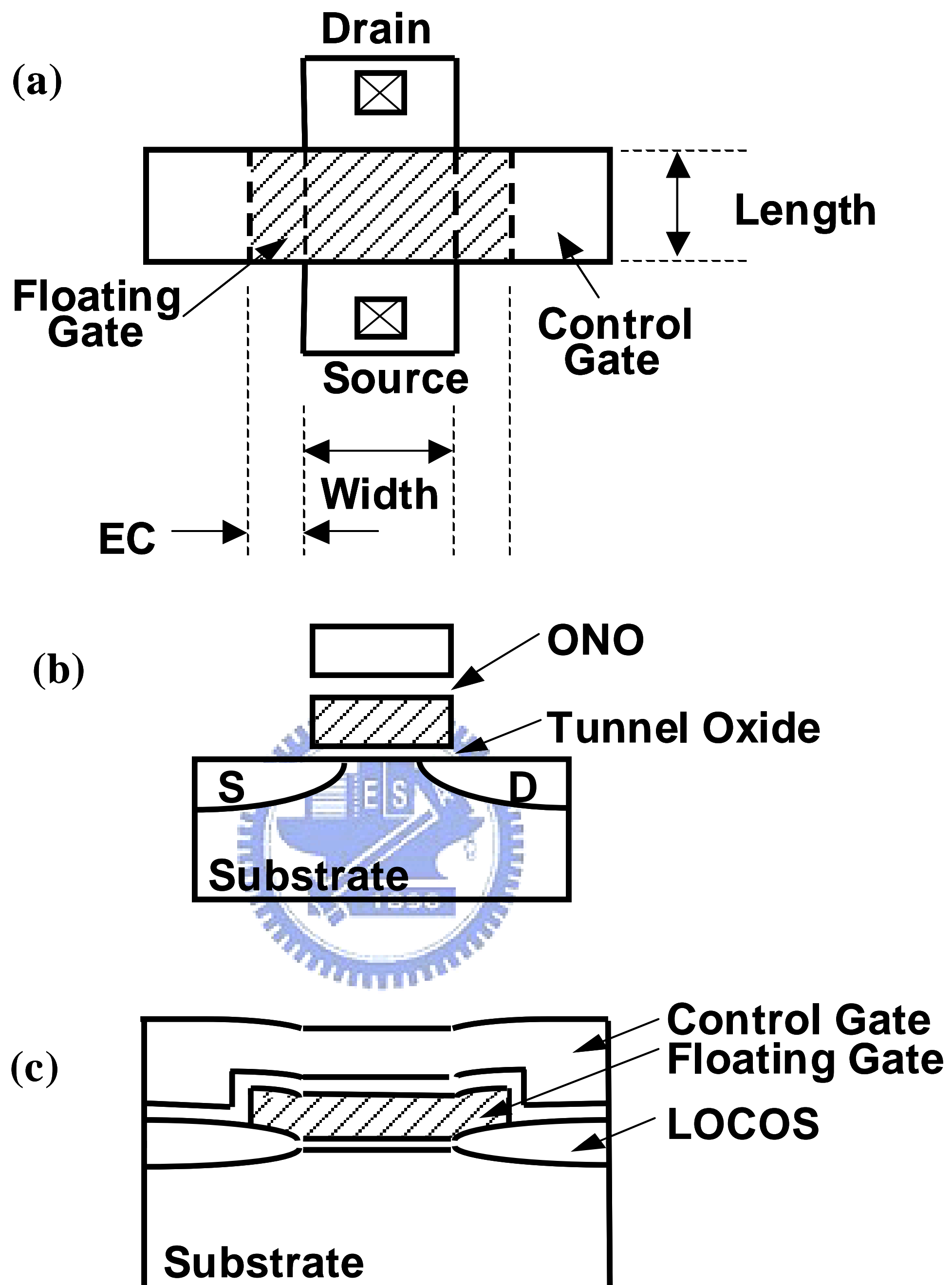


Fig. 2.1 Sketch and cross-sectional view of stacked-gate flash memory. (a) Layout of experimental stacked-gate flash memory. (b) The cross-sectional view of the flash cell in the length direction. (c) The cross-sectional view of the flash cell in the width direction.

(Medium Doped Drain) is used for applying the asymmetric electric field. There are two different doping types of floating gate. The detailed device parameters are summarized in Table 2.1.

2.2 Experimental Setup

The experimental setup for the I-V and transient characteristics measurement of flash memory is illustrated in Fig. 2.2. Based on the PC controlled instrument environment via HP-IB (GP-IB, IEEE-488 Standard) interface, the complicated and long-term characterization procedures during analyzing the intrinsic and degradation behaviors in flash memory cells can be easily achieved. As shown in Fig. 2.2, the characterization equipments, including semiconductor parameter analyzer (HP 4156C), dual channels pulse generator (HP 8110A), low leakage switch mainframe (HP5250A Switching Matrix), and a probe station, provide an adequate capability for measuring the device I-V characteristics and carrying out the flash cell program/erase operations.

The HP 4156C is equipped with four programmable source/monitor units, two source units, and two monitor units for supplying or monitoring the voltage and the current. It provides a high current resolution up to pico-ampere range which facilitates the very low current measurement such as gate current measurement, subthreshold characteristics extraction, DCIV and GIDL measurement for oxide damage characterization. The dual channels HP 8110A with high timing resolution provides two different pulse levels simultaneously for transient and P/E cycling endurance characterization. The Switching Matrix, which is equipped with an 8-input x 12-outoupt switching matrix (with one output module), switches the signals from the HP 4156B and the HP 8110A to device under test (DUT) in probe station, automatically. In addition, the computer language-HP VEE is used to achieve the PC control of these measurement instruments via HP-IB interface.

In order to precisely control the pulse timing of HP 8110A during transient and P/E cycling endurance characterization, we select the triggered pattern mode to achieve this requirement. Fig. 2.3 (a) and Fig. 2.3 (b) show the timing diagrams of the triggered pattern mode method during program and erase operations, respectively. By taking the program timing diagram as an example, the triggered pattern mode method can be explained as follows. In Fig. Fig. 2.3 (a), the VSMU1 of HP 4156B (illustrated in Fig. 2.2) generates a voltage signal which is equal to the low voltage of HP 8110A. This approach can provide a substrate bias during programming and prevent additional stress to device during programming. The patterned pulses (defined as 01000 in Fig. 2.3 (a) with the minimum period of 10ns) from HP 8110A are then sent and programming of a flash memory cell is performed. In our experience with the HP 8110A, we must set the period of pulse larger than 50 ~ 100ns to obtain a good square pulse and the leading or trailing edge larger than 25 to 50 ns to prevent the over-shoot and under-shoot of output pulse.

	P	N
T_{ox} (Å)	90	
T_{ONO} (Å)	T.O/SiN/B.O=65/90/55	
Floating Gate Doping	BF_2 20keV 5E13	In-Situ

Table 2.1 The spilt table of stacked-gate flash memories used in this study.

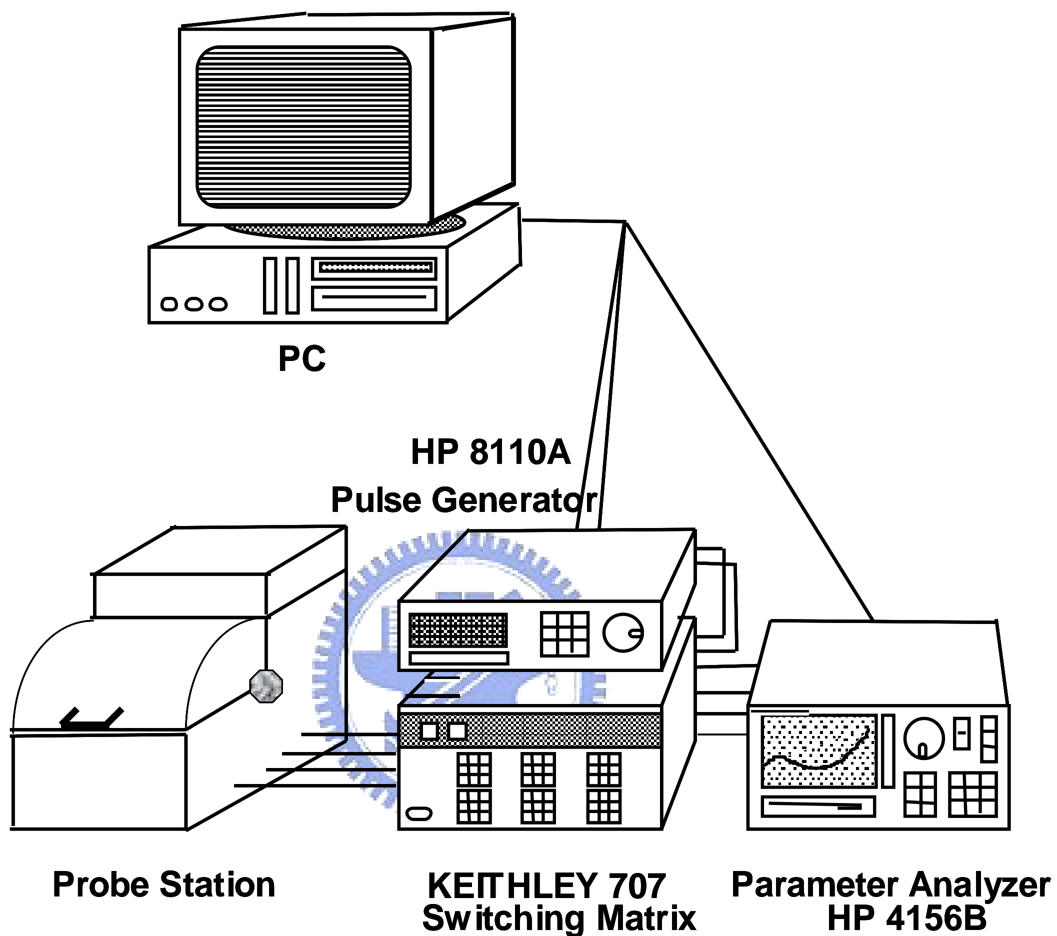
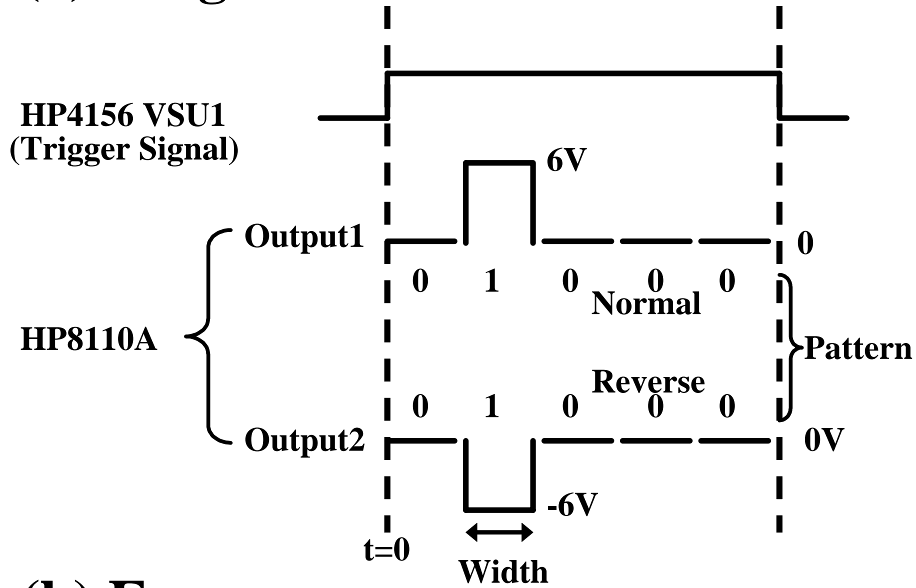


Fig. 2.2 The experimental setup of the current-voltage and the transient characteristics measurement in flash cells. Automatic controlled characterizations system is setup based on the PC controlled instrument environment.



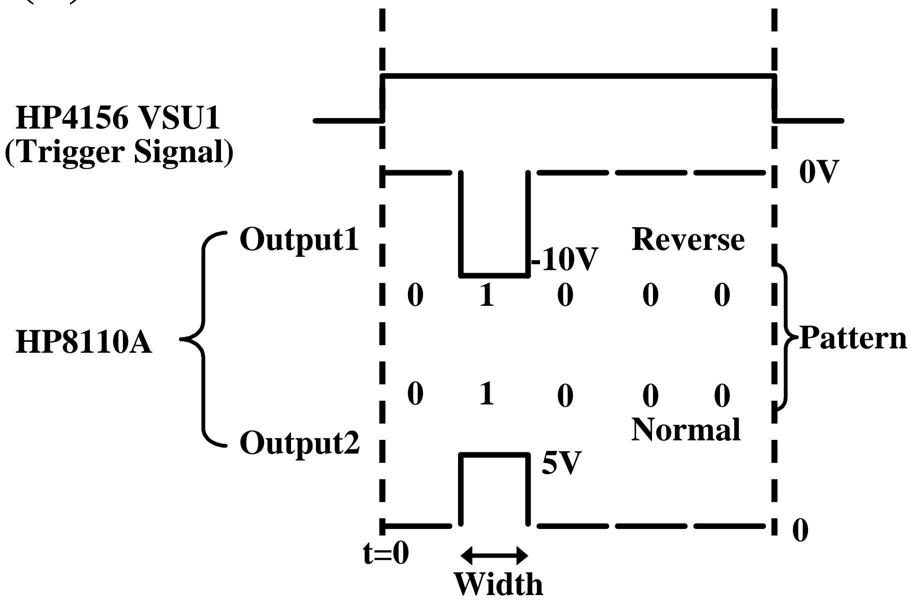
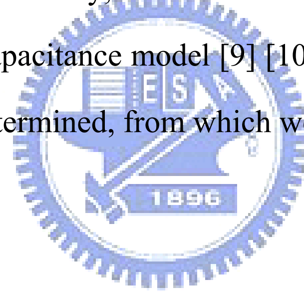


Fig. 2.3 The timing diagram of the trigged pattern mode method during (a) program and (b) erase operation.

Chapter 3
Basic I-V Characteristics and Capacitance Coupling Model of P-channel Flash Memories

This Chapter involves the extraction of MOSFET device basic parameters, threshold voltage, and the coupling model of floating gate memories to extract the parameter. Finally, we show the basic I-V Measurement of flash cells and dummy cells.

In floating gate memory, the floating gate is fully isolated. The voltage of floating gate (V_{FG}) is determined by the coupling of the voltages of other terminals (Control-Gate, Substrate, Source and Drain). Usually, the flash memory cell is modeled by a capacitance equivalent circuit, called the capacitance model [9] [10]. Based on this model, the equation of floating-gate voltage can be determined, from which we can extract the coupling ratio.



3.1 Threshold Voltage Determination

In this study, the threshold voltage V_{TH} is determined from its I_D-V_G characteristic. Based on the following drain current equation [11]:

$$I_{D0} = \frac{W}{L} \mu \frac{\varepsilon_{ox}}{t_{ox}} (V_{GS} - V_{TH}) V_D \, , \tag{3.1}$$

in which we can choose an arbitrary drain current level, and set the gate voltage at this drain current level as the device threshold voltage. However, by taking the device geometry into consideration, we need to normalize the device drain current with its channel length and width. This can be done by transforming Eq. (3.1) into

$$I_D = (I_{D0})/(\frac{W}{L}) = \mu \frac{\varepsilon_{ox}}{t_{ox}} (V_{GS} - V_{TH}) V_D . \quad (3.2)$$

Here, we set the threshold voltage as the gate voltage at which the normalized current (I_{D0}) is 0.1μA.

3.2 Capacitance Model and Extraction of Coupling Ratio

It is required for an accurate prediction of the associated coupling ratio in the design of the flash memory cell. We will describe how to extract these parameters in this section. Besides, we will derive the equations for calculating the floating gate potential.

Fig. 3.1 shows the conventional capacitance model for the flash memory cell. According to this model, the charge balance of floating gate can be derived based on the electrostatic potential Ψ in the various regions of the cell. From

Fig. 3.2, we know that the electrostatic potential Ψ equals to (V+ψ) where V is applied voltage and ψ is flat-band voltage for each terminal, respectively. However, the electrostatic potential Ψ_{FG} of floating gate is equal to the voltage V_{FG}. Therefore, the floating gate voltage can be expressed as the following equation

$$V_{FG} = \alpha_{CG}(V_{FG} - \psi_{CG}) + \alpha_B(V_{FG} - \psi_B) + \alpha_D(V_{FG} - \psi_D) + \alpha_S(V_{FG} - \psi_S) - \frac{Q_{FG}}{C_{TOTAL}} \quad (3.3)$$

where

$$C_{TOTAL} = C_{CG} + C_B + C_D + C_S \quad (3.4)$$

and

$$\alpha_X = \frac{C_X}{C_{TOTAL}} . \quad (3.5)$$

Here, α_{CG} is the gate coupling coefficient, α_B is the substrate coupling coefficient, α_S is the

source coupling coefficient, and α_D is the drain coupling coefficient. Then we assume ($\alpha_{CG}\psi_{CG}$+ $\alpha_B\psi_B$+ $\alpha_S\psi_S$+ $\alpha_D\psi_D$) = V_0 ≒ $\alpha_{CG}\psi_{CG}$ = V_{FB}, which is the flat-band voltage of the flash memory cell, and V_B = 0 as the reference by setting V_S and Q_{FG} to zero, the floating gate voltage cab be rewritten as

$$V_{FG} = \alpha_{CG} V_{CG} + \alpha_D V_D + V_{FB} . \tag{3.6}$$

The coupling coefficients will be determined in Eq. (3.6). Fig. 3.3 shows the subthreshold characteristics of flash and dummy cells at drain voltage of 0.1 and 1V. Dummy cells show the identical subthreshold characteristic but flash cells are different. This is because the effect of coupling between electrodes. By selecting any two different drain current level (such as I_{DS} = 10^{-5} and 10^{-6} A) at V_{DS} = 0.1V, the control gate coupling coefficient can be obtained as

$$\alpha_{CG} = \frac{V_{FG}(I_{DS} = 10^{-5}\,A) - V_{FG}(I_{DS} = 10^{-6}\,A)}{V_{CG}(I_{DS} = 10^{-5}\,A) - V_{CG}(I_{DS} = 10^{-6}\,A)} . \tag{3.7}$$

Since α_{CG} is determined from Eq. (3.7), the value of V_{FB} can also be determined from V_{FG}-$\alpha_{CG}V_{CG}$ at any fixed drain current level of V_{DS} = - 0.1V.

3.3 GIDL Measurement

Gate Induced Drain Leakage Current (GIDL) is one of the major approaches to detect the damage of the oxide and silicon interface in the gate-drain overlap region. [12] [17]. The GIDL current is due to direct band-to-band tunneling in the Off-state MOSFETs. For flash memories, the effect of Band-to-Band tunneling can be also used to inject charges into floating gate [5]. Please refer to section 4.2 for the detailed description of Band-to-Band tunneling.

Fig. 3.4 Shows the GIDL measurements of P-channel flash and dummy cells. For flash cells, the more negative drain voltage causes electrons to be injected into floating gate via Band-to-Band tunneling and then gives rise to the threshold voltage shifts.

The GIDL current is sensitive to trapped charge in the tunneling oxide at the overlap region between the gate and drain, so that we can monitor the qualitative or the quantitative analysis of these Q_{ox} and N_{it}. The analytic model of the GIDL current [18] is shown as follows:

$$I_{GIDL} = A \times E_{surface} \times \exp\left(\frac{-B}{E_{surface}}\right) \tag{3.8}$$

where $E_{surface}$ is the electric field at silicon interface. The parameters A and B are defined in ref. [18], and basically they are the fitting parameters in most measurements.

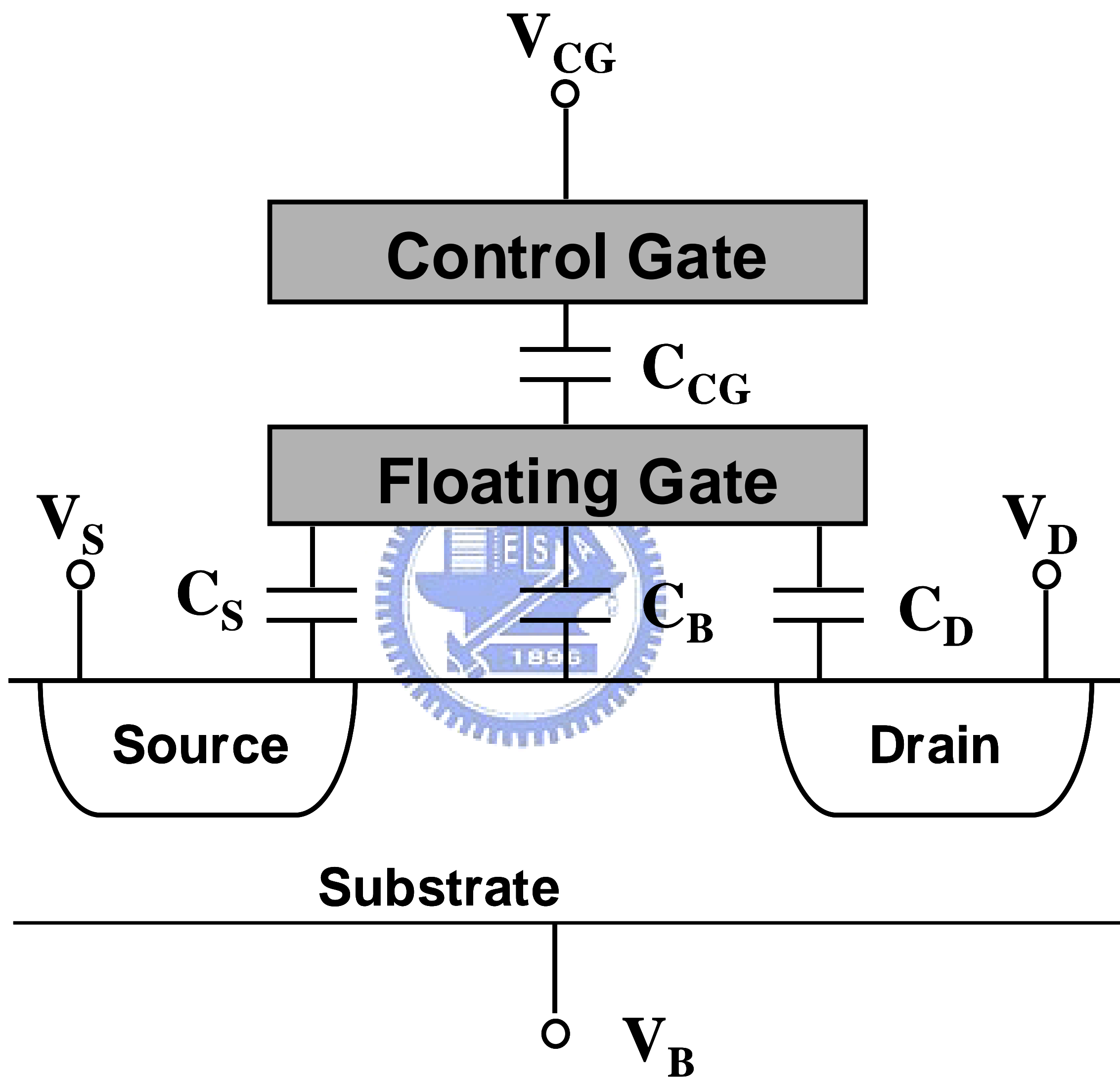


Fig. 3.1 The schematic cross section of a flash cell showing a four-capacitance model.

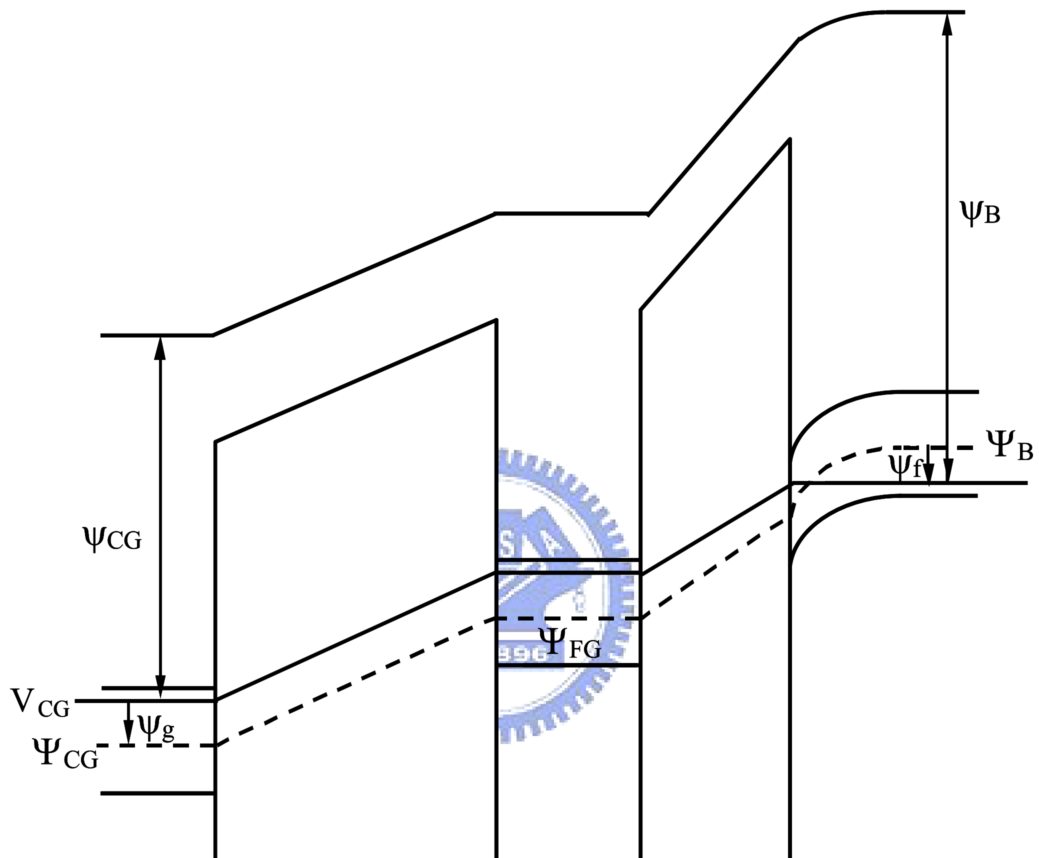


Fig. 3.2 Energy band diagram of a flash cell at the onset of inversion (with no charge in the floating gate).

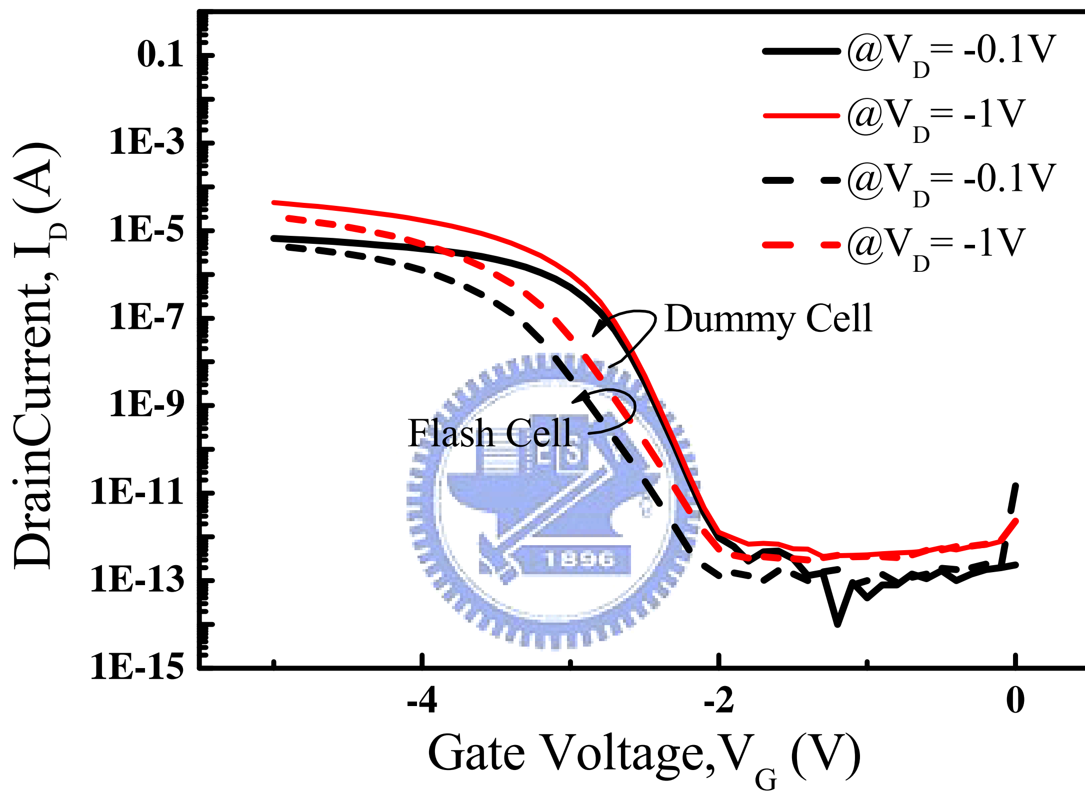


Fig. 3.3 The subthreshold characteristic of p-channel flash cell and dummy cell with different drain voltages.

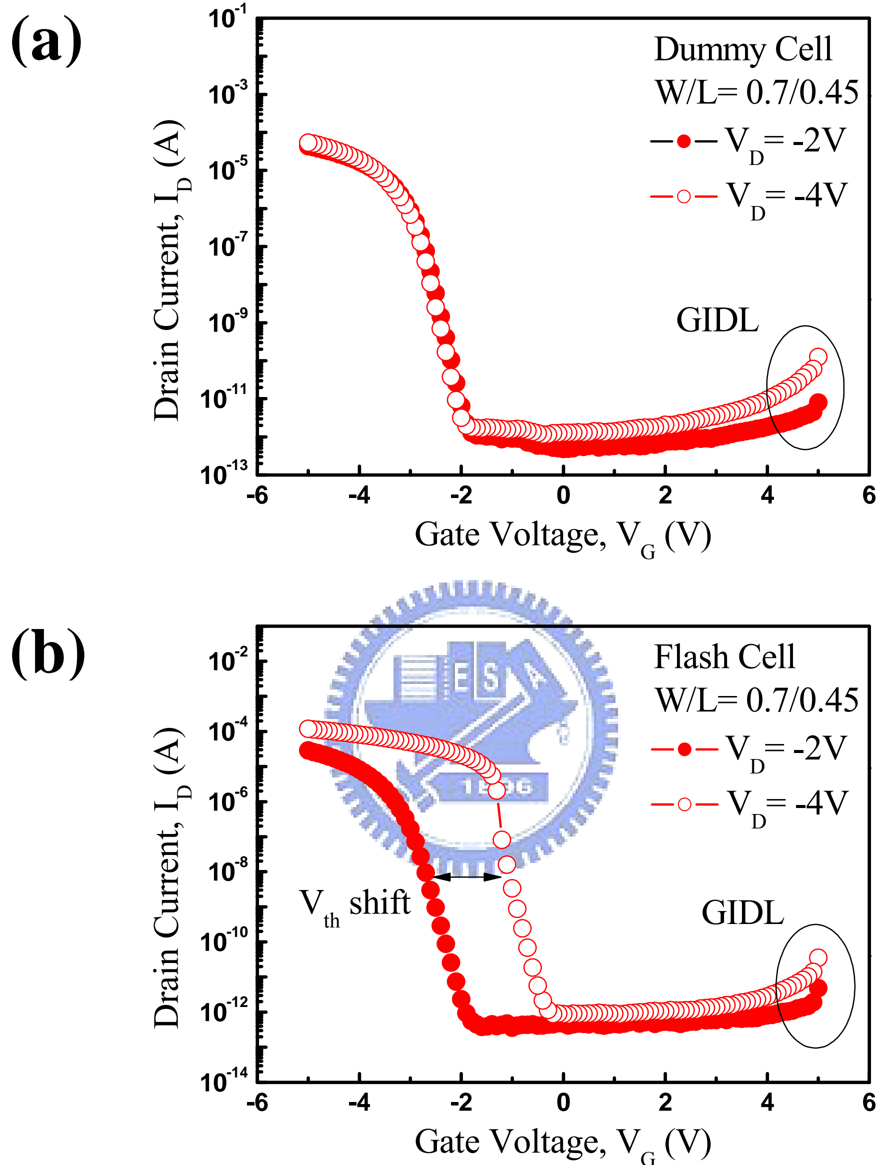
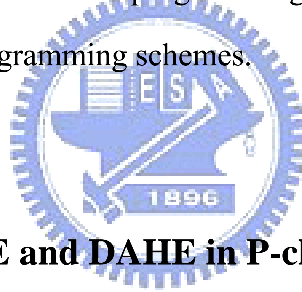


Fig. 3.4 GIDL current of (a) dummy cell and (b) flash cell in p-channel cells. For flash cells, the more negative drain voltage causes electrons inject into floating gate by Band-to-Band tunneling so the threshold voltage shifts.

Chapter 4
Mechanism and Characteristics of Forward Bias Assisted Drain Hot Electron Injection

4.1 Introduction

In this chapter, first we describe the mechanism of familiar program schemes used in p-channel flash cells, Band-to-Band tunneling Hot Electron Injection (BBHE), and Drain Avalanche Hot Electron Injection (DAHE). Then, we propose a new program scheme called Forward Bias Assisted Drain Hot Electron Injection (FBADHE). We will investigate the mechanism and the effect on different programming conditions. Finally, we compare the characteristics of the above programming schemes.



4.2 Mechanism of BBHE and DAHE in P-channel Flash cells

In the past, BBHE and DAHE are the popular methods used for programming in P-channel Flash cells. In this section, we will investigate the difference of mechanism between them.

First, we discuss Band-to-Band tunneling Hot Electron Injection [5]. BBHE is used in P-channel Flash cells. For n-channel cells, if we change the structure of MOSFET, i.e., buried a p+ region in the channel, BBHE can also be achieved [7].

When a negative drain voltage and a positive control-gate voltage are applied, the energy band is bended by the difference of the two side of the insulator and band-to-band tunneling occurs. The electron-hole pair generates in the gate-drain overlap region [12]. Band-to-Band

tunneling means that: in the strong band bending, electrons in the valence band edge can move to conduction band, and holes are left on the valence band so that electron-hole pairs generate. Generated carriers will be accelerated by the horizontal electric field and some of them gain sufficient energy. The energetic electrons then inject into floating gate by the effect of vertical field. Fig. 4.1 shows the schematic illustration and the band diagram of BBHE.

Drain Junction avalanche hot carrier (DAHC) historically denotes the emission of free carriers towards the gate generated by impact ionization in the deep depletion region of p-n junction. In p-channel flash cells, when a large negative drain voltage is applied and substrate is grounded, the drain-substrate p-n junction is in strong reverse-bias, the avalanche breakdown occurs. Electrons with high energy cause more and more impact ionization when moving from p region to n region and generate a large amount of electron-hole pairs. Carriers are accelerated by the strong electric field in depletion region so that the energy of these carriers is high enough to surmount the oxide barrier and inject into floating gate even without the assistance of vertical electric field. Fig. 4.2 shows the illustration and band diagram of DAHE. Vertical field determines which carrier will be injected. We measure the relationship of gate voltage and gate current at the condition of drain avalanche breakdown, as shown in Fig. 4.3. We can easily appear that only in the strong negative gate bias is applied (~-8V) that hot holes will inject into floating gate. In other condition, electron current is dominated. So the DAHE Programming Scheme is easily to be achieved in p-channel flash cells. DAHE is a high speed, self-convergent programming scheme [20]. Besides, if a positive gate voltage and a strong drain voltage are applied, a large electric field is generated in the gate-drain overlap region. Band-to-Band tunneling will occurs and electron will inject into floating gate. It is called Gate Induced Drain Leakage (GIDL) current [12].

From the above description, it is realized that the mechanism of BBHE and DAHE is quite different. First, the generation region of carriers (electron-hole pair) of BBHE is in the gate-drain overlap region, and the DAHE is in the deep depletion region of drain-substrate

junction. Second, the energy of generated carriers of BBHE is lower than those in DAHE. Thus the stronger vertical field is needed for BBHE in order to surmount the oxide barrier. The advantage of BBHE is high injection efficiency (I_G/I_D). Fig. 4.4 shows the relationship of I_G, I_D and V_D under the BBHE and DAHE condition respectively. In Fig. 4.4 (a), the Gate current rises at the drain junction breakdown (~-7V). It is realized as DAHE injection. In Fig. 4.4 (b), the Gate current involves F-N tunneling current and band-to-band tunneling current. The Gate current rises as the increasing of band bending. Because the F-N tunneling current is fixed, the increasing part is band-to-band tunneling current. Fig. 4.5 shows the injection efficiency (I_G/I_D) of BBHE and DAHE. The efficiency of BBHE ($\sim 10^{-3}$) is high than DAHE ($\sim 10^{-7}$). It is noticed that in the injection efficiency has a maximum at V_D= 0V. It is because there is a little tunneling current ($\sim 10^{-11}$@ V_G= 10V) but no drain current ($\sim 10^{-13}$@ V_G= 10V). So, the magnitude of gate and drain current should be also notified when measuring the injection efficiency. The higher efficiency means the less power consumption. It is noticed than the definition of injection efficiency (I_G/I_D). For CHEI Programming in N-Channel Flash cells, The MOS is turned on, so the I_D is drain current. On the other hand, using either BBHE or DAHE on p-channel flash cells, the MOS isn't turn on, so the I_D is drain leakage current. The other advantage of BBHE is the better drain disturb characteristics. The serious problem in reliability of p-channel flash is drain disturb [21], which impedes its practical applications for mass production purpose. Larger drain voltage causes more serious drain disturbance. The comparison of BBHE and DAHE is listed in Table 4.1.

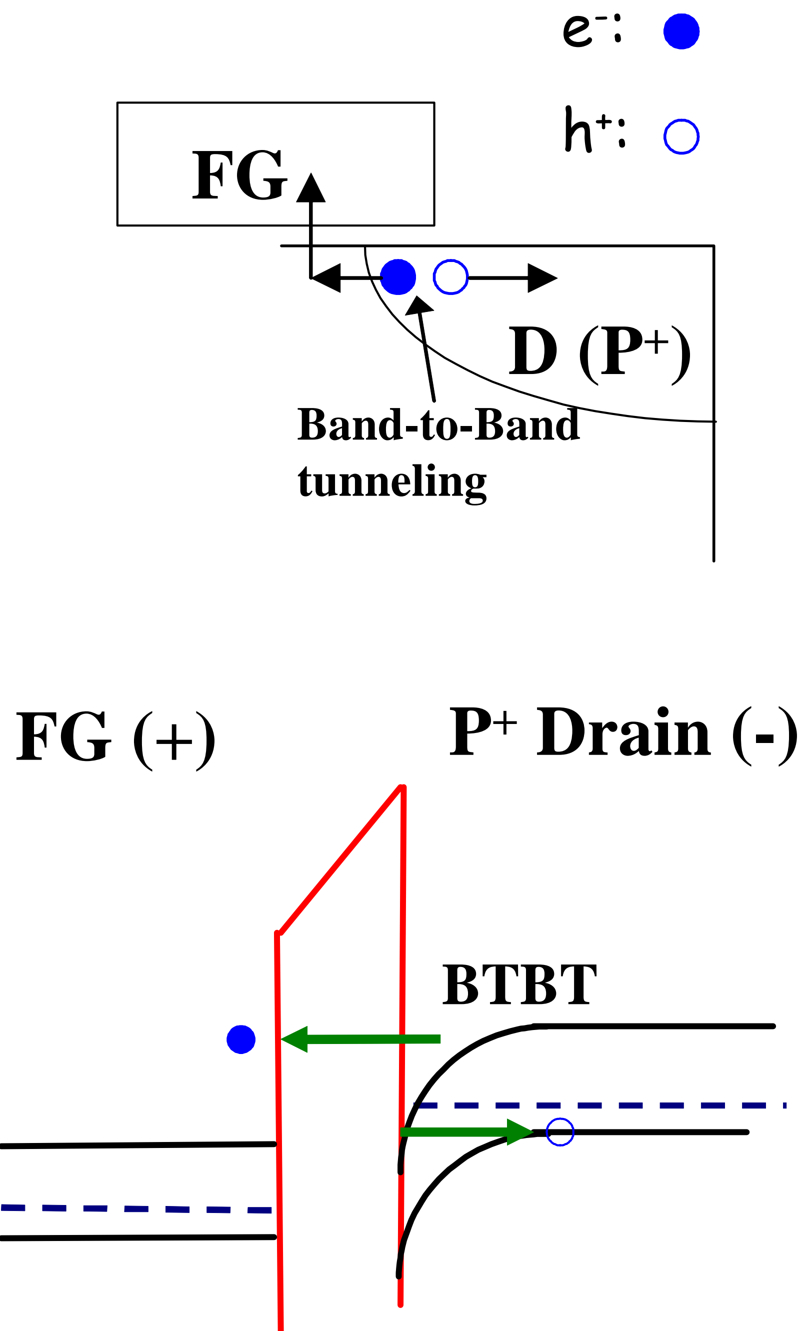


Fig. 4.1 Schematic illustration and Band Diagram of BBHE on the p-channel flash cells. Electrons in the valence band edge can move to conduction band, and holes are left on the valence band so that electron-hole pairs generate.

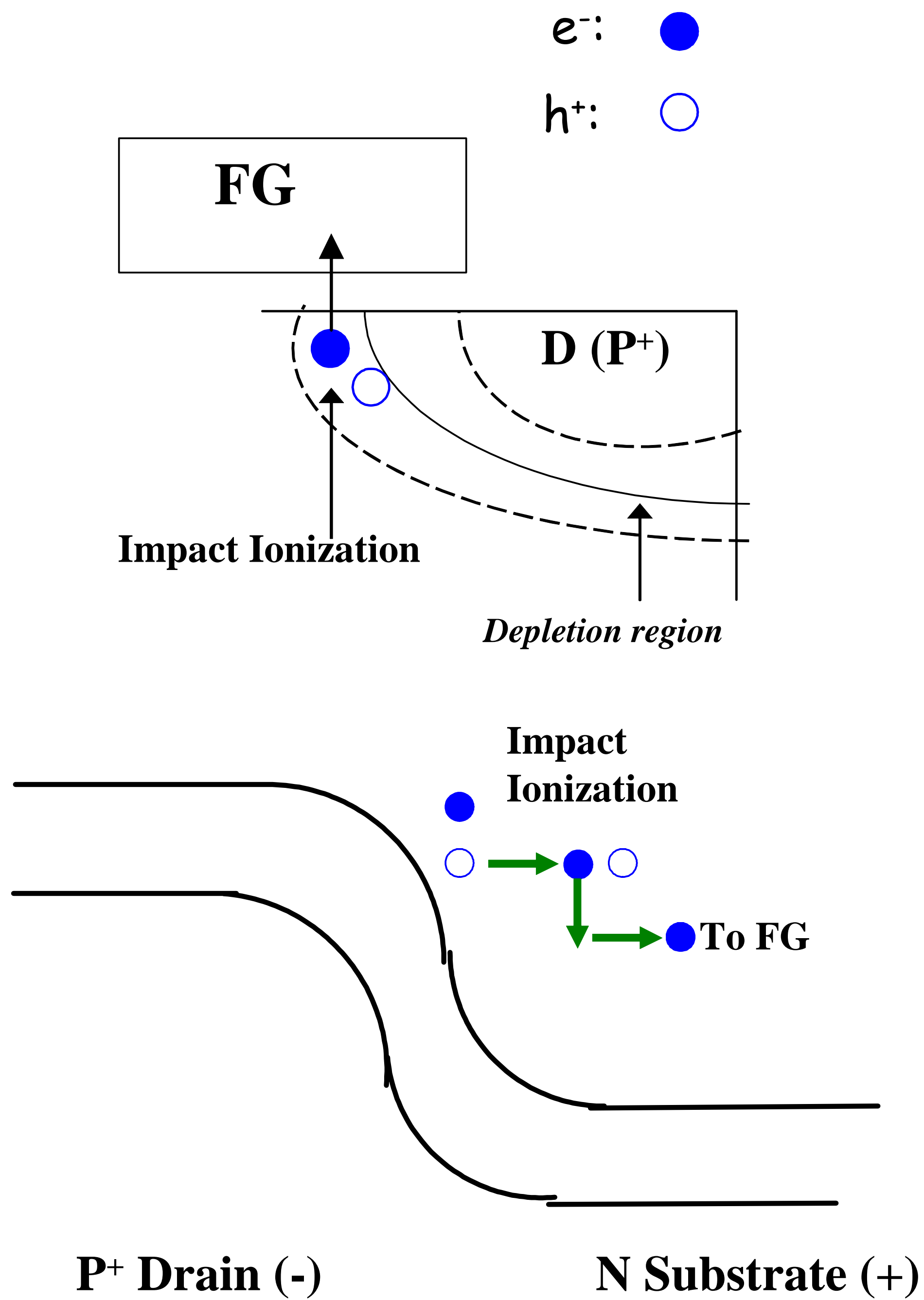


Fig. 4.2 Schematic illustration and band diagram of DAHE on the p-channel flash cells. Electrons with high energy cause more and more impact ionization when moving from p (Drain) region to n (Substrate) region and generate a large amount of electron-hole pairs.

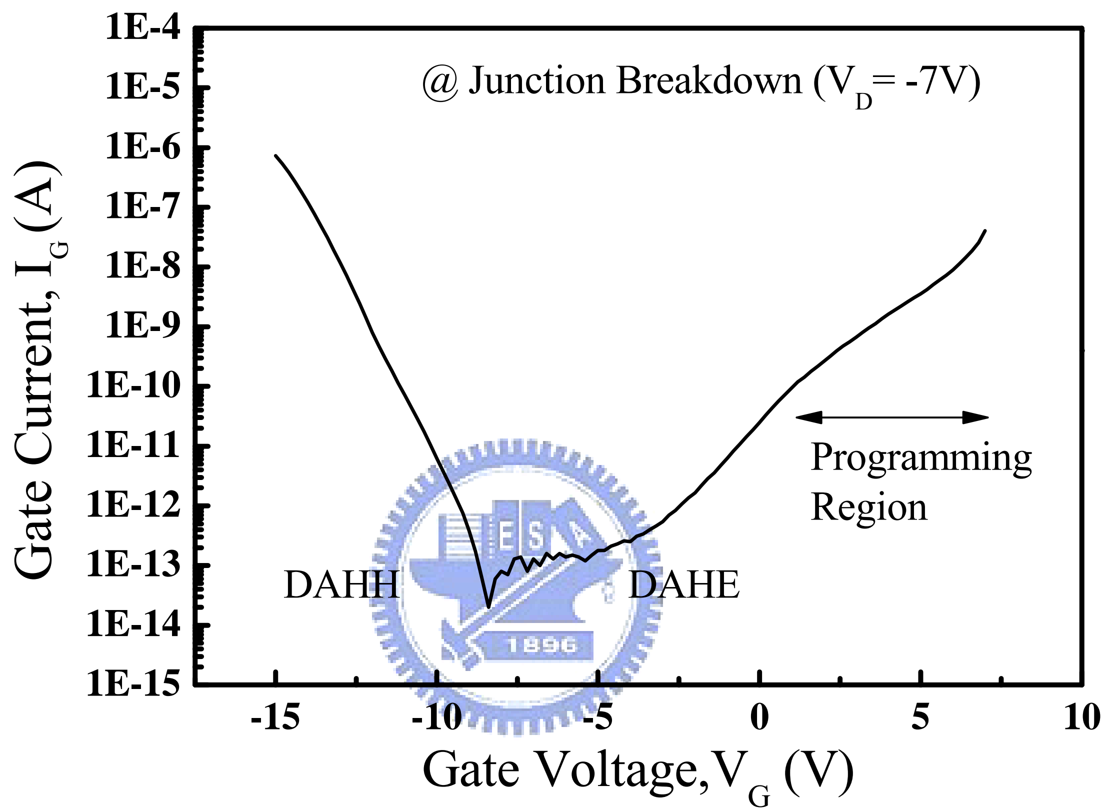


Fig. 4.3 The relationship of gate voltage and gate current under the condition of drain junction breakdown on p-channel flash cells.

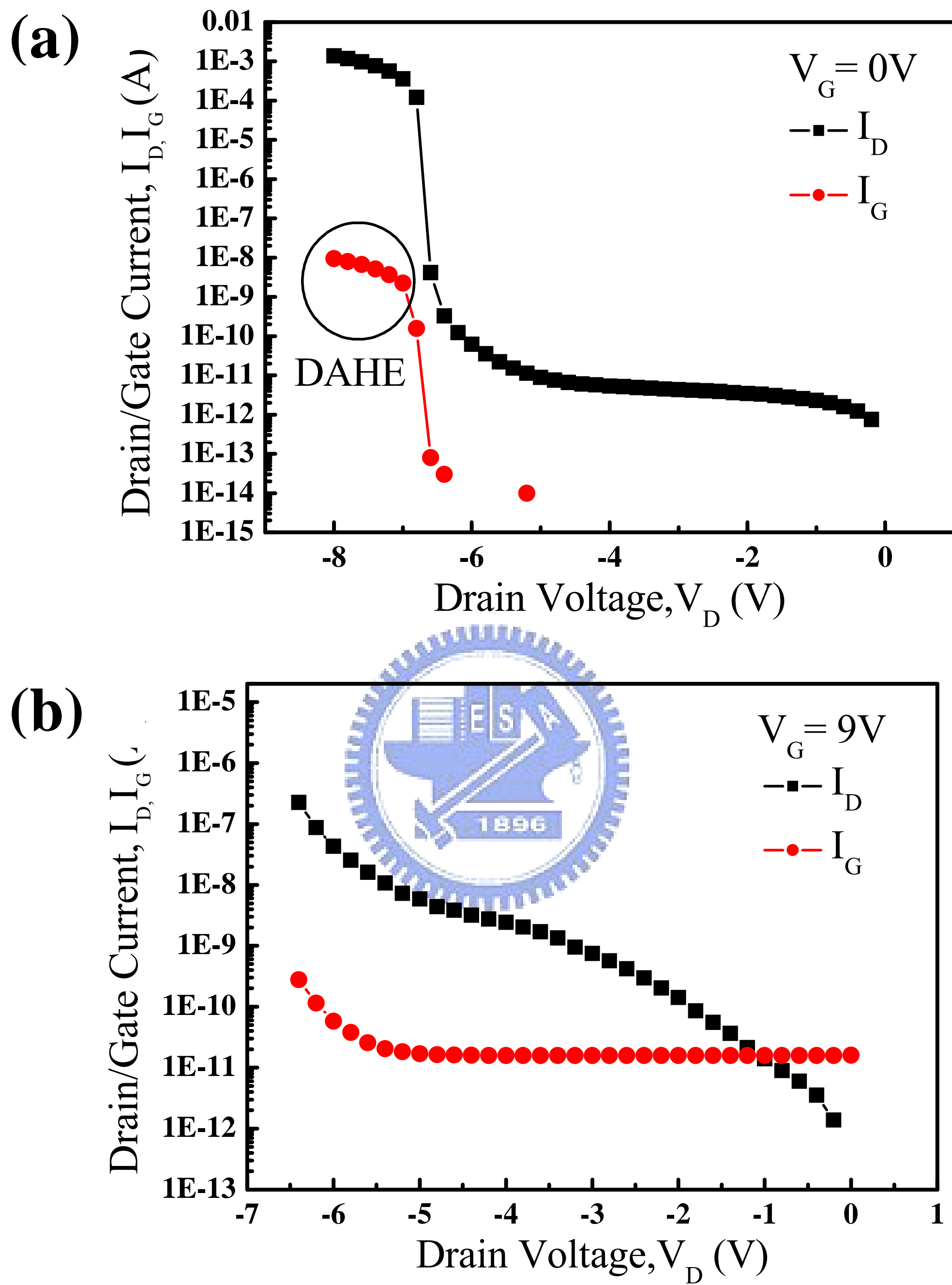


Fig. 4.4 The relationship of drain current, gate current and drain voltage under different gate voltage ((a) V_G= 0V and (b) 9V respectively).

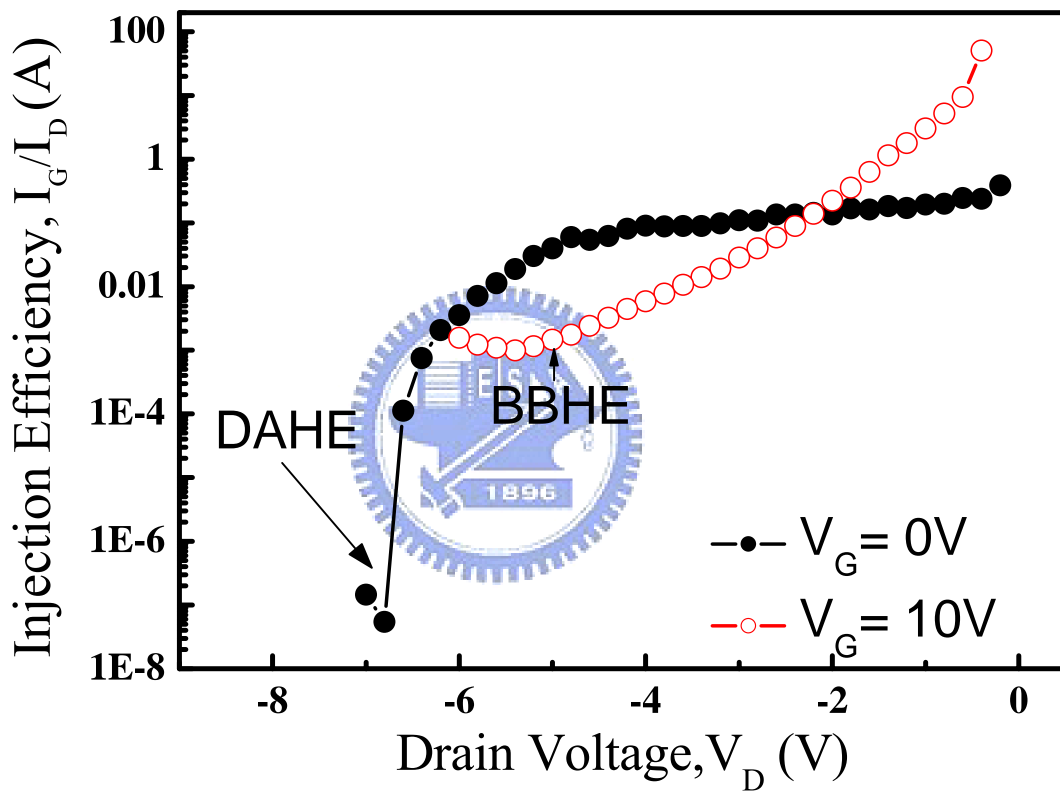


Fig. 4.5 Injection efficiency (I_G/I_D) with different gate voltages. The injection efficiency of BBHE (~10^{-3}) is better than DAHE (~10^{-7}).

	BBHE	**DAHE**
Program Speed	Slow	Fast
Fast Self-convergence	No	Yes
Gate Disturb	Not serious	No
Drain Disturb	Worse	Worst

Table 4.1 Comparison of BBHE and DAHE in p-channel flash cells.

4.3 Mechanism of Forward Bias Assisted Drain Hot Electron Injection in N-Channel Flash Cell

The concept of forward bias assisted programming is proposed by Z. Liu in 1999 [13]. The sample is N-Channel floating gate cells. This program scheme is called "Pulse Agitated Substrate Hot Electron Injection" (PASHEI). The illustration of Program setup is shown in Fig. 4.6. The programming pulse is divided into two parts, emitting time and collect time. The emitting is used to make the substrate-drain diode in forward bias (a negative drain voltage V_e is applied) and the diffusing current is taking place. Then, changing the diode to reverse bias mode (a positive drain voltage V_{wr} is applied), and the electrons which are flowing into substrate will be injected into floating gate. The gate voltage V_G keeps positive during emitting and collecting time. Please refer to Fig. 4.7 for the injection mechanism. The advantage of this programming technique is only a small programming voltage (~5V as gate voltage and 4V as drain writing voltage) is needed. The disadvantage is a multiple pulse is needed in a program procedure, such as: 100 pulse per programming. It makes the long period of programming time and a complicated circuit design is needed in order to apply this continuous pulse.

We modify the pulse pattern in this study. As shown in Fig. 4.8. We call the new scheme "Forward Bias Assisted Substrate Electron Injection" (FBASEI). The difference is that the gate voltage is zero during the emitting period and is positive during collecting period. The mechanism is similar as PASHEI, so the injection current is from the substrate too. The performance comparison of CHE, PASHEI and FBASEI is shown in Fig. 4.9. The characteristic of FBASEI is also pulse count dependent, pulse period independent, and low operating voltage is achieved [13]. The reason of poorer programming speed of FBASEI (comparing with PASHEI) is that the less gate pulse period makes fewer electrons injected into the floating gate.

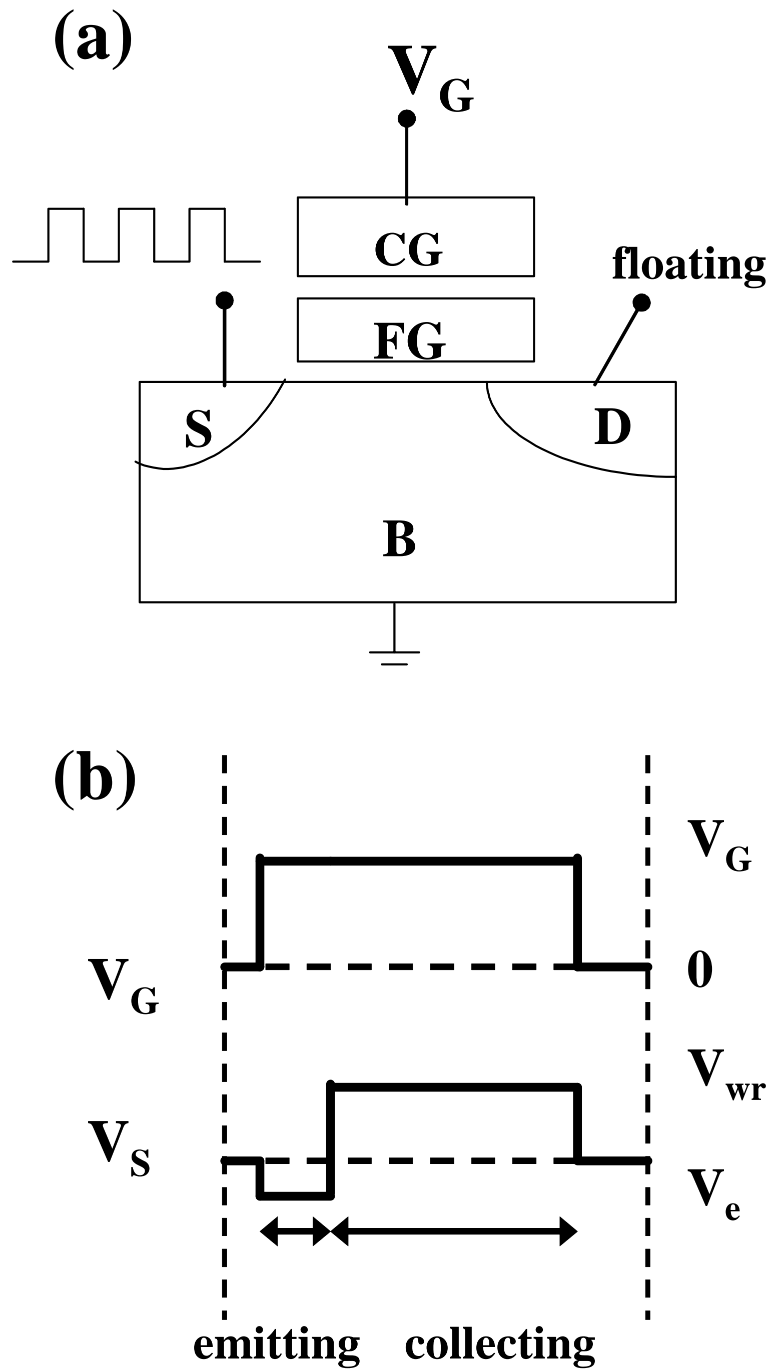


Fig. 4.6 The programming setup and pulse pattern of PASHEI in n-channel flash cells.

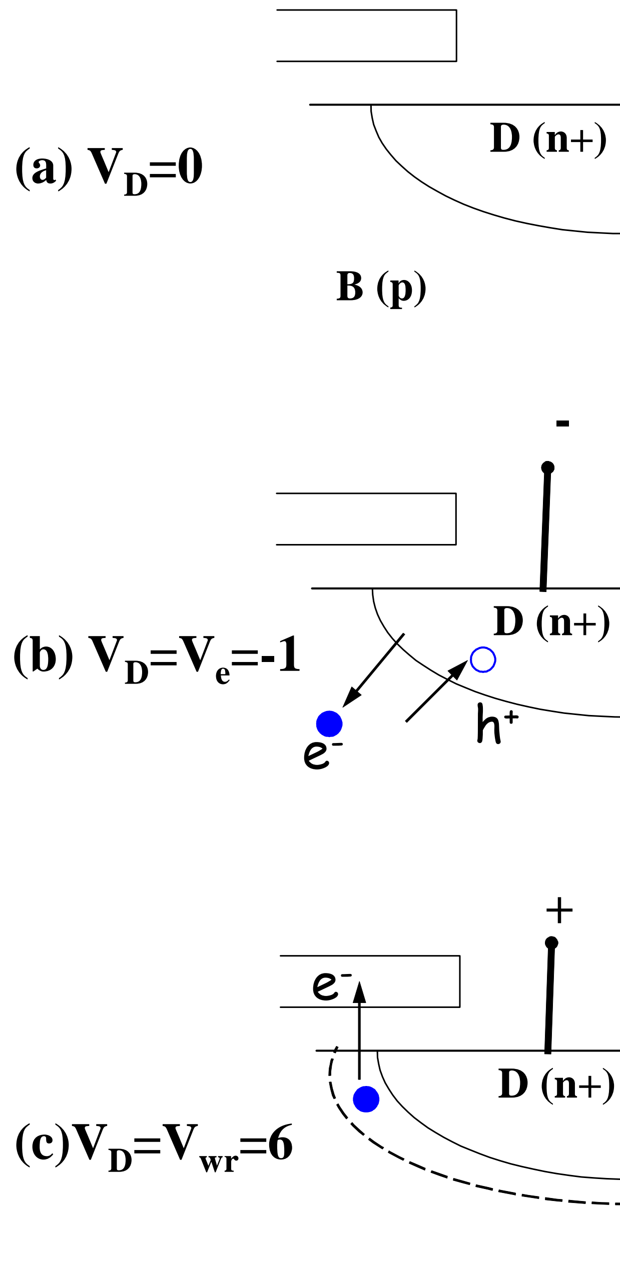


Fig. 4.7 The injection mechanism of PASHEI in n-channel flash cells. Substrate electron current inject to floating gate by the effect of vertical electric field.

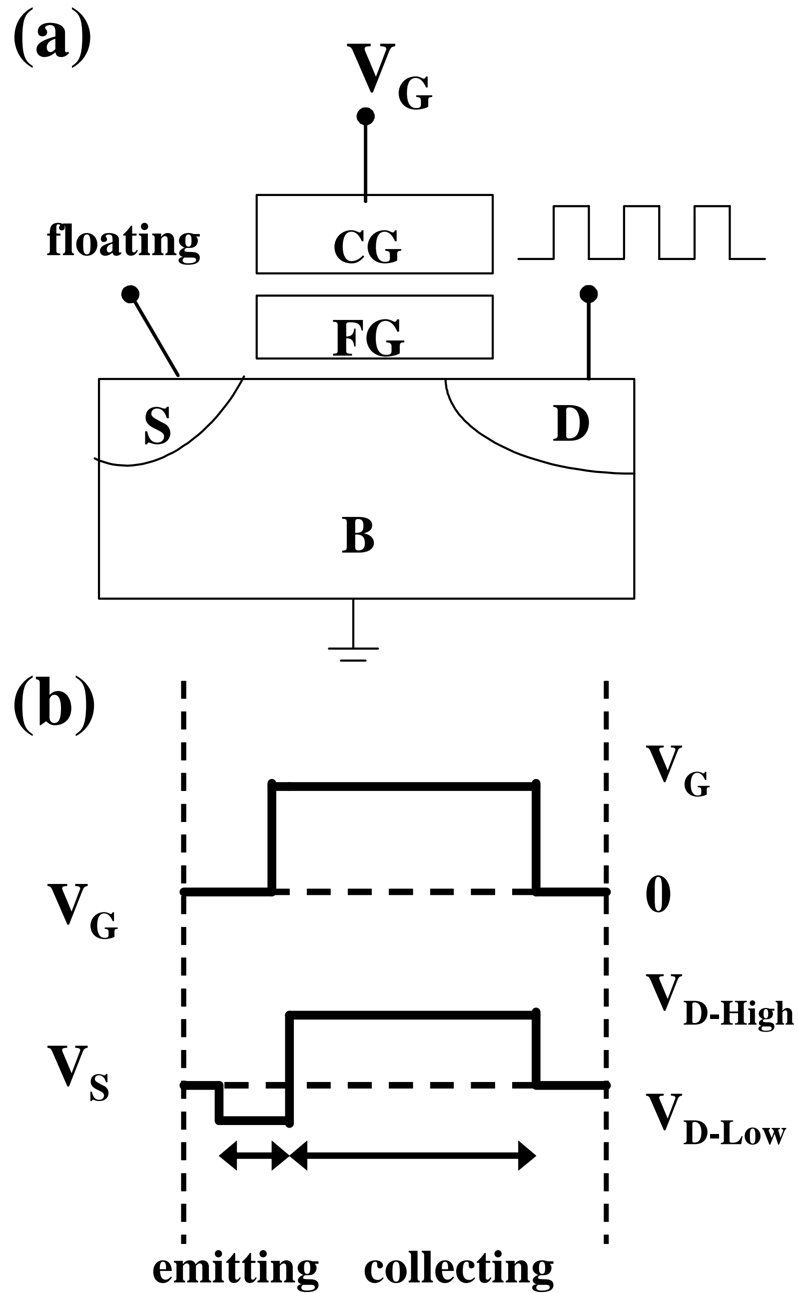


Fig. 4.8 The programming setup and pulse pattern of FBASEI in n-channel flash cells. The substrate (p)-drain (n) diode in emitting period is in forward bias and is in reverse bias in collecting time.

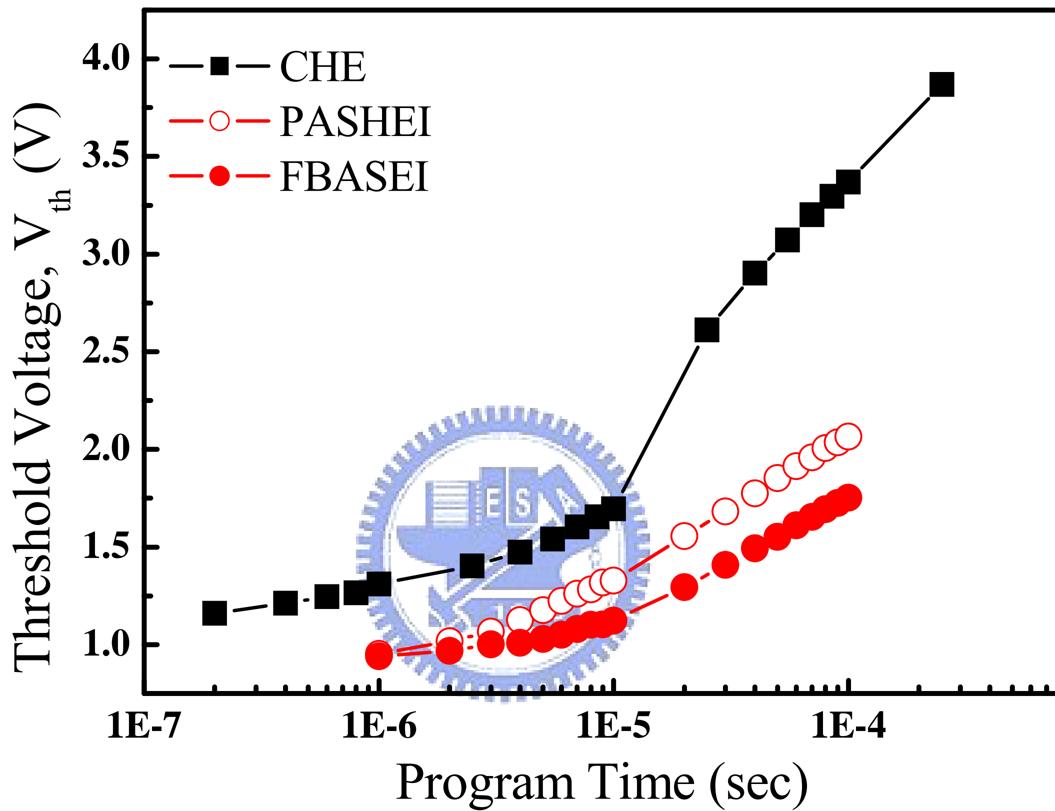


Fig. 4.9 The characteristics of transient program of CHE, PASHEI and FBASEI in n-channel flash cells.

4.4 Mechanism of Forward Bias Assisted Drain Hot Electron Injection in P-channel Flash Cell

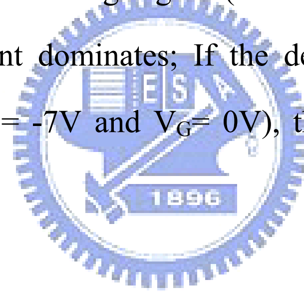
In this section, we will first discuss the difference of mechanism of forward bias assisted hot electron injection in P and N Channel cells. Second, we define the gate current components under different program schemes. Finally, we compare the characteristics with the assistance of forward bias at different operation conditions.

The injection mechanism and performance of the forward bias assisted programming scheme in P-channel cells is quite different from the in N-Channel cells. Fig. 4.10 (a) shows the programming setup and pulse pattern of this program scheme. The concept is similar to N-Channel ones. First, a moderate positive bias is applied on the P-type drain side, and the junction is under forward bias. When changing the mode to reverse bias, deep depletion region is created and carriers in this region suffer strong electric field. Impact Ionization occurs and much energetic carrier is generated and injected into the floating gate. Please refer to Fig. 4.10 (b) for the detailed mechanism. The difference of the scheme in N-Channel ones is the carrier generated in the drain depletion region, so it is Drain Electron Injection and can be called Forward Bias Assisted Drain Hot Electron Injection (FBADHE).

Fig. 4.11 shows the operational region (the combination of gate voltage and drain voltage) with different programming schemes. Before discussing the injection mechanism of the above schemes, we should first clearly define the regions of breakdown, sub-breakdown and Band-to-Band tunneling region. The I-V characteristic of drain-substrate diode is shown in Fig. 4.12. When the drain voltage is larger than 0.7V, the junction is in forward bias (drain current >0); on the contrary, when the drain voltage is below than 0.7V, the junction is in reverse bias (drain current <0). Significant reverse bias leakage current rises as V_D= -5.5V (~10^{-11} A), and increases rapidly to ~10^{-4} A at V_D= -7V. The point (I) (V_D= -5.5V) is called Zener Breakdown. This is due to Band-to-Band tunneling which electrons tunnel from the

p-side valance band to n-side conduction band [23]. The point (II) (V_D= -7V) is avalanche breakdown. It is due to impact ionization induced generated electron-hole pair. The original and generated electrons are both swept to the n side of the junction, and the holes are swept to the p-side [23]. The drain voltage in FBADHE (~6V) is between (I) and (II), both breakdown mechanism may also occur. Avalanche breakdown is the main breakdown mechanism in p-n junction [24], so in this region, we believe the avalanche breakdown dominates. In the following description of this thesis, we call this region sub-breakdown region. It is noted that the breakdown voltage is related to the doping concentration of the drain and substrate.

Next, we discuss the mechanism of injected electrons in different operational conditions. In FN-PGM regions, of course the injected current is F-N tunneling current; If the device operates in the Band-to-Band tunneling region (i.e. V_G-V_D= 15V, V_D= -5V, V_G= 10V), the Band-to-Band tunneling current dominates; If the device operates in the drain avalanche breakdown condition (i.e. V_D = -7V and V_G= 0V), the injection charge is drain avalanche induced hot electrons.



The drain leakage current at V_G= 0V and V_D in sub-breakdown region in MOSFETs is realized as Band-to-Band tunneling current from the gate-drain overlap region [12] [19]. But in flash EEPROMs, the thicker oxide makes the Band-to-Band tunneling current injection at zero gate voltage impossible. So the applied voltage must be adjusted, i.e., applying an appropriate gate voltage to achieve enough band bending. Furthermore, if we operate the device in the sub-breakdown region plus a small gate voltage (i.e. V_D = -6V and V_G= 6V), the composition of gate current is more complicated. The amount of drain breakdown (involves Zener breakdown and avalanche breakdown and avalanche breakdown dominates) induced carriers in sub-breakdown region is small but some of them may gain sufficient energy from vertical field. Besides, with the influence of vertical electric field caused by the positive gate voltage, the potential difference between gate and drain is large enough to induce more Band-To-Band Tunneling current. Thus, the injection charge will involve drain avalanche

induced hot electrons current and Band-To-Band Tunneling current.

It is hard to separate the gate current components in different operational conditions accurately by measurement only because the current component substantially changes as the change of the combination of the drain voltage and gate voltage. Fig. 4.13 shows the relationship of gate current and drain current in p-channel dummy cells. By varying the value of gate voltage, we can roughly separate the gate current components. The gate current difference between V_G= 0V and V_G= 9V at V_D= 0V is F-N tunneling current. The triangle line (V_G= 9V) rises as V_D= 5.5V is due to Band-to-Band tunneling. The square line (V_G= 0V) rises sharply at V_D= 7V is drain avalanche breakdown current. The gate current of circle line (V_G= 5V) rises at V_D= 5V is initially Band-to-Band tunneling current. And then as the increasing of drain voltage (to ~6V), some of the avalanche breakdown induced electrons gain enough energy from vertical field and become a part of gate current.

Another simple method to determine the gate current components is shown in Fig. 4.14. The potential difference of gate and drain is fixed as 12V (solid circle) and 7V (hollow circle) respectively. It represents the band bending is the same at the whole line so that the amount of injection carrier due to BBHE is the identical. It helps us to find the other gate current components with different combinations of gate/drain voltage. When the drain voltage is -2V, electron F-N tunneling current is the main component. At drain voltage equals to -6V, sub-DAHE generates, as previous description. At drain voltage equals to -7V, there is a lot carriers injecting due to drain avalanche breakdown. Besides, comparing the two spot at V_D= -7V, we can see the small gate voltage (5V) enhance the injection of electrons by an order. This is because the injection of Band-to-Band hot electrons and additional drain avalanche breakdown hot electrons.

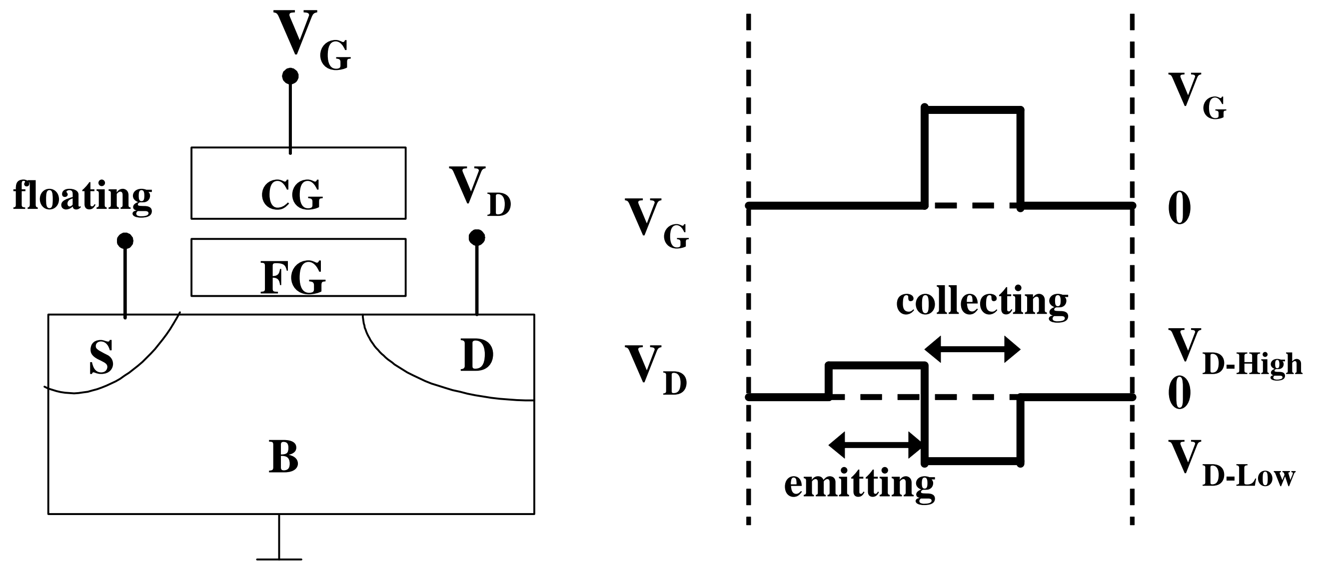
Fig. 4.15 shows the transient programming characteristics of BBHE, DAHE and sub-breakdown regions. We find that the characteristic of DAHE and sub-breakdown A is fast

self-convergent. So, in these 2 conditions the Drain Avalanche breakdown current is the main injection component. The injection component of BBHE is Band-To-Band Tunneling current. The threshold voltage of sub-breakdown B is almost fixed. This is because both the band bending and drain avalanche is not large enough to generate apparent injected current. The measurement result conforms to our explanation.

In summary, the injected gate current under sub-breakdown region is drain avalanche induced hot electrons and Band-To-Band tunneling current, and from Fig. 4.15, drain avalanche induced hot electrons is the main component.

Now we investigate the influence of the drain forward bias in the above operational regions. Table 4.2 shows the operational condition of P/E schemes in this study. Fig. 4.16 shows the original and forward-bias-assisted characteristics of transient program under different operating conditions. From this measurement, we clearly identify the mechanism of the assistance of forward bias. At BBHE (circle line), there is no difference after assistance of forward bias, this is because the forward bias doesn't influence the Band-to-Band tunneling current. The performance with forward bias shows the emphatically difference on DAHE (square line) and sub-breakdown region (triangle line). This is because the forward bias makes more impact ionization at the Drain-Substrate junction so that more energetic carriers surmount the oxide barrier and inject into floating gate. This result is consistent with our proposal at the second paragraph of this section.

Since the assistance of forward bias works in drain avalanche breakdown and sub-breakdown region, considering the reliability issue: drain disturb, the larger drain bias during programming causes more serious drain disturb, so we select the sub-breakdown region as our operational condition in this thesis..



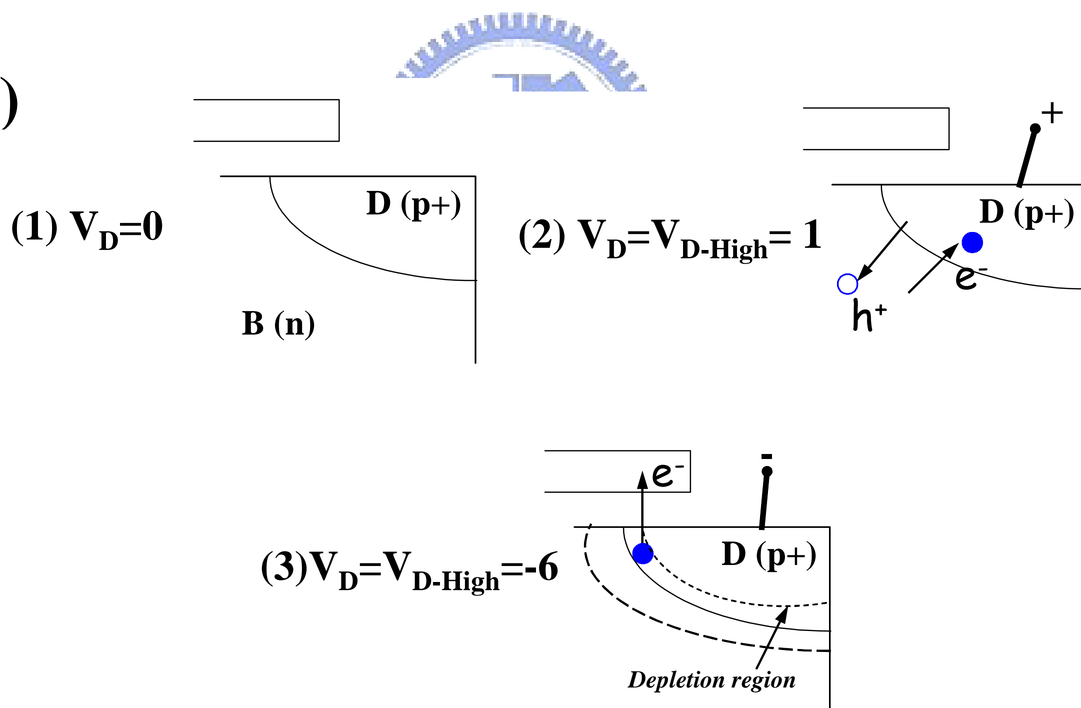


Fig. 4.10 The (a) program setup, pattern mode and (b) injection mechanism of PASHEI in p-channel flash cells.

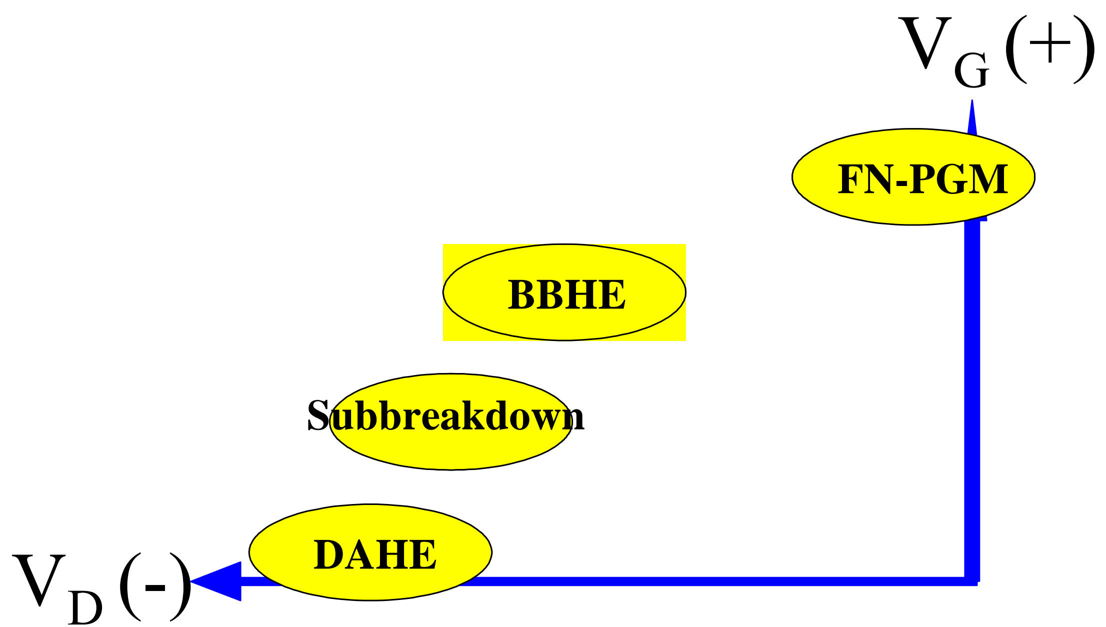


Fig. 4.11 The combination of gate voltage and drain voltage of different programming schemes.

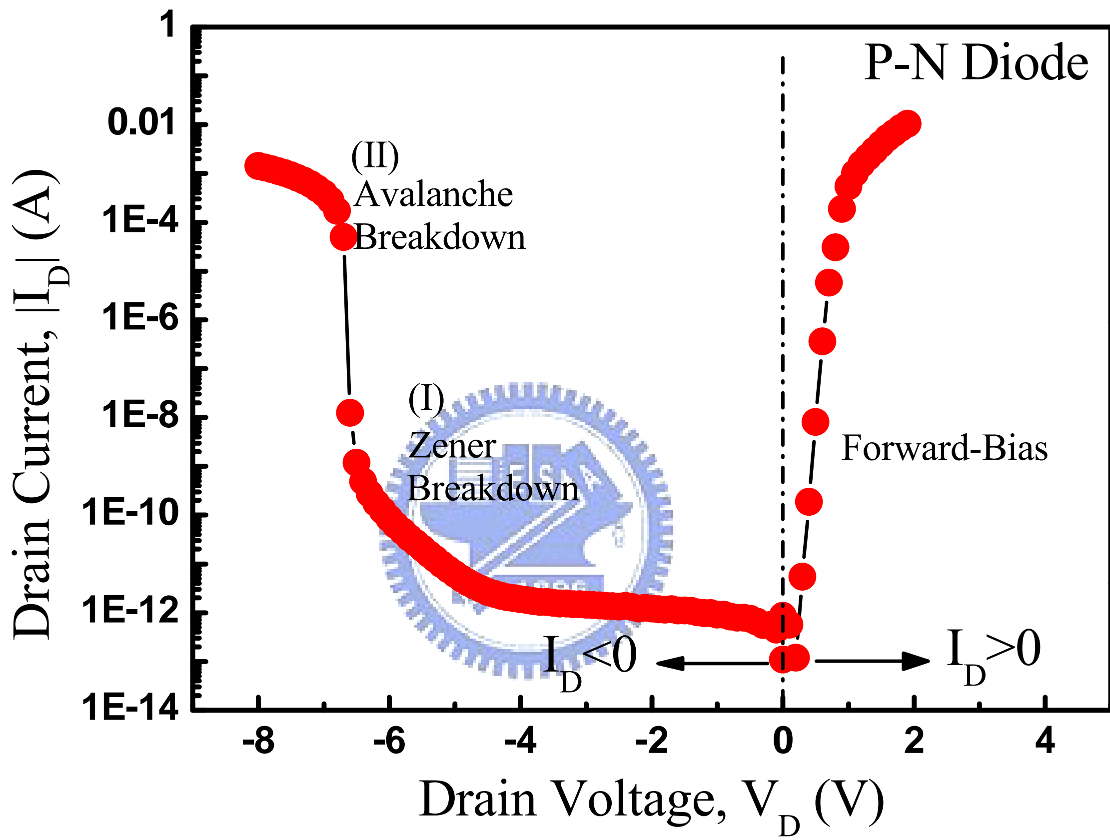


Fig. 4.12 I-V characteristics of drain (p)-substrate (n) diode.

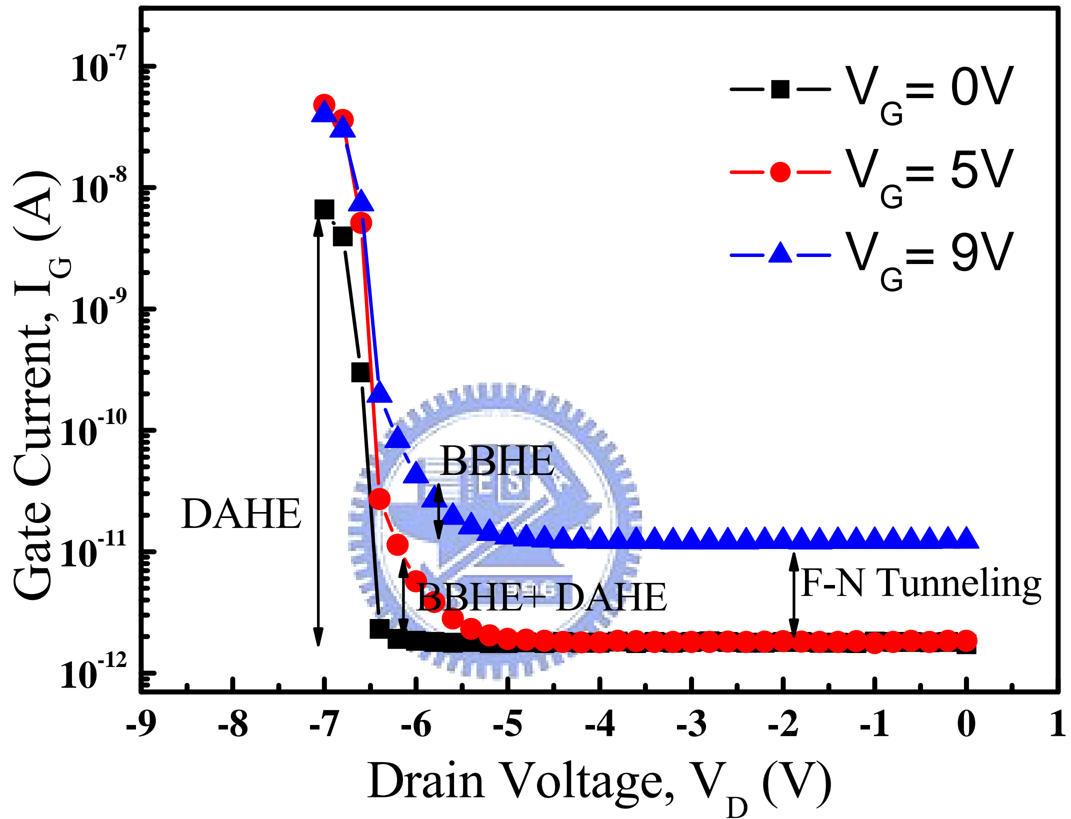


Fig. 4.13 The I_G-V_D characteristics with different gate voltage in p-channel dummy cells. From this figure, we can roughly separate the gate current components by the difference of gate voltage.

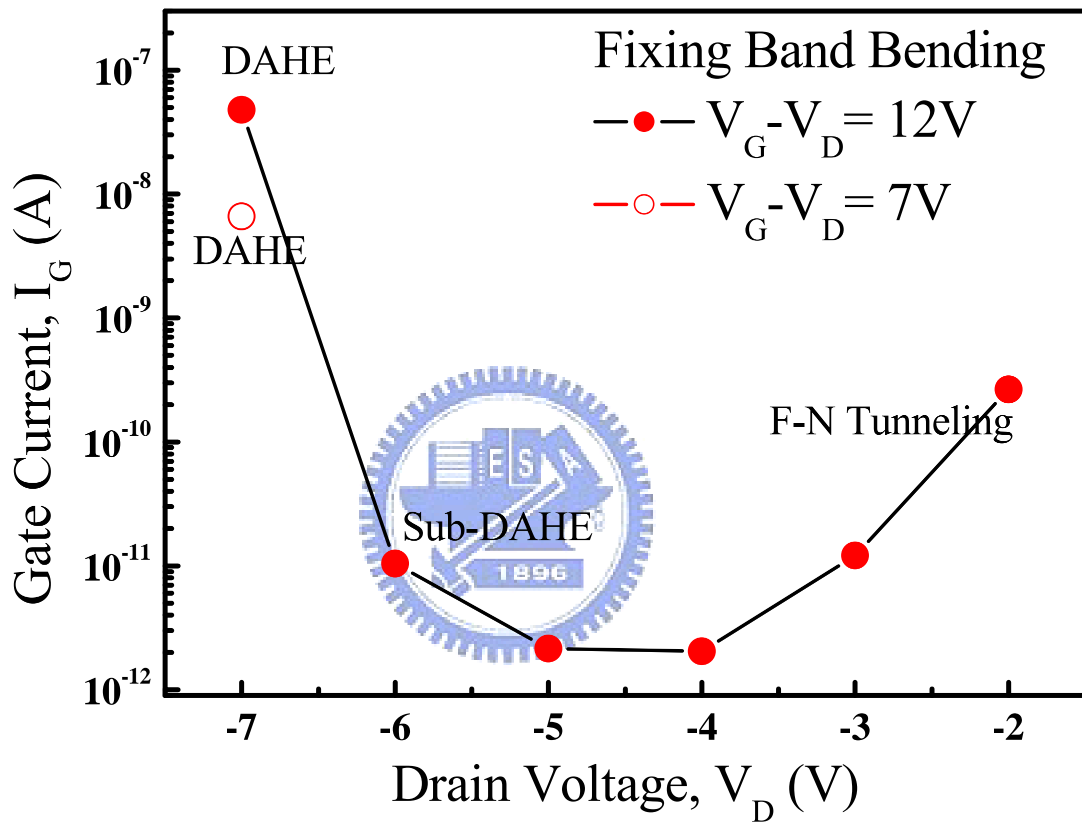


Fig. 4.14 The gate current components by fixing the potential difference between gate and drain terminal.

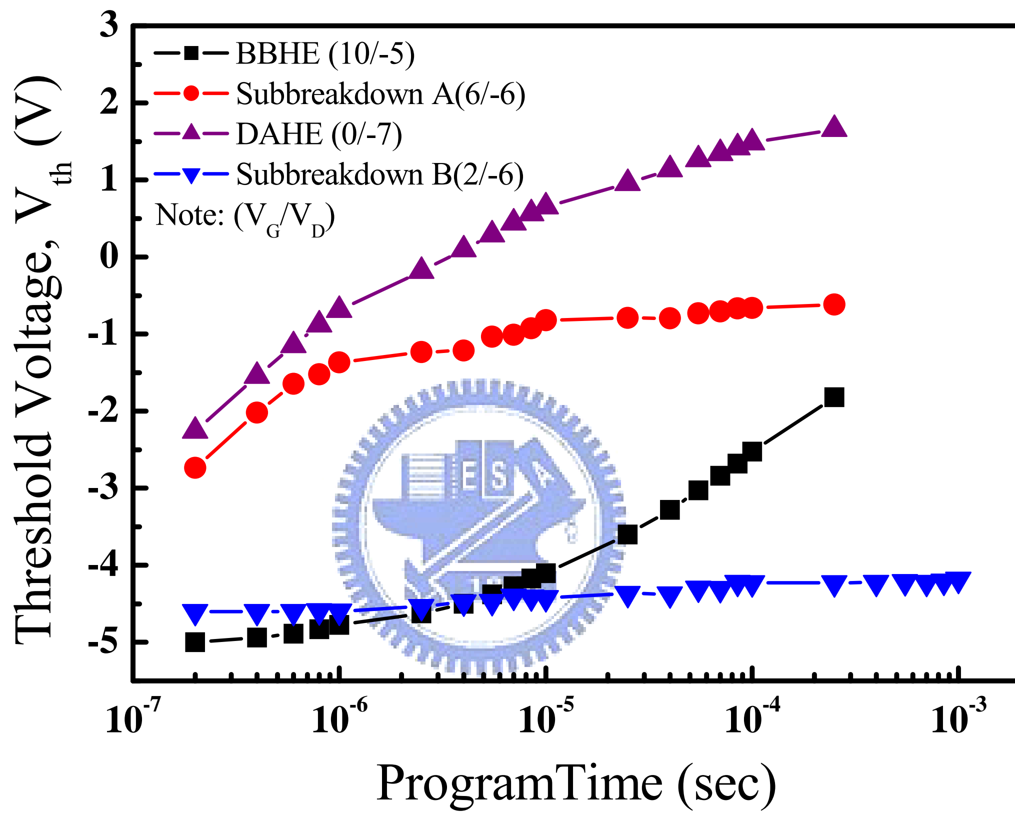


Fig. 4.15 Transient program characteristics under different operational conditions.

Operate		V_G	V_D	V_B	V_S
PGM	**BBHE**	11	-5	0	f
	DAHE	0	-7	0	f
	FBADHE	6	$V_{D\text{-}High}$=1 $V_{D\text{-}Low}$=~ -6*	0	f
ERS	**Channel-FN**	-10	f	5	f
Read		-3	-1	0	0

Table 4.2 Operational table of P/E schemes in this study.

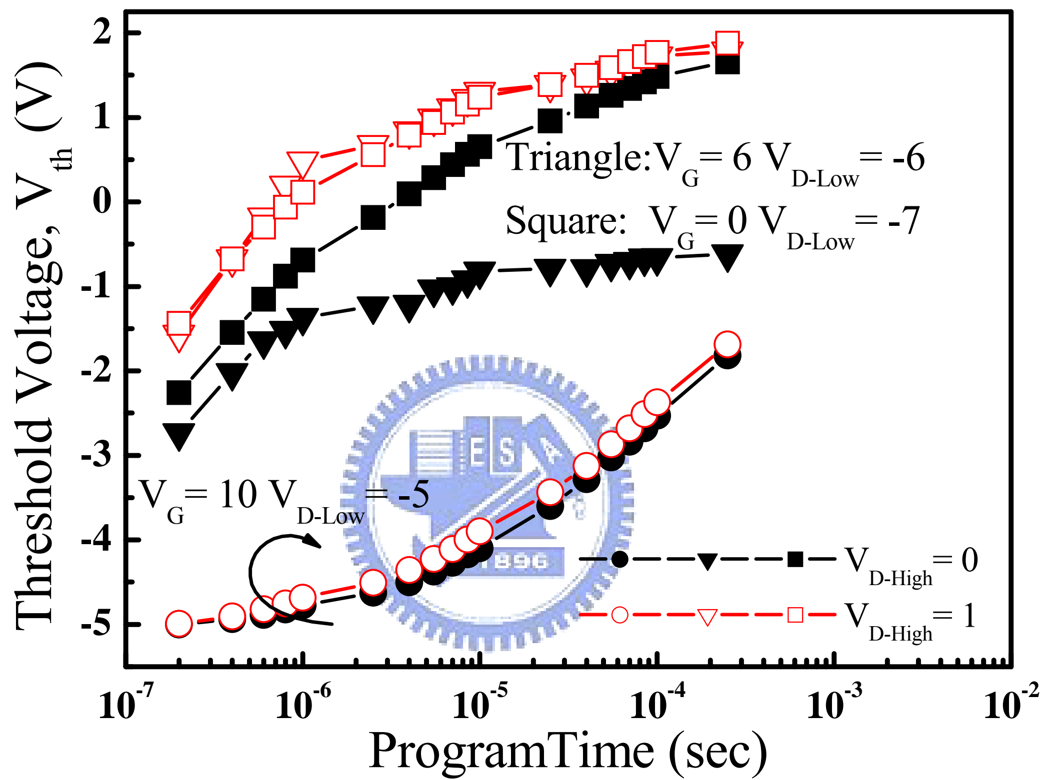


Fig. 4.16 Transient program characteristics with the assistance of drain forward bias under different operating conditions.

4.5 Study of FBADHE at Sub-breakdown Region in P- Channel Flash Cell

In this section, we will investigate the characteristics of Forward Bias Assisted Drain Hot Electron Injection at sub-breakdown region. The related symbols are shown in Fig. 4.10 (a).

First, we discuss the dependence of drain emitting voltage (drain forward bias, $V_{D\text{-}High}$), as shown in Fig. 4.17 (a). The threshold voltage shift has a maximum at $V_{D\text{-}High}$= 1V. It can be explained that the large forward bias (> 1V) will increase the momentum of diffusing carriers and reduce the effect of impact ionization. Fig. 4.17 (b) shows the impact of the drain collecting voltage (drain reverse bias $V_{D\text{-}Low}$). As predicted, the larger drain bias makes the more threshold voltage shift. This is because the larger drain bias causes more impact ionization in the deep depletion region.

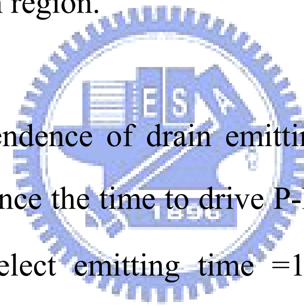
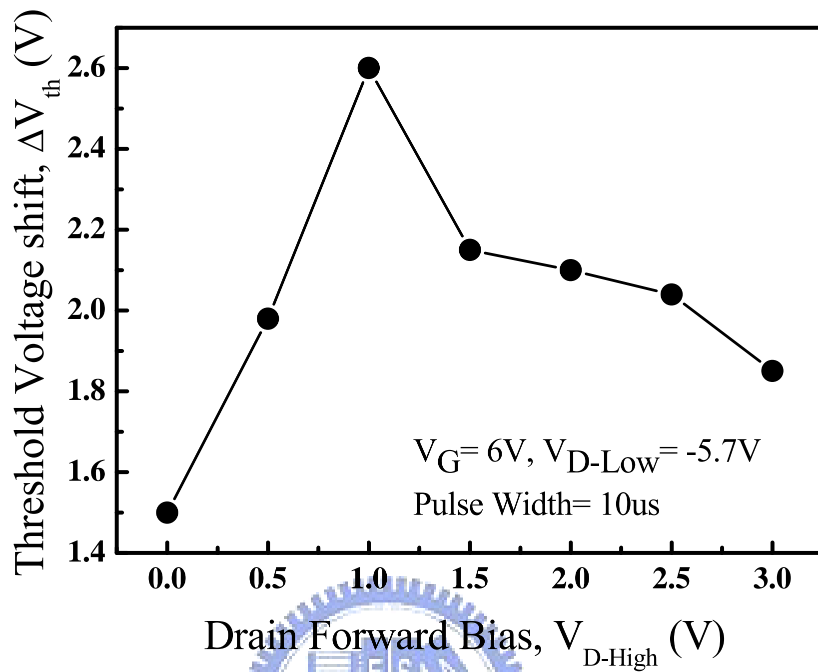


Fig. 4.18 shows the dependence of drain emitting time. We can find that there is no dependence of emitting time since the time to drive P-N diode in active area is very short (<1 μs). For convenience, we select emitting time =1 μs on following measurement. The dependence of collecting time (pulse width) is shown in Fig. 4.19. This figure shows that the FBADHE is pulse width independent, because the same program time with different width shows the different result. The characteristic is different from other programming schemes, i.e., CHE, F-N and BBHE. Let me redraw this figure by replacing the X axis to programming pulse count, as shown in Fig. 4.20. We find that FBADHE is indeed pulse width dependent. This characteristic is due to a large amount of carriers are generated by the path of Sub-drain avalanche breakdown and Band-to-Band tunneling in a very short time (<1 μs), and increasing the programming width is useless to enhance the amount of generated charges.



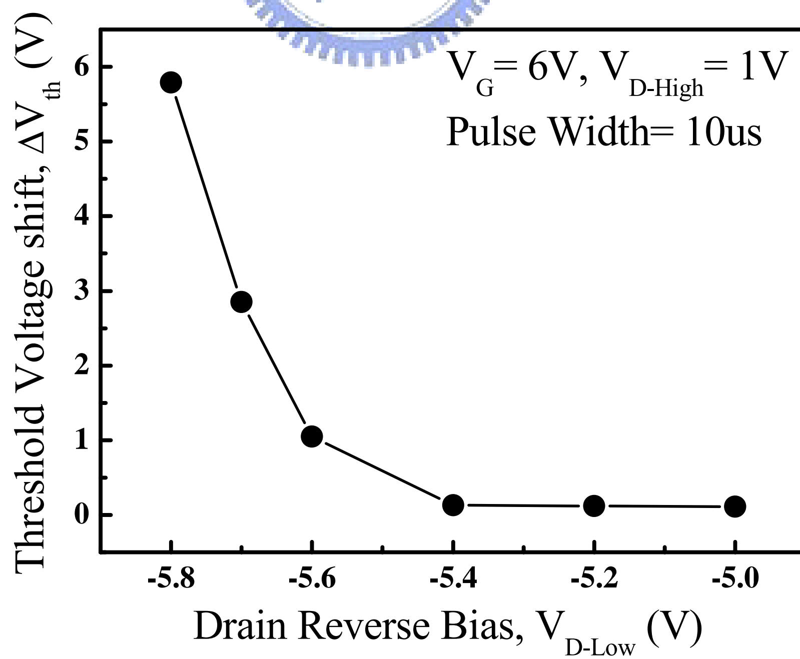


Fig. 4.17 The relationship of drain (a) forward bias ($V_{D\text{-}High}$) and (b) reverse bias ($V_{D\text{-}Low}$) on FBADHE in p-channel flash cells.

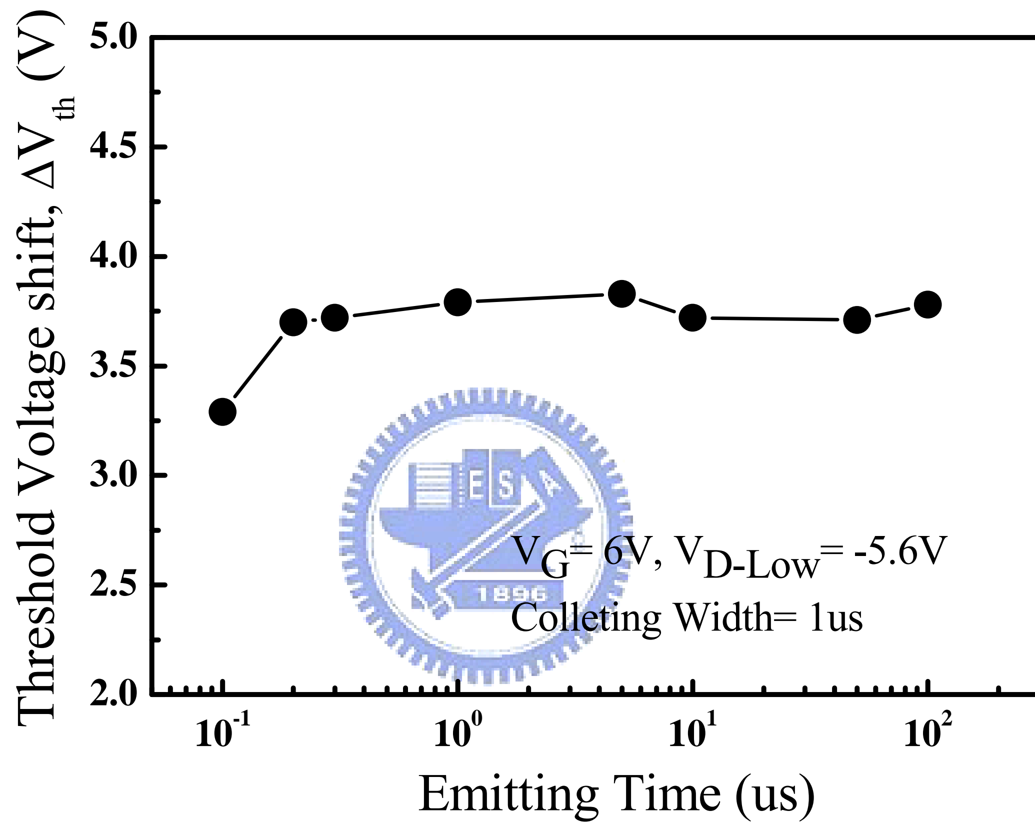


Fig. 4.18 The relationship of drain emitting time on FBADHE in p-channel flash cells.

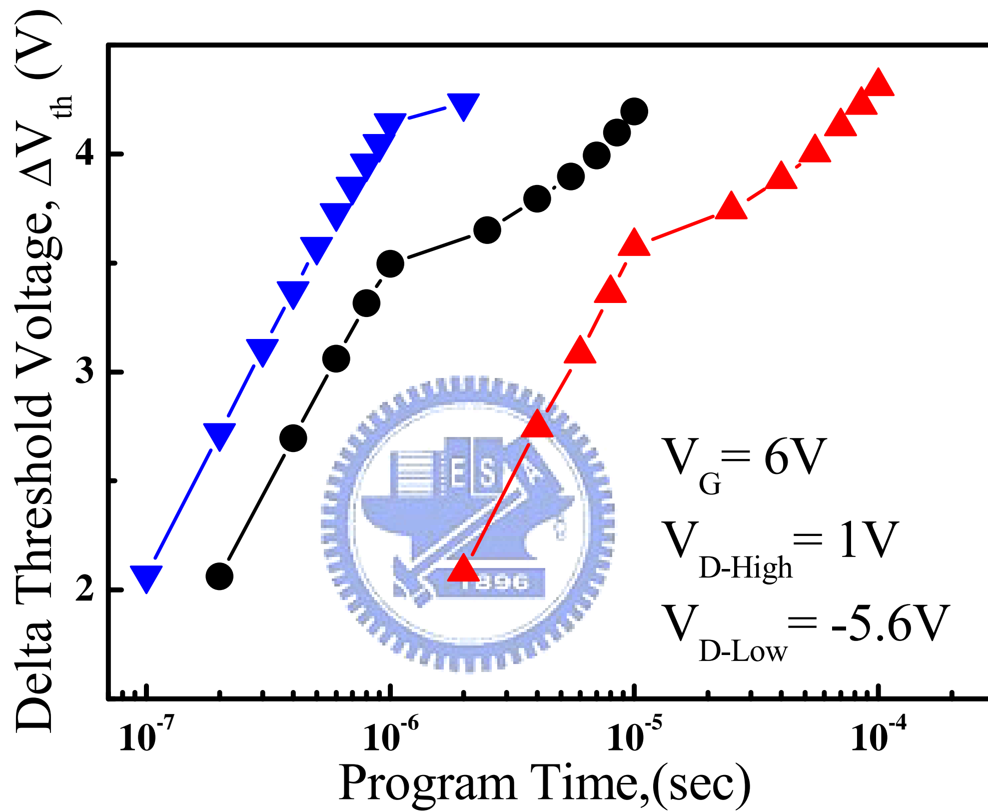


Fig. 4.19 The relationship of pulse count and threshold voltage shift in p-channel flash cells.

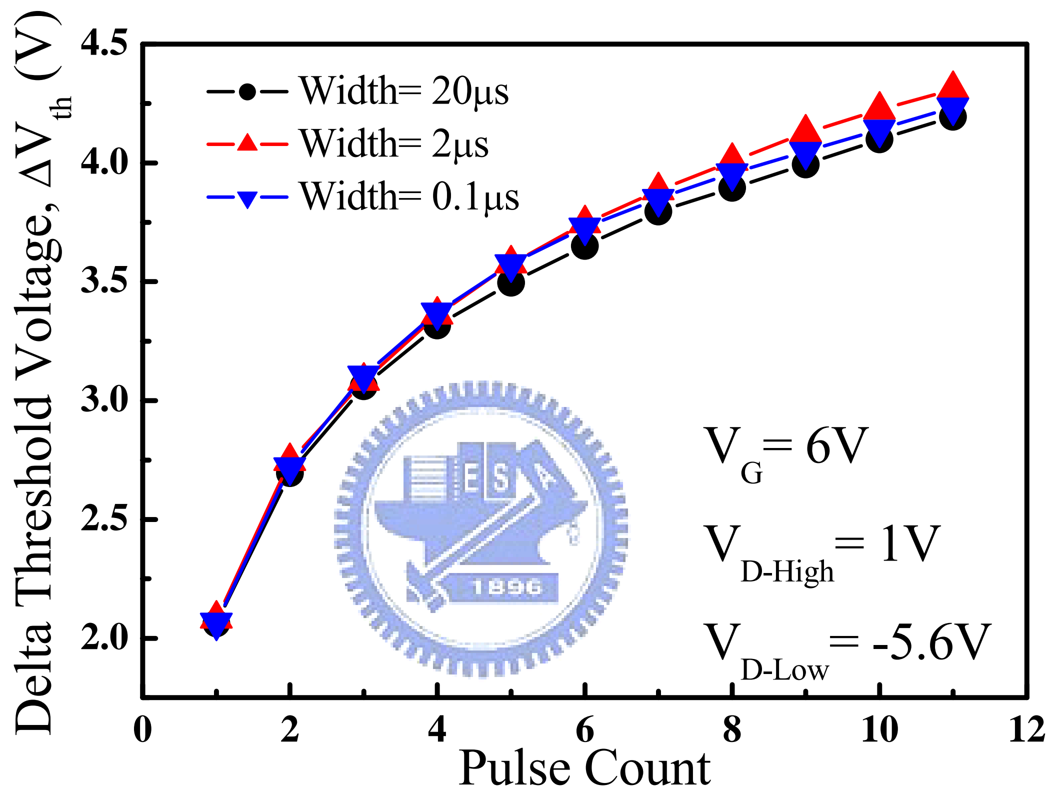


Fig. 4.20 The relationship of pulse count and threshold voltage shift in p-channel flash cells.

4.6 Erase Operation in P- Channel Flash Cell

The purpose of erase operation is to remove the electrons form floating gate to the device and reduce its threshold voltage. Folwer-Nordheim (F-N) tunneling is used to erase the flash cells. However, F-N operation requires high voltage across the tunnel oxide. Under strong electric field, electrons tunnel from floating gate and inject into channel or source region. This F-N tunneling effect is one of the main reliability issues in MOS devices, but is used to operate in flash devices.

Folwer-Norheim tunneling current occurs when the electric field crossing the tunnel oxide is large enough to make electrons whose energy is less than the barrier height can tunnel through the oxide to substrate. The F-N tunneling current I_{FN} can be expressed as a function of the electric field：

$$I_{FN} \ \alpha \ E_{OX}^{2} \exp(\frac{-B}{E_{OX}}) \quad (4.1)$$

where B is a constant depending on effective mass and the barrier height.

There are two kinds of Folwer-Nordheim tunneling Erase scheme used in the flash cells: Channel F-N [14] [15] and edge F-N [1] [16]. The channel F-N is to remove the electrons uniformly from the whole channel, and the Edge F-N is from the source edge. For N-Channel cells, the negative Control Gate voltage and positive Substrate voltage is applied during Channel F-N; Control Gate grounded and positive Source voltage is applied during Edge F-N. The two erase techniques have their own advantage. The reliability of F-N Erase is hole injection into the tunnel oxide (Q_{ox}). Channel F-N Erase generates positive oxide charge in the oxide above the channel and the Edge F-N erase will generate positive charge in the tunnel oxide. Q_{ox} will damage the oxide and cause the data loss of the memory cell [22].

However, in p-channel Flash cells, only Channel F-N can be used to remove the

electrons form floating gate. This is because the doping type of source (P) and substrate (n). In order to keep the diode in reverse bias, the potential of source must not be larger than substrate potential so edge F-N is unsuitable in P-channel cells. Fig. 4.21 shows the setup and band diagram of Channel F-N erase. Source and drain are floating during erase process. Fig. 4.22 shows the erase characteristic of P-channel Flash cells. Fig. 4.22 (a) indicates the larger electric field between the tunnel oxide causes more threshold voltage shift. Fig. 4.22 (b) indicates that the same electric field makes the same threshold voltage shift, whether the voltage is applied on gate or substrate.

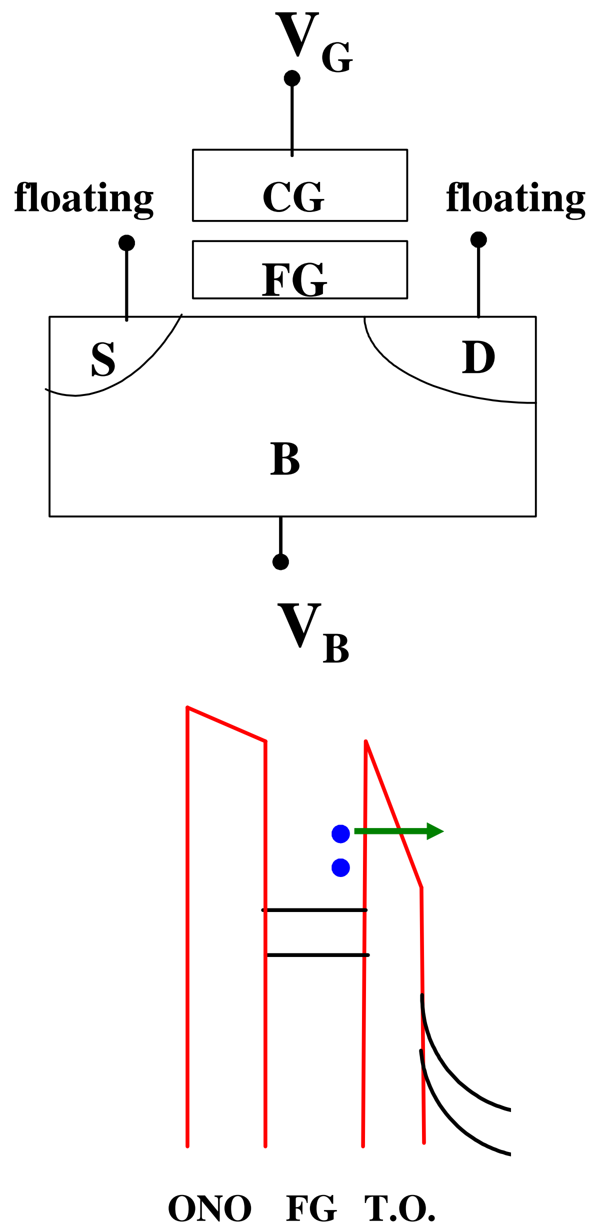
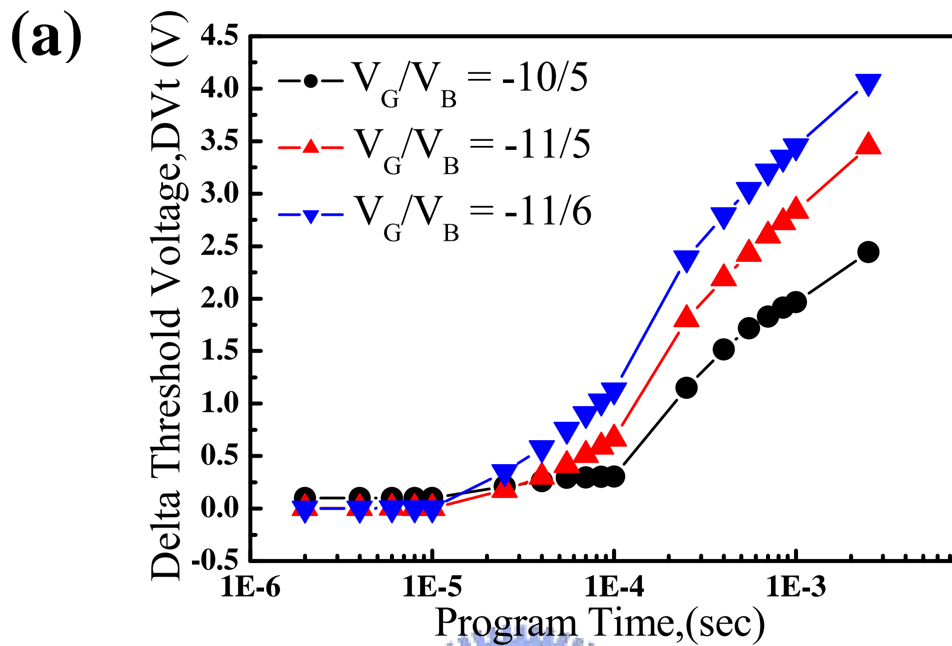


Fig. 4.21 Experimental setup and band diagram of Channel F-N erase in p-channel flash cells.



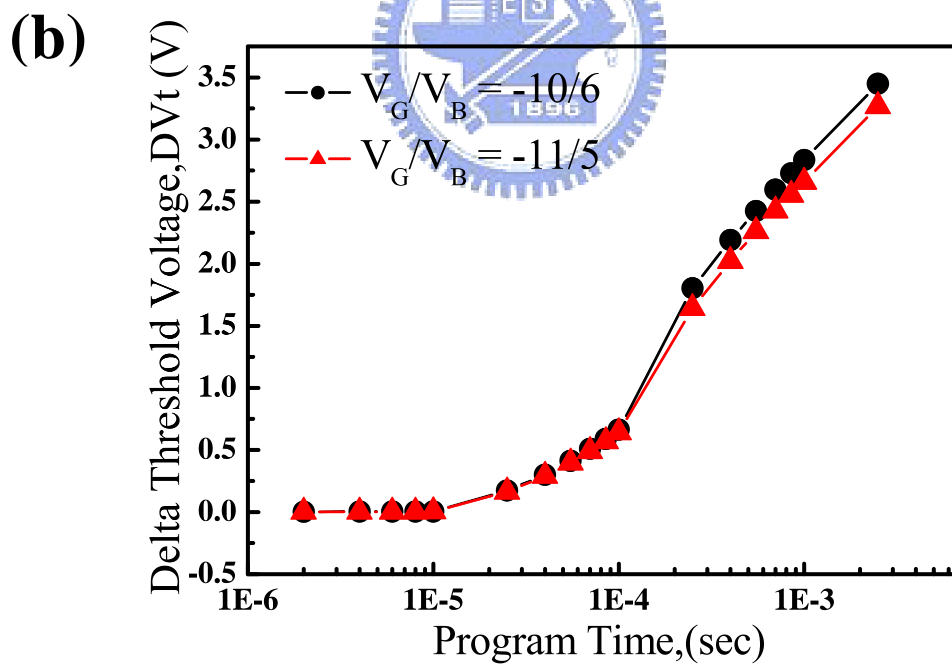
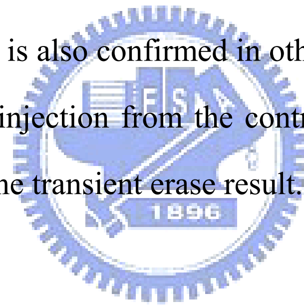


Fig. 4.22 Transient characteristics of Channel F-N erase in p-channel flash cells

4.7 Comparison of Program/Erase Performance with Different Doped Poly Gate

Which poly gate has better performance is an interesting issue to investigate, and is an easier way to improve the performance of flash cells. Fig. 4.23 shows the transient program characteristics with difference doped poly gate. In this figure, in order to identify which has larger threshold voltage shift, we operate the devices with smaller drain voltage, -5.8V for reverse bias. We can see the p+ gate show the better performance. It is because the work function difference of p+ doped gate is larger than n+ ones. It is noted that the initial threshold voltage of above two devices is adjusted to be the same. So that for the given gate and drain voltage, the electric field is also the same. The larger work function difference of p+ doped gate makes electrons easier t to be stored in the floating gate and make the larger threshold voltage shift. The phenomenon is also confirmed in other study [15] [25]. For erase operation, p+ gate suppress the electron injection from the control gate and result the larger threshold voltage shift. Fig. 4.24 shows the transient erase result.



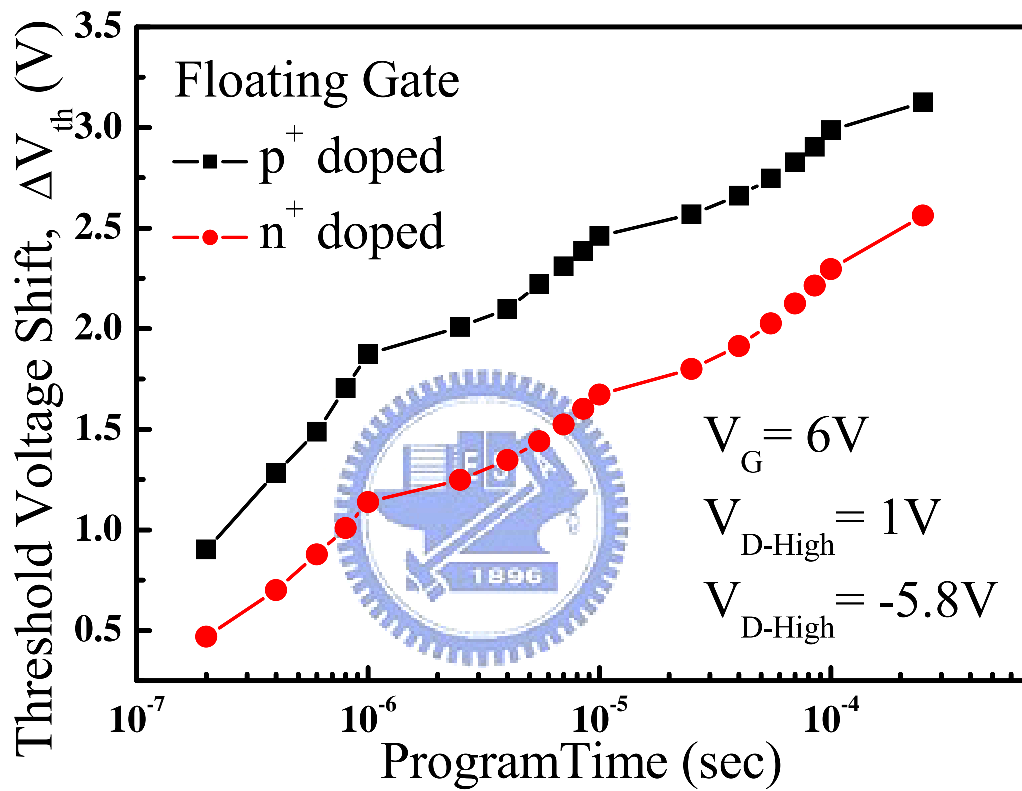


Fig. 4.23 Transient program characteristics of FBADHE with different doped poly gate.

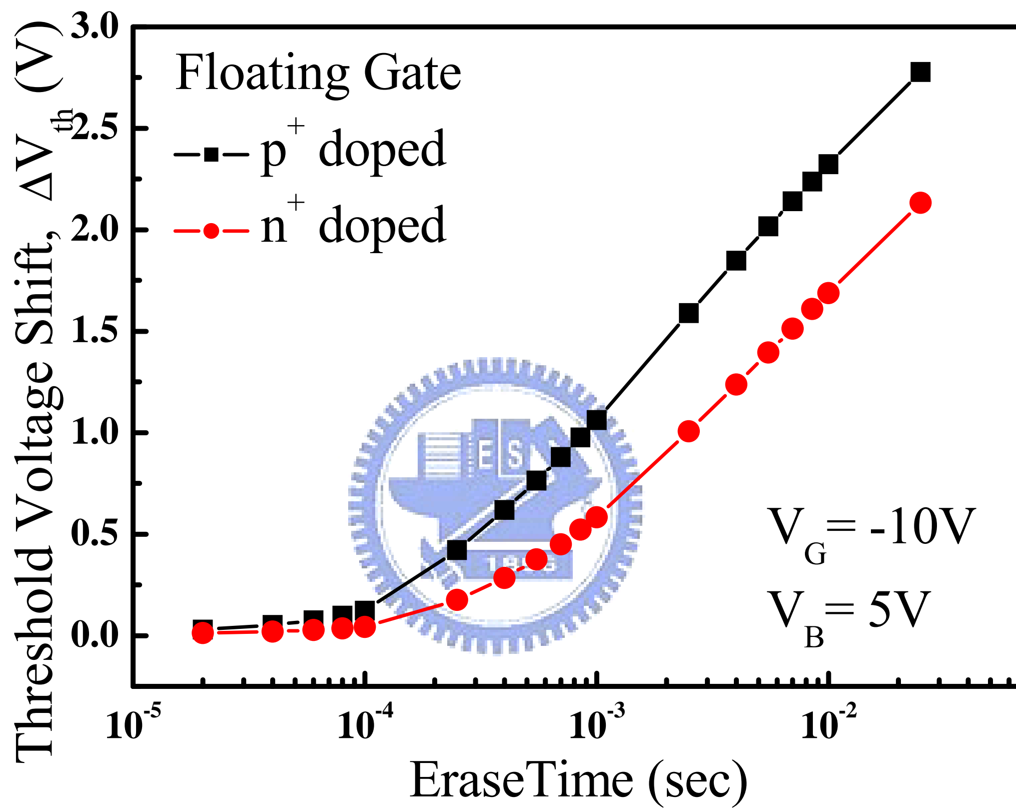


Fig. 4.24 Transient erase characteristics of FBADHE with different doped poly gate.

Chapter 5
Reliability of FBADHE in P-channel Cells

5.1 Introduction

In this Chapter, we will investigate the reliability of FBADHE. First, we compare the P/E cycling endurance between BBHE and FBADHE. Both of the erase method is Channel F-N erase. And then, we measure the charge loss at room temperature and 80°C respectively to observe its data retention ability. Finally, the Gate/Drain/Read disturb issues are investigated.

ES
1896

5.2 Program/Erase Cycling Endurance

In this section, we compare the cycling endurance of BBHE and FBADHE programming scheme. Both of the erase schemes are Channel-FN. The drain disturb of DAHE is too serious so that we do not compare it with the other programming schemes.

Fig. 5.1 shows the measured endurance of FBADHE and BBHE in P-channel flash with p+ doped gate. The operational condition of BBHE is: V_G= 11V, V_D=-5V and programming time is 20μs. For FBADHE is: V_G= 6V, V_D=1 to -5.9V and programming time is 1μs. The erase setup is: V_G= -10V, V_B=-6V and erase time is 10ms. The window doesn't change after cycling (~2.3V) but the threshold voltage is slightly decreasing during cycling. An explanation is the pileup of positive charge in the tunnel oxide during erase operation, and the cycling induced interface state (N_{it}) or oxide trap charge (Q_{ox}) causes the degradation of flash cells. In this sample, the endurance performances of FBADHE is a little bit better than BBHE due to the less threshold voltage shift in both program and erase states.

Next, we compare the endurance of FBADHE with different poly gate doping, as shown in Fig. 5.2. The endurance in n+ gate flash cell is better than p+ gate ones. The program/erase window of FBADHE with p+ poly gate is a little larger than n+ ones with different poly gate and doesn't show much significant difference. It is because the work function difference, as described in Section 4.6. The main injection mechanism is avalanche induced hot electron and the inter-poly ONO and tunnel oxide is very thick (165Å and 90Å respectively). Thus the electric field difference due to potential difference caused by poly gate doping (~1eV) across the oxide is small so that the performance with different doped gate is close. Furthermore, in our experience, the uniformity of devices also plays an important role on programming/erase window.

Many measurements can be used to investigate the generation of interface state (N_{it}) and oxide trapped charge (Q_{ox}), such as, Charge pumping, GIDL, subthreshold characteristics and Gated Diode. Fig. 5.3 shows the subthreshold characteristics of different program schemes. Interface state (N_{it}) can be estimated by measuring the subthreshold slope. The slope doesn't show significant difference. Previous researches identified that the injection of hot hole is the major factor to degrade the device performance. During FBADHE or BBHE, impact ionization or Band-to-Band tunneling induced hot holes are attracted by the drain, which is applied an appropriate negative voltage, and flow to the p+ drain. At the same time, the gate voltage is positive and will stop hot holes to inject into the floating gate. So that in P-channel flash cells, program operation can eliminate the injection of holes and has better reliability.

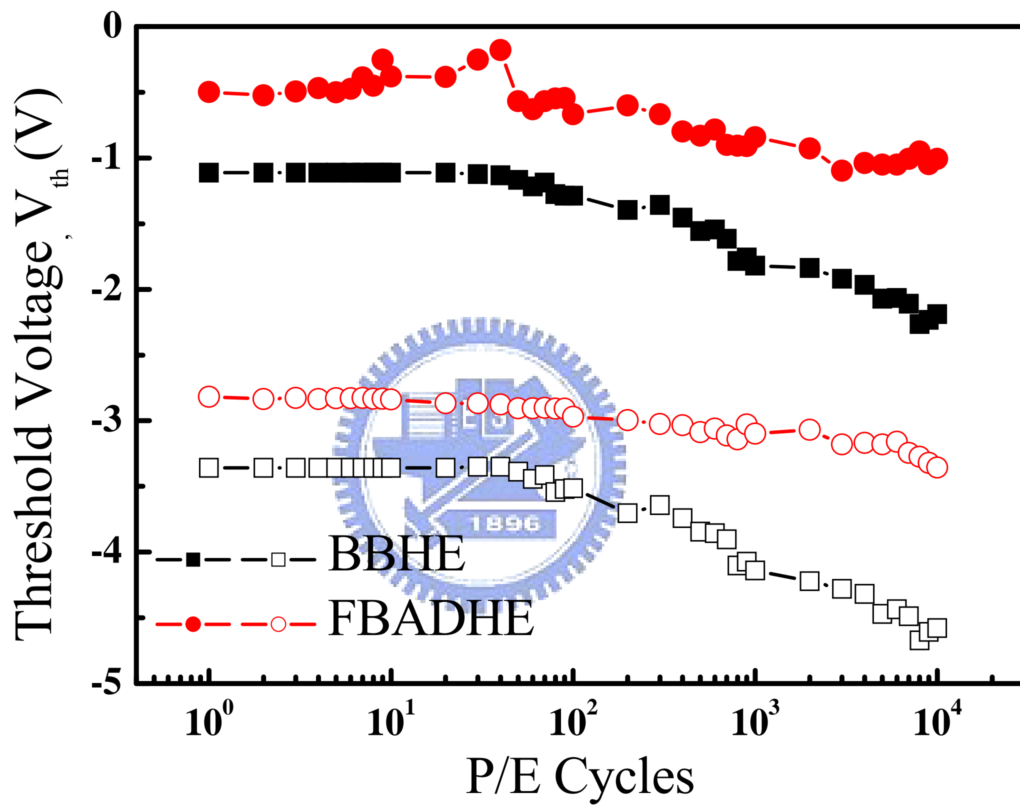


Fig. 5.1 Cycling endurance of BBHE and FBADHE in p-channel flash cells. Experimental setup: BBHE: V_G= 11V, V_D=-5V, programming time is 20 μs; FBADHE is: V_G= 6V, V_D=1 to -5.9V, programming time is 1 μs; Erase setup is: V_G= -10V, V_B= 6V and erase time is 10ms.

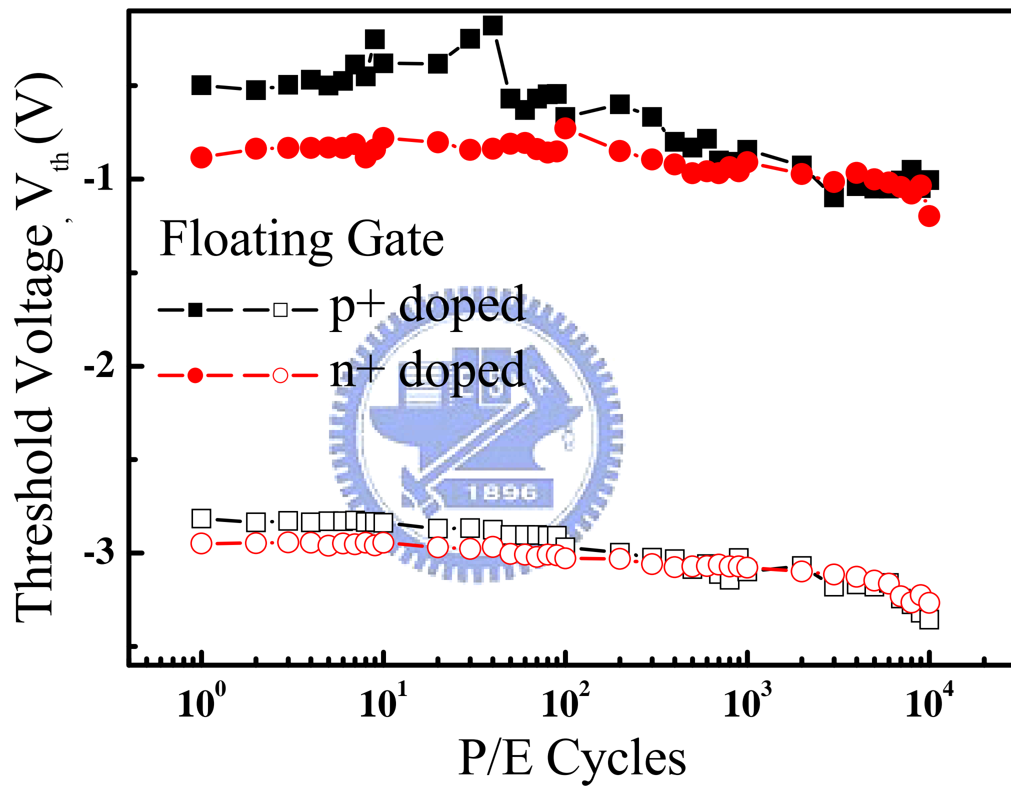


Fig. 5.2 Cycling endurance of FBADHE with different doped gate in p-channel flash cells. Experimental setup: FBADHE is: V_G= 6V, V_D=1 to -5.9V, 1 μs; Erase setup is: V_G= -10V, V_B=5V, 10ms

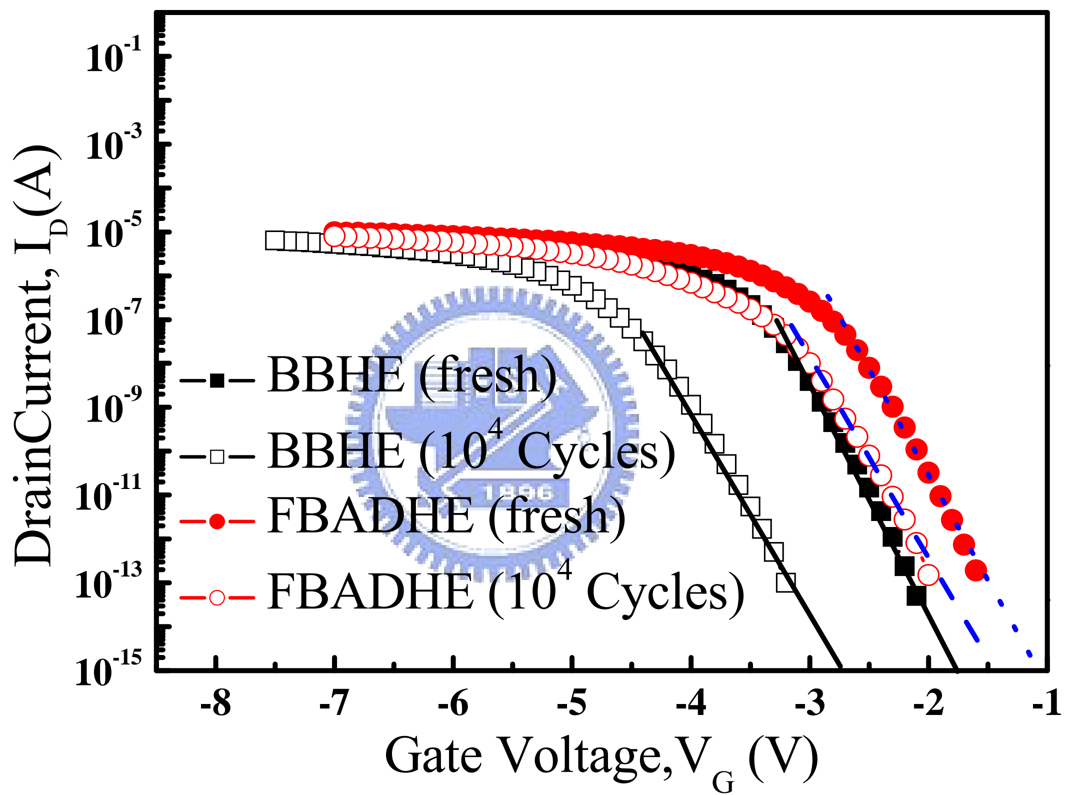
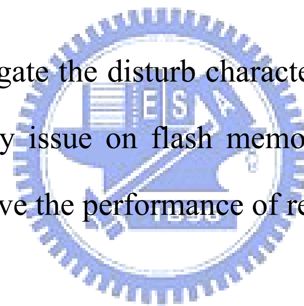


Fig. 5.3 Subthreshold characteristics comparison of fresh and cycled cells with different programming scheme. The subthreshold slope doesn't show significant difference.

5.3 Data Retention Characteristics

The other important reliability issue of flash memory cells is data retention ability. Fig. 5.4 shows the data retention characteristics of BBHE and FBADHE respectively at program state at the temperature of 80°C. With the benefit of thick inter-poly ONO and tunnel oxide, the data retention ability is excellent.

5.4 Disturb Characteristics



In this section, we investigate the disturb characteristic before and after cycling. Disturb is a one of the main reliability issue on flash memory cells. We will measure the disturb characteristics and try to improve the performance of reliability to achieve high performance.

There are several kinds of disturbing phenomenon in flash memory cells. Fig. 5.5 shows the type of disturbance effect in a conventional NOR array. WL is Word Line, which connects the gates, and BL is Bit Line, which connects the drains. In order to save the space of chip, the adjacent two cells along the bit line direction share a source line. When we program the cell of the cross-over point of WL1 and BL1, the cells along the WL1 will suffer gate disturb (solid circle in this figure), and the cells along the BL1 suffer gate disturb (dashed circle in this figure). Besides, during read operation, read disturb also occurs. The origin of read and gate disturb failures is the stress induced leakage current [27], and the mechanism of drain disturb is due to Band-to-Band tunneling (BBHE) or Channel Hot Hole (CHH) induced hot electron injection [21]. All of the disturb phenomenon will give rise to the threshold voltage shift, so it is a major concern of the reliability of flash memories.

Fig. 5.6 shows the gate disturb characteristics before and after 10k cycling respectively. At this figure, we compare the disturb characteristic of BBHE (V_G= 11V) and FBADHE (V_G= 6V). At BBHE operation, performance before and after cycling shows more significant difference than FBADHE operation. In BBHE, threshold voltage shift of cycled cell after 1000 second stress is about 0.45V. On the contrary, the gate disturb performance of FBADHE shows much less threshold voltage shift, even after 10k cycled. It is an advantage of FBADHE. Fig. 5.7 shows the read disturb characteristic of 10k cycled cells. At read condition (V_G= -3V and V_D= -1V), threshold voltage doesn't shift even at a long stress time. It is because the inter-poly ONO and tunnel oxide is thick enough

The gate disturb makes the threshold voltage shift but is not the major concern of reliability because the gate voltage during programming is small and doesn't make significant threshold voltage shift. Besides, Drain Disturb is well known of the most important reliability issue in p-channel Flash memories [21]. Fig. 5.8 shows the drain disturb characteristics of different program schemes. First we compare the difference with difference drain voltage. The V_{th} shift at BBHE operation (V_D= -5V) is much smaller than is avalanche operation (V_D= -6V) by about 2 orders. This is because the larger drain voltage causes the more generated carriers due to impact ionization and band-to-band tunneling and has sufficient energy to inject into floating gate. Besides, comparing the influence of forward bias, we find that the cells with forward bias assisted suffers serious drain disturb. The reason is the same as above description. More impact ionization and band-to-band tunneling fastens the programming speed, but the same mechanism degrades the reliability of cells. Fig. 5.9 shows the comparison of drain disturb with fresh and cycled cells of BBHE and FBADHE respectively. The conclusion is the same as Fig. 5.8: the drain disturb of BBHE is much better than FBADHE. It is a trade-off between reliability and programming speed. In above two figures, with FBADHE operation, threshold voltage start to shift at stress time= 10us in fresh cell and 1us in 10k cycled cell. It is unacceptable for industrial requirement so we must find other solution to solve this reliability issue.

A possible solution is to change the array architecture. Since Drain Disturb is the major issue on reliability of P-channel Flash cells in most of the programming schemes (i.e., BBHE, DAHE, FBADH etc), some researches propose new array architecture to improve the drain disturb performance. DINOR structure is one of the solutions used to eliminate drain disturb [26] [21] [5]. During programming operation, by setting all the other select transistors off except for the one which is in the programming mode, maximum drain disturb time can be reduced to 7 times the programming time from the cells connecting to the same sub bit line. But at FBADHE operation, Drain Disturb is so severe at cycled cells (please refer to Fig. 5.9) that drain disturb can not be solved even DINOR is adopted. Other structure must be developed.

Since the main injection mechanism of FBADHE is forward bias induced impact ionization of the p-n junction, to reduce the impact ionization of junction is the simplest way to mitigate the drain disturb. During programming, -6V is applied on the whole bit line, so we propose to apply an additional moderate negative substrate bias (~ -3V) to the cells along the bit line. It can reduce the junction reverse bias from -6V (V_D= -6V; V_B= 0V) to -3V (V_D= -6V; V_B= -3V). Junction breakdown is prevented such that drain disturb is improved. Therefore, triple well is needed to provide individual substrate bias. Fig. 5.10 shows the improved structure of triple well NOR array. The well line is parallel to the word line. When programming the cell on the node of BL1 and WL1, well line 1 is grounded and -3V is applied at the other well lines. Thus the reverse bias of unselected cells is moderate and won't cause impact ionization. On the other hand, due to the triple well interconnection, additional isolation and contact area is needed to apply substrate bias and enlarges the cell size. Besides, additional substrate bias induces the other disturb phenomena: substrate disturb. Fig. 5.11 shows the substrate disturb characteristic of unselected cells. Fortunately, moderate substrate bias does not make threshold voltage shift even under long time stress such that the reliability issue of FBADHE operation can be solved by the improved structure. On the other hand, this

structure will make drain-substrate junction along the well line at forward bias mode and causes additional power consumption. So it is a tradeoff between reliability and power consumption. As a result, we suppose the DINOR-like structure (maybe four cells with a select transistor) is the better solution in this operation scheme.

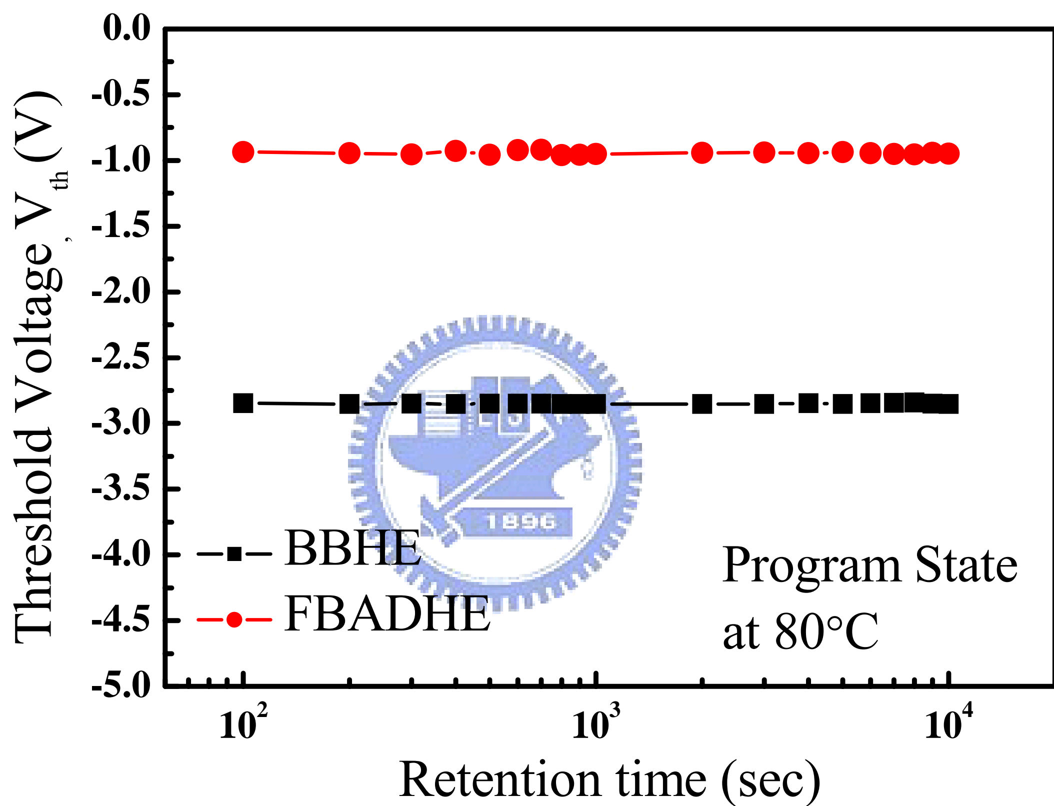


Fig. 5.4 Data retention characteristics of FBADHE and BBHE at 80°C in p+ poly gate flash cells. By the benefit of thick interpoly ONO and tunnel oxide, the data retention ability is excellent even in room temperature or high temperature (80°C).

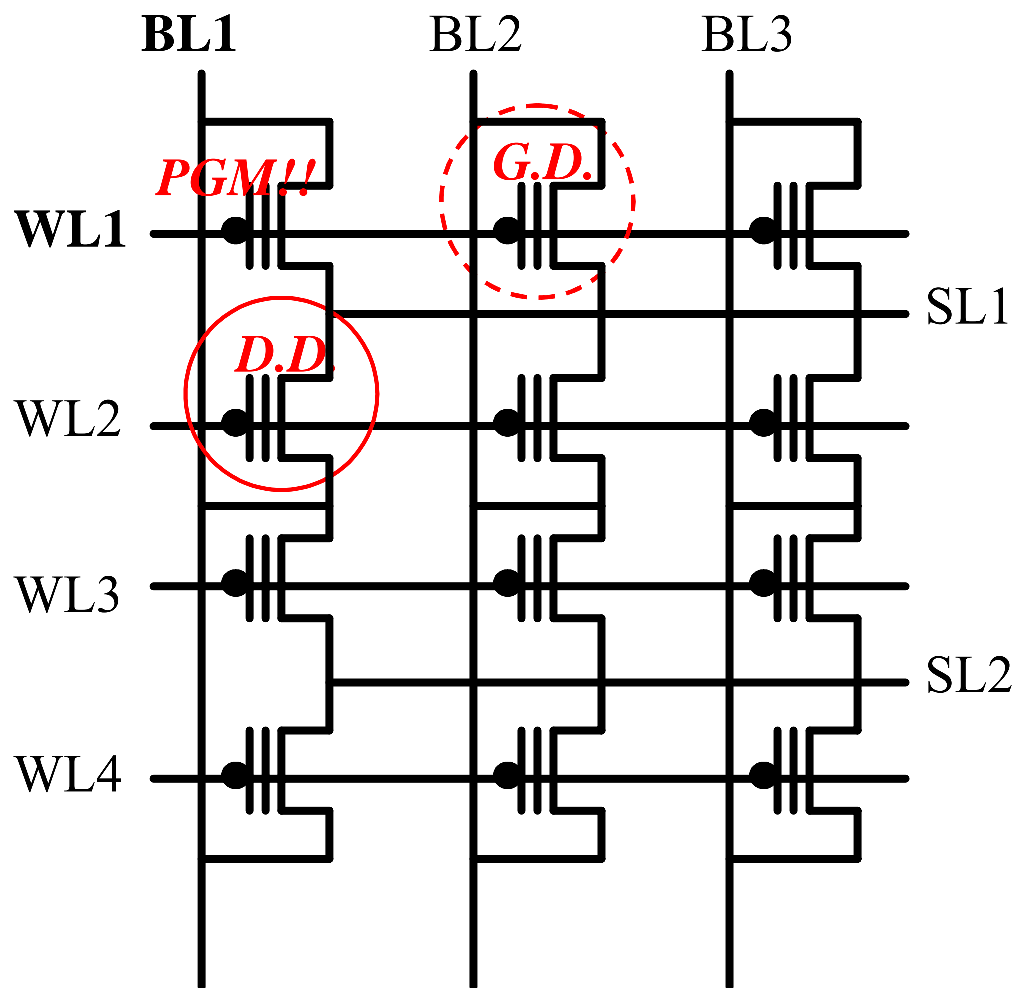


Fig. 5.5 Types of disturb in conventional NOR array. D.D. represents the drain disturb, which is along the Bit Line (BL1). G. D. represents the gate disturb, which is along the Word Line (WL1).

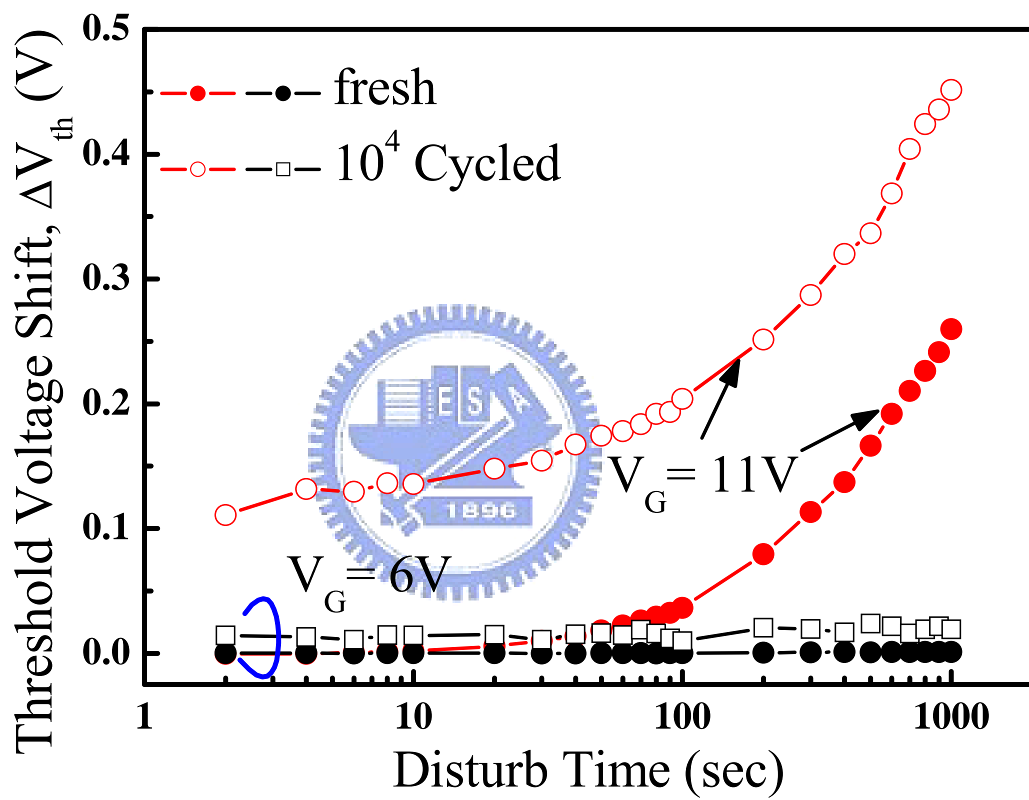


Fig. 5.6 Gate disturb characteristics with different gate voltage in p-channel flash cells.

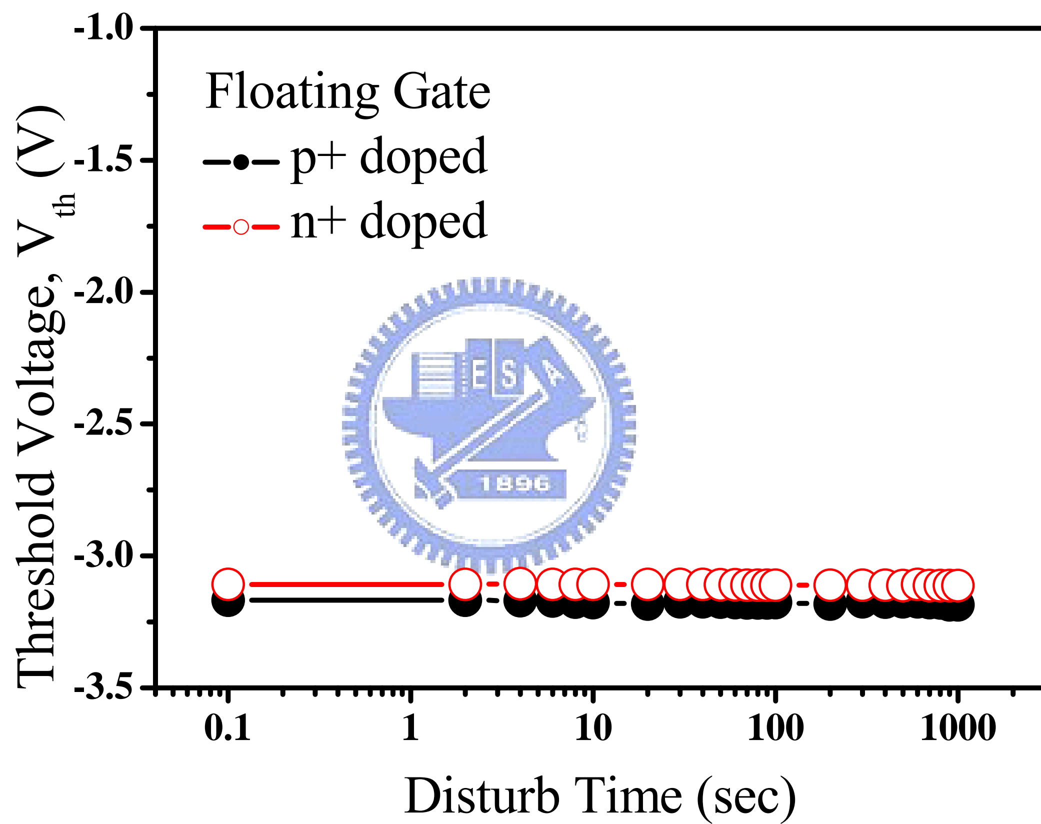


Fig. 5.7 Read disturb characteristics with different doped floating gate in p-channel flash cells after 10k cycling. Read disturb is very small due to the thick dielectric layer.

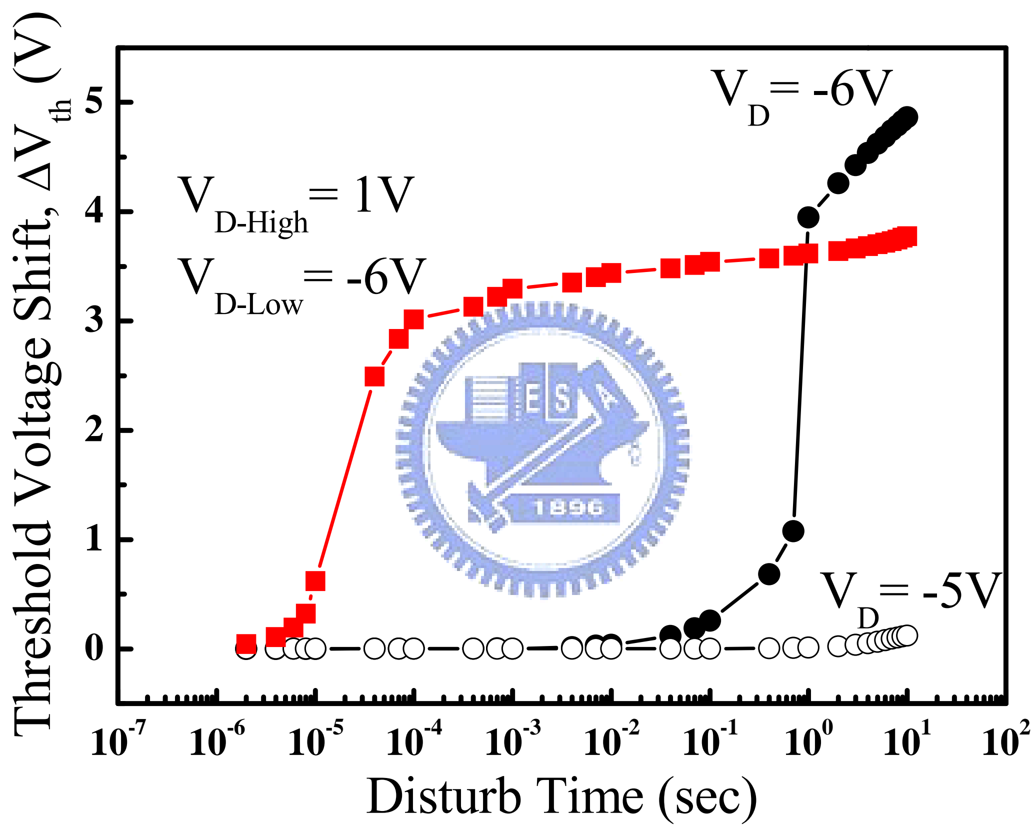


Fig. 5.8 Drain disturb characteristics with different gate voltage in fresh p-channel flash cells.

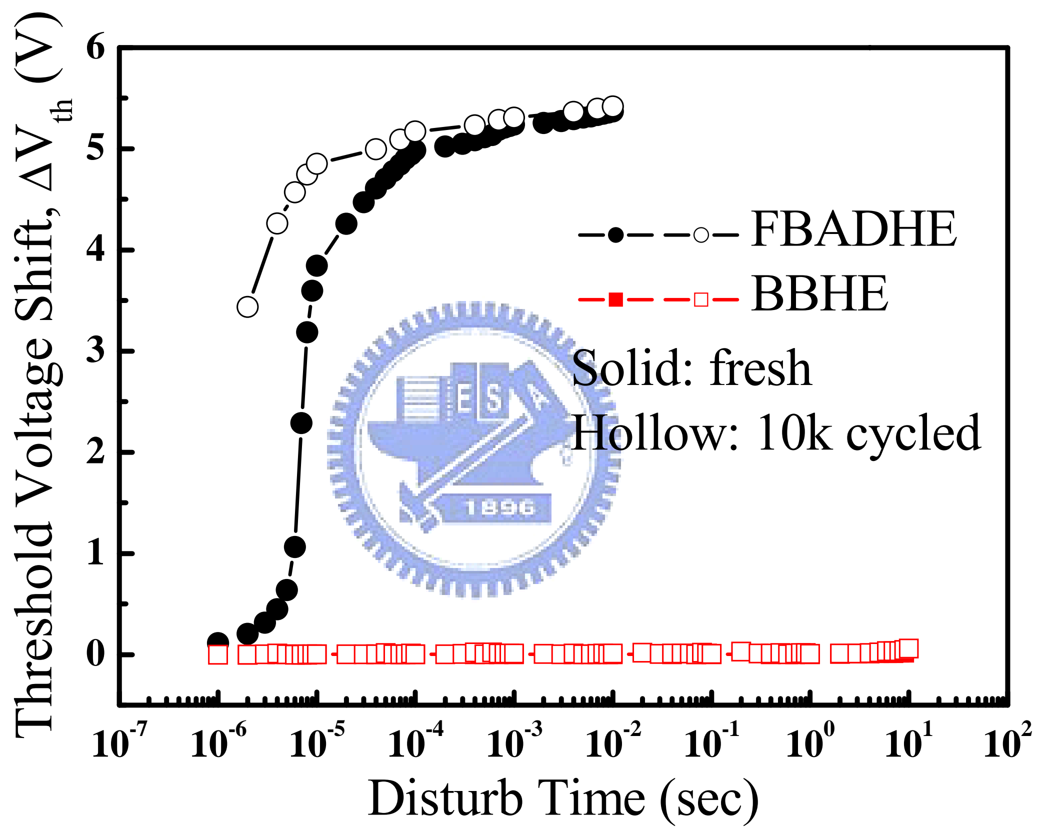


Fig. 5.9 Drain disturb characteristics with different drain voltage before and after 10k cycling in p-channel flash cells.

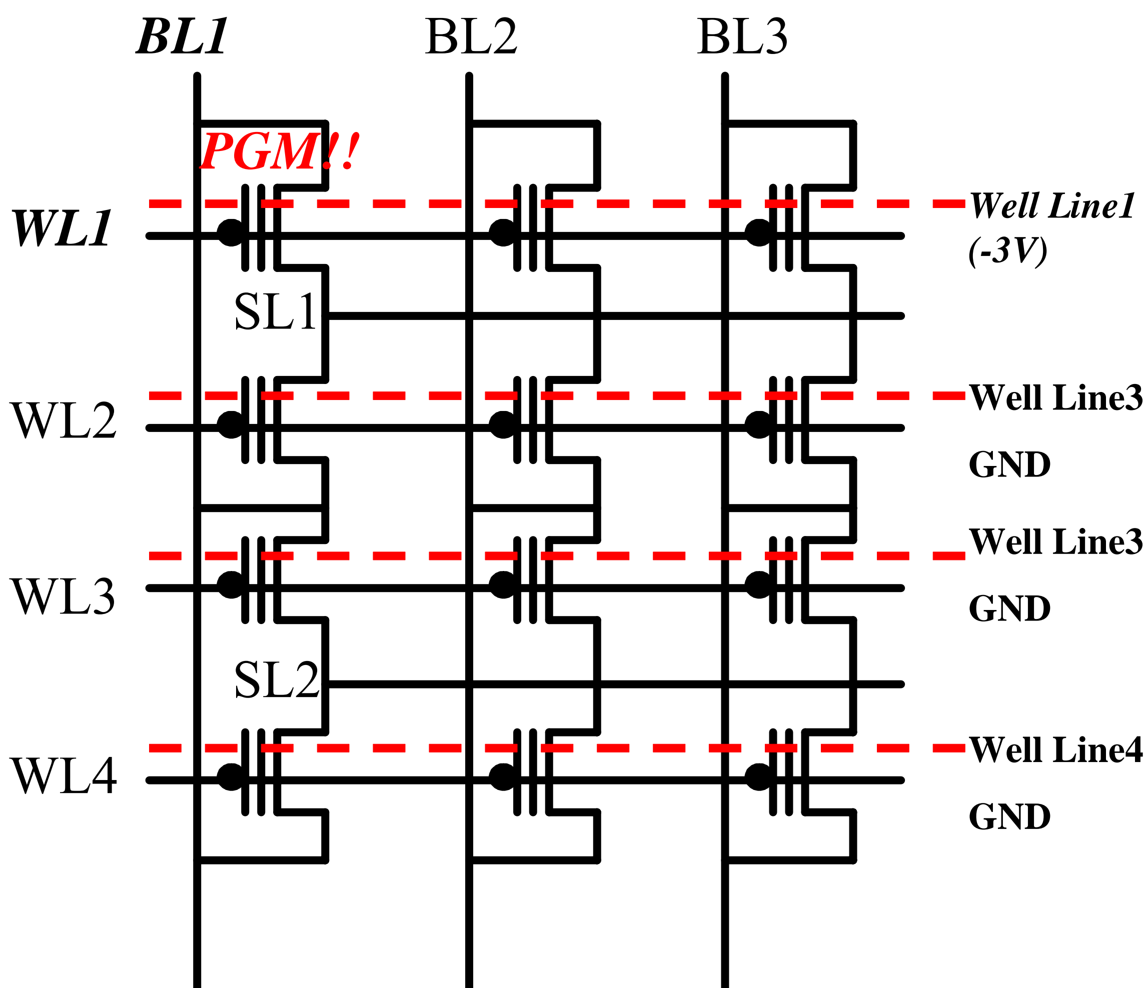


Fig. 5.10 Improved NOR array with triple well technology for applying the substrate bias.

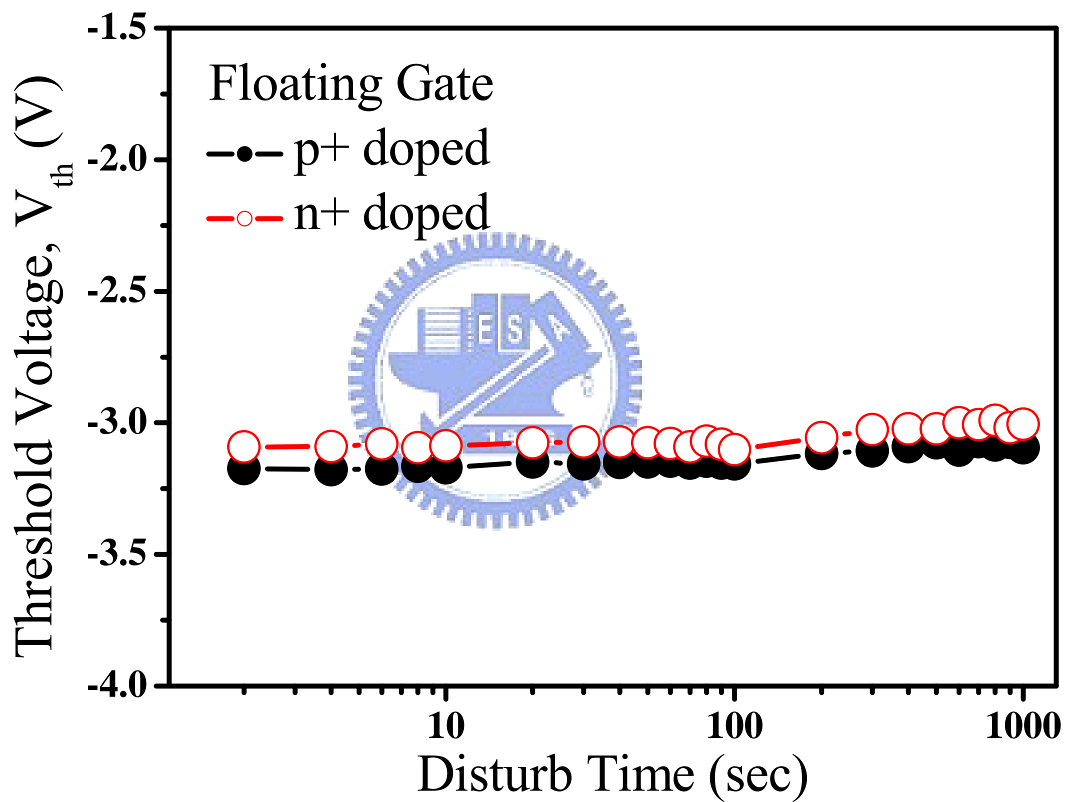


Fig. 5.11 Substrate disturb characteristics with different doped floating gate after 10k cycling in p-channel flash cells.

Chapter 6
Summary and Conclusions

Based on the experimental results, several issues have been addressed in this thesis. A novel programming scheme used in p-channel flash memories is proposed. i.e., Forward Bias Assisted Drain Hot Electron injection (FBADHE). The mechanism is to apply a forward bias on drain-substrate junction and then change the mode of p-n junction to reverse bias. Forward bias induced carrier flowing will generate more impact ionization under the reverse bias condition and more energetic electrons will be injected into the floating gate.

Then, we compare the characteristic of FBADHE. BBHE and DAHE. DAHE has the fastest programming speed, and BBHE has the lowest drain voltage. By comparing the transient characteristic, we can clarify the injection mechanism of FBADHE which involves band-to-band tunneling and drain avalanche induced hot electron, and the later one is the main injection component.

The advantage of FBADHE is fast programming speed (~1 μs) and low operating voltage (~6V). Even in flash cells with thick tunnel oxide and inter-poly, low operating voltage can achieve large window. It means that the data retention performance will be excellent. Comparing to BBHE, the higher gate voltage (~10V) limits the programming speed by the gate disturb phenomenon. The disadvantage of FBADHE is the severe drain disturb. It is also the major reliability concern of p-channel flash memories. We can improve the drain disturb performance by adopting a new array structure, which is described in Chapter 5. Although larger cell area is needed in new structure, the low voltage and high speed programming are suitable for the applications.

References

[1] F. Masuoka, M. Asano, H. Iwahashi, T. Lomuro and S. Tanaka, "A New flash EEPROM cell using triple polysilicon technology," in *IEDM Tech. Dig.*, pp. 464-467, 1984.

[2] N. Tsuji, N. Ajika, K. Yuzuriha, Y. Kunori, M. Hatanaka, and H. Miyoshi, "New Erase Scheme for DINOR Flash Memory Enhancing Erase/Write Cycling Endurance Characteristics," in *IEDM Tech. Dig.*, pp. 53-56, 1994.

[3] C. Y. Hu, D. L. Kencke, and S. K. Banerjee, "Substrate-current-induced hot electron (SCIHE) injection: a new convergence scheme for flash memory," in *IEDM Tech. Dig.*, pp. 283-286, 1995.

[4] J. D. Bude, M. Mastrapasqua, M. R. Rinto, R. W. Gregor, P. J. Kelley, R. A. Kohler, C. W. Leung, Y. Ma, R. J. McPartland, P. K. Poy, and R. Singh, "Secondary Electron Flash – a High Performance, Low Power Flash Technology for 0.35um and Below," in *IEDM Tech., Dig.*, pp. 279-282, 1997.

[5] T. Ohnakado, K. Mitsamaga, M. Nunoshita, H. Onoda, K. Sakakibara, N. Tsuji, N. Ajika, M. Hatanada, and H. Miyoshi, "Novel Electron Injection Method Using Band-to-Band Tunneling Induced Hot Electron (BBHE) for Flash Memory with a P-channel Cell," in *IEDM Tech. Dig*., pp. 279-282, 1995.

[6] B. Eitan, P. Pavan, I. Bloom, E. Aloni, A Frommer, and D. Finzi, "NROM: A novel localized trapping, 2-bit nonvolatile memory cell," in *IEEE Electron Device Letters*, Vol. 21, no. 11, 2000.

[7] I. C. Chen, D. J. Coleman, and C.W. Tang, "Gate current injection initiated by electron band-to-band tunneling in MOS devices," in *IEEE Electron Device Letters*, vol. 10, no. 7, pp. 297, 1989.

[8] K. Komori, K. Kuroda, S. Meguro, K. Nagasawa, M. Fukuda, K. Uchibori and T. Hagiwara, "A high performance memory cell technology for mega bit EEPROMs," in *IEDM Tech. Dig*., pp. 627-630, 1985.

[9] P. Pavan, R. Bez, P. Olivo and E. Zanoni, "Flash Memory Cells-An Overview," *Proceedings of the IEEE*, vol. 85, pp.1248, 1997.

[10] P. Pavan, L. Larcher and A. Marmiroli, Floating Gate Devices: Operation and Compact Modeling, *Kluwer Academic*, 2004.

[11] S. M. Sze, Physics of semiconductor devices, 2nd Ed. *Wiley-Interscience*, New York, p. 440, 1981.

[12] T. Y. Chan, J. Chen, P. K. Ko and C. Hu, "The Impact of Gate-Induced Drain Leakage Current on MOSFET Scaling," in *IEDM Tech. Dig*., pp. 718, 1987.

[13] Z. Liu and T. P. Ma, "A New Programming Technique for flash memory devices," in *VLSI TSA*, pp.195-198, 1999.

[14] T. Endoh, H. Iizuka, S. Aritome, R. Shirota and F. Masuoka, "New write/erase operation technology for flash EEPROM cells to improve the read disturb characteristics," in *IEDM Tech. Dig*., pp. 603-606, 1992.

[15] H. W. Tsai, P. Y. Chiang, S. S. Chung, D. S. Kuo and M. S. Liang, "The performance and Reliability of ETOX P-channel Flash EEPROM Cell With P-Doped Floating-Gate," in *VLSI TSA*, pp.36-39, 2003.

[16] A. T. Wu, T. Y. Chen, P. K. Ko and C. Hu,, "A source-side injection erasable programmable read-only-memory (SI-EEPROM) device," *IEEE Electron Device Letters*, vol. 7, no. 9, pp.540-542, 1986.

[17] T. E. Chang, C. Huang, and T. Wang, "Mechanism of Interface Trap-Induced Drain Leakage Current in Off-State N-MOSFET's" *IEEE Trans. on Electron Devices*, vol. 42, no. 4, pp738-743, 1995.

[18] R. S. Scott and D. J. Dumin, "The Charging and Discharging of High-Voltage Stress-Generated Traps in Thin Silicon Oxide," *IEEE Trans. Electron Devices*, vol. 43, pp. 130-136, 1996.

[19] J. Chen, T. Y. Chen, I. C. Chen, P. K. Ko and C. Hu, "Subbreakdown Drain Leakage Current in MOSFET," *IEEE Electron Device Letter*, vol. 8, no. 11, pp.515-517, 1987.

[20] S. Yamada, T. Suzuki, E. Obi, M. Oshikiri, K. Naruke, and M. Wada, "A

self-convergence erasing scheme for a simple stacked gate flash EEPROM," in *IEDM Tech. Dig*., pp. 307-310, 1991.

[21] S. Chung, S. T. Liaw, C. M. Yih, Z. H. Ho, "N-Channel versus P-channel Flash EEPROM- Which one has better reliabilities," in *Proc. IRPS*, pp. 67-72, 2001.

[22] K. T. San, C. Kaya, and T. P. Ma,"Effects of Erase Source Bias on Flash EEPROM Device Reliability," *IEEE Trans. Electron Devices*, vol. 42, no. 1, pp. 150-159, 1995.

[23] Ben. G. Streetman, Sanjay Banerjee, Solid State Electronic Devices, *Prentice Hall*, 1995.

[24] D. A. Neaman, Semiconductor Physics and Devices: Basic Principles, *McGraw Hill*, 1997.

[25] Victor. C. W. Kuo, C. M. Chao, C. K. Kang, L. E. Liu, L. T. Kuo, S. H. Chen, H. C. Wei, H. P. Hwang, and S. Pittikoun, "Detailed Comparisons of Program, Erase and Data Retention Characteristics between P+- and N+-Poly SONOS NAND Flash Memory," in *MTDT*, pp. 77-79, 2006.

[26] H. Onoda, Y. Kurori, S. Kobayashi, M. Ohi, A Fukumoto, N. Ajika, and H. Miyoshi, "A novel cell structure suitable for a 3V volt operation, sector erase flash memory," in *IEDM Tech. Dig*., pp. 599-602, 1992.

[27] K. Naruke, S. Taguchi, and M. Wada, "Stress Induced Leakage Current limiting to scale down EEPROM tunnel oxide thickness," in *IEDM Tech. Dig*., pp. 560-563, 1987.

簡　　歷

姓名：黃耀賢

性別：男

生日：69 年 7 月 2 日

籍貫：台灣省台南縣

地址：台北市文山區辛亥路五段 124 巷 20 號 4 樓

學歷：國立中正大學機械工程學系

　　　國立交通大學微電子奈米科技產業研發碩士班

碩士論文題目：

低電壓且高速操作的 P 通道快閃式記憶體元件性能及可靠性研究

Performance and Reliability Evaluation of a Low Voltage and High Speed P-channel Floating Gate Flash Memory