國 立 交 通 大 學

電子工程學系電子研究所

博 士 論 文

先進元件中原子尺寸缺陷之研究

Study on the Atomic-sized Traps in Advanced Devices

研 究 生 ： 呂明霈

指導教授 ： 陳明哲 博士

中華民國九十五年七月

先進元件中原子尺寸缺陷之研究

Study on the Atomic-sized Traps in Advanced Devices

研 究 生 ： 呂明霈　　Student : Ming-Pei Lu

指導教授 ： 陳明哲 博士　　Advisor : Dr. Ming-Jer Chen

國 立 交 通 大 學

電子工程學系　電子研究所

博 士 論 文

A Dissertation
Submitted to Institute of Electronics
College of Electrical & Computer Engineering
National Chiao Tung University
in Partial Fulfillment of the Requirement
for the Degree of Doctor of Philosophy
in
Electrical Engineering

July 2006
Hsinchu, Taiwan, Republic of China

中華民國九十五年七月

先進元件中原子尺寸缺陷之研究

研究生: 呂明霈　　指導教授: 陳明哲博士

國立交通大學 電子工程學系 電子研究所

摘要

隨著電子元件尺寸縮小的未來趨勢，將會面對更多的單電子效應，尤其以隨機擾動的電子訊號 (RTS) 更爲之重要。研究小尺寸元件中，經由單一一個氧化層缺陷所造成的雜訊行爲，可以提供更不一樣的當元件操作過程中氧化層退化的訊息。因此，一個單一缺陷是否爲中性或是帶電性，可經由電流的擾動行爲來觀察判斷。低頻雜訊 (又稱 1/f noise) 可以視爲一個在頻率域把所有不同的 RTS 訊號的貢獻都加起來的電子訊號，低頻雜訊可以當成研究半導體和絕緣層之間的介面性質的有效方法。

在本篇論文中最主要的目的是，更深入的研究探討雜訊與擾動在超薄氧化層元件、擁有高介電常數之絕緣層 (high-k) 的元件以及受製程應力 (process strained) 之元件。根據在不同元件上的研究，本篇論文的架構如下所描述。

首先，第一章是針對雜訊與擾動的介紹。接下來，在超薄氧化層元件中 RTS 現象的研究將在第二章中會有所描述，庫倫能量 (Coulomb energy) 可以視爲對一個奈米尺寸的缺陷充電過程中所不可忽略的要素，在本論文，我們首次發表實驗上在 1.7 奈米厚度的氧化層元件所做的研究分析，發現更深入氧化層的缺陷將會面對到更高的庫倫能量，另一個成果是，利用 multiphonon 理論成功的解釋電子被缺陷抓取以及釋放的能量問題。相對應的能量配置座標圖也建立在本論文中。在第三章中，我們將探討經由 RTS 振幅大小所淬取出庫倫散射在二維電子氣中所造成的影響，基於對一個單一缺陷所造成的 RTS 現象的相關實驗，發現到更

深的氧化層缺陷將會造成更小的庫倫散射現象。但是當在強反轉 (strong inversion) 範圍，庫倫散射對在表面的缺陷而言，會造成更強烈的變化現象，一個更強烈的遮蔽效應反映在庫倫電位是造成這這現象的主因，庫倫散射和缺陷的位置的關係將是未來奈米元件中更需要重視的問題。

接下來在第四章中，我們將研究有關高介電的絕緣層元件，打開-關閉切換的行爲或是兩個級別的隨機擾動雜訊都是在低電壓下量測 N 型通道超薄閘介電層(1 奈米的氧化層 + 1 奈米的氮化矽)金氧半電晶體的邊緣直接穿透電流。起因是由於製程所造成的缺陷可以說是局部的閘極堆疊變薄(或是等效於具傳導性的漏電流蘇)。在這個非固有的狀況中，電流中電子被陷阱抓住或是電子被陷阱釋放出來的理論可以適當的解釋我們所量到的數據，尤其隨機擾動雜訊振幅的大小比例高達百分之十八。電流對電壓的特性曲線很直覺的關聯著某些個缺陷點，展現和氧化層變薄的事實十分的一致。RTS 可以用來觀察電流流通過一個奈米線狀的缺陷點的有效工具，一個敏感的監督製程的角色就像是展示出缺陷發生的機率及位置所在。

之後在第五章中，低頻雜訊拿來使用在監控受不同的製程應力程度下的氧化層介面品質，在低頻雜訊在承受製程應力的金氧半電晶體中的量測中，發現靠近表面的缺陷密度隨著通道寬度而變化。這個發現可以解釋爲在矽和氧化矽的介面間，因爲晶格長度不匹配所造成的 P_b 中心可視爲靠近表面的缺陷的主要來源。在低頻雜訊的實驗中，對於通道寬度的縮減，相對應於應力的提高，也降低晶格長度不匹配的程度。最後，我們把所有所做過的研究結論放在第六章。

Study on the Atomic-sized Traps in Advanced Devices

Student: Ming-Pei Lu Advisor: Dr. Ming-Jer Chen

Department of Electronics Engineering & Institute of Electronics

National Chiao-Tung University

Hsinchu, Taiwan, R.O.C.

Abstract

The continuous shrinking in the feature dimensions of metal-oxide -semiconductor field-effect transistors (MOSFETs) brings into prominence the single electron effects, among which the most important is the Random Telegraph Signals (RTS). Studies on noise from individual oxide traps in small structures can supply new information of device operation as well as degradation phenomena. Thus, individual traps can be observed in their neutral or charged state and, as a consequence, the current fluctuates between two discrete levels. The low-frequency noise (so-called 1/f noise) can be considered as the superposition of several random telegraph signals (RTS) in frequency domain. The 1/f noise can be used as a potential tool for studying the interface between the semiconductor and insulator.

The main purpose of this dissertation is to deeply investigate the fluctuations and noise in ultrathin oxide devices, high-k devices and process strained devices. Based on the study of different devices, the organization of this dissertation is described below.

First, an introduction to the RTS and noise is described in Chapter 1. The study of the RTS phenomenon in ultrathin oxide devices is demonstrated in Chapter 2 of the dissertation. Coulomb energy is essential to the charging of a nanometer-scale trap in the oxide of a metal-oxide-semiconductor (*MOS*) system. In this dissertation, we present for the first time experimental evidence from a 1.7-nm oxide: substantial enhancements in Coulomb energy due to the existence of a deeper trap in the oxide. Other corroborating evidence is achieved on a multiphonon theory, which can adequately elucidate the measured capture and emission kinetics. The corresponding configuration coordinate diagrams are established. Then, Chapter 3 presents the study on Coulomb scattering in a two-dimensional electron gas (2DEG) system through the relative amplitude of RTS. Experiments on an individual nanoscale trap in the oxide responsible for random telegraph signals lead to remarkable results. In this work, we demonstrated a study for relationship between the capture time, emission time, and the relative amplitude. Initially, the deeper trap in oxide corresponds to weaker Coulomb scattering in a 2DEG. However, as the 2DEG enters into the strong inversion regime, the amount of Coulomb scattering with an interface trap drops with a faster rate than the deep trap. A stronger screening potential confinement is shown to be the physical origin of this effect. The near-distance effect is expected to remain a challenging issue in the area of nanoscale devices.

Second, the study of the high-k devices is described in Chapter 4. On-off switching behaviors or two-level RTS are measured in the low voltage direct tunneling currents in ultrathin gate stack (10 Å oxide + 10 Å nitride) tunneling diode. The plausible origin is the process-induced defects in terms of localized gate stack thinning (or equivalently the conductive filament). In such extrinsic case, the current trapping-detrapping theories can adequately elucidate the data, particularly the RTS magnitude as large as 18%. The current-voltage (I-V) characteristic associated with a

certain defective spot is assessed straightforwardly, showing remarkable compatibility with existing oxide thinning case. The current tunneling through the wire-like weakened spot can be probed by RTS. The role as a sensitive process monitor is demonstrated in terms of the occurrence probability of the defects as well as their locations.

Third, the 1/f noise used to monitor the quality of oxide interface with different tensile stress is presented in Chapter 5. Low-frequency noise measurement in process tensile-strained n-channel metal-oxide-semiconductor field-effect transistors yields the density of the interface states, exhibiting a decreasing trend while decreasing the channel width. This finding corroborates the group of P_b centers caused by the lattice mismatch at (100) Si/SiO_2 interface as the origin of the underlying interface states. The present noise experiment therefore points to the enhancement of the tensile strain in the presence of channel narrowing, which in turn reduces the lattice mismatch. Finally, we summarize the conclusion of our works in Chapter 6.

致謝

首先，這篇論文能完成，必須要感謝我的指導教授陳明哲博士，他不僅是在學業上給予我耐心的指導，也在我的研究陷入迷惘的時候，指引我正確方向，以及教導我做人處世的態度。他不僅僅是我求學生涯的導師，也是我人生學習的對象，在與其相處的這幾年，這所有一切都是我永難忘懷的人生經驗。另外感謝師母侯錦珊博士，感謝她在工作繁忙之時，還能對於實驗室每一份子付出那份溫暖的關懷。

再來要感謝在我求學過程中，幫助過我，關心過我的學長們，感謝矽介觀系統實驗室的每一份子，感謝你們對實驗室的付出，也才有這篇論文的完成。在實驗室這幾年，因爲有你們的陪伴，才讓我在孤獨的研究路上走完一個階段。感謝國立交通大學的生物科技學系的楊裕雄教授研究團隊以及蕭程允同學，感謝楊教授提供研究上的資源以及合作的機會，也感謝蕭同學在我研究過程中提供不同方向的研究想法以及實驗上的幫助，也感謝在研究過程中，提供意見的研究團隊其他成員。感謝台積電的李文欽博士提供量測元件以及研究上的指導，感謝工研院的魏拯華博士提供奈米碳管元件，感謝所有幫助過我，關心過我的人。

最後，感謝這幾年來，若不是我家人朋友全心全力的支持，讓我無後顧之憂，可以放心的朝向自己想要的目標前進，也因爲有他們，才有今天的我，感謝奶奶、爸爸、媽媽、敏瑜、明穎、明諺、舒婷，也以這篇論文，獻給在我博士班求學過程中，去世的爺爺。

Table of Contents

Figure Captions

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

List of Symbols

Chapter 2

τ_c mean time for an electron captured by a single trap

τ_e mean time for an electron emission from a single trap

q elementary charge

t_{ox} oxide thickness

z_T trap depth

A_{DB} effective Debye screening area

C_{ox} gate oxide capacitance per unit area

C_g trap to anode (near the gate) capacitance per unit area

C_c trap to cathode (near the channel) capacitance per unit area

C_{inv} inversion layer capacitance per unit area

C_{dep} silicon depletion layer capacitance per unit area

C_{eff} equivalent capacitance per unit area

Q_G image charge developed on the gate electrode

k_B Boltzmann's constant

T absolute temperature

E_F Fermi energy level

E_T trap energy level

ν_{th} carrier thermal velocity

n_s inversion-layer electron density per unit area

z_{qm} average thickness of the inversion layer

σ capture cross section

E_0 energy level of the lowest subband for unprimed valley

$S\hbar\omega$ the lattice relaxation energy

ΔE_S the storage energy

Chapter 3

I_D drain current

ΔI_D drain current fluctuation

Q_{inv} image charge developed on the channel

A channel area

N_s carrier density per unit area

μ total mobility

μ_{ox} mobility due to the charged oxide trap

μ_n mobility due to the other mechanisms

N_{ox} number of charged oxide trap per unit area

α Coulomb scattering coefficient

m^* the effective mass

ΔE_C barrier height of the conduction-band discontinuity

$\varepsilon(z)$ position-dependent dielectric permittivity

ε_{si} permittivity of silicon

N_i inversion carrier density for subband *i*

E_f Fermi energy level

Chapter 4

t_{EOT} gate stack's equivalent oxide thickness

t_{ox} total effective oxide thickness

ε_{ox} permittivity of oxide

ε_{ni} permittivity of nitride

τ_{on} mean time for an electron captured by a single trap

τ_{off} mean time for an electron emission from a single trap

Δt_{ox} gate dielectric thinning thickness

J_n tunneling current density associated with the gate stack thickness t_{ox}

A_{EDT} edge direct tunneling area

ΔJ_n tunneling current density for a net thickness of t_{ox} - Δt_{ox}

Chapter 5

S_{vg} input-referred noise voltage spectral density

C_{eff} equivalent capacitance per unit area

q elementary charge

k_B Boltzmann's constant

T absolute temperature

λ tunneling distance (~ 0.1 nm)

W channel width of active area

L channel length of active area

Chapter 1

Introduction to Random Telegraph Signals and 1/f Noise

1.1 Overview

The continuous shrinking in the feature dimensions of metal-oxide -semiconductor field-effect transistors (MOSFETs) brings into prominence the single electron effects, among which the most important is the Random Telegraph Signals (RTS). These signals are generally considered as carrier trapping–detrapping from traps situated in the silicon oxide [1-6]. Studies on noise from individual oxide traps in small structures can supply new information of device operation as well as degradation phenomena. When the channel area of the devices is small enough, it is possible that only one trap or few traps may be close to the surface Fermi level over the entire channel [1-5]. Thus, individual trap can be observed in their neutral or charged state and, as a consequence, the current fluctuates between two discrete levels. Taking an acceptor-like trap as one example, the high-level current corresponds to the trap in a neutral state while the low-level current corresponds to the negatively charged state. Fig. 1.1 and Fig. 1.2 show the two-level current fluctuations and multi-level current fluctuations due to a single trap and few traps, respectively. Therefore, the RTS properties are completely determined by the up and down times and its amplitude. Through the examination of the mean captured and emission times, the trap depth and trap energy level can be obtained [1,3-5]. However, there is the

lack of a clear physical model including power dissipation and quantum effect. Different models trying to explain RTS amplitude have been proposed [3,7-8], although the physical grounds of RTS phenomenon are not fully understood yet. Therefore, any results, whether experimentally or theoretically, that can shed light on this subject will undoubtedly be welcome. The low-frequency noise (so-called 1/f noise) can be considered as the superposition of several random telegraph signals (RTS) in frequency domain [3,9-11]. The noise characteristic is shown in Fig. 1.3. The 1/f noise can be used as a potential tool for studying the underlying states near the interface of semiconductor and insulator. The characteristic of the 1/f noise can be related to the property of only a few traps existing in the nanoscale device. We hence believe that the 1/f noise serves as a useful means to analyzing the properties of the interface for the nanoscale devices.

1.2 Dissertation Organization

The main purpose of the dissertation is to deeply investigate the fluctuations and noise in ultrathin oxide devices, high-k devices and strained devices. Based on different devices, the organization of this dissertation is described below.

First, an introduction to the RTS and noise is described in Chapter 1. Chapter 2 of the dissertation is focused on the RTS phenomenon in ultrathin oxide devices. Coulomb energy is essential to the charging of a nanometer-scale trap in the oxide of a metal-oxide-semiconductor (*MOS*) system. In this dissertation, we present for the first time experimental evidence from a 1.7-nm oxide: substantial enhancements in Coulomb energy due to the existence of a deeper trap in the oxide. Other corroborating evidence is achieved on a multiphonon theory, which can adequately elucidate the measured capture and emission kinetics. The corresponding configuration coordinate diagrams are established. Thus, Chapter 3 presents the study

on Coulomb scattering in a two-dimensional electron gas (2DEG) system through the relative amplitude of RTS. Experiments on an individual nanoscale trap in the oxide responsible for random telegraph signals lead to remarkable results. Initially, the deeper trap in oxide corresponds to weaker Coulomb scattering in a 2DEG. However, as the 2DEG enters into the strong inversion regime, the amount of Coulomb scattering with an interface trap drops with a faster rate than the deep trap. A stronger screening potential confinement is shown to be the physical origin of this effect. The near-distance effect is expected to remain a challenging issue in the area of nanoscale devices.

Second, the study of the high-k devices is described in Chapter 4. On-off switching behaviors or two-level RTS are measured in the low voltage direct tunneling currents in ultrathin gate stack (10 Å oxide + 10 Å nitride) tunneling diode. The plausible origin is the process-induced defects in terms of localized gate stack thinning (or equivalently the conductive filament). In such extrinsic case, the current trapping-detrapping theories can adequately elucidate the data, particularly the RTS magnitude as large as 18%. The current-voltage (I-V) characteristic associated with a certain defective spot is assessed straightforwardly, showing remarkable compatibility with existing oxide thinning case. The current tunneling through the wire-like weakened spot can be probed by RTS. The role as a sensitive process monitor is demonstrated in terms of the occurrence probability of the defects as well as their locations.

Third, the 1/f noise used to monitor the quality of oxide interface with different tensile stress is presented in Chapter 5. Low-frequency noise measurement in process tensile-strained n-channel metal-oxide-semiconductor field-effect transistors yields the density of the interface states, exhibiting a decreasing trend while decreasing the channel width. This finding corroborates the group of P_b centers caused by the lattice

mismatch at (100) Si/SiO_2 interface as the origin of the underlying interface states. The present noise experiment therefore points to the enhancement of the tensile strain in the presence of channel narrowing, which in turn reduces the lattice mismatch. Finally, in Chapter 6 we summarize the conclusion of our works.

References

[1] K. S. Ralls, W. J. Skocpol, L. D. Jackel, R. E. Howard, L. A. Fetter, R. W. Epworth and D. M. Tennant, "Discrete Resistance Switching in Submicrometer Silicon Inversion Layers: Individual Interface Traps and Low-Frequency (1/*f*?) Noise," *Phys. Rev. Lett.*, vol. **52**, pp. 228-231, 1984.

[2] K. R. Farmer, C. T. Rogers, and R. A. Buhrman, "Localized-State Interactions in Metal-Oxide-Semiconductor Tunnel Diodes," *Phys. Rev. Lett.*, vol. **58**, pp. 2255-2258, 1987.

[3] M. J. Kirton and M. J. Uren, "Noise in solid-state microstructures: a new perspective on individual defects, interface states and low-frequency (1/f) noise," *Adv. Phys.*, vol. **38**, pp. 367-468, 1989.

[4] M. Schulz, "Coulomb energy of traps in semiconductor space-charge regions," *J. Appl. Phys.*, vol. **74**, pp. 2649-2657, 1993.

[5] H. H. Mueller, D. Wörle, and M. Schulz, "Evaluation of the Coulomb energy for single-electron interface trapping in sub-μm metal-oxide-semiconductor field-effect transistors," *J. Appl. Phys.*, vol. **75**, pp. 2970-2979, 1994.

[6] M. J. Chen and M. P. Lu, "On–off switching of edge direct tunneling currents in metal-oxide-semiconductor field-effect transistors," *Appl. Phys. Lett.*, vol. **81**, pp. 3488-3490, 2002.

[7] K. K. Hung, P. K. Ko, C. Hu and Y. C. Cheng, "Random telegraph noise of deep-submicrometer MOSFETs," *IEEE Electron Device Lett.*, vol. **11**, pp. 90-92, 1990.

[8] E. Simoen, B. Dierickx, C. L. Claeys, and G. J. Declerck, "Explaining the amplitude of RTS noise in submicrometer MOSFETs," *IEEE Trans. Electron Devices*,

vol. **39**, p. 422, 1992.

[9] K. K. Hung, P. K. Ko, C. Hu, and Y. C. Cheng, "A unified model for the flicker noise in metal-oxide-semiconductor field-effect transistors," *IEEE Trans. Electron Devices*, vol. **37**, pp. 654-665, 1990.

[10] M. J. Chen, T. K. Kang, Y. H. Lee, C. H. Liu, Y. J. Chang, and K. Y. Fu, "Low-frequency noise in *n*-channel metal-oxide-semiconductor field-effect transistors undergoing soft breakdown," *J. Appl. Phys.*, vol. **89**, pp. 648-653, 2001.

[11] E. Simoen, G. Eneman, P. Verheyen, R. Delhougne, R. Loo, K. De Meyer, and C. Claeys, "On the beneficial impact of tensile-strained silicon substrates on the low-frequency noise of *n*-channel metal-oxide-semiconductor transistors," *Appl. Phys. Lett.*, vol. **86**, 223509, 2005.

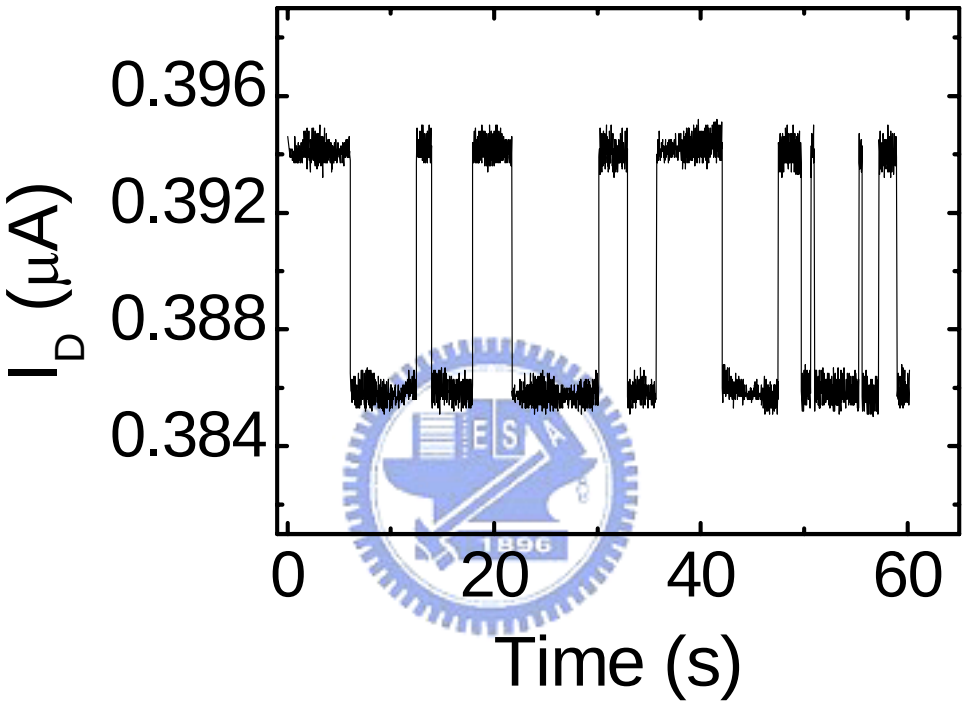

Fig. 1.1 The two-level RTS in drain current measured from nMOSFET with 1.7 nm gate oxide.

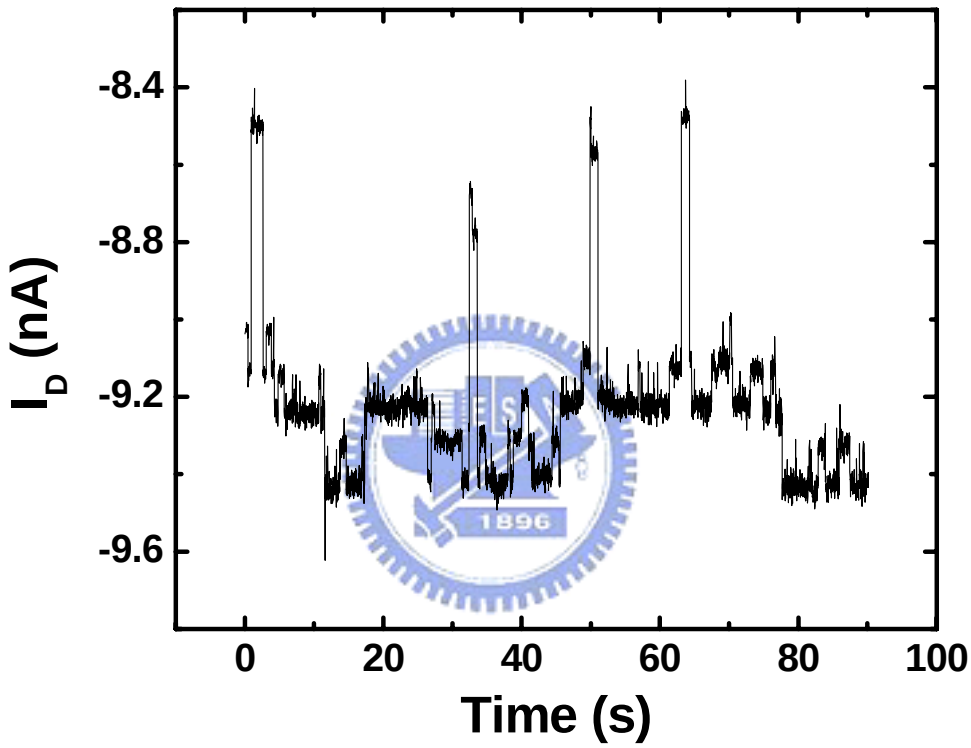

Fig. 1.2 The multi-level RTS in drain edge direct tunneling current measured from nMOSFET with ultrathin gate stack (10 Å oxide + 10 Å nitride).

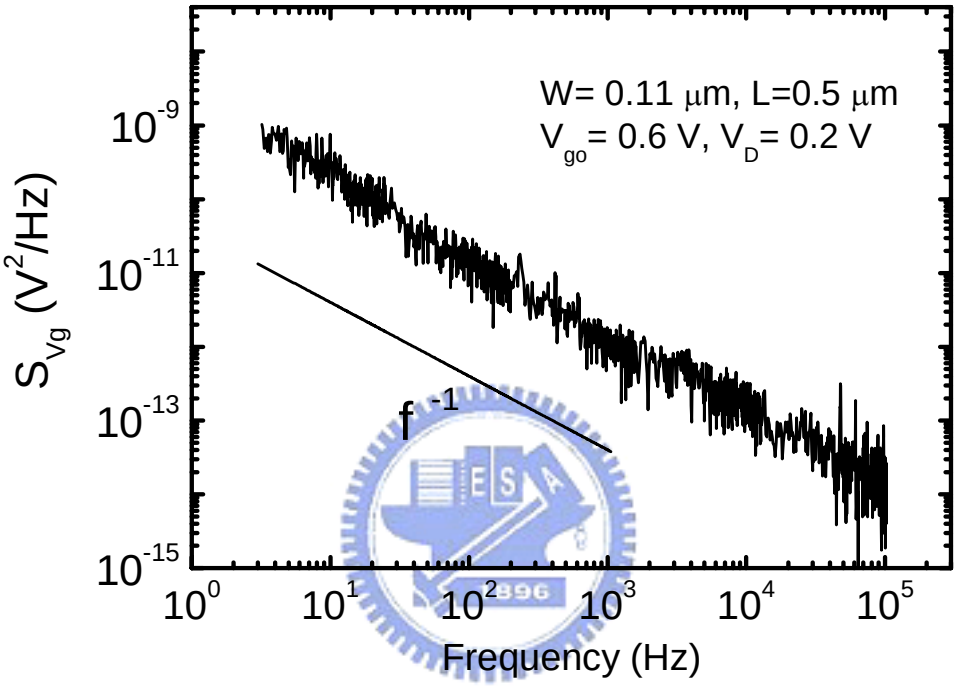

Fig. 1.3 The low-frequency (1/f) noise measured from the device.

Chapter 2

Oxide Trap Enhanced Coulomb Energy in a Metal-Oxide-Semiconductor System

Coulomb energy is essential to the charging of a nanometer-scale trap in the oxide of a metal-oxide-semiconductor (*MOS*) system. Traditionally the Coulomb energy calculation was performed on the basis of an interface-like trap. In this chapter, we present for the first time experimental evidence from a 1.7-nm oxide: substantial enhancements in Coulomb energy due to the existence of a deeper trap in the oxide. Other corroborating evidence is achieved on a multiphonon theory, which can adequately elucidate the measured capture and emission kinetics. The corresponding configuration coordinate diagrams are established. We further elaborate on the clarification of the Coulomb energy and differentiate it from that in memories containing nanocrystals or quantum dots in the oxide. Some critical issues encountered in the work are addressed as well.

2.1 Introduction

In a metal-oxide-semiconductor (*MOS*) system, a Coulomb barrier arises during the charging of a nanometer-scale trap in the oxide. Thus, a critical energy to overcome the barrier, namely Coulomb energy, plays a vital role in the capture kinetics.[1,2] Traditionally the Coulomb energy was calculated on the basis of an

interface-like trap. This treatment essentially remains valid if the oxide used is much thicker. However, with the currently aggressive downscaling of the oxide thickness, the oxide trap is likely situated deeper into the oxide from the SiO_2/Si interface and therefore, the Coulomb energy is expected to be affected due to enhanced image charge. However, little work has been done in this direction since the introduction of the Coulomb energy concept [1-2]. On the other hand, it is noteworthy that the definition of the Coulomb energy in the case of oxide trap [1-2] is significantly different from that in memories containing nanocrystals or quantum dots in the oxide [3-6]. However, such confusing issue has not been clarified yet.

In this dissertation, we exhibit for the first time experimental evidence for the Coulomb energy enhancement in the presence of a deeper oxide trap. The other corroborating evidence is achieved based on a multiphonon theory with the configuration coordinate diagrams taken into account. We further elaborate on the clarification of the Coulomb energy in a *MOS* system containing a nanometer-scale trap in the oxide and differentiate it from that in a *MOS* memory containing a nanocrystal or dot in the oxide, followed by a concrete discussion on the critical issues encountered in the work.

2.2 Experimental

The n-channel metal-oxide-semiconductor field-effect transistors (*MOSFETs*) with varying channel lengths and widths (60 nm to 600 nm) were fabricated in a state-of-the-art manufacturing process. The key process parameters as obtained by capacitance-voltage (*C-V*) fitting were n^+ polysilicon doping concentration = 1.3×10^{20} cm^{-3}, gate oxide thickness = 1.7 nm, and channel doping concentration = 8×10^{17} cm^{-3}. To detect a potential oxide trap with fluctuating occupancy, the random telegraph signals (*RTS*) measurement is a good means [1-2, 7-9]. The *RTS* measurement

equipment and method used were the same as that described elsewhere [10]. The operating conditions at room temperature were V_D = 10 mV and with V_G ranging from 0.2 V to 0.4 V. The purpose of the low voltage operation is twofold: (i) it can ensure no extra trap created during the long-term *RTS* measurement; and (ii) the devices under study can readily reduce to a near-equilibrium *1-D MOS* system. We conducted extensively *RTS* measurement across the whole wafer and found that as expected, the occurrence probability of *RTS* events in underlying devices is extremely low. For those devices identified with *RTS*, it was found that (i) the same abrupt transitions between two distinct states in drain current also simultaneously occur in source current; and (ii) no such noticeable changes can be observed in gate or bulk current, opposed to the recent literature [11] with a smaller oxide thickness (~1.3 nm). Therefore, the *RTS* events encountered in our work are due to the transfer of a single electron between a certain process-induced defect in the oxide and the underlying conductive channel layer. The capture time associated with the upper level of *RTS* current and the emission time associated with the lower current level both were exponentially distributed. The mean of the capture time distribution, designated τ_c, divided by the mean of the emission time distribution, τ_e, is given in Fig. 2.1 against gate voltage for two devices labeled Trap *A* and Trap *B*. The inset of Fig. 2.1 shows the corresponding time evolutions of RTS drain current at a certain gate voltage. Fig. 1 reveals that while initially the τ_c/τ_e ratio is comparable between Trap *A* and Trap *B*, with gate voltage increasing further, the Trap B's τ_c/τ_e drops with a faster rate than Trap *A*.

2.3 Analysis and Physical Interpretations

The size of the trap under study must be significantly less than the oxide thickness used (1.7 nm) since no noticeable change in the gate current was observed.

Hence, the trap responsible for the measured RTS in drain current is a nanometer-scale trap. To explore the measured τ_c/τ_e, it is necessary to know in advance the amount of the image or induced charge on the gate as a single electron is inserted into the oxide trap. First of all, it is well recognized that once a single electron is inserted into the oxide, the Debye screening length of a single electron (~70 nm) [4, 6] develops laterally around a ***negatively charged*** nanometer-scale trap in the oxide. Here, the Debye screening length is the effective size of the "cloud" of the induced charges on the electrodes. Thus, only within the Debye screening length can the plate capacitor approximation readily apply, leading to a capacitive coupling equivalent circuit as shown in Fig. 2.2. The capacitance model accounts for the effect of the trap depth and the charge sharing between gate, inversion layer, and silicon depletion region. Owing to the insertion of one electron into a depth z_T from the SiO_2/Si interface, the gate oxide capacitance per unit area C_{ox} associated with the oxide thickness t_{ox} can be separated into two distinct components: the trap to anode (near the gate) capacitance per unit area $C_g = C_{ox}t_{ox}/(t_{ox} - z_T)$ and the trap to cathode (near the channel) capacitance per unit area $C_c = C_{ox}t_{ox}/z_T$. The other capacitances such as the inversion-layer capacitance per unit area C_{inv} and the silicon depletion capacitance per unit area C_{dep} can be quantified using a self-consistent Schrödinger-Poisson equations solver with the process parameters mentioned above as input. Fig. 2.3 shows the simulated results of the key capacitance components versus gate voltage. The proposed capacitance model exactly reduces to that by Schulz[1] for the case of $z_T = 0$. Indeed, the calculated results on a 17-nm oxide are consistent with those in the literature [2].

While a single electron is inserted into the trap, the potential change, ΔV, in the trap reads as $\Delta V = q/(A_{DB} \times C_{eff})$ where A_{DB} is the effective Debye screening area and C_{eff}, the equivalent capacitance per unit area seen from the trap to the ground, can be

derived from the model. Then the image charge (positive) Q_G developed on the gate electrode can be expressed as $Q_G = \Delta V \times (A_{DB} \times C_g)$. Combining both equations while eliminating the common factor (i.e., Debye screening area), one achieves Q_G (= qC_g/C_{eff}):

$$Q_G = q \times \frac{z_T \times (C_{inv} + C_{dep}) + C_{ox} t_{ox}}{t_{ox} C_{ox} + t_{ox} (C_{inv} + C_{dep})}. \quad (1)$$

The calculated gate image charge as depicted in Fig. 2.4 remains constant until a *2DEG* (2-D electron gas) layer critically appears (at $V_G \approx 0.1$ V), and then due to increasing screening by the inversion-layer charge, the gate image charge decreases with increasing gate voltage. Specifically, the figure reveals that an increase in the trap depth can substantially increase the gate image charge. In the presence of a *2DEG* layer, the source and drain are electrically tied together and thereby the Coulomb energy can readily be written as $\Delta E \approx Q_G V_G$ [1-2]. The calculated Coulomb energy is together plotted in Fig. 2.4, showing that the Coulomb energy associated with the interface trap increases with gate voltage until encountering a certain peak. However, such peak point disappears in the case of non-zero trap depth and the Coulomb energy instead piles up over the conventional value.

According to the principle of detailed balance with the Coulomb energy included, the τ_c/τ_e ratio can read as [1]

$$\frac{\tau_c}{\tau_e} = e^{\frac{E_T - E_F + \Delta E}{k_B T}}. \quad (2)$$

In (2), the trap level E_T relative to the quasi-Fermi level E_F is a function of gate

voltage and can readily be quantified using the Schrödinger-Poisson solver. The band diagram indicating the Fermi level, trap energy level and trap depth is shown in Fig. 2.5. The best fitting results achieved using (2), with z_T = 0.7 nm and $E_{OX} - E_T$ = 3.2 eV for Trap A and z_T = 0 nm and $E_{OX} - E_T$ = 3.3 eV for Trap *B*, are shown in Fig. 2.1. Here E_{OX} denotes the oxide conduction band edge. Evidently, the fitting quality is fairly good. The extracted $E_{OX} - E_T$ values are close to the SiO_2/Si interface barrier height, as expected due to the low voltage operation. It is hence argued that an interface trap exists in Trap *B* device while a 0.7-nm deep trap in the oxide prevails in Trap *A*. In other words, the conventional Coulomb energy appears to work well for the Trap *B* device but leads to poor quality in fitting Trap *A*'s data. Such remarkable difference in τ_c/τ_e between Trap *A* and Trap *B* can therefore serve as experimental evidence of the Coulomb energy enhancement.

Other corroborating evidence can be obtained through the fitting of the measured mean capture time versus gate voltage as shown in Fig. 2.6. Since the capture kinetics involve the thermal activation process at room temperature of operation, a multiphonon emission theory was utilized to calculate the capture time:

$$\frac{1}{\tau_c} = \sigma v_{th} \frac{n_s}{z_{qm}} e^{-\frac{\Delta E}{k_B T}} \tag{3}$$

where v_{th} is the carrier thermal velocity ($\approx 1.23 \times 10^5$ m/s), n_s is the inversion-layer electron density per unit area and z_{qm} is the average thickness of the inversion layer. σ is the multiphonon capture cross section and can be written as

$$\sigma = \sigma_0 e^{-\frac{E_B}{k_B T}}. \tag{4}$$

The pre-factor σ_0 involves the interaction between the trap state and free electron wave function. E_B is the thermal activation barrier height and according to multiphonon emission theory the thermal activation barrier height at high temperature ($k_BT > \hbar\omega/2$) can reduce to [12-13]

$$E_B = \frac{(E_0 - E_T - S\hbar\omega)^2}{4S\hbar\omega} \tag{5}$$

where E_0 is the energy level of the lowest subband for unprimed valley and $S\hbar\omega$ is the lattice relaxation energy (S is the Huang-Rhys factor). Fitting the τ_c data in Fig. 2.6 to (3) yielded the lattice relaxation energy $S\hbar\omega$ of 1.2 eV and 0.025 eV for Trap A and Trap B, respectively; and σ_0 of 2.03×10^{-23} m^2 and 3.66×10^{-22} m^2 for Trap A and Trap B, respectively. The fitting quality is again good and the same parameters readily reproduced the τ_e data as depicted in Fig. 2.6. Specifically, the extracted σ_0 values are physically reasonable from the viewpoint of the penetration of the wave function into the oxide: the capture cross section decreases with increasing trap depth from the SiO_2/Si interface. The extracted values of the lattice relaxation energy also correctly reflect the status of the trap: a deeper trap (i.e., Trap A in our work) is accompanied with a higher lattice relaxation energy [14-15]. Using above extracted results, we constructed a configuration coordinate diagram of the underlying electron-lattice system as schematically shown in Fig. 2.7 for both devices. Also plotted in Fig. 2.7 are the *MOS* energy band diagrams (removing the polysilicon part) in flatband case, showing the spatial distance and energetic level of the trap. The calculation results show that the thermal activation barrier E_B of Trap A is substantially smaller than Trap B, as clearly indicated in Fig. 2.7.

2.4 Further Considerations

A. On the Definition of Coulomb Energy in Trap Case

Good reproduction of the measured time constants over gate voltage range, such as those in Fig. 2.1 and 2.6, is essential and crucial in the areas of *MOSFET RTS*. This means that the Coulomb energy involved must quantitatively follow that in Fig. 2.4. The corresponding Coulomb energy lies between 120 meV and 280 meV, comparable with that (250 meV) in the similar RTS measurements by Schulz [1].

As a single electron is inserted into the oxide trap, the total energy of the *MOS* system will change. The change in energy of the system can be divided into two parts: one is the storage energy and the other is the work done by the voltage source. The change in the storage energy term is

$$\Delta E_S = \frac{q^2}{2 \times A_{DB} \times C_{eff}} \tag{6}$$

ΔE_S was calculated to have a value of around 1 meV for the Debye screening length of 70 nm, which is negligibly small in magnitude. This means that the Coulomb energy in terms of the work ($\approx Q_G V_G$) done by external voltage source dominates. Therefore, the definition of $\Delta E \approx Q_G V_G$ as adopted in the areas of *MOSFET RTS* [1,2] is valid.

B. On the Nanocrystals Case

There are several fundamental differences between a MOS system with a nanometer-scale trap in the oxide and a MOS system with a nanocrystal or dot in the oxide. Firstly, the self capacitance of a nanocrystal dot in the oxide can be well linked

to the actual dot diameter (this promises applications as a nanoscale floating gate) whereas from the *MOS* electrostatics point of view, it is the Debye screening length prevailing in the trap case. Secondly, in our *RTS* measurement the gate voltage was fixed such as to ensure a quasi-equilibrium *MOS* system; and different gate voltages under such quasi-equilibrium conditions produced different RTS data. However, during typical Coulomb Blockade experiments on nanocrystalline memories, the gate voltage must continuously change in order to produce a series of Coulomb Blockade events. Thirdly, once captured, the electrons essentially remain in the dots (unless a potential leakage is present or the retention time is exceeded); however, this is not the case for the oxide trap, as evidenced by the fluctuating occupancy.

The experimentally determined Coulomb energy in the nanocrystalline dots memories [3-6] ranged from 46 meV to 168 meV. However, the definition of the Coulomb energy is significantly different from that in Ref. [1] and [2]. Instead, an alternative treatment on the basis of the Coulomb Blockade theory was widely adopted in the areas of nanocrystalline dots memories. For example, the product of the gate voltage shift between two subsequent Coulomb Blockade events and the gate-to-dot coupling coefficient can be directly connected to the critical energy required to overcome the barrier due to the single electron storage energy and the quantum confinement induced energy separation. The single electron storage energy is defined the Coulomb energy $\Delta E \approx q^2/2C_{dot}$ where C_{dot} is the self capacitance of the dot. Obviously, different situations encountered can lead to different definitions on the Coulomb energy.

2.5 Critical Issues

A. Screening Length

Due to the usage of a heavily-doped n^+ polysilicon gate, one may consider the

Thomas-Fermi screening length instead as employed in the metal case. However, a self-consistent Schrödinger-Poisson solving over the range of gate voltage under study reveals a band bending across a poly depletion region near the oxide. The corresponding electron density at the interface is found to be about one order of magnitude less than the immobile positively charged impurity concentration. Hence, in the presence of the poly depletion in our work, the Debye-Hueckel screening length considerably applies, which should be much larger than the Thomas-Fermi screening length (of the order of 1 nm) in the metal gate case. To further support this argument, from the measured RTS relative amplitude at Vg = 0.2 V, we estimate the amount of the affected area to be *at least* 28 nm and 35 nm across the charged trap for z_T = 0.7 nm and z_T = 0 nm, respectively. Thus, the cited 70 nm for the Debye screening length remains reasonable. Even the replacement with a lower value of 28 nm or 35 nm causes little error.

B. Silicon Depletion Charge

The Coulomb energy also includes the contribution by the charge induced at the edge of the semiconductor depletion region. The corresponding amount of energy is the product of the induced charge at the edge of the semiconductor depletion region times the difference (~ 0.07 eV) between Fermi level and valence band edge at the bulk part of the substrate. The depletion image charge at Vg = 0.2 V is found to be 0.1e and 0.07e, respectively, for z_T = 0 and 0.7 nm, and each decreases with increasing gate voltage. As a result, the Coulomb energy due to the depletion image charge becomes of the order of a few meV and drops with increasing gate voltage. Obviously, the role of the charge induced at the edge of the semiconductor depletion region is so insignificant that the depletion image charge can be neglected in the present work.

C. Electron Tunneling

First of all, a deeper oxide trap may not always dictate a longer time. According to the configuration coordinate diagrams that describe the electron-lattice coupling, our data point to the opposite case: a deeper oxide trap produces a smaller time constant. This is reasonable since all the extracted parameters can find their physical origins as detailed above. If the electron tunneling were involved only, then the capture time would be the sum of the tunneling time from the channel conduction band edge to certain oxide depth z_T plus the subsequent multiphonon emission time such as to lower the energy of the tunneling electrons to the same level as the trap. One can estimate the tunneling time of around 10^{-9} sec across z_T of 0.7 nm [16] and can reasonably hypothesize that the multiphonon emission time is a spontaneous event (as can be easily understood from the configuration coordinate diagrams in Fig. 2.6; the hypothesis also works well for the areas of the trap assisted tunneling), leading to a capture time of the order of 10^{-9} sec. Obviously, the possibility of the electron tunneling must in principle be removed since the measured capture times fall within 0.5 to 6 sec. On the other hand, once trapped the electrons may instantly tunnel to the gate electrode, contributing to the gate current. In other words, under such situations, no RTS in drain or gate current can be detected due to the extremely slow detection process in measurement setup. Moreover, in our work the gate current was found to be several orders of magnitude less than the drain current, indicating the absence of the electron tunneling in determining the experimental RTS drain current.

Note that the high and low levels of RTS current represent the different stable states as denoted the free and bound state in the configuration coordinate diagrams in Fig. 2.6. The detailed balance essentially applies only to such two states, rather than the abrupt transitions between the two. The capture and emission time constants represent the critical times required to overcome the barrier height and reach the

crossing point, then instantly entering into the other stable state.

Eventually, the measured discrete switching RTS drain current indicates that the transit time between the high and low level is substantially less than the integration time in measurement setup [10]. In other words, the abrupt transition between two stable states represents a spontaneous event with respect to the measurement setup. Hence, the corresponding transient displacement current through the gate electrode may escape detection. This explains why we saw only a flat gate current level (with typical thermal or shot fluctuations around it) over the whole observation time.

2.6 Conclusion

We have presented experimental evidence concerning the Coulomb energy enhancement in a MOS system with a nanometer-scale oxide trap. Other corroborating evidence based on a multiphonon theory has elucidated the measured capture and emission kinetics. The corresponding configuration coordinate diagrams have been established. We have further elaborated on the clarification of the Coulomb energy and have differentiated it from that in memories containing nanocrystals as a floating gate. Some critical issues encountered in the work have been addressed as well.

References

[1] M. Schulz, “Coulomb energy of traps in semiconductor space-charge regions,” *J. Appl. Phys.*, vol. **74**, pp. 2649-2657, 1993.

[2] H. H. Mueller, D. Wörle, and M. Schulz, “Evaluation of the Coulomb energy for single-electron interface trapping in sub-μm metal-oxide-semiconductor field-effect transistors,” *J. Appl. Phys.*, vol. **75**, pp. 2970-2979, 1994.

[3] S. Tiwari, F. Rana, H. Hanafi, A. Hartstein, E. F. Crabbe, and K. Chan, “A silicon nanocrystals based memory,” *Appl. Phys. Lett.*, vol. **68**, pp. 1377-1379, 1996.

[4] L. Guo, E. Leobandung, and S. Y. Chou, “A Silicon Single-Electron Transistor Memory Operating at Room Temperature,” *Science*, vol. 275, pp. 649-651, 1997.

[5] M. Saitoh, N. Takahashi, H. Ishikuro, and T. Hiramoto, “Large Electron Addition Energy above 250 meV in a Silicon Quantum Dot in a Single-Electron Transistor,” *Jpn. J. Appl. Phys.*, vol. **40**, pp. 2010-2012, 2001.

[6] S. Huang, S. Banerjee, R. T. Tung, and S. Oda, “Electron trapping, storing, and emission in nanocrystalline Si dots by capacitance–voltage and conductance–voltage measurements,” *J. Appl. Phys.*, vol. **93**, pp. 576-581, 2003.

[7] K. S. Ralls, W. J. Skocpol, L. D. Jackel, R. E. Howard, L. A. Fetter, R. W. Epworth and D. M. Tennant, “Discrete Resistance Switching in Submicrometer Silicon Inversion Layers: Individual Interface Traps and Low-Frequency (1/*f*?) Noise,” *Phys. Rev. Lett.*, vol. **52**, pp. 228-231, 1984.

[8] K. R. Farmer, C. T. Rogers, and R. A. Buhrman, “Localized-State Interactions in Metal-Oxide-Semiconductor Tunnel Diodes,” *Phys. Rev. Lett.*, vol. **58**, pp. 2255-2258, 1987.

[9] M. J. Kirton and M. J. Uren, “Noise in solid-state microstructures: a new

perspective on individual defects, interface states and low-frequency (1/f) noise," *Adv. Phys.*, vol. **38**, pp. 367-468, 1989.

[10] M. J. Chen and M. P. Lu, "On–off switching of edge direct tunneling currents in metal-oxide-semiconductor field-effect transistors," *Appl. Phys. Lett.*, vol. **81**, pp. 3488-3490, 2002.

[11] A. Avellan, W. Krautschneider, and S. Schwantes, "Observation and modeling of random telegraph signals in the gate and drain currents of tunneling metal–oxide–semiconductor field-effect transistors," *Appl. Phys. Lett.*, vol. **78**, pp. 2790-1792, 2001.

[12] D. V. Lang, *Deep Centers in Semiconductors*, edited by S. T. Pantelides (Gordon and Breach, Yverdon, 1992).

[13] B. K. Ridley, *Quantum Processes in Semiconductors*, 3rd ed.(Clarendon, Oxford, 1993).

[14] A. Palma, A. Godoy, J. A. Jiménez-Tejada, J. E. Carceller and J. A. López-Villanueva, "Quantum two-dimensional calculation of time constants of random telegraph signals in metal-oxide–semiconductor structures," *Phys. Rev. B*, vol. **56**, pp. 9565-9574, 1997.

[15] W. B. Fowler, J. K. Rudra, M. E. Zvanut and F. J. Feigl, "Hysteresis and Franck-Condon relaxation in insulator-semiconductor tunneling," *Phys. Rev. B*, vol. **41**, pp. 8313-8317, 1990.

[16] I. Lundström and C. Svensson, "Tunneling to traps in insulators," *J. Appl. Phys.*, vol. **43**, pp. 5045-5047, 1972.

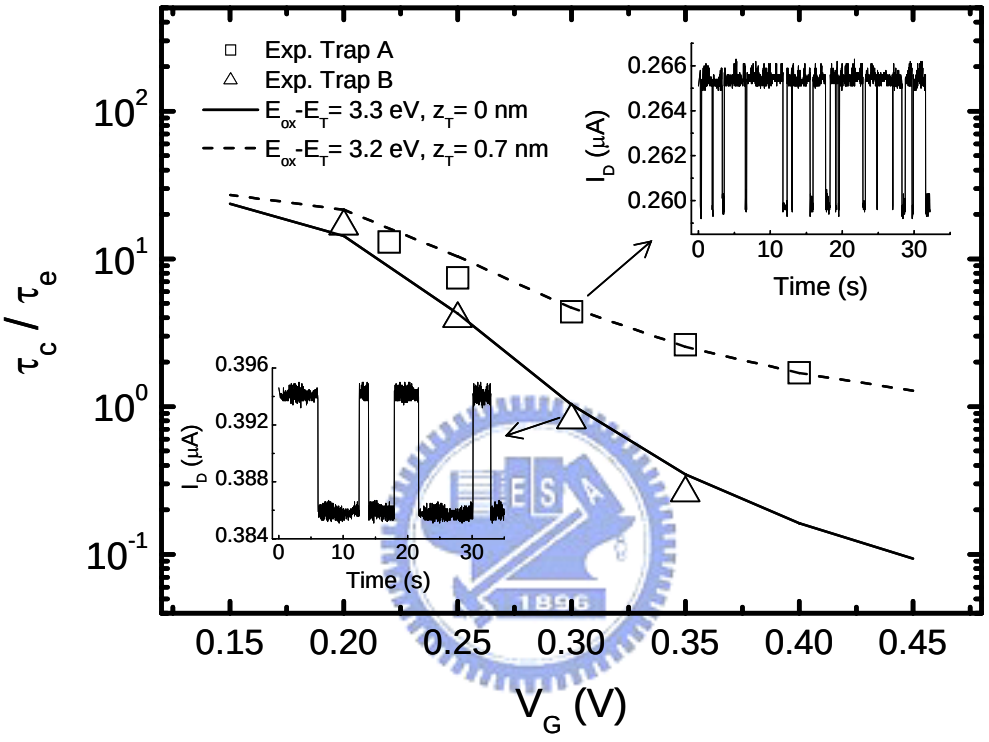

Fig. 2.1 Measured mean capture time to mean emission time ratio versus gate voltage for two devices labeled Trap A and Trap B. The inset shows the time records of RTS drain current at a fixed gate voltage of 0.3 V. The fitting lines from (2) are also shown.

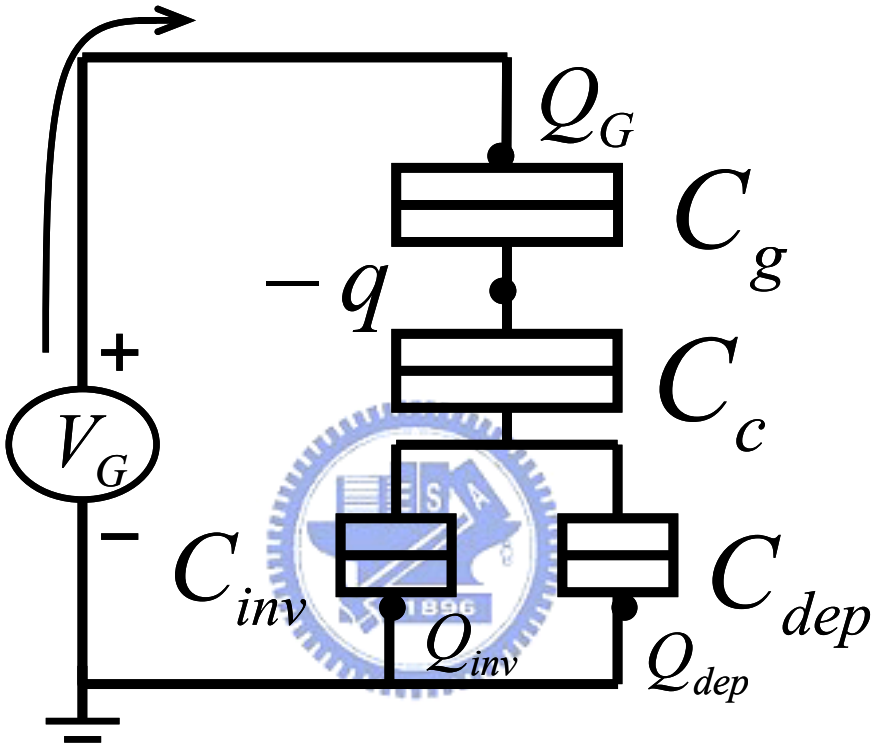

Fig. 2.2 Capacitive coupling equivalent circuit, accounting for the effect of the trap depth and the charge sharing between gate, inversion layer, and silicon depletion region.

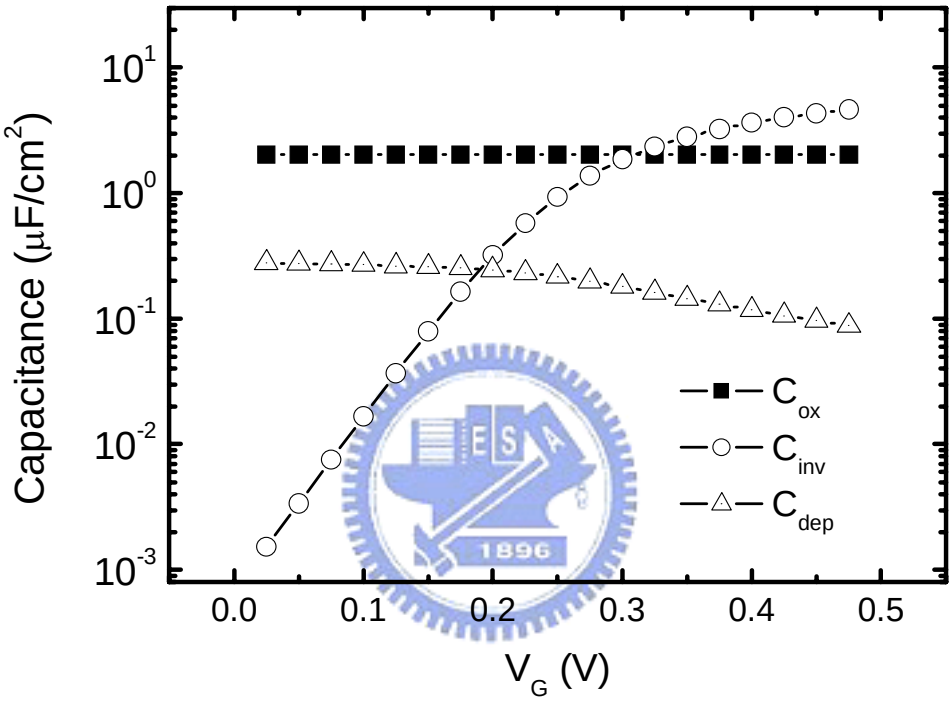

Fig.. 2.3 Simulated results of the key capacitance components versus gate voltage.

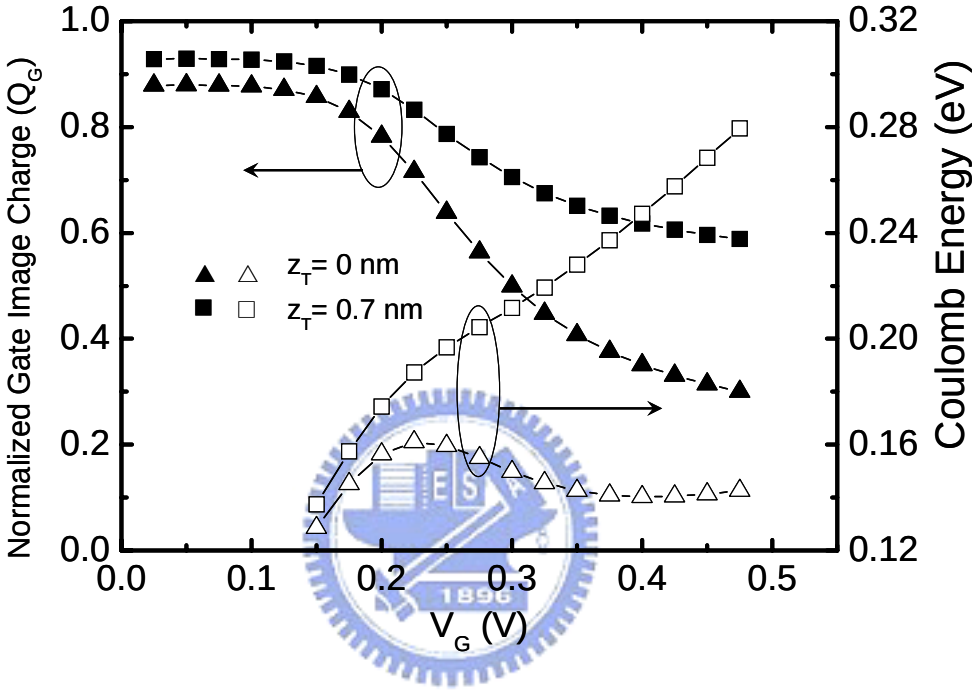

Fig. 2.4 Calculated gate image charge and Coulomb energy versus gate voltage for two trap depths in the oxide.

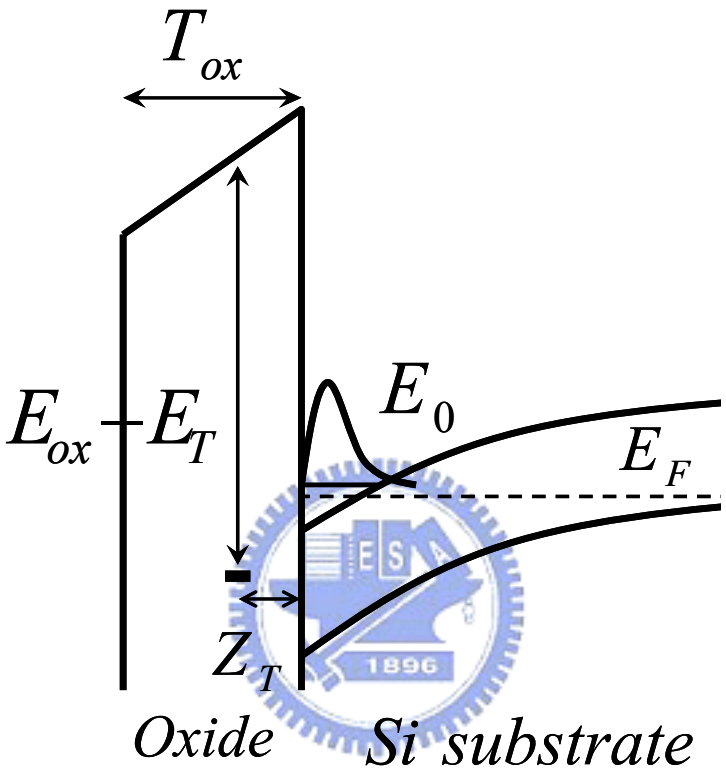

Fig. 2.5 The MOS energy band diagram schematically indicating the Fermi level, the trap energy level, and the trap depth.

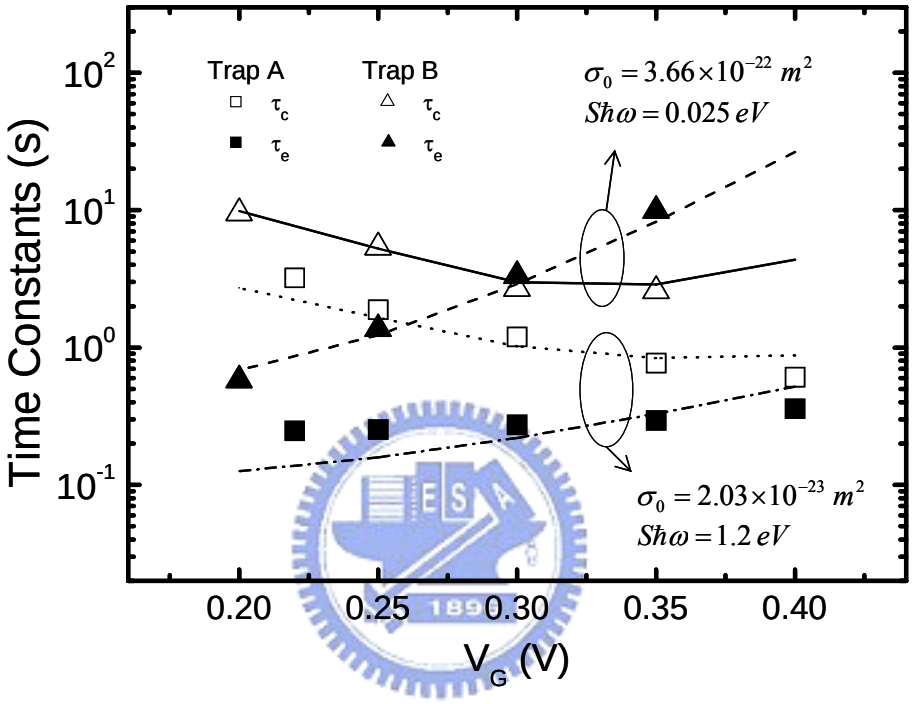

Fig. 2.6 Comparison of the measured and calculated capture time constants and emission time constants versus gate voltage.

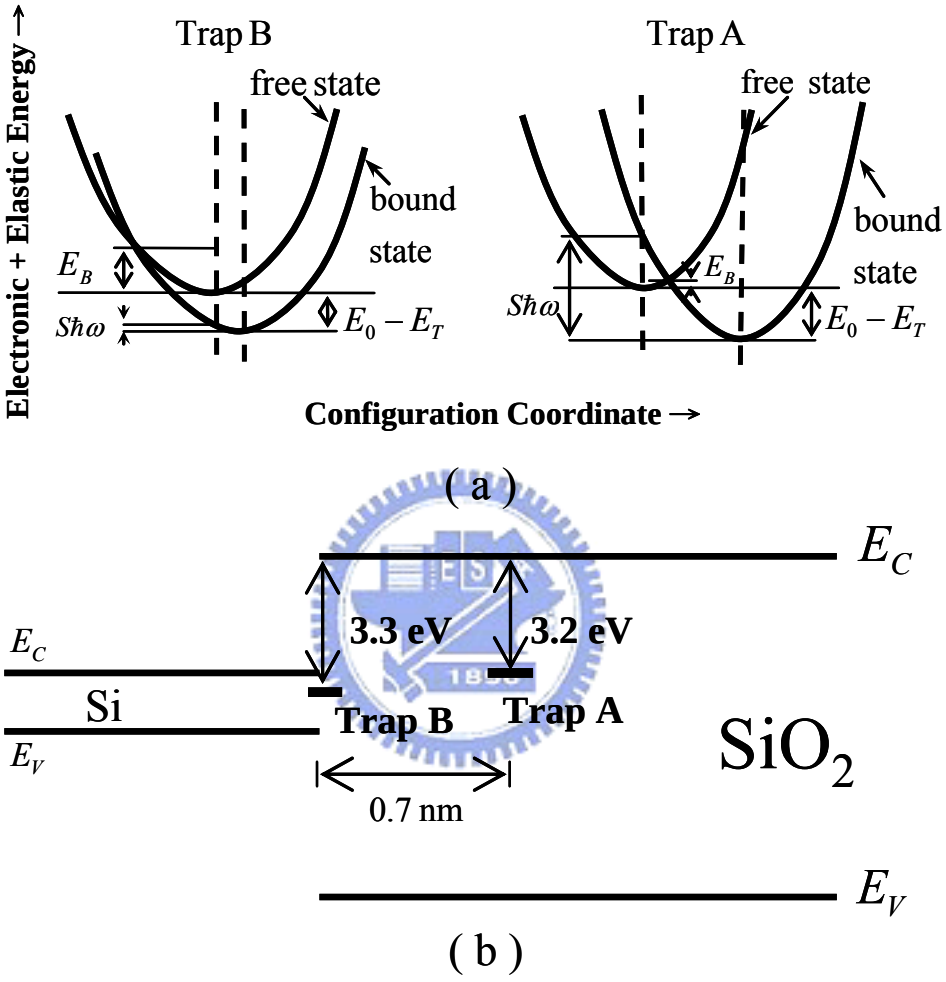

Fig. 2.7 Schematic configuration coordinate diagrams used for a phenomenological description of the capture and emission kinetics encountered in Trap A and Trap B. The corresponding energy band diagrams in flatband conditions are also given, schematically showing the trap depth and its energetic level in the oxide.

Chapter 3

Near-distance Effect of a Charged Oxide Trap on Coulomb Scattering in a Two-dimensional System

In this chapter, experiments on an individual nanoscale trap in the oxide responsible for random telegraph signals lead to remarkable results. Initially, the deeper trap in oxide corresponds to weaker Coulomb scattering in a two-dimensional electron gas (2DEG). However, as the 2DEG enters into the strong inversion regime, the amount of Coulomb scattering with an interface trap drops with a faster rate than the deep trap. A stronger screening potential confinement is shown to be the physical origin of this effect. The near-distance effect is expected to remain a challenging issue in the area of nanoscale devices.

3.1 Introduction

A two-dimensional electron (or hole) gas beneath the gate oxide of metal-oxide-semiconductor field-effect transistors (MOSFETs) can determine the device's performance. In this two-dimensional system, carrier scattering via the Coulomb force plays a vital role. Over the past few decades, Coulomb scattering caused by charged centers, such as ionized impurities in the underlying silicon [1-3], charged traps in a nearby oxide or dielectric layer [4], and ionized impurities in remote depleted poly-silicon [5], have been extensively investigated. However, the

charged centers involved [1-5] dealt with *a number of* charged centers, rather than an individual nanoscale charged center. Thus, if a study of Coulomb scattering is done with an individual nanoscale charged center, then it can be helpful in the area of nanoscale-sized devices. The measurement of random telegraph signals (RTS) is a good means by which to achieve this goal since it can provide the opportunity to communicate directly with an individual charged center. So far, there have been a large number of literature articles devoted to the area of MOSFETs RTS [6-9]; however, not all the data were addressed in a systematical way (*consistently related between* time constants and relative magnitude, for example), nor was a *rigorous* analysis done (that is, the role of *Coulomb energy* in capture kinetics [10] was lacking). Coulomb energy enhancements in the presence of a charged oxide trap situated away from the SiO_2/Si interface have recently been demonstrated based on the measured RTS time constants [11]. In this letter, the RTS time constants and relative magnitude are both closely linked. The result is remarkable in terms of the near-distance effect: Coulomb scattering in the two-dimensional system is a strong function of the trap depth from the SiO_2/Si interface.

3.2 Experimental Procedure

The devices under study were the same as those detailed elsewhere [11]: 1.7-nm gate oxide n-channel MOSFETs with varying channel lengths and widths (60 nm to 600 nm). The measured RTS events in the source and drain current under a 10-mV drain voltage indicate that an individual nanoscale trap naturally (it is unlikely that it is caused by electrical stressing during the RTS measurement) exists in the oxide or at the interface between silicon and oxide. Two distinct trap depths, one of 0 and one of 0.7 nm, were rigorously determined by accounting for the RTS emission and capture time data with the Coulomb energy enhancements included (see Ref.[11,12]

for details). The discrete current fluctuations in time domain along with the definitions for I_D and ΔI_D are displayed in Fig. 3.1. The high-current level and low-current level represent the empty trap state and filled trap state, respectively. The corresponding current fluctuations in terms of the relative magnitude $\Delta I_D/I_D$ (normalized with respect to 1 μm^2 for fair comparison) versus gate voltage for both traps is shown in Fig. 3.2. It can be seen that (i) the $\Delta I_D/I_D$ decreases with increasing gate voltage; (ii) the interface trap (trap depth z_T = 0 nm) produces a drop in $\Delta I_D/I_D$ with a faster rate than the oxide trap away from the interface (z_T = 0.7 nm); and (iii) the $\Delta I_D/I_D$ is higher for the case of the interface trap until a crossover occurs. It is well recognized that once a single electron is captured by and released from the oxide trap, the conductance of the device fluctuates with two effects involved: one is the carrier number fluctuations and the other is referred to as the mobility fluctuations (via Coulomb scattering in the case of the charged oxide trap as studied here). The current fluctuations can hence be written as [13]

$$\frac{\Delta I_D}{I_D} = \left(\frac{\Delta I_D}{I_D}\right)_{number} + \left(\frac{\Delta I_D}{I_D}\right)_{coulomb} \tag{1}$$

The number fluctuation term, occurring when a single electron is captured by the oxide trap, is a function of the image (polarized) charge Q_{inv} (in a normalization form) induced in the two-dimensional electron gas (2DEG), the 2DEG carrier density N_s per unit area, and the channel area A (= 1 μm^2 as explained above):

$$\left(\frac{\Delta I_D}{I_D}\right)_{number} = \frac{Q_{inv}}{AN_s} \tag{2}$$

Until now it has been difficult in the open literature to distinguish between the contributions from the number fluctuations and from the Coulomb scattering ones. This problem can be considerably overcome by means of a capacitive coupling model [11]. The resulting image charge against gate voltage is given in Fig. 3.3 with the oxide trap depth as a parameter. The capacitance model utilized is inserted in Fig. 3.3. Obviously, moving a trap toward the underlying 2DEG can induce more image charge on 2DEG layer. The magnitude of the number fluctuations in the devices used was then achieved as shown in Fig. 3.4. The experimental $\Delta I_D/I_D$ was straightforwardly decoupled according to Eq.(1), leading to the Coulomb scattering counterpart as plotted together in Fig. 3.4. Fig. 3.4 clearly reveals that the Coulomb scattering dominates in most of the gate voltage range, and thereby is primarily responsible for the measured $\Delta I_D/I_D$.

The Coulomb scattering fluctuations originate from the mobility modulation caused by the charged oxide trap. To facilitate the procedure, a Coulomb scattering coefficient (α) is introduced based on Matthiessen's rule:

$$\frac{\mathbf{1}}{\mu} = \frac{\mathbf{1}}{\mu_n} + \frac{\mathbf{1}}{\mu_{ox}} = \frac{\mathbf{1}}{\mu_n} \pm \alpha N_{ox} \tag{3}$$

where μ is the total mobility, μ_{ox} is the mobility due to the charged oxide trap, μ_n is the mobility due to the other mechanisms, and N_{ox} is the number of charged oxide trap per unit area. The positive/negative sign depends on the state of the trap when it is filled (the positive sign applies in our case). Consequently, the Coulomb scattering fluctuations term can be expressed by [13]

$$\left(\frac{\Delta I_D}{I_D}\right)_{coulomb} = \frac{\alpha\mu}{A} \tag{4}$$

By substituting the measured total mobility and extracted Coulomb scattering fluctuations in Fig. 3.4 into Eq.(4), one obtains the values of α as depicted in Fig. 3.5 against the 2DEG carrier density N_s. The resultant α for both traps exhibits a weak dependence on N_s in the low carrier density regime while being a decreasing function of N_s in the high carrier density regime due to screening effect, in agreement with the literature [13]. Further observation reveals that a deeper oxide trap corresponds to a lower Coulomb scattering in weak inversion, as expected; however, as the channel conductivity increases further, the amount of Coulomb scattering with an interface trap drops with a faster rate. Indeed, Fig. 3.5 corroborates the existence of the near-distance effect: moving a trap toward the underlying strongly inverted 2DEG layer can dramatically reduce the carrier Coulomb scattering. The near-distance effect can also critically give rise to a crossover point as observed in Fig. 3.5. The same argument readily applies to Fig. 3.2 and 3.4.

3.3 Analysis and Physical Interpretations

A plausible explanation of the above near-distance effect is that the Coulomb potential associated with an interface trap is confined in a faster rate, relative to the deep trap case, due to a stronger screening effect in the high carrier density regime. Corroborating evidence can be obtained from the wavefunction penetration calculation; that is, for a finite barrier height of the conduction-band discontinuity at the Si/SiO_2 interface, the wavefunction will penetrate into the oxide layer with a critical length called the penetration depth. The penetration depth, designated λ, by

definition is

$$\lambda = \sqrt{\frac{\hbar^2}{2m^* \Delta E_C}} . \tag{5}$$

where m^* is the effective mass and ΔE_C is the barrier height of the conduction-band discontinuity. The calculated penetration depth was around 0.1 nm in our work. The trap center at the interface (z_T = 0 nm) therefore is surrounded by the mobile carriers with a local density in the presence of the wavefunction penetration. However, this is not the case for a deeper trap (z_T = 0.7 nm). As a result of the different wavefunction penetrations encountered, the amount of Coulomb scattering for z_T = 0 nm drops faster than that for z_T = 0.7 nm in the high mobile carrier density region. In other words, the screening potential due to a single charged trap at z_T = 0 nm is much more confined than z_T = 0.7 nm in the strong inversion region.

Other corroborating evidence can be found by means of a widely accepted theory of Coulomb scattering. According to the cylindrical geometry of the underlying system, the Coulomb potential around a single charged oxide trap can be expanded in a Fourier-Bessel form: $\phi(r,z) = \int_0^\infty J_0(kr) A_k(k,z) k\, dk$, where $J_0(kr)$ is the zeroth order Bessel function, r is the coordinate parallel to the SiO_2/Si interface, and z is the coordinate perpendicular to it. The two-subband Poisson equation with a single charged center reduces to the following form [1,2,14]:

$$\left(\frac{\partial}{\partial z}\varepsilon\frac{\partial}{\partial z} + \varepsilon\frac{\partial^2}{\partial z^2} - \varepsilon(z)k^2\right) A_k(k,z) - 2\sum_{i=0,0'} \varepsilon_{si} S_i g_i(z) \int_0^\infty A_k(k,z) g_i(z)\, dz = -\rho_{ext}(k, z_T) \tag{6}$$

where $i = 0$ means the lowest subband of a 2-fold degenerate valley and $i = 0'$ is the lowest subband of 4-fold degenerate valley, $g_i(z)$ is the inversion charge distribution for subband i, z_T is the position of the Coulomb scattering center (charged center), ε(z) is the position-dependent dielectric permittivity, and ε_{si} is the permittivity of silicon. S_i in Eq.(5) is the two-dimensional screening constant, defined by [1, 2, 14]

$$S_i = \frac{q^2}{2\varepsilon_{si}} \frac{\partial N_i}{\partial E_f} \tag{7}$$

where N_i is the inversion carrier density for subband i and E_f is the Fermi energy level. The electron-electron interaction can be neglected in our work because of the non-degenerate system (corresponding to applied gate voltages less than 0.5 V). However, it should be kept in mind that a hypothesis behind Eq. (7), termed the weak potential approximation, exists; that is, the scattering center must be located far away from the wavefunction range of the two-dimensional electron gas, where the Coulomb potential slowly varies. Under such situations, the carrier distribution function $g_i(z)$ does not have a significant interaction with the Coulomb potential of the charged center. Consequently, Eq.(7) itself does not account for the aforementioned near-distance effect, as will be explained slightly later. The approximated boundary conditions [1,2] for solving of Eq. (6) were cited, followed by a calculation of the differential cross section based on the Born approximation for a Coulomb scattering center located at z_T:

$$\sigma_i(\theta) = \frac{2\pi m_d{}^i e^2}{\hbar^3 \upsilon} \left| \int_0^\infty A_k(k,z) g_i(z)\, dz \right|^2 \tag{8}$$

where m_d^i is the density of states mass for subband i ($m_d^0 = 0.19m_0$ and $m_d^{0\prime} = 0.417m_0$), θ is the scattering angle, and υ is the carrier velocity. Eq. (6) is the momentum-relaxation-time model constructed by Stern [1] on the basis of the elastic, intrasubband scattering process. A correction factor is therefore needed to account for the other processes, such as inelastic Coulomb scattering and the intersubband scattering. Use of the boundary conditions also affects the correction factor [2]. The result is displayed in Fig. 3.5. The corresponding correction factor was adjusted to a certain value (relatively not large), producing a good match in the weak inversion region for both traps. Another correction factor with a comparable value was also utilized to reproduce the measured total mobility. It can be seen from Fig. 3.5 that a deviation appears in the large N_s regime for both traps. However, relative to interface trap, the discrepancy for the deeper trap is not so large. Thus, it is reasonably argued that the weak potential approximation remains valid for the charged trap having a large depth. As a result, the Coulomb scattering theory mentioned above can describe the screening effect in the deep trap case adequately. As to the trap at the interface, the discrepancy substantially worsens with increasing N_s. This reflects that a stronger screening potential confinement is encountered. In other words, the weak potential approximation essentially fails in the case of z_T = 0 nm in the strong inversion conditions. Once again, we confirm that the charged center at the interface can cause a strong interaction of Coulomb scattering with the mobile electron carriers in the channel and, as a result, the underlying carrier distribution function $g_i(z)$ will change accordingly. The origin of the discrepancy in Fig. 3.5 for z_T = 0 nm, therefore, is that the two-dimensional screening constant in Eq. (6) applies only to the weak potential case, rather than the strong potential one which is actually encountered.

3.4 Conclusion

We believe that the above mentioned near-distance Coulomb scattering effect is not going to be absent in the area of nanoscale devices in the presence of one or a few traps. First, in such nanoscale active dimensions, different depths of charged traps may produce different Coulomb scatterings in the two-dimensional system, which, in turn, give rise to different mobilities. Second, the RTS due to the trapping-detrapping process can be substantially enhanced while scaling devices down toward the nanometer regime. Different energetic levels and different spatial depths of the underlying traps in the dielectric layer can produce different RTS characteristics through different capture time constants, different emission time constants, and different relative magnitudes. Finally, it is well recognized that a single RTS in the time domain can have the frequency equivalent of a Lorentzian noise. If there are one or a few traps, then the superposition of the corresponding Lorentzian type components lead to low-frequency or flicker noise. Once again, according to our work, the different energetic levels and different spatial depths of the underlying traps in dielectric layer can together produce different noise spectra.

The undertaken study makes a great contribution to the areas of MOSFETs RTS. In addition, a near-distance effect is experimentally confirmed, which is that Coulomb scattering in a two-dimensional system is sensitive to the depth of a nanoscale oxide trap. Specifically, in strong inversion conditions, the amount of Coulomb scattering with an interface trap drops with a faster rate than the deep trap, indicating a stronger screening potential confinement. The near-distance effect is expected to prevail in next-generation nanoscale devices involving the mobility, the RTS, and the low-frequency noise.

References

[1] F. Stern, “Two-subband screening and transport in (001)silicon inversion layers,” *Surf. Sci.*, vol. **73**, pp. 197-206, 1978.

[2] T. Ando, A. B. Fowler and F. Stern, “Electronic properties of two-dimensional systems,” *Rev. Mod. Phys.*, vol. **54**, pp. 437-672, 1982.

[3] M. V. Fischetti and S. E. Laux, “Monte Carlo study of electron transport in silicon inversion layers,” *Phys. Rev. B*, vol. **48**, pp. 2244-2274, 1993.

[4] F. Gámiz, J. A. López-Villanueva, J. A. Jiménez-Tejada, I. Melchor and A. Palma, “A comprehensive model for Coulomb scattering in inversion layers,” *J. Appl. Phys.*, vol. **75**, pp. 924-934, 1994.

[5] F. Gámiz and M. V. Fischetti, “Remote Coulomb scattering in metal–oxide–semiconductor field effect transistors: Screening by electrons in the gate,” *Appl. Phys. Lett.*, vol. **83**, pp. 4848-4850, 2003.

[6] K. S. Ralls, W. J. Skocpol, L. D. Jackel, R. E. Howard, L. A. Fetter, R. W. Epworth and D. M. Tennant, “Discrete Resistance Switching in Submicrometer Silicon Inversion Layers: Individual Interface Traps and Low-Frequency (1/*f*?) Noise,” *Phys. Rev. Lett.*, vol. **52**, pp. 228-231, 1984.

[7] K. R. Farmer, C. T. Rogers, and R. A. Buhrman, “Localized-State Interactions in Metal-Oxide-Semiconductor Tunnel Diodes,” *Phys. Rev. Lett.*, vol. **58**, pp. 2255-2258, 1987.

[8] M. J. Kirton and M. J. Uren, “Noise in solid-state microstructures: a new perspective on individual defects, interface states and low-frequency (1/f) noise,” *Adv. Phys.*, vol. **38**, pp. 367-468, 1989.

[9] M. J. Chen and M. P. Lu, “On–off switching of edge direct tunneling currents in

metal-oxide-semiconductor field-effect transistors," *Appl. Phys. Lett.*, vol. **81**, pp. 3488-3490, 2002.

[10] M. Schulz, "Coulomb energy of traps in semiconductor space-charge regions," *J. Appl. Phys.*, vol. **74**, pp. 2649-2657, 1993.

[11] M. P. Lu and M. J. Chen, "Oxide-trap-enhanced Coulomb energy in a metal-oxide-semiconductor system," *Phys. Rev. B*, vol. **72**, p. 235417, 2005.

[12] Essentially, Coulomb energy and Coulomb scattering are different each other. Coulomb energy is the work done by system against the induced charge while inserting a single electron into the trap. Coulomb energy can be related to RTS capture time and emission time. On the other hand, Coulomb scattering involves the mobile carrier transport in the channel while suffering the collision events due to the perturbation caused by Coulomb potential. Coulomb scattering can be related to RTS relative magnitude.

[13] K. K. Hung, P. K. Ko, C. Hu and Y. C. Cheng, "Random telegraph noise of deep-submicrometer MOSFETs," *IEEE Electron Device Lett.*, vol. **11**, pp. 90-92, 1990.

[14] There are at least 95% of the inversion-layer carriers occupying the lowest subband of the 2-fold and 4-fold degenerate valley for the gate voltages demonstrated (2 V to 4.5 V). The two-subband elastic momentum-relaxation-time model is therefore suitable to describe the Coulomb scattering process in our case.

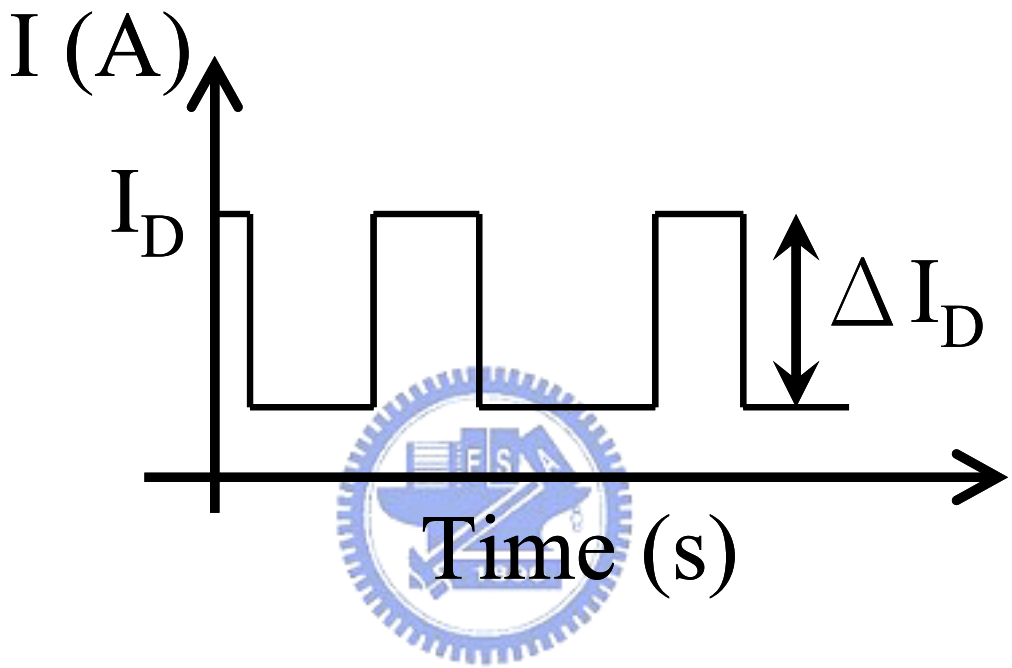

Fig. 3.1 The discrete current fluctuations in time domain along with the definitions for I_D and ΔI_D. The high-current level and low-current level represent the empty trap state and filled trap state, respectively.

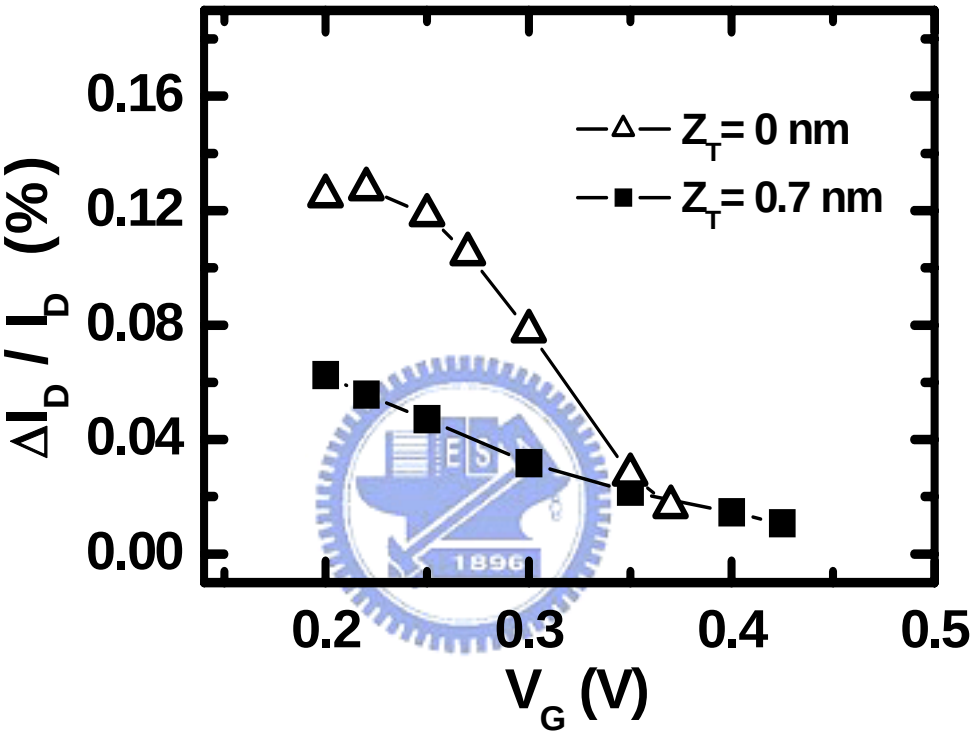

Fig. 3.2 The normalized RTS relative amplitude measured versus gate voltage for two different trap depths.

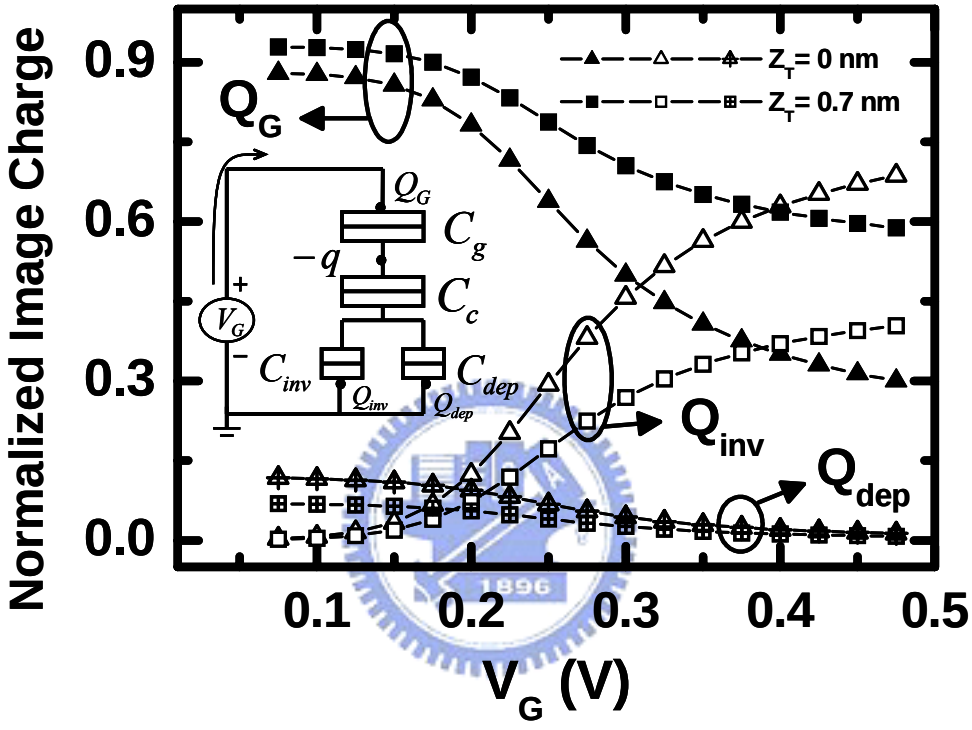

Fig. 3.3 Calculated image (polarized) charge on the gate (Q_G), the inversion layer (Q_{inv}), and the depletion region (Q_{dep}) versus gate voltage for different trap depths. The inset shows corresponding capacitance equivalent circuit [11], taking into account the trap depth and the charge sharing between the gate, inversion layer and silicon depletion region.

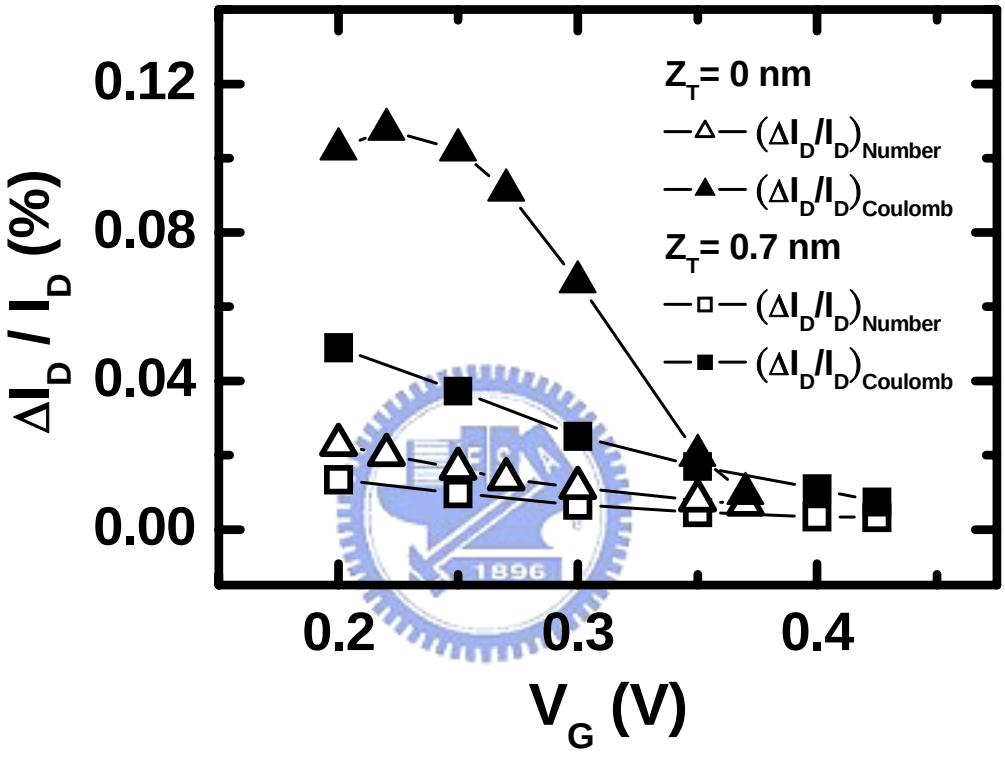

Fig. 3.4 Separated relative current magnitude due to number fluctuations and relative current magnitude due to mobility fluctuations, plotted against gate voltage.

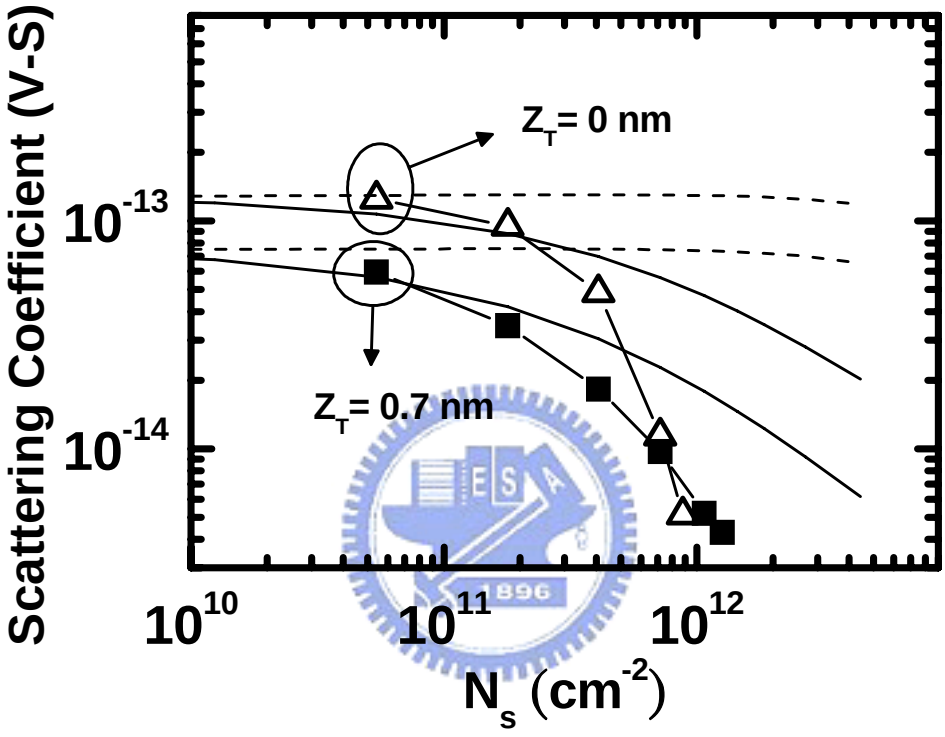

Fig. 3.5 Experimental (symbols) and calculated (lines) Coulomb scattering coefficients versus inversion-layer carrier density for two different trap depths. The calculated results with and without screening effect are plotted in terms of the solid lines and dash lines, respectively.

Chapter 4

Observation of RTS in Direct Tunneling Regime for Local Wire-like Spot with High-k Dielectric

The statistical switching of the electrical resistance of a device between two discrete values is termed “random telegraph signals” (RTS) because the current through such a device resembles a telegraph signal. It is well recognized that in the case of MOS devices, characterization of RTS behaviors is very useful: (i) it can explore microscopic aspects of bulk/interface traps in gate stack dielectric; (ii) it can sensitively monitor potential defects in a manufacturing process; and (iii) it can estimate performance limits in nanodevices due to noise concern.

In this chapter, on-off switching behaviors or two-level random telegraph signals (RTS) are measured in the low voltage (-1.40 V < V_G < - 0.88 V) edge direct tunneling (EDT) currents in ultrathin gate stack (10 Å oxide + 10 Å nitride) n-channel metal-oxide-semiconductor field-effect transistors. The plausible origin is the process-induced defects in terms of localized gate stack thinning (or equivalently the conductive filament). In such extrinsic case, the current trapping-detrapping theories can adequately elucidate the data, particularly the RTS magnitude as large as 18%. The current-voltage (I-V) characteristic associated with a certain defective spot is assessed straightforwardly, showing remarkable compatibility with existing oxide thinning case. The currents tunnel through the wire-like weakened spot can be probed by RTS. The role as a sensitive process monitor is demonstrated in terms of the

occurrence probability of the defects as well as their locations.

4.1 Introduction

For ultrathin gate stack metal-oxide-semiconductor field-effect transistors, the edge direct tunneling of electrons or holes from polysilicon to underlying silicon diffusion region has very recently been investigated [1]-[3]. According to Quantum Mechanical simulation [2],[3], the EDT spans from the gate edge a range of about 6 nm wide, which exactly falls within the gate-to-diffusion overlap region. Thus, exploration of the noise or fluctuations via EDT currents is highly probable in the presence of such a nanoscale dimension. In this article we exhibit a two-level RTS phenomenon for the first time measured from the EDT currents at low voltages.

On-off switching behaviors or two-level random telegraph signals are measured in the low voltage (-1.40 V < V_G < - 0.88 V) edge direct tunneling currents in ultrathin gate stack (10 Å oxide + 10 Å nitride) n-channel metal-oxide-semiconductor field-effect transistors. The plausible origin is the process-induced defects in terms of localized gate stack thinning (or equivalently the conductive filament). In such extrinsic case, the current trapping-detrapping theories can adequately elucidate the data, particularly the RTS magnitude as large as 18%. The current-voltage (I-V) characteristic associated with a certain defective spot is assessed straightforwardly, showing remarkable compatibility with existing oxide thinning case. The role as a sensitive process monitor is demonstrated in terms of the occurrence probability of the defects as well as their locations.

4.2 Experimental Details

The n-channel MOSFETs with the gate width and length of 10 um and 0.5 um respectively were manufactured in a standard process technology. During the gate

stack formation, a 10 Å thick SiO_2 layer was thermally grown and underwent Rapid Plasma Nitridation (RPN) treatment, then on which a 10 Å thick nitride film was grown, followed by NO annealing at 950 ^{o}C. The gate stack's e*quivalent oxide thickness* (EOT) is about 15 Å, which was extracted from the capacitance–voltage (C – V) measurement using a quantum-mechanical (QM) scheme. To calculate the tunneling leakage current, one needs to know the physical thickness of both the nitride and the oxide layers. Here the bottom oxide thickness was determined directly from a split wafer, and thereby the nitride EOT thickness was obtained by subtracting the oxide thickness from the total EOT of the N/O stack. The physical thickness of the nitride layer can then be calculated as

$$t_{EOT} = (\epsilon_{ox} / \epsilon_{ni}) t_{ni} \quad (1)$$

where the nitride dielectric constant ϵ_{ni} was determined to be 6.9 from the single-layer nitride control set and the oxide dielectric constant ϵ_{ox} was taken as 3.9. The key process parameters as obtained by capacitance-voltage (*C-V*) fitting were n^{+} polysilicon doping concentration = 6×10^{19} cm^{-3}, and the equivalent oxide thickness (EOT) is about 16 Å that is similar to the ideal value of the EOT. The devices were characterized by means of a Semiconductor Parameter Analyzer HP4156B with source, drain, and bulk tied to ground. Experimental setup is shown in Fig. 4.1. The measured steady-state terminal currents versus negative gate voltage characteristics, as depicted in Fig. 4.2 for the measured device were found to be nearly the same for all samples. From Fig. 4.2, we drew an EDT region of interest throughout the work: -1.40 V < V_G < - 0.88 V. In such low voltage range, the tunneling mechanism occurs in gate edge overlap region. An EDT current I_S flowing from source to gate and another EDT current I_D from drain to gate both dominantly constitute the gate current

I_G. Fig. 4.2 further point out that the source and drain currents are comparable each other in the sense of steady-state characterization. To strengthen RTS measurement precision, the integration time in HP4156B for data sampling was changed from a default value of 1 ms down to equipment limit of 8 us.

4.3 Observation of RTS

The measured fluctuations in drain, gate, and source currents for the sample in Fig. 4.2 are displayed in Fig. 4.3, looking like those from a two-terminal tunnel gate stack with a percolation path [4-7].We usually consider the high current state as "on" and low current state as "off", and the mean time spent in "on" state as $\bar{\tau}_{on}$ while the mean time spent in "off" state as $\bar{\tau}_{off}$. Fig. 4.4 shows the definitions of measured time constants τ_{on} and τ_{off}. For two-level switching in n-MOSFETs, the gate voltage dependence of the gate tunneling current fluctuations can be utilized to identify which level represents the case when the trap is full and which level represents when the trap is empty. The abrupt transitions between two distinct states occur in gate and source currents whereas being absent in the drain current, suggesting that (i) significant defects exist in the gate-to-source overlap part rather than gate-to-drain overlap; and (ii) they should be localized such as to match above steady-state characterization. Fig. 4.3 also reveals identical RTS parameters between source and gate currents: the same time duration τ_{on} and τ_{off} in the high and low current state I_{on} and I_{off}, respectively, and the same state current change $\Delta I_{tunneling}$ (= $I_{on} - I_{off}$). This identity confirms consistently the existence of a two-terminal EDT path.

Fig. 4.5 shows experimental I_{on} and I_{off} of the source current against gate voltage. We also found that (i) the statistical distributions of τ_{on}'s and τ_{off}'s indeed obey an

exponential behavior with the mean $\bar{\tau}_{on}$ and $\bar{\tau}_{off}$, respectively; and (ii) $\bar{\tau}_{on}$ and $\bar{\tau}_{off}$ each exponentially decrease with increasing magnitude of gate voltage. Finally, repeated RTS measurements yielded the same RTS parameters, indicating that no stress induced defects are encountered, as expected due to low voltage operation.

4.4 Discussion

The above on-off switching phenomenon can be likely ascribed to the manufacturing process induced defects [8,9] as schematically shown in Fig. 4.6 in terms of localized gate stack thinning, or equivalently the conductive filament[9] defined by the thinning thickness Δt_{ox} and the occupied area ΔA (shaded region; shown only partly). Such defective spot can be thought to be a conductive filament connected in series with the remainder (with the net thickness of t_{ox} - Δt_{ox}) of the gate stack. The conductive filament is also a trap-rich region in nature as characterized by the trapping-detrapping probabilities [7] in terms of both the emission time τ_{off} and the capture time τ_{on} as mentioned above. According to current trapping-detrapping theories [7], as a single electron is captured in the trap-rich region, the caused Coulomb repulsive force repels subsequently tunneling electrons, thus effectively turning off the influenced area around the trapped electron. This area can be represented by a capture cross-section area [10] denoted πr^2 in Fig. 4.6. Consequently, the EDT current in the filled trap mode can be written as (analogous to Simoen et al.'s work [10])

$$I_{off} \approx J_n \times (A_{EDT} - \pi r^2) \quad (2)$$

where J_n is the tunneling current density associated with the gate stack thickness t_{ox}

and A_{EDT} is the EDT area. The region πr^2 can be recovered to on state in the empty trap mode. The high current state I_{on} can thereby be expressed as

$$I_{on} \approx J_n \times (A_{EDT} - \Delta A) + \Delta J_n \times \Delta A \quad (3)$$

where ΔJ_n is the tunneling current density for a net thickness of $t_{ox} - \Delta t_{ox}$, indicating that $\Delta J_n >> J_n$. for considerable Δt_{ox}. The mentioned "localization" implies that ΔA can be of comparable order with πr^2 [9]. As a result, (3) minus (2) yields $\Delta I_{tunneling} \approx \Delta J_n \times \Delta A$ quantitatively for the two-terminal I-V characteristic of a defective spot. In Fig. 4.5 the fractional change $\Delta I_{tunneling}/I_{off}$ turns out to be between 12% and 18%. Also plotted in Fig. 4.5 is a line from an empirical formula of direct tunneling for oxide thinning $\Delta I_{tunneling} = a\,|V_G|^b$ [6], showing not only agreements with experiment but also reasonable values of a and b compared with the citation [6]. Fig. 4.7 shows that the scheme of $\Delta I_{tunneling}$ tunneling through the local wire-like spot. The currents tunneling through the wire-like spot can therefore be probed by RTS.

Regarding the role as a sensitive process monitor, we carried out a very extensive RTS measurement across the wafer. It was turned out that about 36% of the samples exhibit two-level RTS only in EDT source currents; 12% only in EDT drain currents; and no noticeable RTS was observed for the rest as well as for both EDT source and drain currents. Therefore, it is claimed that characterization of RTS via EDT currents can furnish process-concerned information in terms of the occurrence probability of the defects as well as their locations.

5.5 Conclusion

A new on-off switching behavior is measured via EDT currents in the overlap part of the gate stack. We have shown that RTS can be a useful tool for studying

process-induced trap. It can find the trap location and trap type by analyzing RTS data. We also find that the process-induced trap is located near interface, or around the middle of the dielectric layer. With process defects as the plausible origin, experimental data can be adequately described by electron trapping-detrapping theories. Also highlighted straightforwardly are: assessing the two-terminal current-voltage (I-V) characteristic associated with a defective spot; finding the occurrence probability of the defects as well as their locations in a manufacturing process; etc. Oxide reliability is a major concern for deep-submicron technologies as the dielectric thickness and channel area scale down. We can thereby infer that the RTS is a promising tool to addressing the dielectric quality.

Appendix:

The measured RTS data were sent to Prof. Varotsos's group, University of Athens, Greece. As a result, they published in Physical Review B, vol. 73, 054504, 2006. Following are the acknowledgments in this paper.

ACKNOWLEDGMENTS

We express our sincere thanks to Dr. Rinke J. Wijngaarden for sending us the $YBa_2Cu_3O_{7-x}$ and rice pile data. We also thank **Professor Ming-Jer Chen** for providing us the MOS data.

References

[1] N. Yang, W. K. Henson, and J. J. Wortman, "Analysis of tunneling currents and reliability of NMOSFETs with sub-2 nm gate oxides," *International Electron Devices Meeting (IEDM), Technical Digest*, pp. 453-456, 1999.

[2] K. N. Yang, H. T. Huang, M. J. Chen, Y. M. Lin, M. C. Yu, S. M. Jang, C. H. Yu, and M. S. Liang, "Edge direct tunneling (EDT) induced drain and gate leakage in ultrathin gate oxide MOSFETs," Extended Abstracts of *International Conference on Solid State Devices and Materials (SSDM)*, p. 208, 2000.

[3] K. N. Yang, H. T. Huang, M. J. Chen, Y. M. Lin, M. C. Yu, S. M. Jang, C. H. Yu, and M. S. Liang, "Edge hole direct tunneling in off-state ultrathin gate oxide p-channel MOSFETs," *International Electron Devices Meeting (IEDM), Technical Digest*, pp. 679-682, 2000.

[4] K. R. Farmer, C. T. Rogers, and R. A. Buhrman, "Localized-state interactions in metal-oxide-semiconductor tunnel diodes," *Physical Review Letters*, vol. **58**, pp. 2255-2258, 1987.

[5] B. Neri, P. Olivo, and B. Riccò, "Low-frequency noise in silicon-gate metal-oxide-silicon capacitors before oxide breakdown," *Applied Physics Letters*, vol. **51**, pp. 2167-2169, 1987.

[6] E. Miranda, J. Suñé, R. Rodríguez, M. Nafría, and X. Aymerich, "Switching behavior of the soft breakdown conduction characteristic in ultra-thin (< 5 nm) oxide MOS capacitors," *International Reliability Physics Symposium (IRPS) Proceedings*, p. 42, 1998.

[7] I. Polishchuk, T. J. King, and C. Hu, "Physical origin of SILC and noisy breakdown in very thin silicon nitride gate dielectric," *Device Research Conference*

(DRC), Conference Digest, p. 20, 2001.

[8] K. F. Schuegraf and C. Hu, "Hole injection SiO_2 breakdown model for very low voltage lifetime extrapolation," *IEEE Trans. Electron Devices*, vol. **41**, pp. 761-767, 1994.

[9] H. T. Huang and M. J. Chen, "A novel sphere-based statistical model for local oxide thinning induced gate oxide breakdown," Extended Abstracts of *International Conference on Solid State Devices and Materials (SSDM)*, p. 246, 2000.

[10] E. Simoen, B. Dierickx, C. L. Claeys, and G. J. Declerck, "Explaining the amplitude of RTS noise in submicrometer MOSFETs," *IEEE Trans. Electron Devices*, vol. **39**, pp. 422-429, 1992.

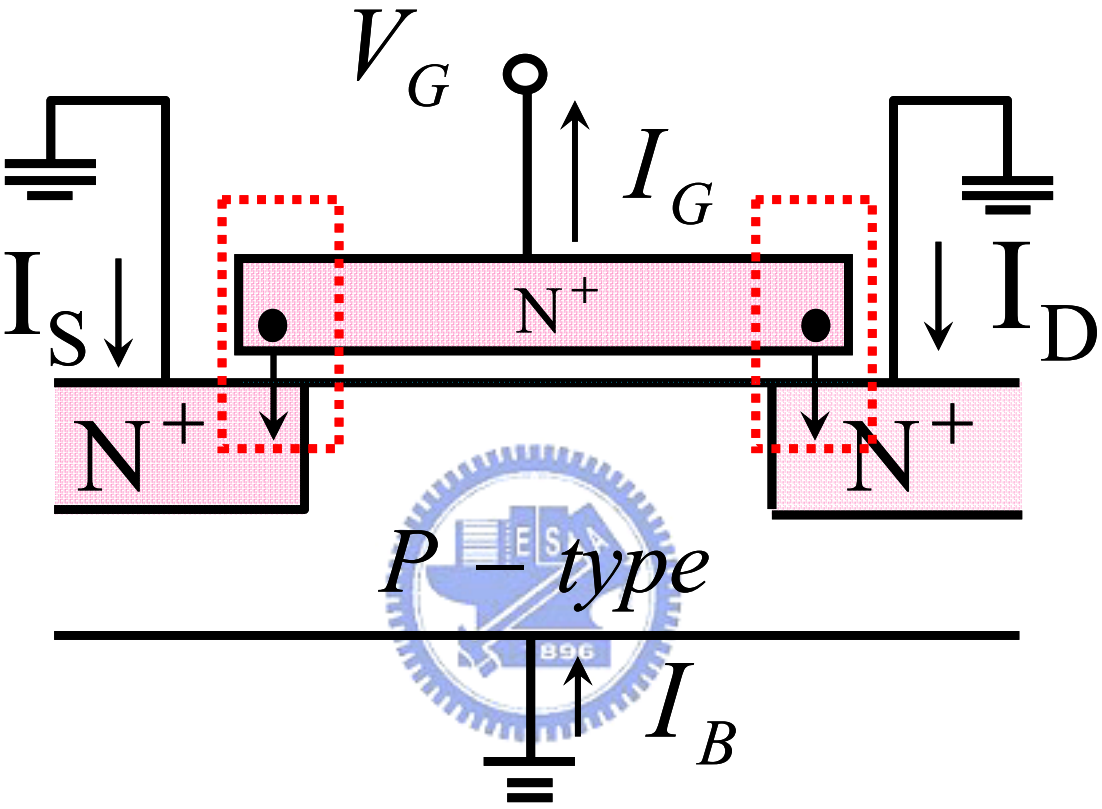

Fig. 4.1 The experimental setup.

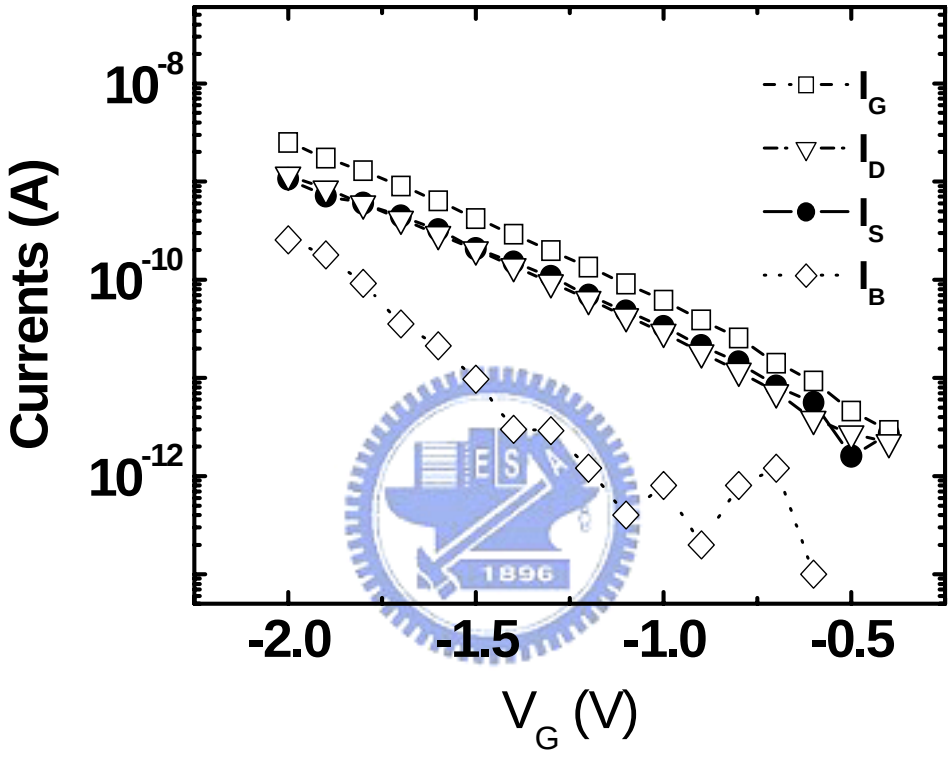

Fig. 4.2 Measured steady-state terminal currents versus negative gate voltage.

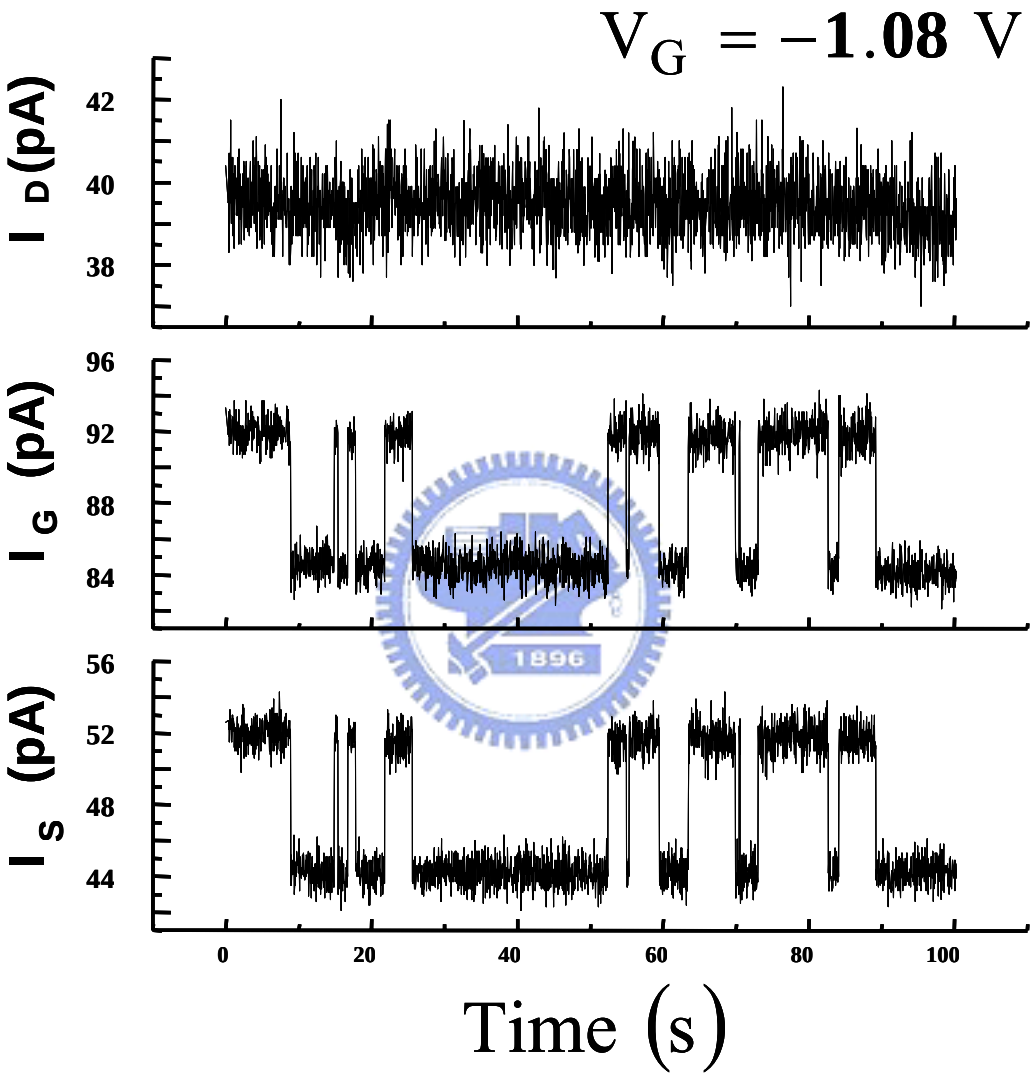

Fig. 4.3 Time records of drain, gate, and source currents at V_G = -1.08 V.

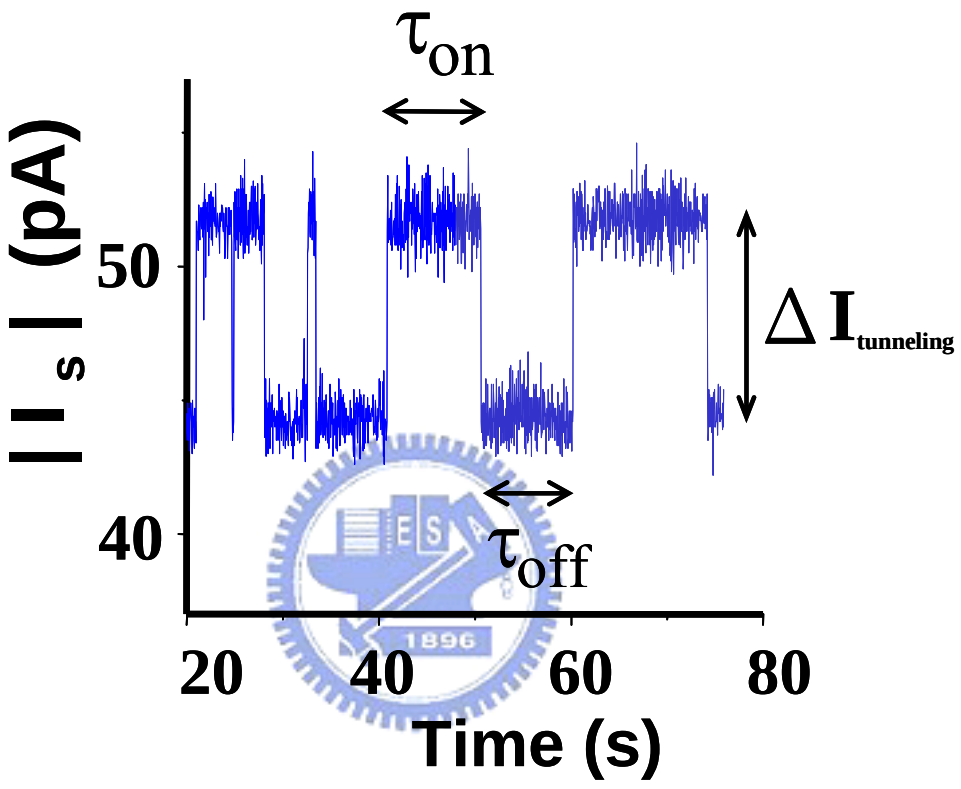

Fig. 4.4 The definitions of time constants τ_{on}, τ_{off} and $\Delta I_{tunneling}$.

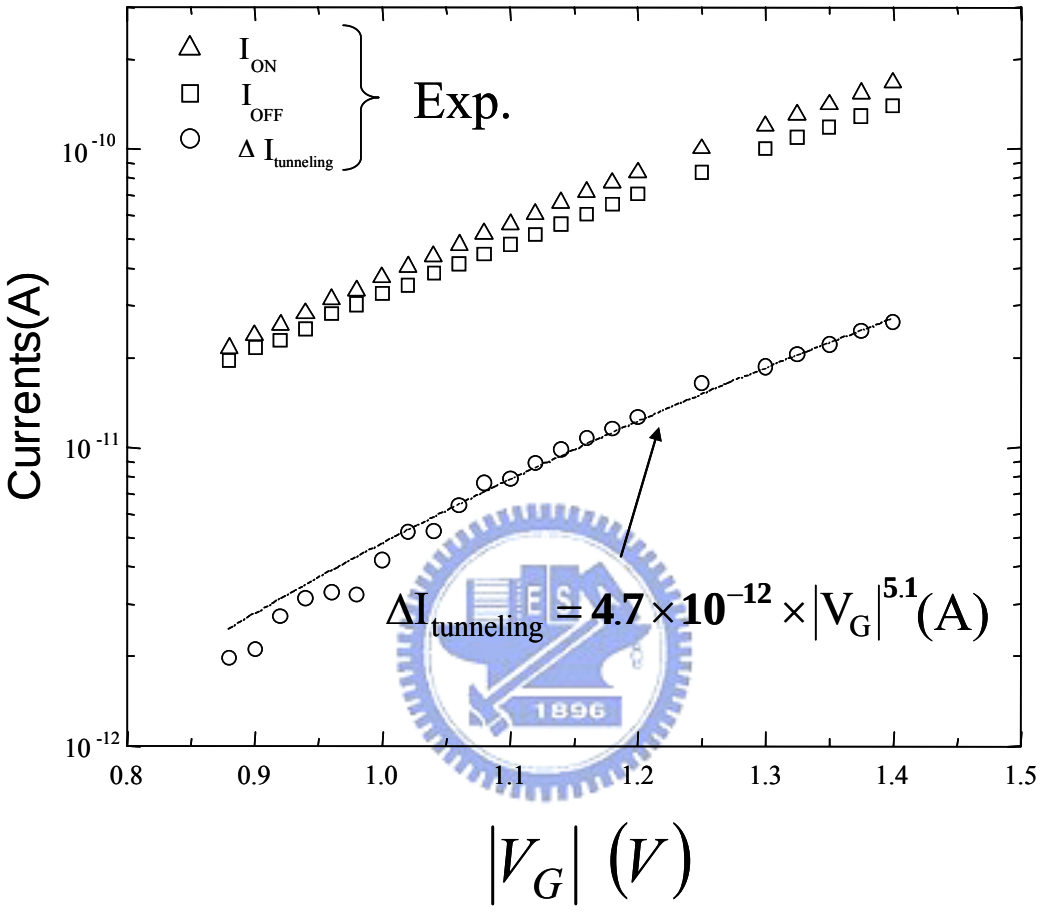

Fig. 4.5 Experimental high and low state currents versus magnitude of gate voltage for the measured device. Also shown is a line for comparison.

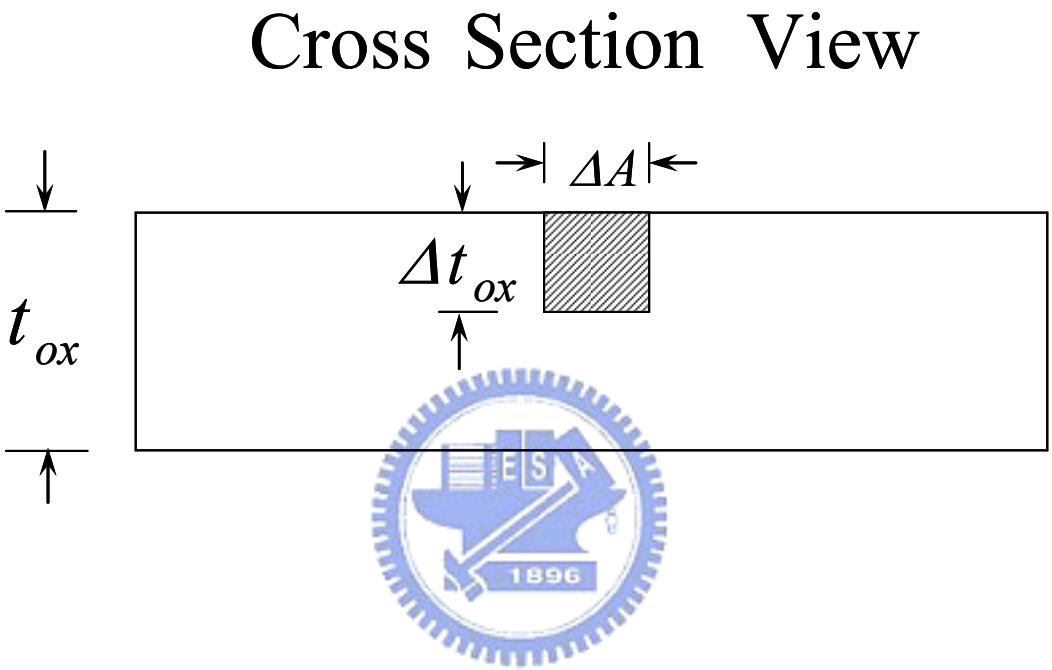

Fig. 4.6 (a) Cross section views of a gate stack overlap part devoted to EDT, defined by the thickness t_{ox} A localized gate stack thinning or conductive filament is drawn in terms of the thinning thickness Δt_{ox} and the occupied area ΔA.

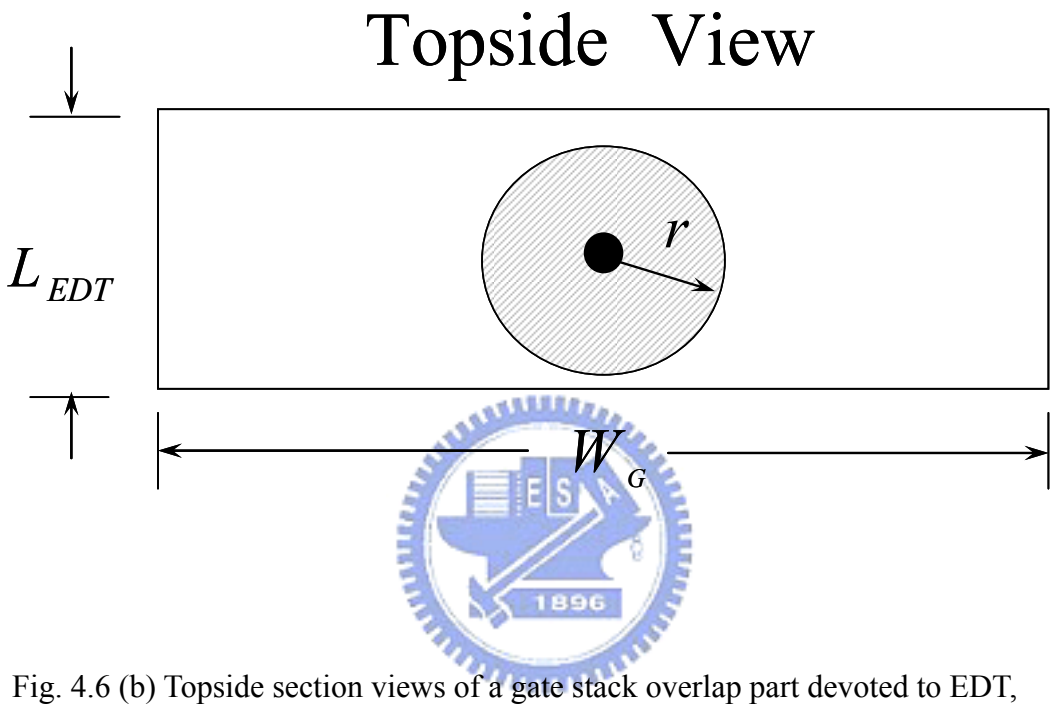

Fig. 4.6 (b) Topside section views of a gate stack overlap part devoted to EDT, defined by the gate width W_G, the effective tunneling size L_{EDT}(~6 nm). A shaded circle with the trapped electron as the origin is drawn with a radius r.

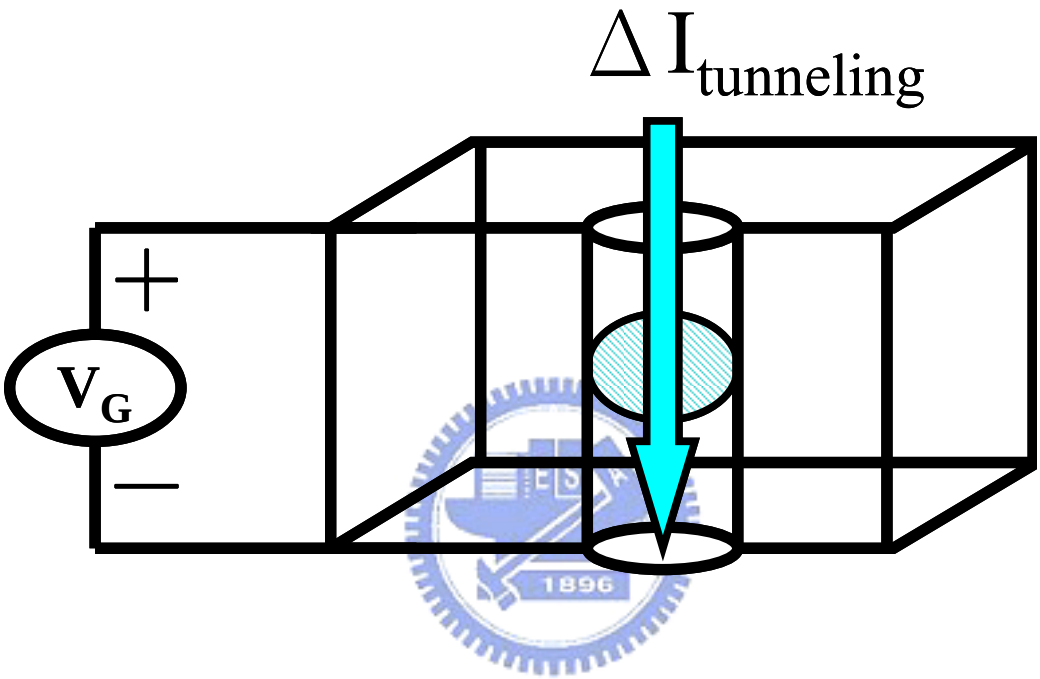

Fig. 4.7 The scheme of the $\Delta I_{tunneling}$ tunneling through a local wire-like spot.

Chapter 5

Channel-width Dependence of Low-frequency Noise in Process Tensile-strained N-channel metal-oxide-semiconductor Transistors

Low-frequency noise measurement in process tensile-strained n-channel metal-oxide-semiconductor field-effect transistors yields the density of the interface states, exhibiting a decreasing trend while decreasing the channel width. This finding corroborates the group of P_b centers caused by the lattice mismatch at (100) Si/SiO_2 interface as the origin of the underlying interface states. The inverse narrow width effect appears to be insignificant, substantially confirming the validity of the noise measurement. The present noise experiment therefore points to the enhancement of the tensile strain in the presence of channel narrowing, which in turn reduces the lattice mismatch.

5.1 Introduction

Channel strain engineering is currently recognized as an indispensable performance booster in producing next-generation metal-oxide-semiconductor field-effect transistors (MOSFETs) [1]. To achieve this goal, two fundamentally different methods have been proposed [1]: (i) strained silicon (SSi) on a relaxed SiGe buffer layer; and (ii) process strained silicon (PSS) through the trench isolation,

silicide, and cap layer. On the other hand, in the areas of unstrained counterparts, low-frequency noise has been extensively utilized since it can provide the opportunity to examine the interfacial physics [2-4]. Thus, it is a challenging issue for the low-frequency noise measurement to find further potential applications in the strain case. Recently, one such study [5] has been demonstrated that an improved noise performance can be achieved on biaxial tensile-strained substrates. In the present work, we conduct a channel-width-dependent low-frequency noise experiment on a process tensile-strained n-channel MOSFET. The resulting noise data are useful in addressing the effect of enhanced tensile strain in the channel width direction.

5.2 Experimental

The device under test was an n-channel MOSFET fabricated using the concept of process tensile strain, mainly through the trench isolation [6]. The physical gate oxide thickness was 1.4 nm as determined by capacitance-voltage fitting. The channel length was 0.5 μm while the channel width spanned a wide range of 0.11, 0.24, 0.6, 1, and 10 μm. Here, a reduction in channel width means an enhancement in tensile strain in the channel width direction. Fig. 5.1 displays the measured drain current per unit channel width versus drain voltage with the gate overdrive as a parameter. This can be easily understood by means of the current drive enhancement factor against channel width as shown in Fig. 5.2. As expected, the drain current per unit channel width increases as the channel width is decreased. The increased drain current can be well related to the mobility enhancement; that is, the tensile stress causes subbands energy shift, which in turn suppresses the intervalley phonon scattering while reducing the effective conductivity mass, and thereby enhances the mobility [1,7].

The low-frequency noise measurement setup used was the same as that detailed

elsewhere [4]. The measurement frequency ranged from 3 Hz to 100 kHz while operating the devices at a drain voltage of 0.2 V. Here, the noise experiment was carried out in terms of the input-referred noise voltage spectral density S_{Vg}. Fig. 5.3 depicts measured S_{Vg} versus frequency for a gate overdrive of 0.6 V, where three devices, as labeled A, B, and C with the same channel width of 0.11 μm, represent three different positions on the wafer. Apparently, a considerable variation of low-frequency noise exists between devices, which can be attributed to statistical fluctuations of the number of the interface traps [8]. Hence, it is argued that the measured noise data essentially can follow the $1/f^{\gamma}$ relationship with the power coefficient γ close to unity.

5.3 Experiment Results and Discussion

Also shown in Fig. 5.4 is the corresponding average threshold voltage shift with respect to the wide structure (i.e., 10 μm) such as to address the possibility of the inverse narrow width effect (INWE) caused by the impurity segregation and the fringing electric field at the isolation sidewall [9,10]. It can be seen that the threshold voltage shift is rather small (< 5%), indicating that the inverse narrow width effect is not a significant issue in the undertaken devices. In addition, since the noise data are obtained in the strong inversion region, the channel current part along the channel edge is not significant relative to the overall channel. As a result, the inverse narrow width effect can further be weakened. Therefore, it is reasonably drawn that the present low-frequency noise is a good tool to monitor the Si/SiO_2 interface over the whole channel area. The Fig. 5.5 displayed that the measured input-referred noise voltage spectral density at a specific frequency of 100 Hz versus channel width for V_{go} = 0.6 V and V_D = 0.2 V. The error bar represents the standard deviation of the

distribution and the data point the mean of the distribution.

The weak dependence of low-frequency noise on gate overdrive (> 0.4 V) as mentioned above suggests that the carrier number fluctuations prevail in the strong inversion mode. In other words, under such situations the Coulomb scattering can be ignored due to the screening of the trapped charge by the gate electrode and the inversion-layer charge. Hence, the following input-referred noise voltage spectral density expression can be adequately cited [3,11]

$$S_{Vg} = \frac{q^2 k_B T \lambda}{C_{eff}^{\ 2}} \frac{N_t}{WL} \frac{\mathbf{1}}{f^{\gamma}} \tag{1}$$

where q is the elementary charge, k_B is Boltzmann's constant, T is the absolute temperature, λ is the tunneling distance (~ 0.1 nm), W is the channel width, L is the Channel length, C_{eff} is the effective gate oxide capacitance per unit area, and N_t is the effective near-interface oxide trap density. With known C_{eff} (≈1.75 μF/cm^2) from the undertaken manufacturing process, fitting of all the S_{Vg} data using Eq. (1) led to the distribution of N_t as shown in Fig. 5.6 versus channel width. Again on the average, the interface state density decreases with decreasing channel width. Specifically, a reduction in channel width by a factor of about 100 produces a 10-fold reduction in interface state density. Once again, a reduction in channel width means an enhancement of tensile strain in that direction; therefore, the present noise experiment points to a reduction in interface state density in the presence of enhanced tensile strain in the channel width direction.

It is interesting to further examine the physical origin of the underlying interface

states. Analogous to the electron-spin resonance (ESR) experiment on a (111) Si/SiO_2 interface [12-14], the interface states investigated in our low-frequency noise work can be attributed to the group of P_b centers caused by the lattice mismatch (or equivalently the dangling-bond defects as characterized in terms of $Si_3{\equiv}Si^{\cdot}$) at the (100) Si/SiO_2 interface. Here, the lattice-mismatch stress is primarily related to the thermal oxidation process since in this process about 2.2 unit volumes of oxide are produced for each unit volume of silicon consumption. Obviously, the lattice mismatch can be reduced using a tensile strain, thereby leading to a reduction in P_b centers [14]. Note that there were few studies on the usage of the ESR technique to detect P_b centers in the case of (100) Si/SiO_2 interface. The noise experiment on the (100) Si/SiO_2 interface in this work again corroborates the action of applying a tensile stress: *Enlarging the Si—Si interatomic distance before the silicon oxidation process is carried out, which leads to reduced lattice-mismatch stress during the subsequent thermal oxidation process.*

5.4 Conclusion

Channel-width-dependent low-frequency noise measurement has been applied on a process tensile strained n-channel MOSFET. One important finding has been straightforwardly created: *enhanced tensile strain in the channel narrowing direction can reduce the lattice-mismatch defects.*

References

[1] S. E. Thompson, M. Armstrong, C. Auth, M. Alavi, M. Buehler, R. Chau, S. Cea, T. Ghani, G. Glass, T. Hoffman, C. H. Jan, C. Kenyon, J. Klaus, K. Kuhn, Z. Ma, B. Mcintyre, K. Mistry, A. Murthy, B. Obradovic, R. Nagisetty, P. Nguyen, S. Sivakumar, R. Shaheed, L. Shifren, B. Tufts, S. Tyagi, M. Bohr, and Y. El-Mansy, "A 90-nm logic technology featuring strained-silicon**,**" *IEEE Trans. Electron Devices*, vol. **51**, pp. 1790-1797, 2004.

[2] M. J. Kirton and M. J. Uren, "Noise in solid-state microstructures: a new perspective on individual defects, interface states and low-frequency (1/f) noise," *Adv. Phys.*, vol. **38**, pp. 367-468, 1989.

[3] K. K. Hung, P. K. Ko, C. Hu, and Y. C. Cheng, "A unified model for the flicker noise in metal-oxide-semiconductor field-effect transistors," *IEEE Trans. Electron Devices*, vol. **37**, pp. 654-665, 1990.

[4] M. J. Chen, T. K. Kang, Y. H. Lee, C. H. Liu, Y. J. Chang, and K. Y. Fu, "Low-frequency noise in *n*-channel metal-oxide-semiconductor field-effect transistors undergoing soft breakdown," *J. Appl. Phys.*, vol. **89**, pp. 648-653, 2001.

[5] E. Simoen, G. Eneman, P. Verheyen, R. Delhougne, R. Loo, K. De Meyer, and C. Claeys, "On the beneficial impact of tensile-strained silicon substrates on the low-frequency noise of *n*-channel metal-oxide-semiconductor transistors," *Appl. Phys. Lett.*, vol. **86**, 223509 (2005).

[6] C. H. Ge, C. C. Lin, C. H. Ko, C. C. Huang, Y. C. Huang, B. W. Chen, B. C. Perng, C. C. Sheu, P. Y. Tsai, L. G. Yao, C. L. Wu, T. L. Lee, C. J. Chen, C. T. Wang, S. C. Lin, Y. C. Yeo, and C. Hu, "Process-strained Si (PSS) CMOS technology featuring 3D strain engineering," *Tech. Dig. - Int. Electron Devices Meet*. 2003, p. 73.

[7] S. Takagi, J. L. Hoyt, J. J. Welser, and J. F. Gibbons, "Comparative study of phonon-limited mobility of two-dimensional electrons in strained and unstrained Si metal–oxide–semiconductor field-effect transistors," *J. Appl. Phys.*, vol. **80**, pp. 1567-1577, 1996.

[8] M. J. Uren, D. J. Day, and M. J. Kirton, "1/*f* and random telegraph noise in silicon metal-oxide-semiconductor field-effect transistors," *Appl. Phys. Lett.*, vol. **47**, pp.1195-1197, 1985.

[9] M. Sugino and L. A. Akers, "Subthreshold current in oxide isolated structures," *IEEE Electron Device Lett*., vol. **4**, pp. 114-115, 1983.

[10]S. K. H. Fung, M. Chen, and P. K. Ko, "Inverse-narrow-width effect of deep sub-micrometer MOSFETs with locos isolation," *Solid-State Electron.*, vol. **41**, pp. 1885-1889, 1997.

[11] H. M. Bu, Y. Shi, X. L. Yuan, J. Wu, S. L. Gu, Y. D. Zheng, H. Majima, H. Ishikuro, and T. Hiramoto, "Random telegraph signals and low-frequency noise in *n*-metal–oxide–semiconductor field-effect transistors with ultranarrow channels," *Appl. Phys. Lett.*, vol. **76**, pp. 3259-3261, 2000.

[12] E. H. Poindexter, P. J. Caplan, B. E. Deal, and R. R. Razouk, "Interface states and electron spin resonance centers in thermally oxidized (111) and (100) silicon wafers," *J. Appl. Phys.*, vol. **52**, pp. 879-884, 1981.

[13] A. Stesmans, "Structural relaxation of P_b defects at the (111)Si/SiO_2 interface as a function of oxidation temperature: The P_b-generation–stress relationship," *Phys. Rev. B*, vol. **48**, pp. 2418-2435, 1993.

[14] A. Stesmans, D. Pierreux, R. J. Jaccodine, M. T. Lin, and T. J. Delph, "Influence of *in situ* applied stress during thermal oxidation of (111)Si on P_b interface defects," *Appl. Phys. Lett.*, vol. **82**, pp. 3038-3040, 2003.

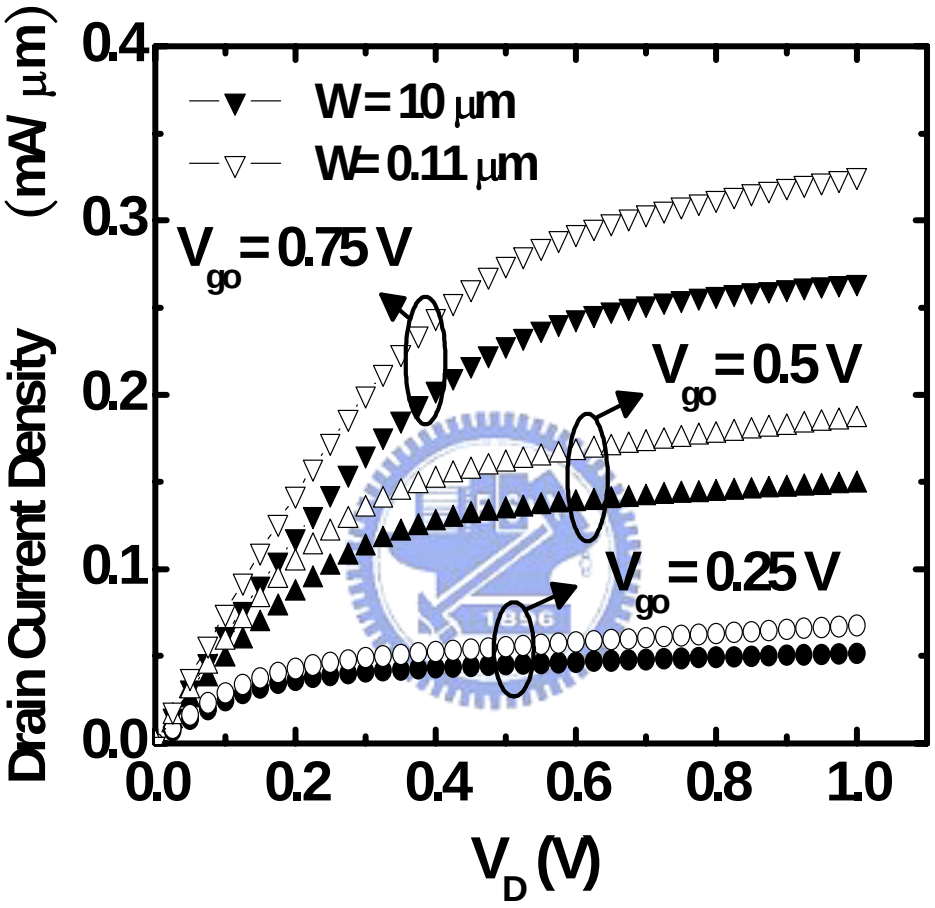

Fig. 5.1 The measured drain current per unit channel width versus drain voltage with gate overdrive as a parameter.

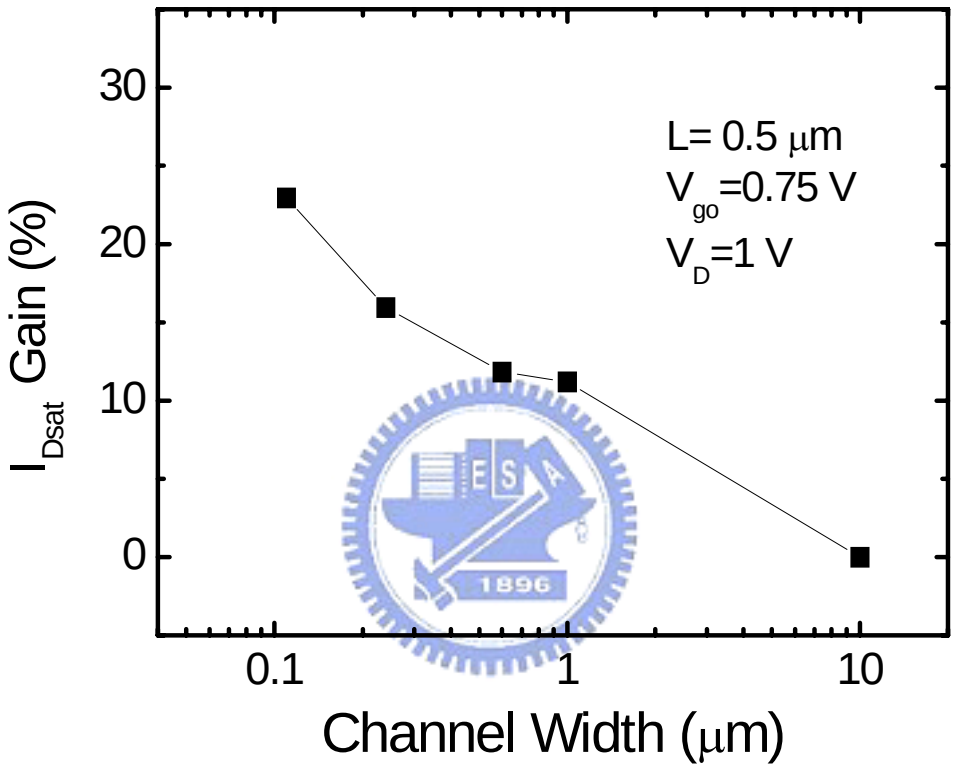

Fig. 5.2 The experimental drain current enhancement factor versus channel width for gate overdrive V_{go} = 0.75 V and drain voltage V_D = 1 V.

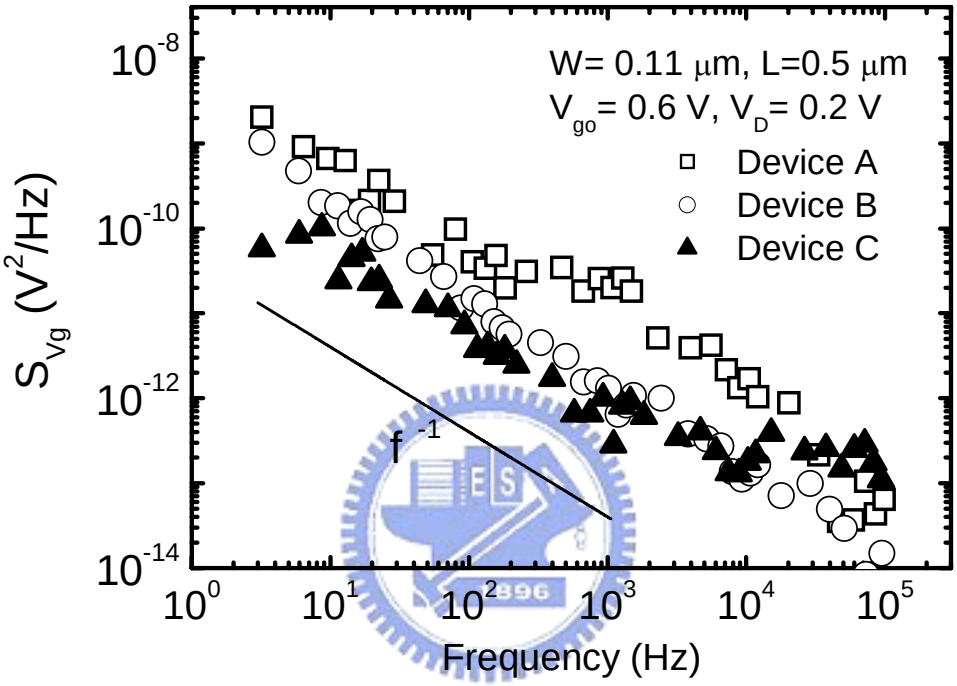

Fig. 5.3 The measured input-referred noise voltage spectral density at V_{go} = 0.6 V and V_D = 0.2 V versus frequency for three different positions on the wafer.

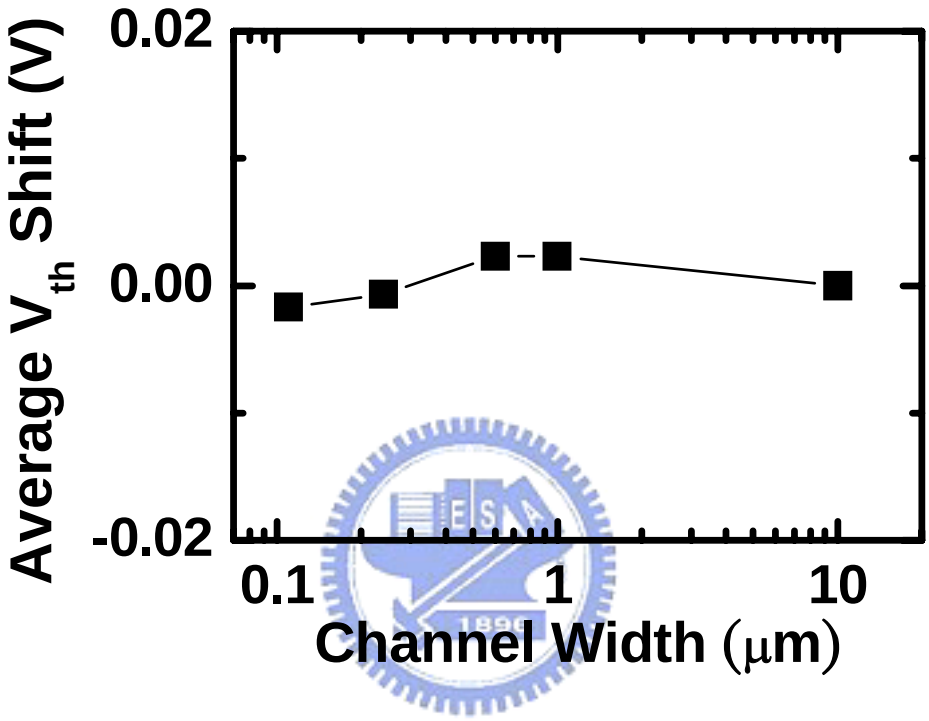

Fig. 5.4 The corresponding average threshold voltage shift versus channel width with the wide structure (i.e., 10 μm) as a reference point. Here, the threshold voltage was determined at V_D = 0.01 V.

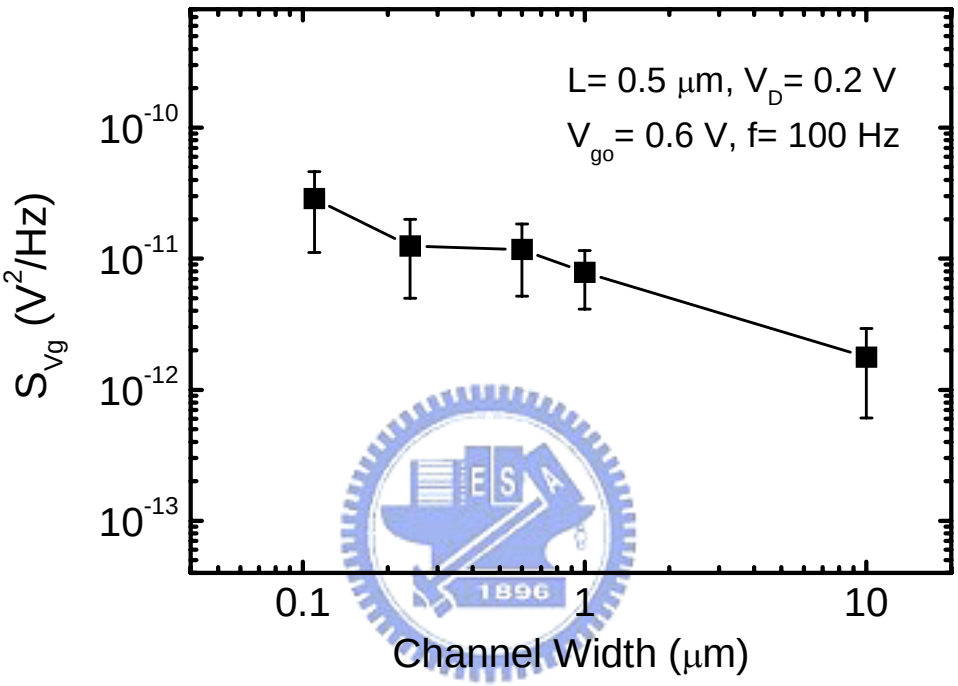

Fig. 5.5 The measured input-referred noise voltage spectral density at a specific frequency of 100 Hz versus channel width for V_{go} = 0.6 V and V_D = 0.2 V. The error bar represents the standard deviation of the distribution and the data point the mean of the distribution.

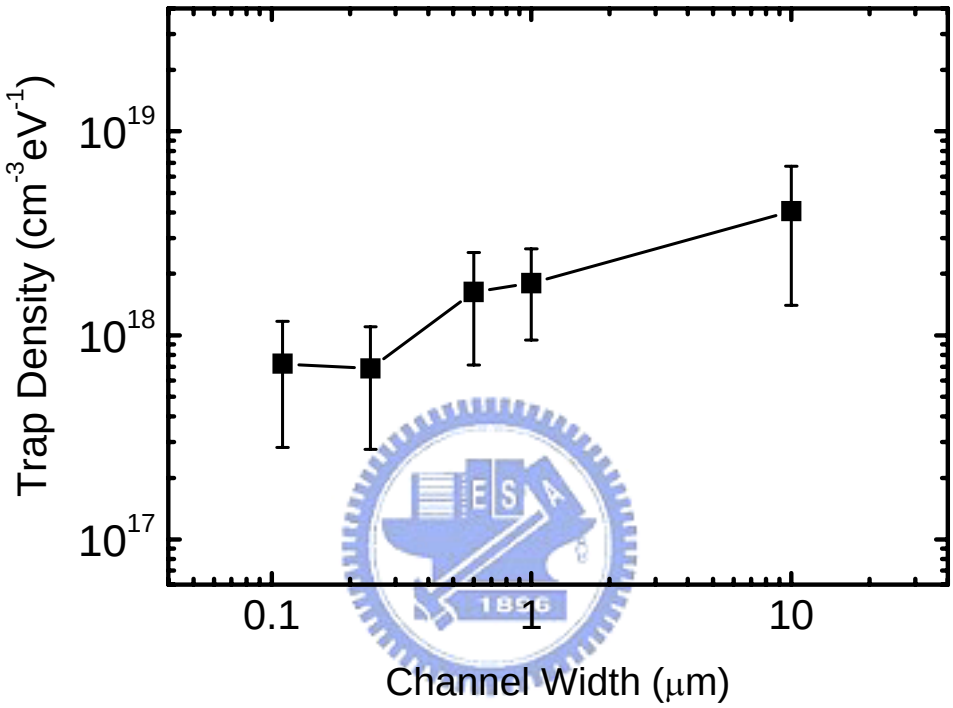

Fig. 5.6 The extracted effective interface state density corresponding to Fig. 5.5.

Chapter 6

Conclusions

This dissertation demonstrates the study of the noise and fluctuations in advanced devices. Conclusions of the work are described as below:

- We have presented experimental evidence concerning the Coulomb energy enhancement in a MOS system with a nanometer-scale oxide trap. Other corroborating evidence based on a multiphonon theory has elucidated the measured capture and emission kinetics. The corresponding configuration coordinate diagrams have been established. We have further elaborated on the clarification of the Coulomb energy and have differentiated it from that in memories containing nanocrystals as a floating gate. Some critical issues encountered in the work have been addressed as well.

- The undertaken study makes a great contribution to the areas of MOSFETs RTS. In addition, a near-distance effect is experimentally confirmed, which is that Coulomb scattering in a two-dimensional system is sensitive to the depth of a nanoscale oxide trap. Specifically, in strong inversion conditions, the amount of Coulomb scattering with an interface trap drops with a faster rate than the deep trap, indicating a stronger screening potential confinement. The near-distance effect is expected to prevail in next-generation nanoscale devices involving the

mobility, the RTS, and the low-frequency noise.

- A new on-off switching behavior is measured via EDT currents in the overlap part of the gate stack. We have shown that RTS can be a useful tool for studying process-induced trap. It can find the trap location and trap type by analyzing RTS data. We also find that the process-induced trap is located near interface, or around the middle of the dielectric layer. With process defects as the plausible origin, experimental data can be adequately described by electron trapping-detrapping theories. Also highlighted straightforwardly are: assessing the two-terminal current-voltage (I-V) characteristic associated with a defective spot; finding the occurrence probability of the defects as well as their locations in a manufacturing process, etc. Oxide reliability is a major concern for deep-submicron technologies as the dielectric thickness and channel area scale down. RTS thereby is a promising tool to address the dielectric quality.

- Channel-width-dependent low-frequency noise measurement has been applied on a process tensile strained n-channel MOSFET. One important finding has been straightforwardly created: enhanced tensile strain in the channel narrowing direction can reduce the lattice-mismatch defects.

博士候選人學經歷表

姓　　名：呂明霈　　Ming-Pei Lu

性　　別：男

出生日期：中華民國 六十六年 一月 八號

出 生 地：台灣雲林縣

住　　址：雲林縣北港鎮光復里光復二路三號

學　　歷：國立交通大學 電子工程學系

(84 年 9 月~ 89 年 6 月)

國立交通大學 電子工程研究所碩士班 固態組

(89 年 9 月~ 91 年 6 月)

國立交通大學 電子工程研究所博士班 固態組

(91 年 9 月 入學)

經　　歷：線性代數助教.

參與 TSMC 與交通大學之產學合作計畫.

與交通大學生物科技所楊裕雄教授研究團隊共同參與工研院電子所的奈米碳管研究計畫.

論文題目：先進元件中原子尺寸缺陷之研究

Study on the Atomic-sized Traps in Advanced Devices

著　　作：(如後附頁)

Publication List

Journals and Letters

1. M. J. Chen and **M. P. Lu**, "On-off switching of edge direct tunneling currents in metal-oxide-semiconductor field-effect transistors," *Applied Physics Letters*, vol. **81**, pp. 3488-3490, 2002.
2. **M. P. Lu** and M. J. Chen, "Oxide Trap Enhanced Coulomb Energy in a Metal-Oxide-Semiconductor System," *Physical Review B*, vol. **72**, 235417, 2005.
3. **M. P. Lu,** C. Y. Hsiao, P. Y. Lo, J. H. Wei, Y. S. Yang, and M. J. Chen, "Semiconducting single-walled carbon nanotubes exposed to distilled water and aqueous solution: electrical measurement and theoretical calculation," *Applied Physics Letters,* vol. **88**, 053114, 2006; Selected Articles in *Virtual Journal of Nanoscale Science & Technology*, Vol. **13**, Issue 6, 2006; Selected Articles in *Virtual Journal of Biological Physics Research*, Vol. **11**, Issue 4, 2006.
4. **M. P. Lu**, W. C. Lee, and M. J. Chen, "Channel width dependence of low-frequency noise in process tensile-strained n-channel metal-oxide-semiconductor transistors," *Applied Physics Letters*, vol. **88**, 063511, 2006.
5. **M. P. Lu** and M. J. Chen, "Near-distance effect of a charged oxide trap on Coulomb Scattering in a Two-dimensional System," submitted to *Physical Review Letters* (under review).

Conference

1. **M. P. Lu** and M. J. Chen, "Anomalous Behaviors of Random Telegraph Signals in Ultra-thin Gate Oxide MOSFETs," ***SSDM (Tokyo)***, pp. 408-409, 2004.