國立交通大學

電機學院微電子奈米科技產業研發碩士班

碩士論文

針對高度微縮金氧半場效電晶體的參數精確萃取方法

Accurate Parameter Extraction Methods for Highly Scaled MOSFETs

研 究 生：陳彥銘

指導教授：陳明哲 教授

中華民國九十七年七月

針對高度微縮金氧半場效電晶體的參數精確萃取方法
Accurate Parameter Extraction Methods for Highly Scaled MOSFETs

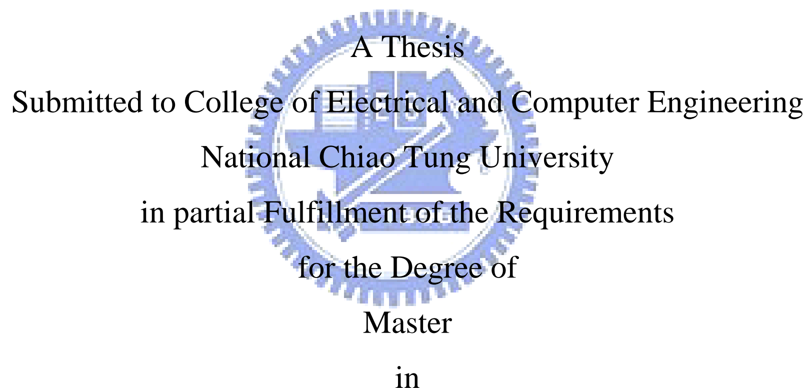
研 究 生：陳彥銘　　　Student：Yen-Ming Chen

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

國 立 交 通 大 學

電機學院微電子奈米科技產業研發碩士班

碩 士 論 文



A Thesis

Submitted to College of Electrical and Computer Engineering

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Industrial Technology R & D Master Program on

Microelectronics and Nano Sciences

July 2008

Hsinchu, Taiwan, Republic of China

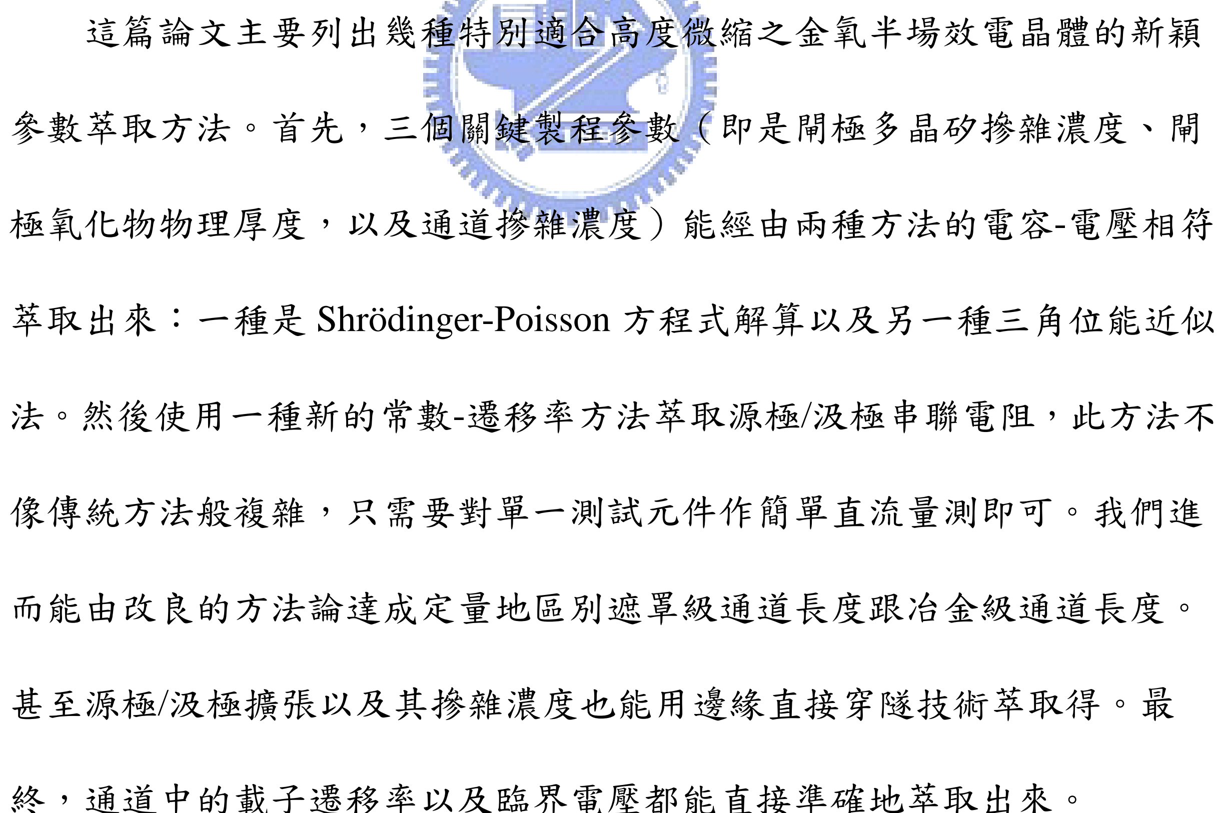
中華民國九十七年七月

針對高度微縮金氧半場效電晶體的參數精確萃取方法

學生：陳彥銘　　　　　　　　　　　　　　　　　指導教授：陳明哲博士

國立交通大學電機學院產業研發碩士班

摘　　要



這篇論文主要列出幾種特別適合高度微縮之金氧半場效電晶體的新穎參數萃取方法。首先，三個關鍵製程參數（即是閘極多晶矽摻雜濃度、閘極氧化物物理厚度，以及通道摻雜濃度）能經由兩種方法的電容-電壓相符萃取出來：一種是 Shrödinger-Poisson 方程式解算以及另一種三角位能近似法。然後使用一種新的常數-遷移率方法萃取源極/汲極串聯電阻，此方法不像傳統方法般複雜，只需要對單一測試元件作簡單直流量測即可。我們進而能由改良的方法論達成定量地區別遮罩級通道長度跟冶金級通道長度。甚至源極/汲極擴張以及其摻雜濃度也能用邊緣直接穿隧技術萃取得。最終，通道中的載子遷移率以及臨界電壓都能直接準確地萃取出來。

Accurate Parameter Extraction Methods for Highly Scaled MOSFETs

Student：Yen-Ming Chen Advisors：Dr. Ming-Jer Chen

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

Abstract

This thesis mainly addresses several novel parameter extraction methods that are particularly suitable for highly scaled MOSFETs. First of all, the three key process parameters (namely, the gate polysilicon doping concentration, the gate oxide physical thickness, and the channel doping concentration) are extracted via C-V fitting by means of the two methods: one of the Shrödinger-Poisson equation solving and one of the triangular potential approximation. Then a new constant-mobility method is adopted to extract the source/drain series resistance Rsd, which, unlike the conventional counterparts, requires only simple DC measurements on a single test device. Once Rsd is extracted, we can quantitatively distinguish between the gate length at the mask level and the channel metallurgical length, which is achieved with the improved methodology. Even the source/drain extension and its doping concentration can be extracted using the edge direct tunneling technique. Finally, the carrier mobility in the channel, as well as the threshold voltage, can be straightforwardly extracted.

誌　　謝

在碩士兩年多的生涯裡，從碩一專注修課的歲月、碩二遭逢幾位親人相繼辭世哀慟的日子，以迄畢業之前步入半工半讀過渡的時光，當中要感謝很多人的鼓勵、協助、輔導與支持。能夠完成這篇論文，首先由衷感謝我的指導教授陳明哲老師，在我研究毫無頭緒時的方向指引、遭遇瓶頸時的循循善誘、慌張軟弱時的體察開導；甚至准假參與親人悼念習俗時的關懷開釋，在在讓我心裡充滿無盡的感激。

在實驗室做研究時，由於實驗室夥伴的用心引導，讓我從中學到不少珍貴的資產。謝振宇學長在電流量測上的啟蒙以及在論文上「用力」的指導；李建志學長的義氣照顧、教導及關心；許智育學長在電流、電容量測上的經驗傳承；呂立方在處理數據及模擬上的協助與建議；林以唐跟梁惕華在寫程式建立模型上的幫助；周佳弘在儀器出問題時的盡力協助；也要感謝李韋漢、湯侑穎、簡鶴年、宋東壕、陳又正、陳以東等夥伴平時在日常生活上的輔助，讓我沉悶的實驗室生活多了些溫暖。

最後要感謝的是，一路上在背後支持的家人。如果說實驗室是提供我研究上的幫助，而家人就是提供我心理上的建設。在研究過程遇到困難、著急及不知所措時，家人總會給予心理上的安撫，進而提供實際行動上的建議，讓我在跌跌撞撞的研究過程中，總能安然渡過。

謹以此篇短小的感言，表達我心中無止盡的感謝，獻給每一位在我研究生涯中，接觸到的每個人，感謝您們。

Content

Abstract (Chinese) i

Abstract (English) ii

Acknowledgement iii

Content iv

Figure Captions v

Chapter 1 Introduction.. 1

Chapter 2 Effective Channel Length Extraction..................... 4

2.1 Definition of Effective Channel Length.................. 4

2.2 Classical Current-based Leff Extraction Methods....... 4

2.2.1 Channel-Resistance Method................................ 4

2.2.2 Shift-and-Ratio Method..................................... 5

2.3 Capacitive Method... 7

2.3.1 Constant ΔL Method..................................... 8

2.3.2 Individual ΔL Method..................................... 8

2.4 Edge Direct Tunneling and Modeling Description...... 9

Chapter 3 Key Parameters Extraction............................... 14

3.1 Threshold Voltage Extraction............................ 14

3.2 Series Resistance Extraction.............................. 14

3.3 Effective Mobility Extraction............................. 17

Chapter 4 Experimental Data and Interpretations.................. 20

Chapter 5 Conclusion.. 23

Reference 24

Figure Captions

Fig. 1 Schematic diagram showing the edge direct tunneling leakage current path for an nMOSFET device under accumulation conditions........... 27

Fig. 2 Schematic description of the increasingly important R_{sd} portion of the total resistance for MOSFET devices during the scaling. 28

Fig. 3 Schematically shown L_{eff} defined as the distance between the source and drain diffusion extensions at the silicon surface. 29

Fig. 4 Schematic illustration of Channel - Resistance L_{eff} extraction method through a relation between measured linear region resistance R_{tot} and gate length. 30

Fig. 5 The detailed triangular potential profile for nMOSFET devices......... 31

Fig. 6 Schematic energy band diagram for an n^+ poly gate/SiO_2/drain extension structure showing the voltage balance relationship in accumulation. 32

Fig. 7 Schematic of the band diagram for edge direct tunneling from accumulated n^+ poly gate to depleted drain diffusion extension. The accumulated well is modeled with a triangular potential one. The wave function for electrons is partially reflected at each potential discontinuity. 33

Fig. 8 The calculation process flow chart for the EDT model...................... 34

Fig. 9 The relations between the measured μ_{eff} and the E_{eff} at the Si/ SiO_2 interface under different V_{bs} conditions. The μ_{eff} converges toward one universal curve when E_{eff} is sufficiently high. 35

Fig. 10 Schematic illustration of the R_{sd} equivalent circuit along with the cross section of an nMOSFET device. 36

Fig. 11 Depicted flowchart summarizing the procedures of μ_{eff} extraction. 37

Fig. 12 Comparsion of the measured (triangular symbol) and simulated (solid curve) C-V characteristics of an nMOSFET device. 38

Fig. 13 The V_{th} extraction results from ELR method and SD method for different nMOSFET devices. 39

Fig. 14 The L_{eff} extraction result of an nMOSFET device with Channel-Resistance method. 40

Fig. 15 The L_{eff} extraction result of an nMOSFET device with shift-and-ratio method. 41

Fig. 16 The experimental $C_{gc}(V_{gs})$ curves for several nMOSFET devices as plotted in a logarithm scale. 42

Fig. 17 The ΔL extraction result of an nMOSFET device with constant ΔL method from a linear regression on the plot of $C_{gc,max}=f(L)$. 43

Fig. 18 The ΔL extraction result of an nMOSFET device using individual ΔL method with a proportionality relationship................................44

Fig. 19 Comparsion of the measured (triangular symbol) and simulated (solid curve) I_g-V_{gs} characteristics under accumulation conditions of an nMOSFET device...45

Fig. 20 The relations between ΔL of nMOSFET devices resulting from EDT method and gate length. 46

Fig. 21 The relations between ΔL of nMOSFET devices resulting from EDT method and gate width. 47

Fig. 22 The L_{eff} behavior for nMOSFET devices resulting from EDT method. 48

Fig. 23 V_{gs} dependences of R_{sd} under different V_{bs} condtions. Erroneous R_{sd} values appear in the low V_{gs} region because the "constant mobility" criterion is not satisfied when the E_{eff} is not sufficiently high...........49

Fig. 24 Plot of the R_{sd} versus gate length of the nMOSFET devices...............50

Fig. 25 Plot of the R_{sd} versus gate width of the nMOSFET devices................51

Fig. 26 Statistical analysis of the R_{sd} variations with different SDE overlap lengths. 52

Fig. 27 The μ_{eff} curves of the W/L = 1μm/0.1μm nMOSFET device with raw data, L_{eff} corrected, R_{sd} corrected, and L_{eff} & R_{sd} corrected...............53

Fig. 28 Displaying the peak mobility variations induced by ΔL versus gate length of the W = 1 μm nMOSFET devices......................................54

Fig. 29 Statistical analysis of the peak mobility variations induced by ΔL with different gate width devices. 55

Fig. 30 Displaying the μ_{eff} @V_{gs} = 1.5V variations induced by R_{sd} versus gate length of the W = 1 μm nMOSFET devices......................................56

Fig. 31 Statistical analysis of the μ_{eff} @V_{gs} = 1.5V variations induced by R_{sd} with different gate widths..57

Fig. 32 Comparision between the raw and the corrected μ_{eff} @V_{gs} = 1.5V values versus gate length of the W = 1 μm nMOSFET devices. 58

Fig. 33 Comparision between the raw and the corrected μ_{eff} @V_{gs} = 1.5V values with different gate widths. 59

Chapter 1
Introduction

In today's integrated circuit industry, the metal-oxide-semiconductor (MOS) devices are moving toward nanoscale to reduce cost and improve performance. In order to scale MOS devices to smaller dimensions while maintaining good control of the short-channel effect, the conventional transistor scaling theory requires that the gate oxide thickness (t_{ox}) be reduced as the minimum channel length is scaled down. According to the National Technological Roadmap for Semiconductors (NTRS), the scaling trend for gate dielectrics is such that for sub - 100 nm generation devices, an equivalent gate oxide thickness of less than 2.0 nm will be required. This gives rise to several issues such as excess gate leakage current, degradation of mobility, C(V) and Id(V_g) characterization difficulties, etc [1]-[4].

The extraction of key MOS devices parameters such as the effective channel length (L_{eff}), the source/drain (S/D) series resistance (R_{sd}) and the mobility of highly scaled devices are of fundamental importance for the analysis of developed MOSFET technologies. However, with the scaling of MOS devices, the increasing gate leakage current will cause more problems such as C-V measurement, the standby power consumption of a highly integrated chip, etc [1]-[6]. The leakage current flowing across the MOS structure usually includes Fowler-Nordheim (FN) tunneling current and direct tunneling (DT) current. With thinning oxide layer, direct tunneling bocomes the main mechanism for the leakage current [7]. By this reason, a reliable and computation efficient physical model for characterizing the direct tunneling current of devices is necessary.

As schematically shown in Fig. 1, for an n-MOSFET at the accumulation

condition, there is no inversion layer formed between the source and drain diffusion extension under the oxide/silicon-substrate interface. The dominant off-state leakage component of electron from the polysilicon to underlying n-type S/D diffusion extension is called edge direct tunneling (EDT).

In order to study the edge direct tunneling current in MOS devices, there are two approaches such as the self-consistent Shrödinger-Poisson equation and the triangular potential approximation. Under the accumulation conditions, the electrons are confined in the narrow potential well close to the accumulated semiconductor interface. The tight confinement leads to a splitting of the energy levels into subbands for motion in the direction normal to the silicon surface, which must be treated quantum-mechanically (QM). That is, the electronic properties of the accumulation layer behave as two-dimensional electron gas (2DEG) [1], [8]. The EDT model used in this thesis is based on the triangular potential approximation for the first time derived for the oxide field (F_{ox}) at the gate edge by accounting for electron subband in the quantized accumulation poly-silicon surface. This model relates the F_{ox} to gate-to-drain voltage (V_{DG}), oxide thickness (t_{ox}), and doping concentration of drain diffusion extension (N_{DE}) [9]. Once F_{ox} is known (input), EDT I-V data will be reproduced consistently with the model. The tunneling path size ($W \times L_{TN}$) and the N_{DE} will be extracted with properly fitting. Finally, the L_{eff} is gotten with the formula：

$$\mathbf{L_{eff} = L - 2L_{TN}} \quad (1.1)$$

where L_{TN} is the length of the drain diffusion extension.

In MOSFET devices, the S/D series resistance R_{sd} leads to excess potential drop and degraded drive capacity. As MOSFET devices scale, the R_{sd} portion of the total resistance in MOSFET devices increases and consequently the drive current degradation is more serious in highly scaled devices as shown in Fig. 2. The traditional current-based R_{sd} extraction methods based on the hypothesis that the carrier mobility or the channel dopant concentration is independent of the channel length might be unsuitable for today's MOSFET technology because process variations may induce nonuniform channel dopant profiles. In order to extract R_{sd} accurately, the constant-mobility method proposed in [10] that provides immunity against process variation and needs not to know the carrier mobility and L_{eff} is adopted. As the L_{eff} and R_{sd} are extracted accordingly, the effective mobility eventually can be extracted precisely.

Chapter 2
Effective Channel Length Extraction

2.1 Definition and Meaning of Effective Channel Length

The effective channel length (L_{eff}) - defined by the inversion layer length is a measure of the "effective" current carrying capability of the MOSFET device [6], [11]. Fig. 3 shows schematically that L_{eff} is also defined as the distance between the source and drain diffusion extensions at the silicon surface. In addtion, L_{eff} is an important parameter for both process monitor and device design of MOSFET devices.

2.2 Classical Current-based L_{eff} Extraction Methods

2.2.1 Channel-Resistance Method

For MOSFET devices operated in linear region (low-drain bias region), the gate-controlled current is given by

$$\mathbf{I_d} = \boldsymbol{\mu}_{\mathbf{eff}} \times \mathbf{C_{ox}} \times \frac{\mathbf{W}}{\mathbf{L_{eff}}}\left(\mathbf{V_{gs}} - \mathbf{V_{th}} - \frac{1}{2}\mathbf{V_{ds}}\right)\mathbf{V_{ds}} \tag{2.1}$$

When biased in the linear region and slightly below threshold voltage region of MOSFET devices, the formula (2.1) can be approximated to the form

$$\mathbf{I_d} = \boldsymbol{\mu}_{\mathbf{eff}} \times \mathbf{C_{ox}} \times \frac{\mathbf{W}}{\mathbf{L_{eff}}}\left(\mathbf{V_{gs}} - \mathbf{V_{th}}\right)\mathbf{V_{ds}} \tag{2.2}$$

The channel resistance of a MOSFET device in the linear region is therefore proportional to the channel length as given by [12]

$$\mathbf{R_{ch}} = \frac{\mathbf{L_{eff}}}{\mu_{\mathbf{eff}} \times \mathbf{C_{ox}} \times \mathbf{W} \times (\mathbf{V_{gs}} - \mathbf{V_{th}})} \tag{2.3}$$

The total resistance of a MOSFET device reads

$$\mathbf{R_{total}} = \frac{\mathbf{V_{ds}}}{\mathbf{I_d}} = \mathbf{R_{ext}} + \mathbf{R_{ch}} \tag{2.4}$$

where R_{ext} includes the contact resistance, series resistance, and other resistances involved in measurement equipments.

After measuring the I_d-V_{gs} characteristics for MOSFET devices with different channel length, extracting the total resistances with different gate bias and then plotting the relation between total resistance and channel length under different gate biases are shown in Fig. 4. By an extrapolation from the intersection point, we can eventually get the channel length reduction ΔL on the intercept point on the x-axis. And the R_{ext} can be obtained at the intercept on the y-axis as in Fig. 4.

Although this L_{eff} extraction method looks easy, but it may fail for highly scaled MOSFET devices with some issues. First, this method assumes the effective mobility is independent of the channel length. This leads to unrealistic values for the extracted effective channel length in short-channel devices. Second, the literature [13] points out that the lines in Fig. 4 are difficult to intersect at a certain point, thus causing unprecise results.

2.2.2 Shift-and-Ratio Method

This method is based on the same channel resistance concept by assuming the effective mobility to be a common function of (V_{gs}-V_{th}) for MOSFET devices. The equations of total resistance of MOSFET devices are given by [14]

$$\mathbf{R^{o}_{tot}}(\mathbf{V_{gs}}) = \mathbf{R_{sd}} + \mathbf{L^{o}_{eff}}\, \boldsymbol{f}(\mathbf{V_{gs}} - \mathbf{V^{o}_{th}}) \tag{2.5}$$

$$\mathbf{R^{i}_{tot}}(\mathbf{V_{gs}}) = \mathbf{R_{sd}} + \mathbf{L^{i}_{eff}}\, \boldsymbol{f}(\mathbf{V_{gs}} - \mathbf{V^{i}_{th}}) \tag{2.6}$$

where superscript i represents the short-channel device and o refers to the long-channel device. Further differentiating equations (2.5) and (2.6) with respect to V_{gs} and neglecting the R_{sd} part under the assumption that R_{sd} is either independent or a weak function of V_{gs} so its derivation can be ignored, we obtain

$$\mathbf{S^{o}}(\mathbf{V_{gs}}) = \frac{d\mathbf{R^{o}_{tot}}}{d\mathbf{V_{gs}}} = \mathbf{L^{o}_{eff}} \frac{d\boldsymbol{f}(\mathbf{V_{gs}} - \mathbf{V^{o}_{th}})}{d\mathbf{V_{gs}}} \tag{2.7}$$

$$\mathbf{S^{i}}(\mathbf{V_{gs}}) = \frac{d\mathbf{R^{i}_{tot}}}{d\mathbf{V_{gs}}} = \mathbf{L^{i}_{eff}} \frac{d\boldsymbol{f}(\mathbf{V_{gs}} - \mathbf{V^{i}_{th}})}{d\mathbf{V_{gs}}} \tag{2.8}$$

The ratio of (2.7) to (2.8) is

$$\mathbf{r}(\mathbf{V_{gs}}) = \frac{\mathbf{S^{o}}(\mathbf{V_{gs}})}{\mathbf{S^{i}}(\mathbf{V_{gs}})} = \frac{\mathbf{L^{o}_{eff}} \dfrac{d\boldsymbol{f}(\mathbf{V_{gs}} - \mathbf{V^{o}_{th}})}{d\mathbf{V_{gs}}}}{\mathbf{L^{i}_{eff}} \dfrac{d\boldsymbol{f}(\mathbf{V_{gs}} - \mathbf{V^{i}_{th}})}{d\mathbf{V_{gs}}}} \tag{2.9}$$

with equation (2.9), the $S^{o}(V_{gs})$ and $S^{i}(V_{gs})$ become simliar functions if $V_t^{o}=V_t^{i}$ and L_{eff}^{i} could be easily extractd from the ratio $S^{o}(V_{gs})/ S^{i}(V_{gs}) = L_{eff}^{o} / L_{eff}^{i} \cong L/ L_{eff}^{i}$. However, V_t^{o} generally is not equal to V_t^{i}. By this reason, this method defines a shift parameter $\delta_{min} = V_t^{o} - V_t^{i}$ and the shifted ratio is redefined as

$$r\left(\delta_{min}, V_{gs}\right) \equiv \frac{S^{o}(V_{gs})}{S^{i}(V_{gs} - \delta_{min})} \tag{2.10}$$

After meeting the minimum standard direvation of $r(\delta, V_{gs})$ with $\delta = \delta_{min}$, L_{eff}^{i} can be extracted by

$$\langle r \rangle_{min} = \frac{L_{eff}^{o}}{L_{eff}^{i}} \cong \frac{L^{o}}{L_{eff}^{i}} \tag{2.11}$$

This method is questionable for highly scaled MOSFET devices, especially in the low-V_{gs} regime because the series resistance becomes a strong function of V_{gs}. Moreover, the assumption such as the invariance of the effective mobility with the channel length of the MOSFET devices tends to yield unreasonable high values of L_{eff} for highly scaled MOSFET devices [15]. In the citation [16], it is also pointed out that this method fails for MOSFET devices with halo/pocket implants.

2.3 Capacitive Method

This method introduced in [11] has two ways to extract L_{eff}：constant ΔL method and individual ΔL method. They all measure the gate-to-channel capacitance $C_{gc}(V_{gs})$ which is proportional to the effective channel area. The authors in [11] indicate that this $C_{gc}(V_{gs})$ measurement has no requirements for any de-embedding structure to get rid of parastic capacitance during $C_{gb}(V_{gs})$ measurements. We can only set the maximum gate-to-channel capacitance, $\max(C_{gc})$, as a reference point of each curve and start L_{eff} extraction with the following two ways.

2.3.1 Constant ΔL Method

This method assumes that the channel length reduction ΔL is independent of channel length L, which can be extracted by a linear regression on the $max(C_{gc})$ - channel length L plot. Then the value of ΔL can be created at the intercept between the regression linear and the L-axis. The ΔL is determined through the following expression

$$\mathbf{C_{gc} = W \times C_{ox} \times (L - \Delta L)} \tag{2.12}$$

It is easy to extract the ΔL with this method, but there would be an error on L_{eff} which is strongly linked to the relevance of the ΔL linearity assumption. Even though this issue exists in this method, however, it is just the choice when we do not have a long enough MOSFET device as reference.

2.3.2 Individual ΔL Method

This ΔL extraction using the longest MOSFET device as reference and extracting an individual ΔL for each MOSFET device from a proportionality relationship which can be written in the form

$$\mathbf{\Delta L^{*} = L^{*} - L_{eff}^{ref} \times \frac{max(C_{gc}^{*})}{max(C_{gc}^{ref})}} \tag{2.13}$$

The authors in [11] assert this method has high accuracy with long-enough MOSFET devicess as reference (error due to this assumption $L^{ref} \cong L_{eff}^{ref}$ being about 2 % for a 1μm long reference MOSFET device, and 0.2 % for a 10μm long one).

While the capacitive methods in [11] provide high accuracy and simple process with no assumption regarding the mobility, it might encounter difficulties to extract effective channel length with these methods on highly scaled MOSFET devices with ultra-thin oxide due to the huge gate leakage and the equipment detection limit (usually the limited minimum device area ≅ 0.2 μm^2).

2.4 Edge Direct Tunneling and Modeling Description

The EDT (edge direct tunneling) method in this thesis is based on the triangular potential approximation. The detailed triangular potential profile for nMOSFET devices is schematically shown in Fig. 5. When MOSFET devices are operated under accumulation or inversion conditions, corresponding band bending at n^+ poly-gate surface or silicon-substrate surface can be characterized by a triangle-like electrostatic potential well. Replacing the electrostatic potential in Shrödinger wave equation with the triangular potential leads to the Airy equation with solutions, which is called the triangular potential approximation method. We can derive the quantized energy of the first subband directly with this approximation [9], [17]

$$\mathbf{E_1} = \left(\frac{\hbar^2}{2\mathbf{m_z}}\right)^{1/3} \left(\frac{9\pi q \varepsilon_{ox}}{8\varepsilon_{Si}} F_{ox}\right)^{2/3} \tag{2.14}$$

The voltage balance relationship for an nMOSFET device operated under accumulation conditions as illustrated in Fig. 6, can be expressed as

$$\mathbf{V_{dg}} - \mathbf{V_{FB}}(\cong \mathbf{0}) = \mathbf{V}_{ox} + \mathbf{V_{poly}} + \mathbf{V_{DE}} \tag{2.15}$$

where V_{FB} is the flat-band voltage, $V_{ox} = t_{ox}F_{ox}$ is the potential drop across the oxide, V_{poly} is polysilicon depletion potential drop, and V_{DE} is the band bending in the drain extention region. The accumulated charge Q available for the tunnel process mainly populates in the first subband E_1 due to the lowest quantized energy dominating. This sheet charge density Q here is modeled as field induced and related to the number of occupied subband states, which can build up the charge conservation relationship [9]

$$\mathbf{Q} = \varepsilon_{\mathbf{ox}} F_{ox} = q\left(\mathbf{E_{fn}} - \mathbf{E}_1\right)\frac{\eta \mathbf{m_d}}{\pi \hbar^2} \tag{2.16}$$

where E_{fn} is the quasi-Fermi level in the n^+-poly gate and η is the degeneracy factor. By applying the lowest subband approximation to the accumulated n^+-poly gate and the depletion approximation to the drain diffusion extention, we get [9]

$$\mathbf{V_{poly}} \cong \frac{\mathbf{E_{fn}}}{q} = \varepsilon_{\mathbf{ox}} F_{ox} \frac{\pi \hbar^2}{q^2 \eta \mathbf{m_d}} + \frac{\mathbf{E}_1}{q} \tag{2.17}$$

$$\mathbf{V_{DE}} = \frac{\varepsilon_{\mathbf{ox}}^2 F_{ox}^2}{2q\, \varepsilon_{\mathbf{Si}} \mathbf{N_{DE}}} \tag{2.18}$$

where N_{DE} is the dopant concentration of the drain diffusion extention. For <100> oriented n^+-polysilicon grains, m_z=0.98m_o, m_d=0.19m_o, and η=2 were adopted to approximate the band structure [2]. The edge tunneling current here is modeled by incorporating the accumulation layer quantization effect with a finite potential barrier height as the boundary condition and choosing a modified Wentzel-Kramers-Brillouin (WKB) method [2] to calculate the electron tunneling probability. Then within the effective mass approximation, the

tunneling probability can be expressed as

$$\mathbf{T} = \mathbf{T_{WKB}} \times \mathbf{T_R} \tag{2.19}$$

where T_{WKB} is the usual WKB approximation of the tunneling probability for slowly varying potentials, and T_R is the correction factor taking into account the reflections from potential discontinuities. The SiO_2 band-gap dispersion relation is modeled as the Franz-type E-k dispersion relation [18]

$$\frac{\hbar^2 \kappa_{ox}^2}{2m_{ox}} = \gamma = E_{ox}\left(1 - \frac{E_{ox}}{E_g}\right) \tag{2.20}$$

$$v_{ox} = \frac{1}{\gamma'}\sqrt{\frac{2\gamma}{m_{ox}}} \tag{2.21}$$

Here k_{ox} and v_{ox} are the purely imaginary wave vector and group velocity of electrons within the oxide gap energy, respectively. E_{ox} is the magnitude of the electron energy with respect to the oxide conduction band edge, $m_{ox} = 0.61m_o$ is the effective mass in the oxide based on the Franz-type dispersion relationship, and $E_g \cong 9$ eV is the band gap for SiO_2. With these conditions shown in Fig. 7, T_{WKB} is given by [2]

$$\begin{aligned} T_{WKB} &= \exp\left(-2\int_0^{t_{ox}} k_{ox}\,dx\right) \\ &= \exp\left[\frac{E_g\sqrt{2m_{ox}}}{4\hbar q F_{ox}}\left(2\gamma'\sqrt{\gamma} + \sqrt{E_g}\sin^{-1}\gamma'\right)\Bigg|_{E_{ox}=q\phi_{an}}^{E_{ox}=q\phi_{cat}}\right] \end{aligned} \tag{2.22}$$

where $q\phi_{cat}$ and $q\phi_{an}$ are the net barrier heights for the electrons at the cathode

and anode interfaces, respectively. They are expressed as

$$q\phi_{\mathrm{cat}} = q\chi_{\mathrm{c}} - (\mathrm{E}_1 + \mathrm{E}_2) \tag{2.23}$$

$$q\phi_{\mathrm{an}} = q\chi_{\mathrm{c}} - (\mathrm{E}_1 + \mathrm{E}_2) - q\mathrm{t}_{\mathrm{ox}}F_{ox} \tag{2.24}$$

where $q\chi_c \cong 3.15$eV is the discontinuity between the silicon and the SiO_2 conduction bands [2],[3]. The correction factor T_R is obtained by considering reflections from the material interfaces as

$$\mathrm{T_R} = \frac{4\upsilon_{\mathrm{Si},\perp}(\mathrm{E}_1)\nu_{\mathrm{ox}}(q\phi_{\mathrm{cat}})}{\upsilon^2_{\mathrm{Si},\perp}(\mathrm{E}_1) + \nu^2_{\mathrm{ox}}(q\phi_{\mathrm{cat}})} \times \frac{4\upsilon_{\mathrm{Si},\perp}(\mathrm{E}_1 + qF_{ox}\mathrm{t}_{\mathrm{ox}})\nu_{\mathrm{ox}}(q\phi_{\mathrm{an}})}{\upsilon^2_{\mathrm{Si},\perp}(\mathrm{E}_1 + qF_{ox}\mathrm{t}_{\mathrm{ox}}) + \nu^2_{\mathrm{ox}}(q\phi_{\mathrm{an}})} \tag{2.25}$$

where $v_{\mathrm{Si},\perp}(\mathrm{E}_1)$ and $v_{\mathrm{Si},\perp}(\mathrm{E}_1+q\mathrm{V}_{\mathrm{ox}})$ are the interface-normal component group velocities of the electrons incident onto and leaving from the oxide, respectively. With the oxide potential drop V_{ox}, $v_{\mathrm{ox}}(\phi_{\mathrm{cat}})$ and $v_{\mathrm{ox}}(\phi_{\mathrm{an}})$ are the imaginary group velocities of the electrons at the cathode and anode side within the oxide, respectively. The tunneling lifetime of the electrons from the lowest subband can be connected with the tunneling probability, as defined below [3]

$$\tau_1 = \frac{\pi\hbar}{\mathrm{T}_1 \times \mathrm{E}_1} \tag{2.26}$$

With the established relation between the accumulated charge and the tunneling lifetime to oxide field, the edge direct tunneling current can be calculated analytically with the estimated effective edge-tunneling area ($L_{TN} \times W$)：

$$\mathbf{I_{EDT}} = \frac{\mathbf{Q}}{\boldsymbol{\tau}_1} \times (\mathbf{L_{TN}} \times \mathbf{W}) \tag{2.27}$$

where L_{TN} is the estimated value of the drain diffusion extension length and W is the channel width of the MOSFET device. In conclusion, the process flow for this EDT model operation is drawn in Fig. 8. With I_{EDT} fitting to the experimental data, L_{TN} can be evaluated, leading to quantified ΔL：

$$\boldsymbol{\Delta}\mathbf{L} = \mathbf{2L_{TN}} \tag{2.28}$$

Like the capacitive method, this EDT method also has no assumption regarding the mobility. It might be feasible to extract L_{eff} for highly scaled MOSFET devices with the EDT method becauce the EDT method has no apparent equipment detection limit which is the major problem for the capacitive method. In addition, the EDT method was involved with the huge leakage current which is the universal phenomenon for highly scaled MOSFET devices with ultra-thin oxide.

It must be noted that the above-mentioned EDT model is for nMOSFETs because the major objects examined in this thesis are nMOSFETs. However, the framework of the EDT model for pMOSFETs is similar to that for nMOSFETs except for the complicated effective mass of hole which is the major carrier of pMOSFETs and the energy band structure [24].

Chapter 3
Key Parameters Extraction

3.1 Threshold Voltage Extraction

Before extracting the series resistance and effective mobility, the first job is to extract the threshold voltage (V_{th}). For MOSFET devices, the V_{th} is a fundamental parameter which represents the onset of the noticeable drain current flow, and is also recognized as the critical gate voltage at which the transition between weak and strong inversion arises in the MOSFET channel [19]. In correct extraction of V_{th} leads to significant errors in extracted effective mobility as mentioned in [20], and therefore, accurate V_{th} extraction is necessary.

Many V_{th} extraction methods have been developed as introduced in [19]. Those extraction methods are mostly done with a low drain voltage so that the devices operate in the linear region. The extrapolation line in the linear region (ELR) method is the most popular V_{th} method. It is the common practice to find the V_{th} as the gate voltage axis intercept (i.e., $I_d = 0$) of the linear extrapolation of the I_d-V_{gs} characteristics at the maximum transconductance. One of the V_{th} extraction methods developed to avoid the dependence on the series resistances is second-derivative (SD) method. It determines the V_{th} from the peak of $dg_m/dV_{gs} = d^2I_d/dV_{gs}^2$ (the derivative of the transconductances).

3.2 Series Resistance Extraction

When the V_{th} of the MOSFET device is extraced, the next task is to extract the series resistance R_{sd}. The R_{sd} extraction method used here is based on the constant-mobility bias conditions as recently introduced in [10]. Regarding the mobility, it is convenient to express the effective mobility μ_{eff} in terms of either the inversion layer charge or the effective surface vertical field E_{eff} for the basic

surface studies [20]. The typical relationship between the measured μ_{eff} and E_{eff} at the Si/SiO2 interface under different body-voltage (V_{bs}) conditions is schematically shown in Fig. 9. The behavior of μ_{eff} can be elucidated with the E_{eff} expression given by [10]

$$\boldsymbol{E}_{\mathbf{eff}} = \frac{1}{\varepsilon_{si}}\left(\left|\mathbf{Q}_{\mathbf{dep}}\right| + \frac{1}{\eta}\left|\mathbf{Q}_{\mathbf{inv}}\right|\right) \quad (3.1)$$

where ε_{Si} is the silicon permittivity, η is an empirical factor with the value ≅ 2 generally used for electron carriers at room temperature, Q_{inv} is the inversion-layer charge, and Q_{dep} is the depletion-layer charge. The different V_{bs} conditions resulting in Q_{dep} variance can be interpreted with the following formula

$$\mathbf{Q}_{\mathbf{dep}} = \sqrt{2q\varepsilon_{si}\mathbf{N}_{\mathbf{sub}}[\phi_S(\mathbf{V}_g) - \mathbf{V}_{\mathbf{bs}}]} \quad (3.2)$$

where N_{sub} is the doping concentration of silicon substrate estimated by C-V curves fitting along with Poisson-Schrödinger self-consistent simulations, and ϕ_s is the substrate band bending. In low E_{eff} (low-V_g) regime, the quantity of the Q_{dep} is compared with the Q_{inv}, leading to the obvious fluctuations of E_{eff} under different V_{bs} conditions. On the contrary, the variances of the Q_{dep} under different V_{bs} conditions cannot be compared with the Q_{inv} component in high E_{eff} region because the Q_{inv} is much larger than the Q_{dep} in amount. Hence, the μ_{eff} under different V_{bs} conditions has the behaviour that disperses in low E_{eff} region and converges toward one universal curve when E_{eff} is sufficiently high. The other expression of E_{eff} can be described as [10]

$$E_{eff} = \frac{V_{gs} + (\eta - 1)V_{th} - \eta V_{FB} - 2\eta\phi_B}{3\eta t_{ox}} \quad (3.3)$$

where V_{FB} is the flatband voltage, ϕ_B is the potential difference between the Fermi level and the intrinsic Fermi level, and V_{th} is the threshold extracted with I_d- V_{gs} data via g_m-method. For a single device, both V_{FB} and ϕ_B essentially remain intact under different bias conditions. Thus, the formula (3.3) implies that a constant E_{eff} can be preserved from $V_{gs}^{(1)}$ and $V_{th}^{(1)}$ to other biases $V_{gs}^{(2)}$ and $V_{th}^{(2)}$ under different V_{bs} conditions. In other words, we can keep the fixed E_{eff} by adjusting the V_{gs} and V_{th} simultaneously which satisfy the following expression [10]

$$V_{gs}^{(2)} = V_{gs}^{(1)} + (1 - \eta)(V_{th}^{(1)} - V_{th}^{(2)}) \quad (3.4)$$

Eventually, the constant-mobility bias conditions are established on the basis of the above-mentioned theory.

By integrating the constant-mobility criterion into the I_d equation of MOSFET devices operating in the linear region, the results for the two specific bias conditions (individually labeled with the superscript 1 and 2) are

$$I_d^{(1)} = \frac{\mu^{(1)} W_{eff} C_{ox}}{L_{eff}} \left(V_{gs}^{(1)} - V_{th}^{(1)} - \frac{1}{2} V_{ds} \right) \left(V_{ds} - R_{sd} I_d^{(1)} \right) \quad (3.5)$$

$$I_d^{(2)} = \frac{\mu^{(2)} W_{eff} C_{ox}}{L_{eff}} \left(V_{gs}^{(2)} - V_{th}^{(2)} - \frac{1}{2} V_{ds} \right) \left(V_{ds} - R_{sd} I_d^{(2)} \right) \quad (3.6)$$

The R_{sd} expression can be derived by dividing (3.5) by (3.6) under the constant-mobility conditions, i.e. $\mu^{(1)} = \mu^{(2)}$ under a high E_{eff}：

$$\mathbf{R_{sd}} = \left(\frac{\mathbf{B}}{\mathbf{I_d^{(2)}}} - \frac{\mathbf{A}}{\mathbf{I_d^{(1)}}}\right) \frac{\mathbf{V_{ds}}}{\eta(\mathbf{V_{th}^{(1)}} - \mathbf{V_{th}^{(2)}})} \tag{3.7}$$

$$\mathbf{A} = \mathbf{V_{gs}^{(1)}} - \mathbf{V_{th}^{(1)}} - \mathbf{0.5V_{ds}} \tag{3.8}$$

$$\mathbf{B} = \mathbf{V_{gs}^{(1)}} + (\eta - 1)\mathbf{V_{th}^{(1)}} - \eta\mathbf{V_{th}^{(2)}} - \mathbf{0.5V_{ds}} \tag{3.9}$$

After the treatment the effective channel length L_{eff}, the effective channel width W_{eff}, and the inversion gate-oxide capacitance C_{ox} are canceled out because they are identical for a single device. This distinctive property makes this R_{sd} extraction method immune to the process variation for highly scaled MOSFET devices where the explicit definitions of L_{eff}, W_{eff}, and C_{ox} are difficult to procure.

3.3 Effective Mobility Extraction

The effective mobility in MOSFET devices is a key parameter to describe the carrier transport and a also probe to study the electric properties of a two-dimensional carrier system. In order to extract the μ_{eff}, one of the most powerful methods is the combination of split capacitance-voltage (C-V) measurments, which are to deal with the used inversion and depletion charges, and linear I_d-V_{gs} measurements [6]. Fig. 10 schematically shows the equivalent circuit of a MOSFET device. If we take the R_{sd} into account in (2.1), the drain and source voltages are rewritten [21]

$$\mathbf{V_{ds}} - \mathbf{I_d R_{sd}} = \mathbf{V_{ds}}' \tag{3.10}$$

$$\mathbf{V_{gs} - I_d R_s \cong V_{gs} - \frac{1}{2} I_d R_{sd} = V_{gs}'} \tag{3.11}$$

where $R_s \cong 0.5R_{sd}$ because the depletion due to V_{ds} can be negligible when the device is operated in the linear region (V_{gs}-V_{th} >> V_{ds}). The series resistances can be considered nearly the same between the drain and source. With the formulas (3.10) and (3.11), the $C_{ox}(V_{gs}-V_{th}-0.5V_{ds})$ factor in (2.1) is kept constant [21]：

$$\mathbf{V_{gs}' - V_{th} - \frac{1}{2} V_{ds}' = V_{gs} - V_{th} - \frac{1}{2} V_{ds}} \tag{3.12}$$

The inversion charge Q_{inv} under strong inversion condition, which is used in evaluating μ_{eff} and E_{eff}, can be approximately expressed [22]：

$$\mathbf{C_{ox} \times \left(V_{gs} - V_{th} - \frac{1}{2} V_{ds} \right)}$$

$$\mathbf{\cong C_{ox} \times (V_{gs} - V_{th}) \cong Q_{inv}(V_{gs})} \tag{3.13}$$

The approximation made in (3.13) is frequently employed to extract the mobility, but it leads to large errors in extracting mobility around the V_{th}. There are two essential approahes to improve the accuracy of (3.13) in extracting μ_{eff} [20]. One is to measure the gate-to-channel capacitance C_{gc} and obtain the Q_{inv} from the voltage integral of the C_{gc} [4], [6], [23]. The second approach is to perform a split C-V measurement to get C-V characteristics, then comparing the experimental and the simulated C-V characteristics (via Poisson-Schrödinger

self-consistent simulations) to obtain the Q_{inv}.

Futhermore, it must be noted that the V_{th} through calculated Q_{inv} may not be exactly equal with that of $I_d(V_{gs})$, especially for different devices. For this reason, we can extract the μ_{eff} more accurately by shifting Q_{inv} characteristics such as to compensate for the V_{th} discrepancy [4], [20]. Finally, the expression of μ_{eff} can be deduced from (2.1), (3.10), and the obtained Q_{inv}：

$$\mu_{eff} = \frac{L_{eff}}{W}\left(\frac{I_d(V_{gs})}{V_{ds}'}\right)\left(\frac{1}{Q_{inv}(V_g)}\right) \tag{3.14}$$

Chapter 4
Experimental Data and Interpretations

The halo-implanted bulk n-channel MOSFET devices with the gate width of 10μm, 1μm, 0.6μm, and 0.24μm under investigation were fabricated in a state-of-the-art manufacturing process. The major parameter extractions are focused on the nMOSFET devices with the channel length from 0.05μm to 0.1μm. Fig. 11 schematically shows the flowchart summarizing the procedures of μ_{eff} extraction. In the flowchart diagram, the bold, solid line, and dashed dots blocks indicate the expreimental data, extracted parameters, and simulated results during the measurements, respectively.

The objective for the C-V measurement is extracting the oxide thickness t_{ox}, the dopant concentration of poly gate N_{poly} and substrate N_{sub}, and the effective channel length L_{eff} of the devices. The C-V characteristics of an nMOSFET device are measured by means of the HP4284A LCR meter, followed by parameter extraction by comparing the measured and simulated C-V characteristics of an nMOSFET device. With the extraction results shown in Fig. 12, we get t_{ox} = 1.215 nm, $N_{poly} = 4\times10^{19}$ cm^{-3}, and $N_{sub} = 4\times10^{17}$ cm^{-3}.

I-V measurement includes I_d-V_{gs} and I_g-V_{gs} characterizations for an nMOSFET device. The I_d-V_{gs} and I_g-V_{gs} experimental data are obtained by HP4156B semiconductor parameter analyzer, which are adopted for the extractions of series resistance R_{sd}, threshold voltage V_{th}, oxide thickness t_{ox}, effective channel length L_{eff}, and the dopant concentration of the drain extension N_{DE}.

After performing the I-V measurement, the first task is to extract the V_{th} for nMOSFET devices. The comparision between the V_{th} extracted by extrapolation line in the linear region (ELR) method and that by second-derivative (SD)

method is shown in Fig. 13, which indicates more obvious difference of V_{th} values with shorter channel devices. Because the R_{sd} extraction via the V_{th} values in ELR method shows more consistent results than that with the V_{th} values extracted by SD method generally, the subsequently extraction procedures of R_{sd} and μ_{eff} would use the V_{th} values extracted by ELR method.

Fig. 14 and 15 present the L_{eff} extraction results for the W/L = 1μm/0.1 μm nMOSFET device by both the channel-resistance method and the shift-and-ratio (S&R) method. The channel-resistance method fails because there is no explicit intersect point of the R_{tot}-L characteristics for different gate voltages, and the L_{eff} extraction results from the S&R method appear to be overestimated. Thus, we continue to try L_{eff} extraction with the capacitive method. The experimental $C_{gc}(V_{gs})$ curves for several nMOSFET devices with a largest width plotted in a logarithm scale are plotted in Fig. 16. The L_{eff} extracted by the constant ΔL method and the individual ΔL method are displayed in Fig. 17 and 18, respectively. The corresponding L_{eff} values for L = 1um are 61.57nm and 75.45nm. Furthermore, the EDT method is adopted for narrower and shorter devices under which the capacitive method can not work well. Fig. 19 demonstrates the example of EDT method for the W/L = 1μm/0.1 μm nMOSFET device, which leads to L_{eff} = 72nm by comparing the measured (triangular symbol) I_g-V_{gs} characteristics with the simulated (solid curve) under accumulation conditions. Subsequently, the statistical analysis of the relations between ΔL and gate length and width is shown in Fig. 20 and 21, respectively. The figures reveal that the EDT method can preserve considerable accuracy for the narrower and shorter devices. Fig. 22 exhibits the L_{eff} behavior with different gate lengths. It must be noted that the difference between L_{eff} and L becomes larger with shorter channel length.

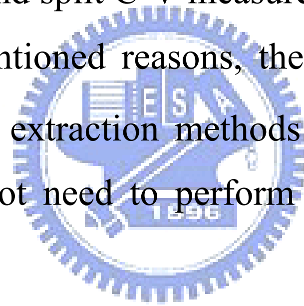
After finishing the L_{eff} extraction, we proceed with the R_{sd} extraction by the constant-mobility method. As demonstrated in Fig. 23, the R_{sd} value for the W/L

= 1μm/0.1 μm nMOSFET device is extracted by estimating the value with the universal curve at high E_{eff}. Fig. 24, 25, and 26 show that the R_{sd} values apparently reduce with shorter devices, and have no visible dependence on the width and overlap length.

When the L_{eff} and R_{sd} are extracted further, we can accurately carry out the μ_{eff} extraction for highly scaled nMOSFET devices accordingly. Fig. 27 displays the resulting μ_{eff} curves of the W/L = 1μm/0.1μm nMOSFET device with raw data also shown are corrected L_{eff} , R_{sd}, and both L_{eff} & R_{sd} corrected. The variations of the peak mobility induced by overlap length for nMOSFET devices with different length and width are shown in Fig. 28 and 29, respectively. They illustrate that the mobility variations caused by overlap length are bigger with short devices but seem to be a weak function of the width. The variations of the mobility induced by R_{sd} at V_{gs} = 1.5V (higher E_{eff}) for nMOSFET devices with different length and width are also illustrated in Fig. 30 and 31. As can be clearly seen, the variations of the mobility result from the R_{sd} seemingly decrease with shorter length because the R_{sd} values reduce slightly with decreasing gate length. In addition, the R_{sd} induced mobility variations are connected with ΔL factor, rather than the width. At the end, the comparision between the raw and the corrected μ_{eff} values at V_{gs} = 1.5V are exhibited in Fig. 32 and 33 for different lengths and widths. The corrected $\boldsymbol{\mu}_{eff}$ value for L = 0.1 μm nMOSFET device arises because the deviation caused by R_{sd} is worse than that of ΔL. However, with device length scaling, the deviation caused by ΔL is larger than that by R_{sd} for highly scaled devices. From Fig. 32 and 33, the $\boldsymbol{\mu}_{eff}$ deviations tend to increase with reduced device length; nevertheless, it appears that the deviations are independent of the device width.

Chapter 5
Conclusion

The novel key parameter extractions for highly scaled MOSFET devices have been systematically executed. First, the R_{sd} extraction with constant-mobility method neither considers the mobility dependence on channel length nor requires the precise values of mobility and channel length. Moreover, it provides immunity against process variation. Second, the problems of the L_{eff} extraction with capacitive method such as the gate leakage issue and the equipment detection limit, have been solved with the EDT method. Furthermore, for more accurate extraction of process parameters such as N_{sub}, N_{poly}, N_{DE}, Q_{inv}, and t_{ox}, the EDT method and split C-V measurement can complement each other. Because of the above-mentioned reasons, the μ_{eff} eventually can be extracted accurately. Moreover, the extraction methods also furnish the convenient and fast approaches that do not need to perform measurements in a large device sample size.



References

[1] S.-H. Lo, D. A. Buchanan, Y. Taur, and W. Wang, “Quantum-Mechanical Modeling of Electron Tunneling Current from the Inversion Layer of Ultra-Thin-Oxide nMOSFET's,” *IEEE Electron Device Letters*, vol. 18, no. 5, pp. 209-211, May 1997.

[2] L. F. Register, E. Rosenbaum, and K. Yang, “Analytic model for direct tunneling current in polycrystalline silicon-gate metal–oxide–semiconductor devices,” *Applied Physics Letters*, vol. 74, no. 3, pp. 457-459, January 1999.

[3] N.Yang,W.K. Henson, J.R.Hauser, and J.J.Wortman, “Modeling Study of Ultrathin Gate Oxides Direct Tunneling Current and Capacitance Measurements in MOS Devices,” *IEEE Transactions on Electron Devices*, vol. 46, no. 7, pp. 1464-1471, July 1999.

[4] F. Lime, C. Guiducci, R. Clerc, G. Ghibaudo, C. Leroux, and T. Ernst, “Characterization of effective mobility by split C(V) technique in N-MOSFETs with ultra-thin gate oxides,” *Solid State Electronics*, vol. 41, pp. 1147-1153 (2003).

[5] Y. T. Hou, M. F. Li, Y. Jin, and W. H. Lai, “Direct tunneling hole currents through ultrathin gate oxides in metal-oxide-semiconductor devices,” *Journal of Applied Physics*, vol. 91, no. 11, pp. 258-264,January 2002.

[6] S. Severi, L. Pantisano, E. Augendre, E. S. Andrés, P. Eyben, and K. D. Meyer, “A Reliable Metric for Mobility Extraction of Short-Channel MOSFETs,” *IEEE Transactions on Electron Devices*, vol. 54, no. 10, pp. 2690-2698, October 2007.

[7] X. Liu, J. Kang, and R. Han, “Direct tunneling current model for MOS devices with ultra-thin gate oxide including quantization effect and polysilicon depletion effect,” *Solid State Communications*, vol. 125, pp. 219-223 (2003)

[8] J. Suiik, P. Olivo, and B. Riccb, “Quantum-Mechanical Modeling of Accumulation Layers in MOS Structure,” *IEEE Transactions on Electron Devices*, vol. 39, no. 7, pp. 1732-1739, July 1992.

[9] K. N. Yang, H. T. Huang, M. J. Chen, Y. M. Lin, M. C. Yu, S. M. Jang, Douglas C. H. Yu, and M. S. Liang, "Characterization and Modeling of Edge Direct tunneling (EDT) Leakage in Ultrathin Gate Oxide MOSFETs," *IEEE Transactions on Electron Devices*, vol. 48, no. 6, pp. 1159-1164,June 2001.

[10] D. W. Lin, M. L. Cheng, S. W. Wang, C. C. Wu, and M. J. Chen, "A Constant-Mobility Method to Enable MOSFET Series-Resistance Extraction," *IEEE Electron Device Letters*, vol. 28, no. 12, pp. 1132-1134, December 2007.

[11] D. Fleury, A. Cros, K. Romanjekl, D. Roy, F. Perriert, B. Dumontt, and H. Brut, "Automatic extraction methodology for accurate measurement of effective channel length on 65nm MOSFET technology and below," *IEEE International Conference on Microelectronic Test Structures*, Tokyo, Japan, pp. 89-92, March 2007.

[12] J. G. J. Chern, P. Chang, R. F. Motta, and N. Godinho, "A New Method To Determine MOSFET Channel Length," *IEEE Electron Device Letters*, vol. EDL-1, no. 9, pp. 170-173, September 1980.

[13] G. Niu, J. D. Cressler, S. J. Mathew, and S. Subbanna, "A Channel Resistance Derivative Method for Effective Channel Length Extraction in LDD MOSFET's," *IEEE Transactions on Electron Devices*, vol. 47, no. 3, pp. 648-650,March 2000.

[14] Yuan Taur, "MOSFET Channel Length：Extraction and Interpretation," *IEEE Transactions on Electron Devices*, vol. 47, no. 1, pp. 160-170, March 2000.

[15] M. Fathipour, E. Fathi, B. Afzal, and A. Khakifirooz, "An improved shift-and-ratio L_{eff} extraction method for MOS transistors with halo/pocket implants," *Solid State Electronics*, vol. 48, pp. 1829-1832 (2004).

[16] C. W. Eng, W. S. Lau, Y. Y. Jiang, D. Vigar, K. C. Tee, L. Chan, V. S. W. Lim, and A. Trigg, "Improving the Accuracy of Modified Shift-and-Ratio Channel Length Extraction Method Using Scanning Capacitance Microscopy," *Japanese Journal of Applied Physics*,

vol. 43, no. 4B, pp. 1869-1872, 2004.

[17] Frank Stern, “Self-Consistent Results for n-Type Si Inversion Layers” *Physical Review B*, vol. 5, no. 12, pp. 4891-4899, June 1972.

[18] J. Maserjian and G. P. Petersson, “Tunneling through thin MOS structure：Depedence on energy (E-k),” *Applied Physics Letters*, vol. 25, no. 1, pp. 50-52, July 1974.

[19] A. O. Conde, F. J. G. Sanchez, J. J. Liou, A. Cerdeira, M. Estrada, and Y. Yue, "A review of recent MOSFET threshold voltage extraction methods," *Microelectronics Reliability* 42 (2002) pp.583–596.

[20] J. R. Hauser, "Extraction of Experimental Mobility Data for MOS Llevices," *IEEE Transactions on Electron Devices*, vol. 43, no. 11, pp. 1981-1988, November 1996.

[21] J.C. Guo, S. S. Chung, and C. C. Hsu, "A New Approach to Determine the Effective Channel Length and the Drain-and-Source Series Resistance of Miniaturized MOSFET's," *IEEE Transactions on Electron Devices*, vol. 41, no. 10, pp. 1811-1818,October 1994.

[22] K. Chen, H. C. Warm, J. Dunster, P. K. Ko, C. Hu, and M. Yoshida, "MOSFET Carrier Mobility Model Based on Gate Oxide Thickness, Threshold and Gate Voltages," *Solid-State Electronics*, vol. 39, no. 10, pp. 1515-1518, 1996.

[23] K. Romanjek, F. Andrieu, T. Ernst, and G. Ghibaudo, "Improved Split C–V Method for Effective Mobility Extraction in sub - 0.1μm Si MOSFETs," *IEEE Electron Device Letters*, vol. 25, no. 8, pp. 583-585,August 2004.

[24] K. N. Yang, H. T. Huang, M. J. Chen, Y. M. Lin, M. C. Yu, S. S. M. Jang, D. C. H. Yu, and M. S. Liang,"Edge Hole Direct Tunneling Leakage in Ultrathin," *IEEE Transactions on Electron Devices*, vol. 48, no. 12, pp. 2790-2795, December 2001.

ES 1896

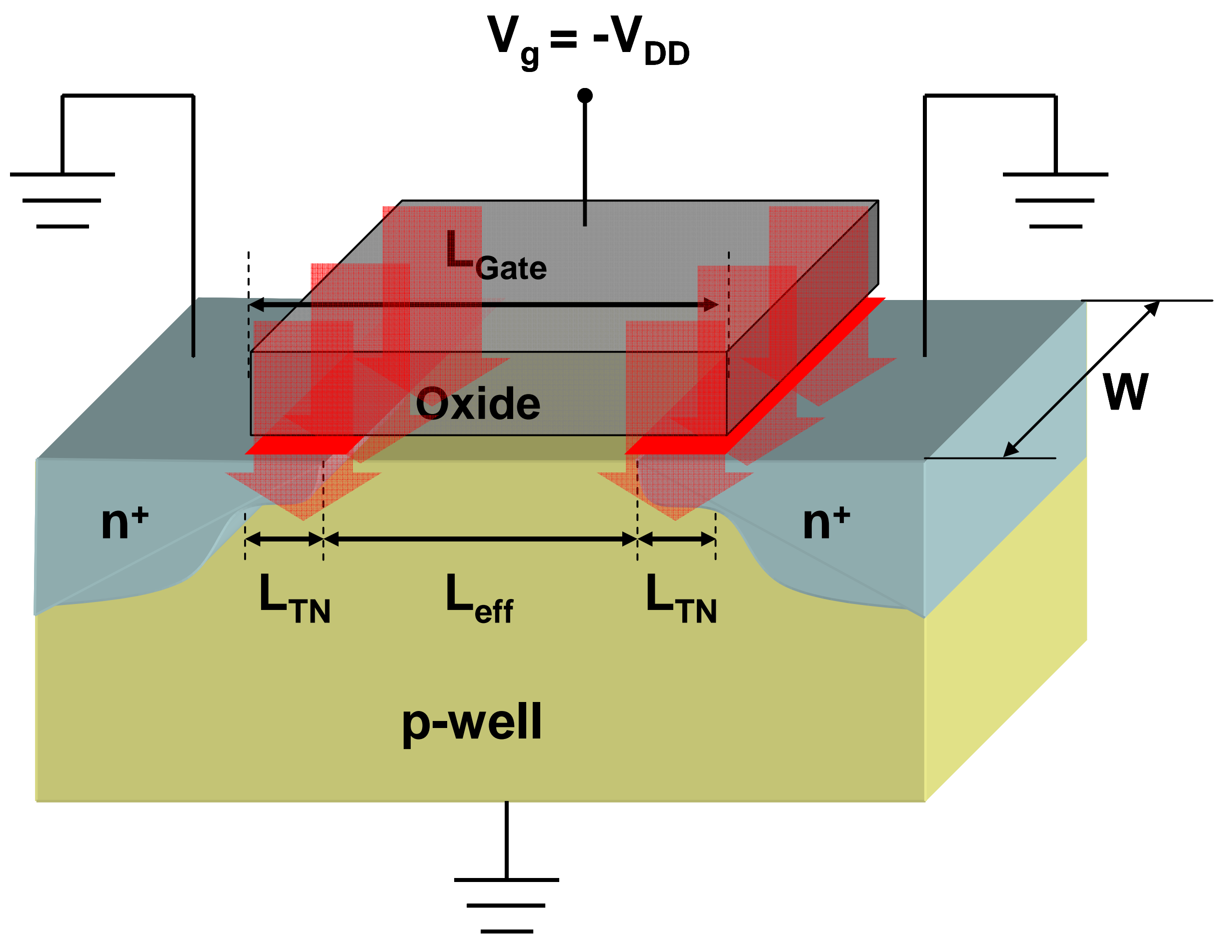


Fig. 1

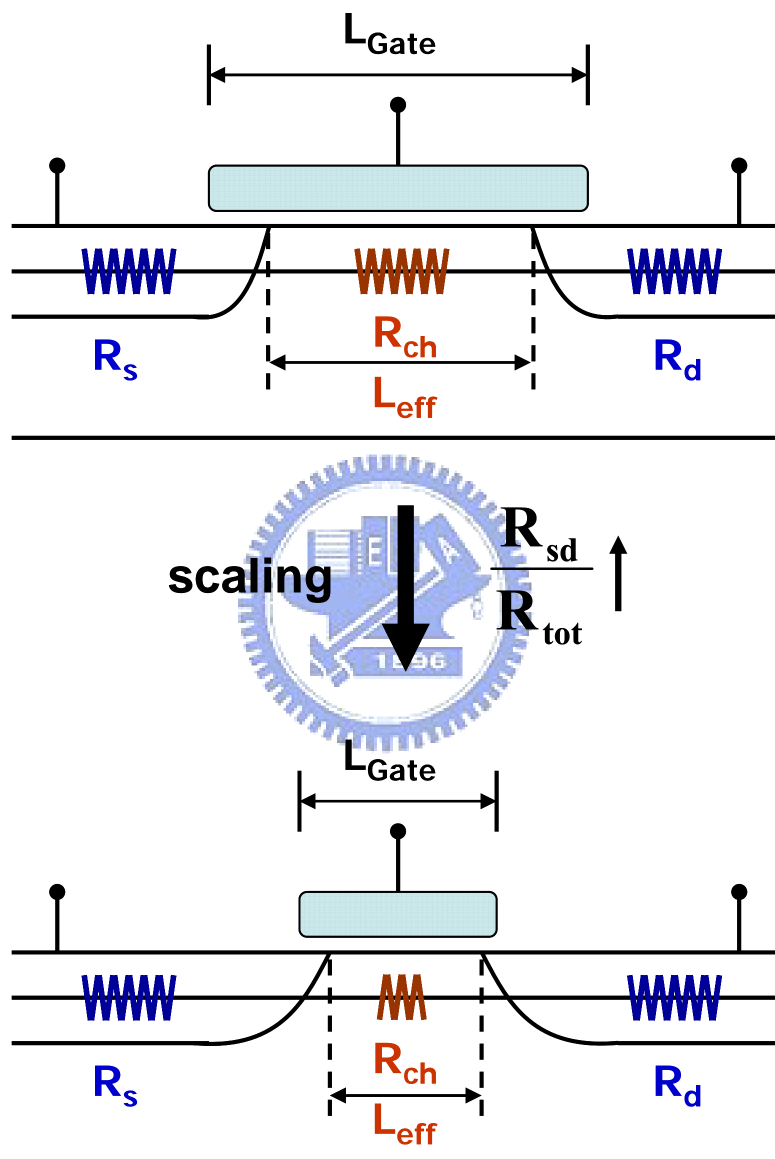


Fig. 2

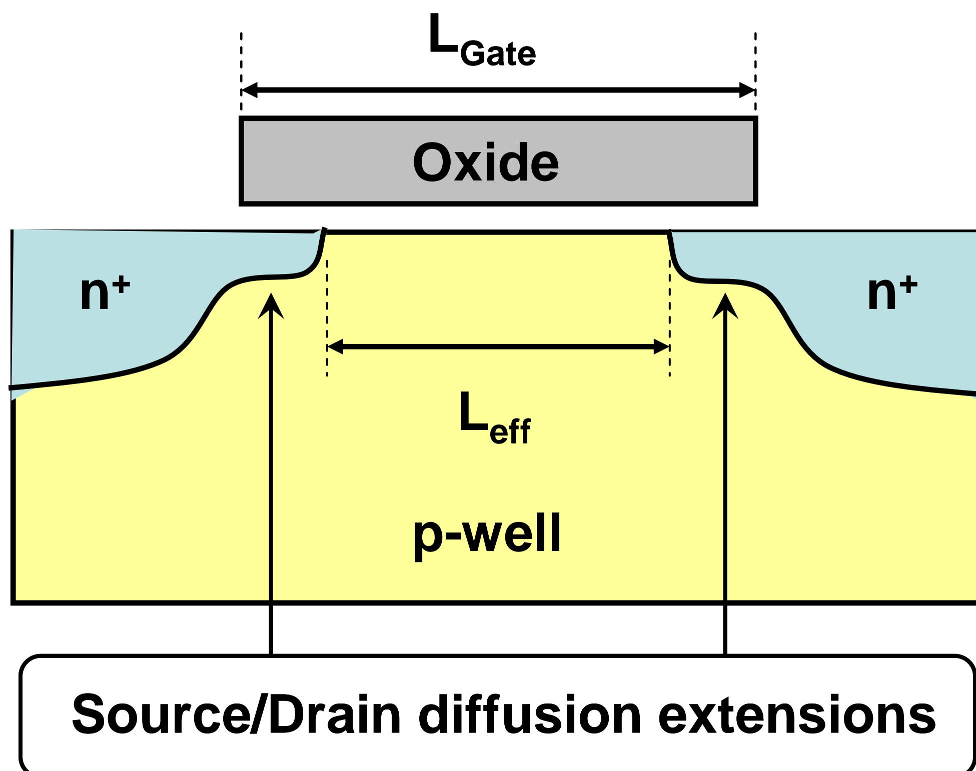


Fig. 3

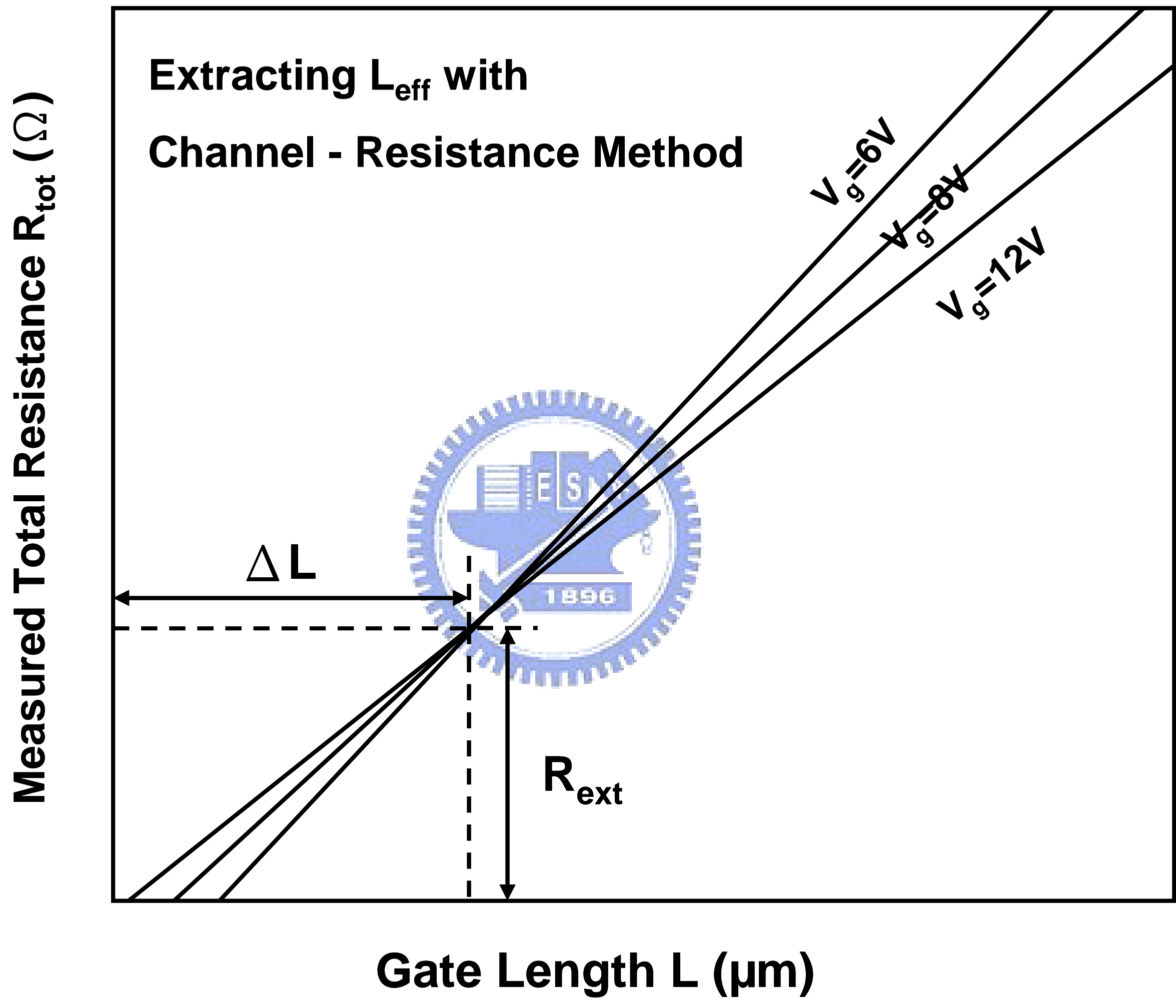


Fig. 4

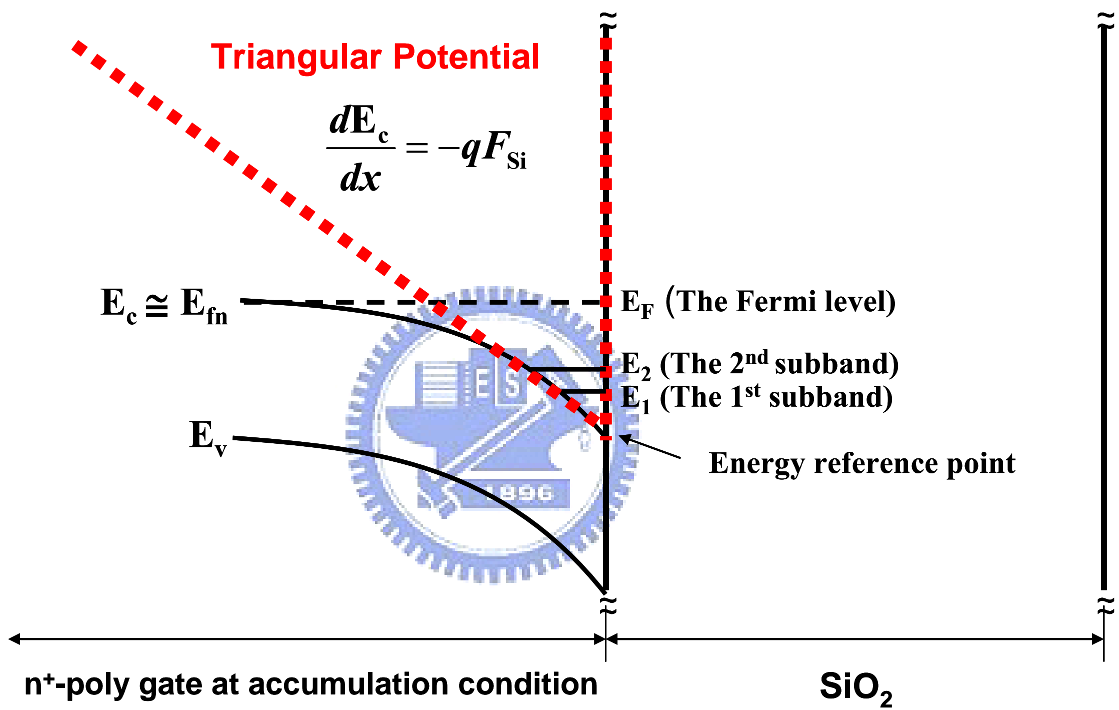


Fig. 5

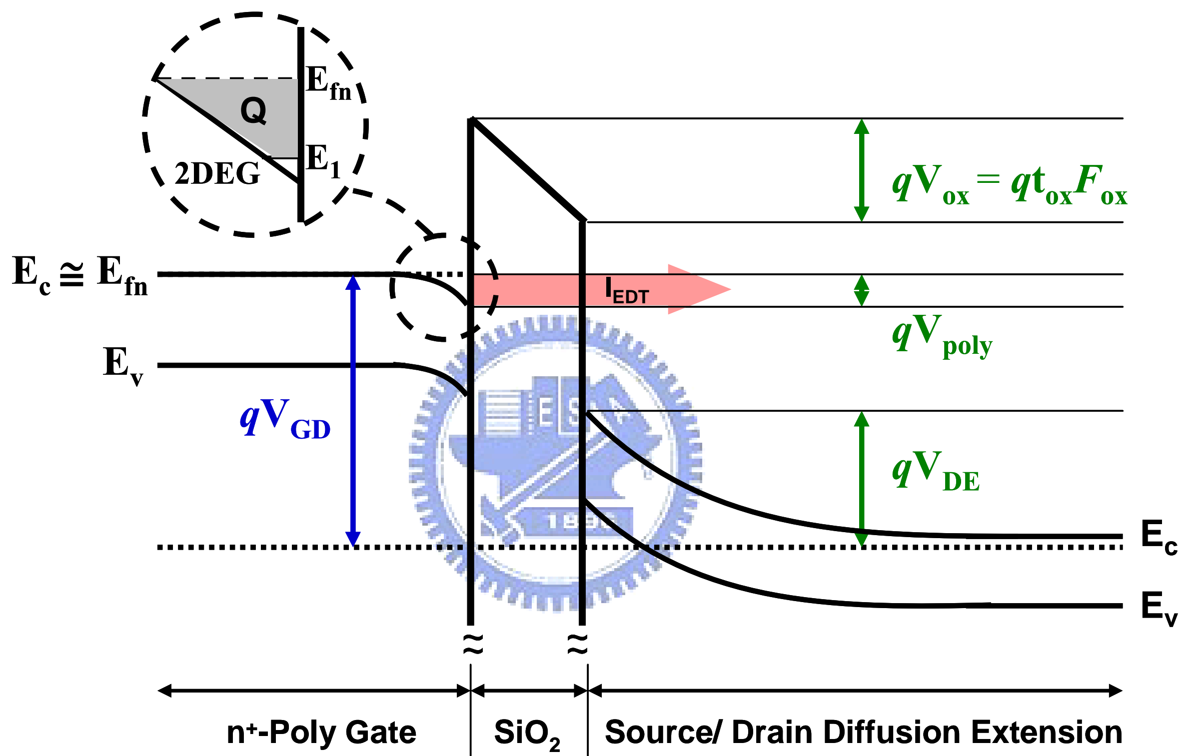


Fig. 6

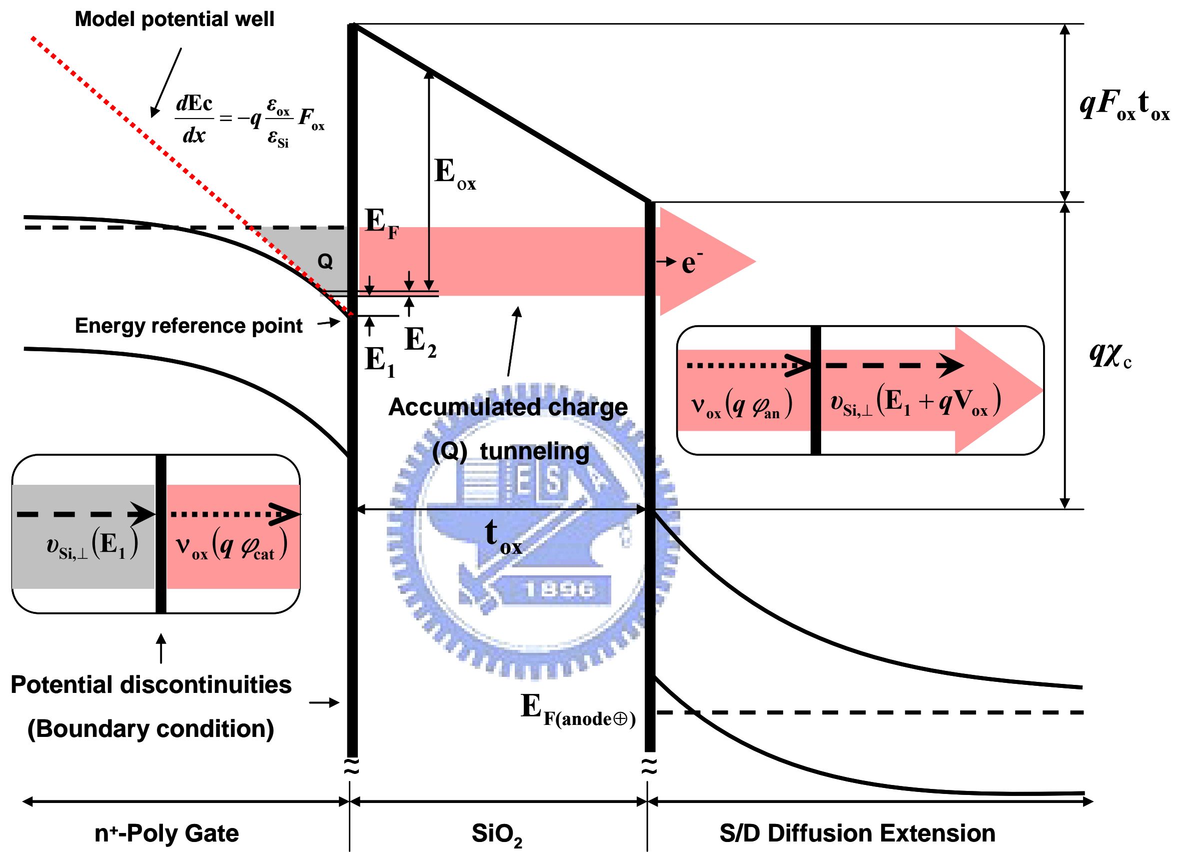


Fig. 7

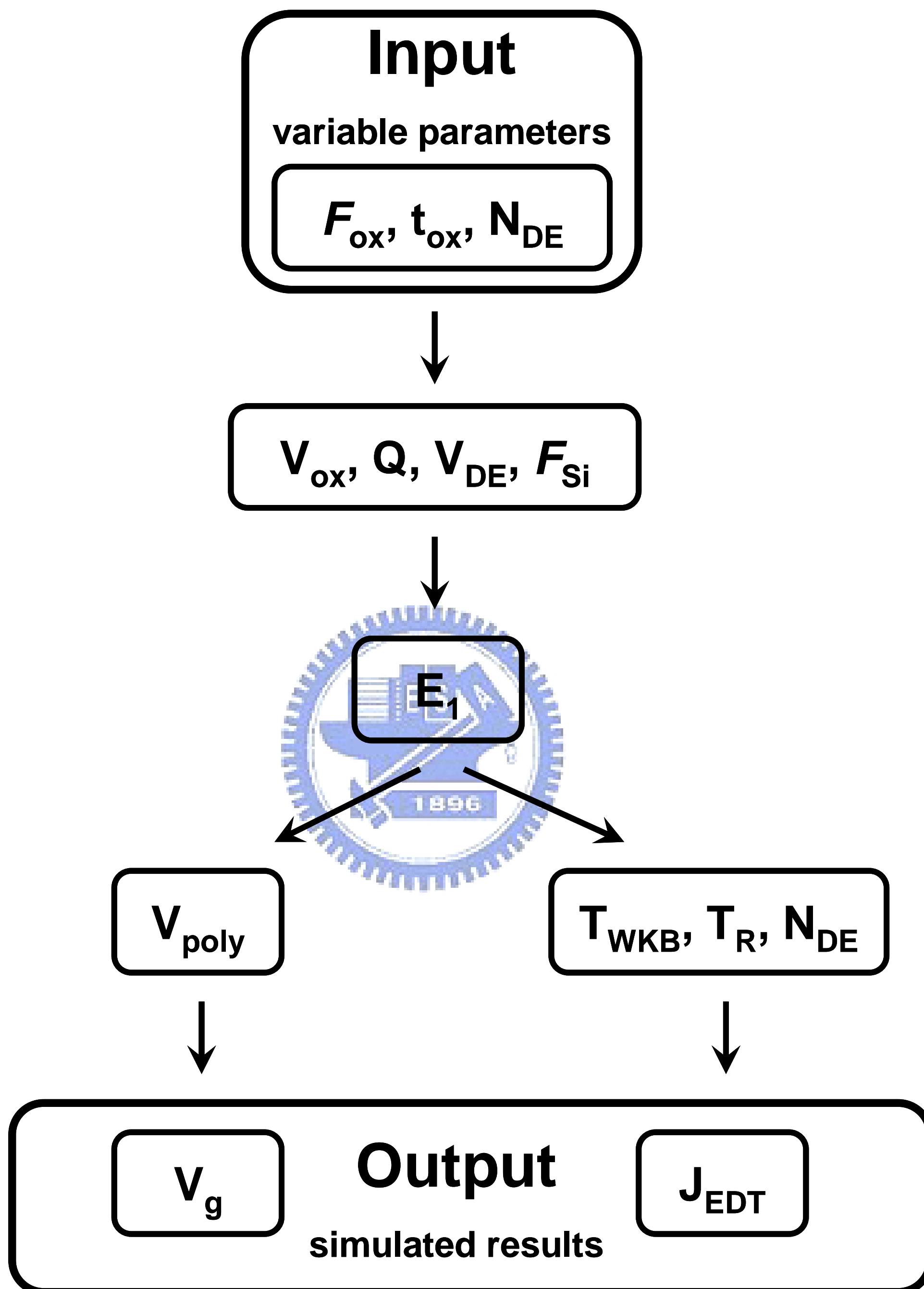


Fig. 8

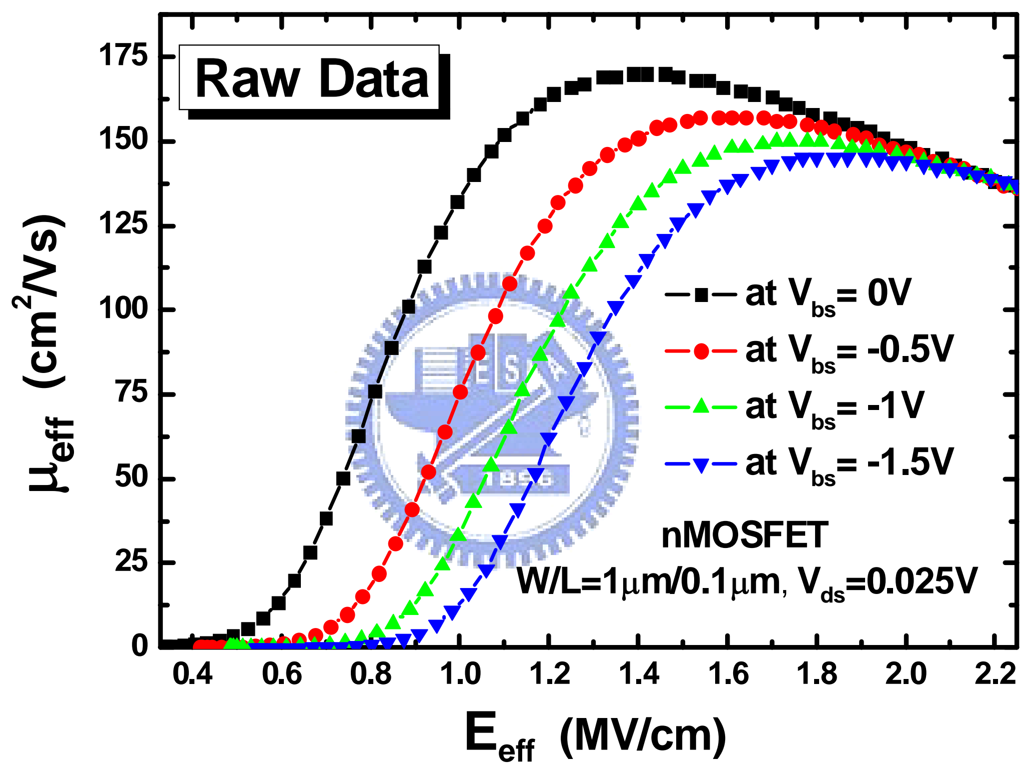


Fig. 9

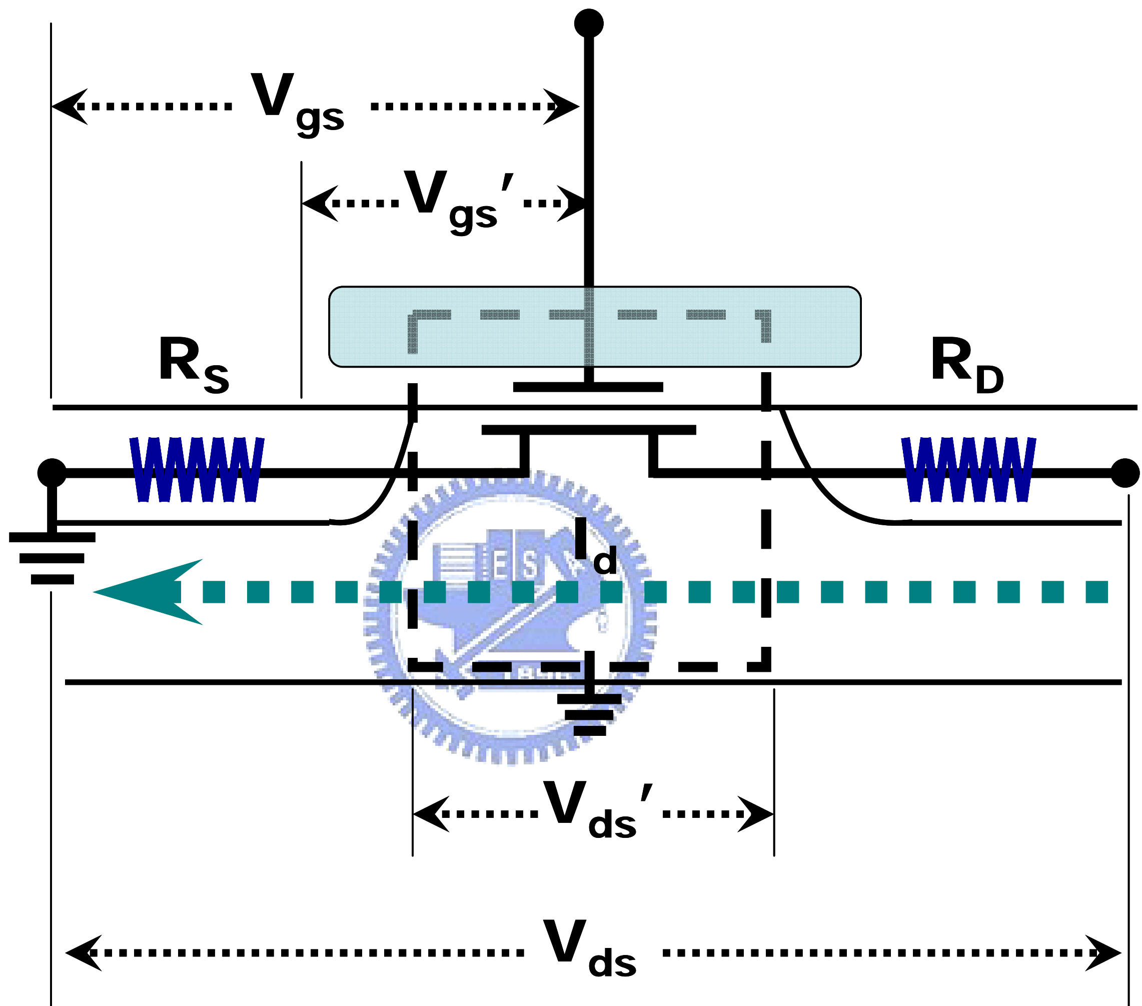


Fig. 10

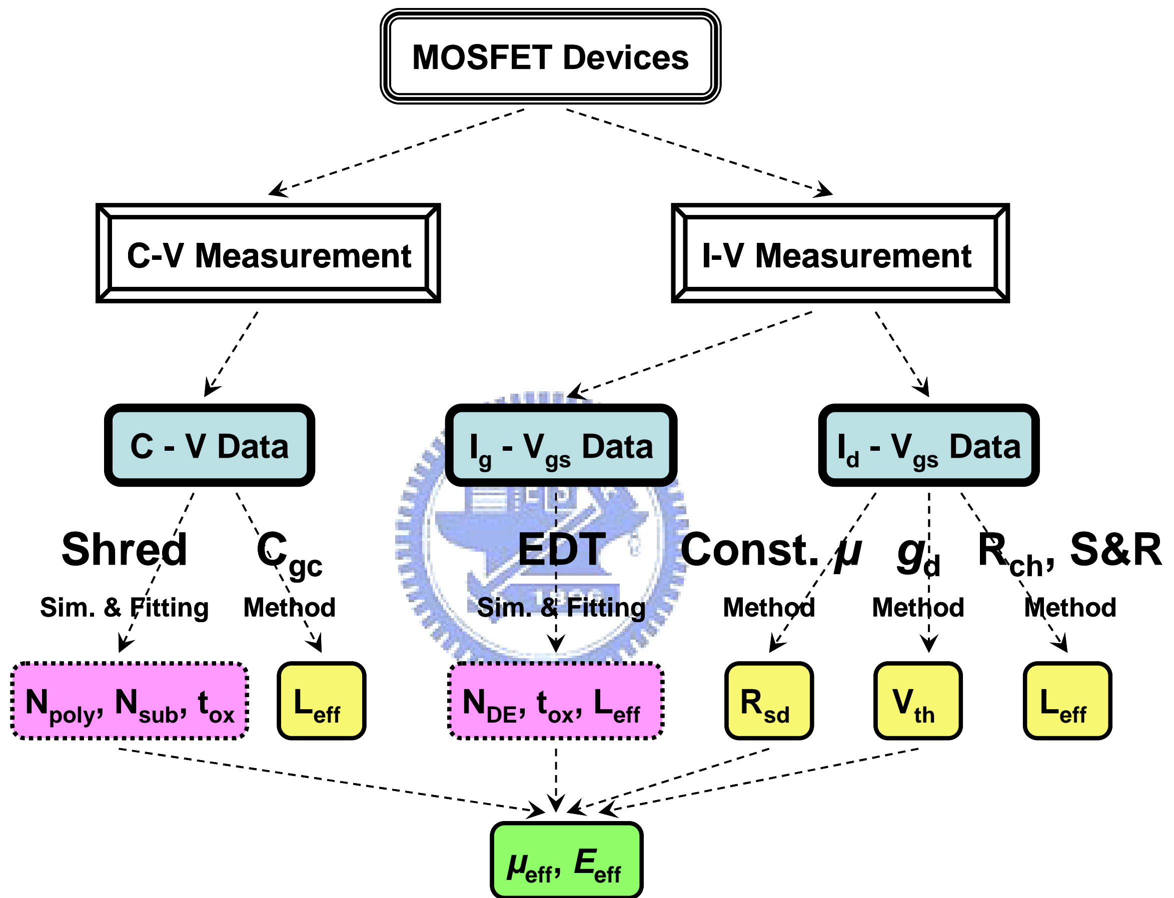


Fig. 11

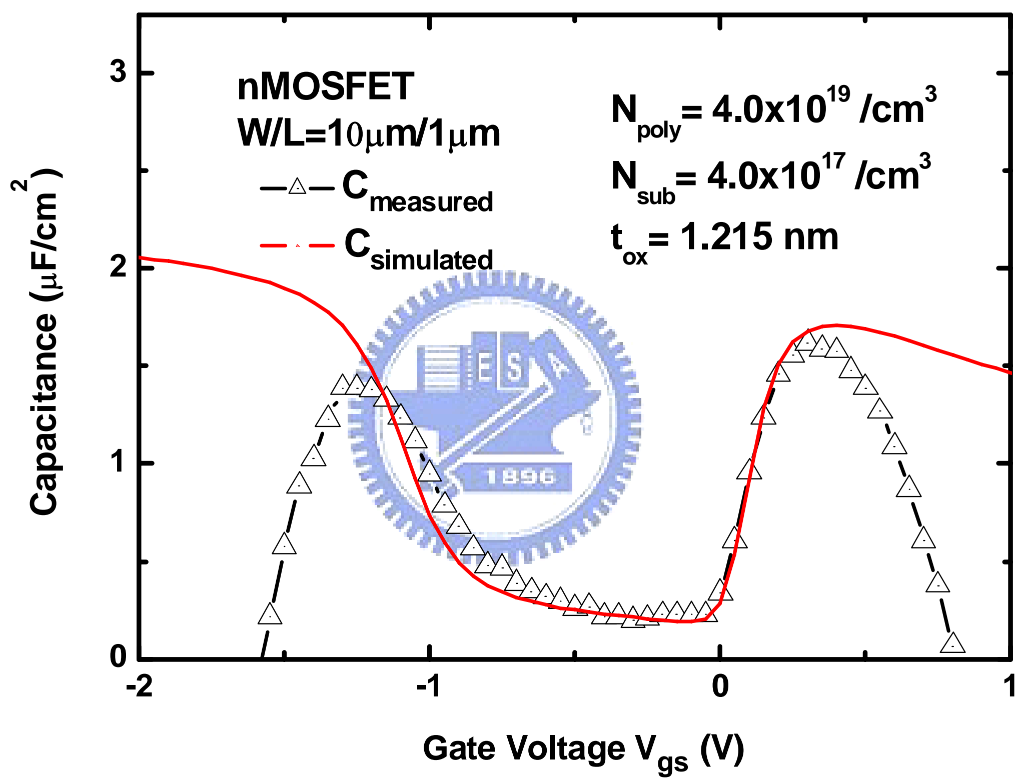


Fig. 12

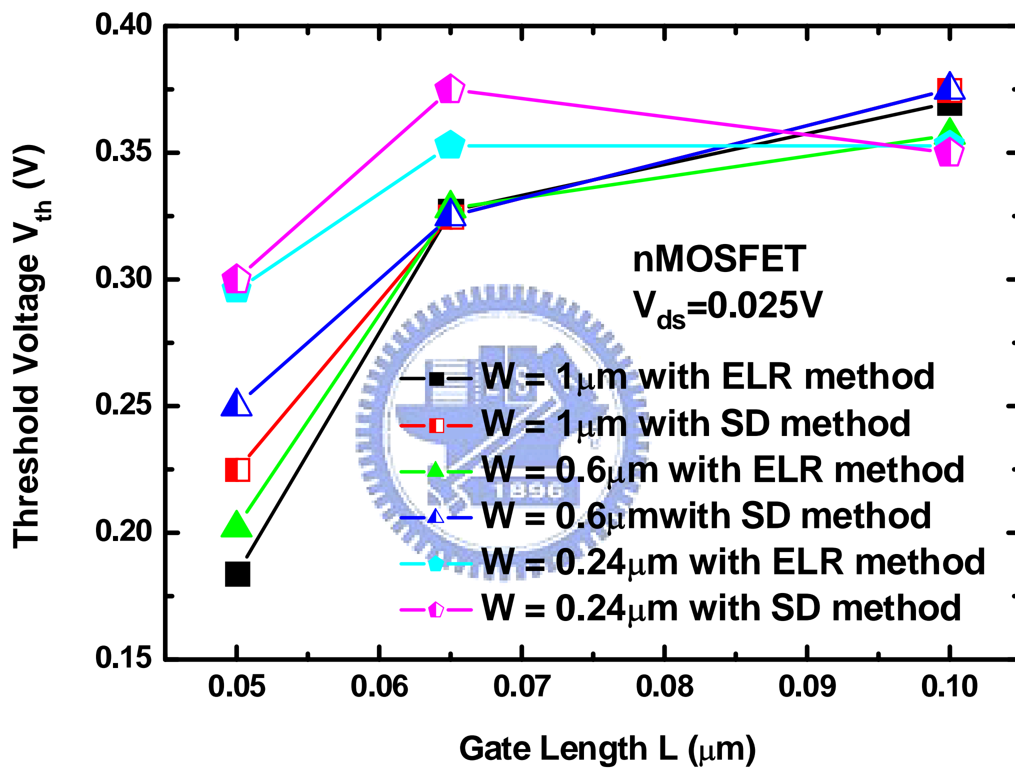


Fig. 13

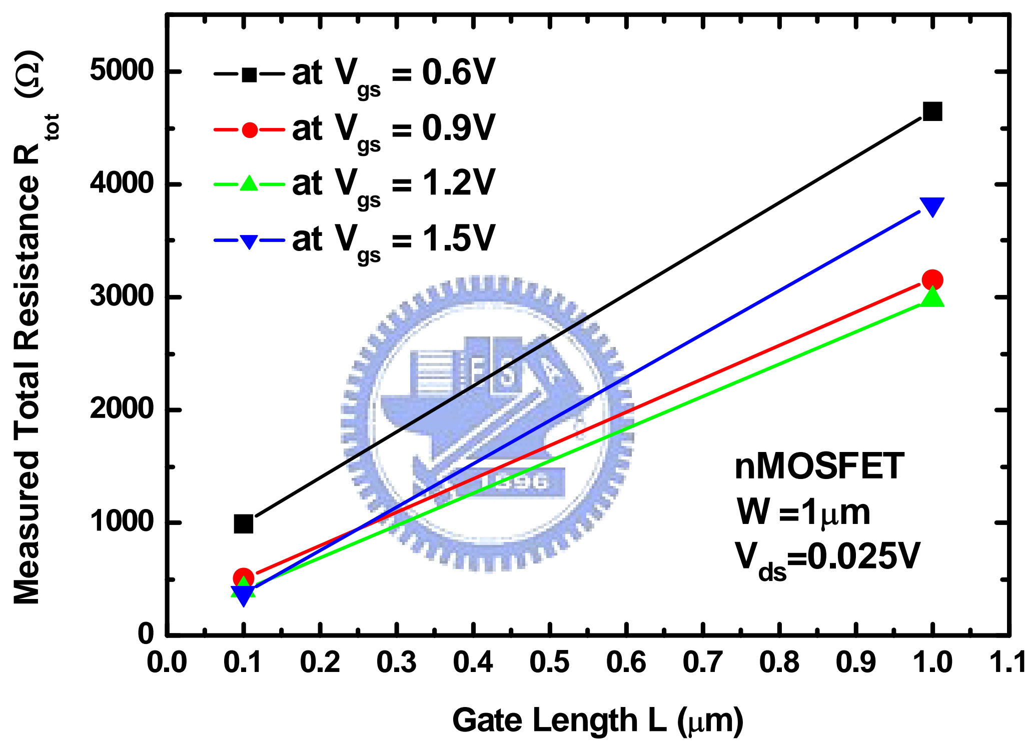


Fig. 14

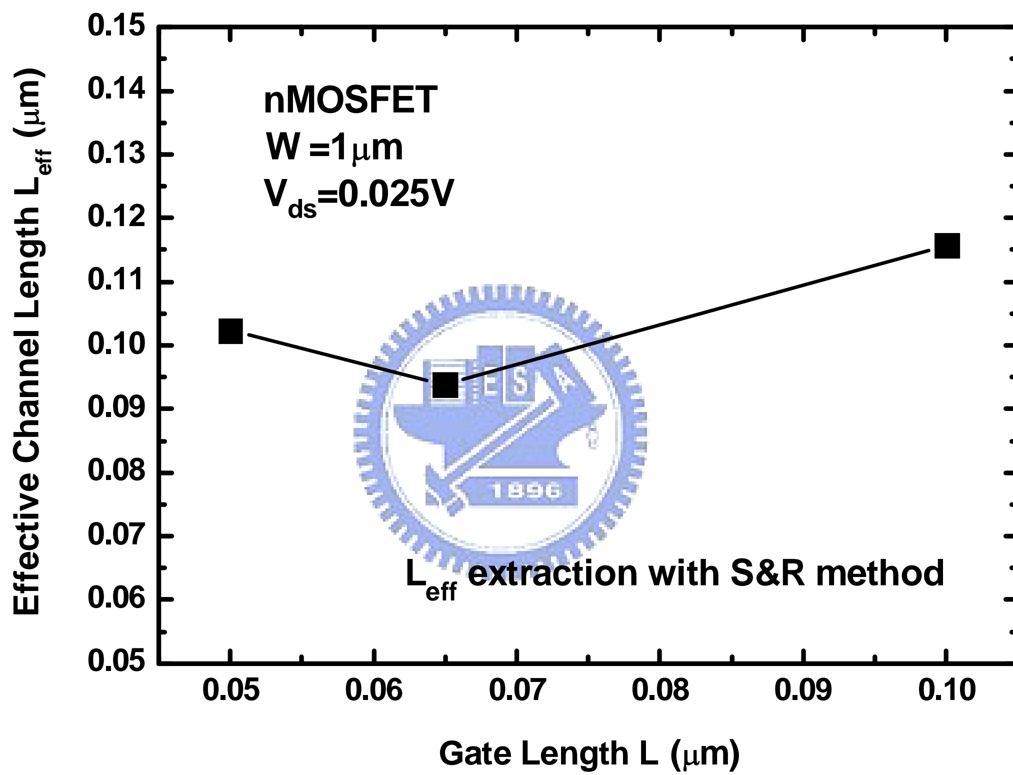


Fig. 15

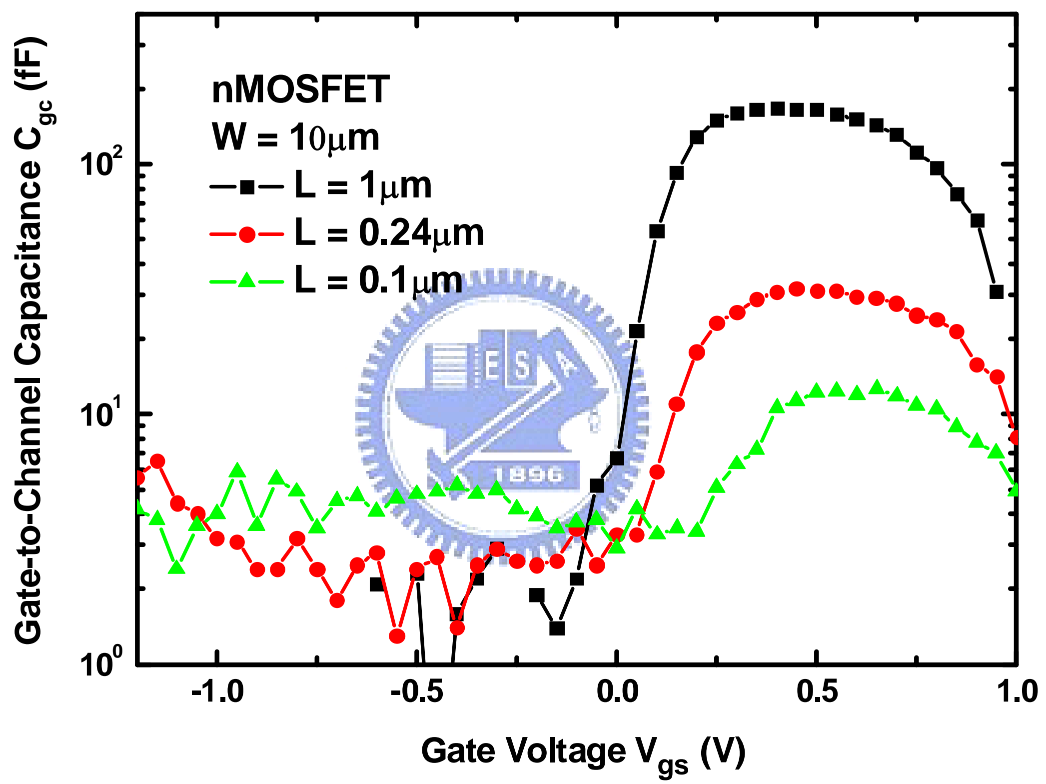


Fig. 16

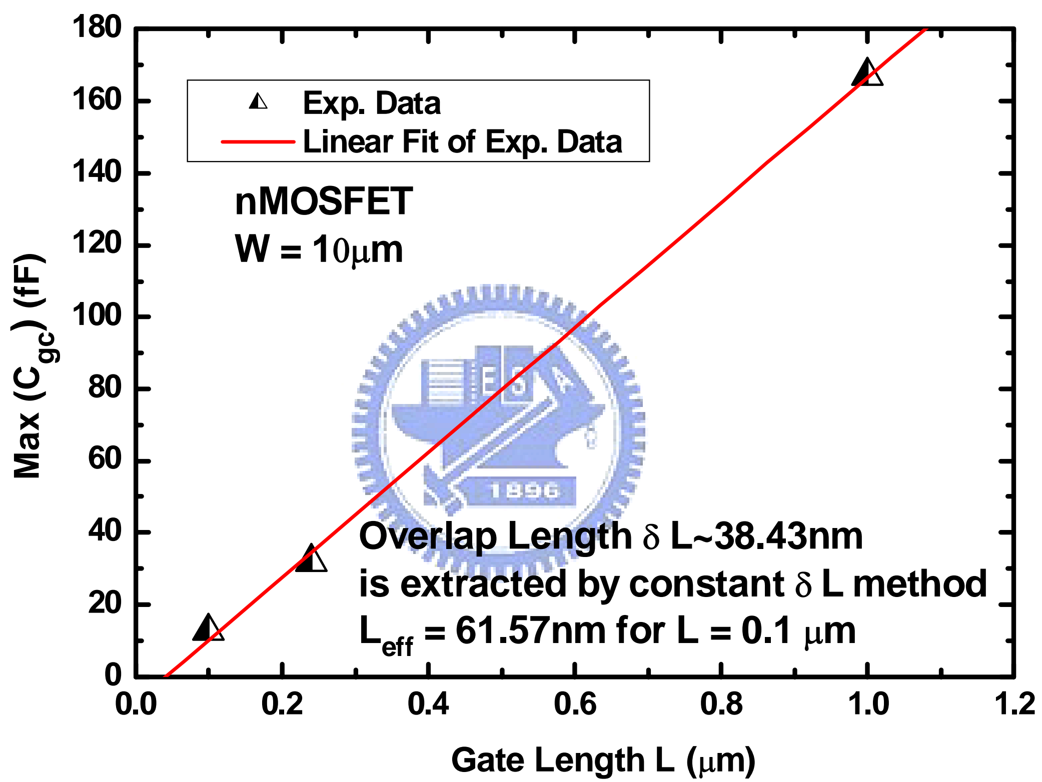


Fig. 17

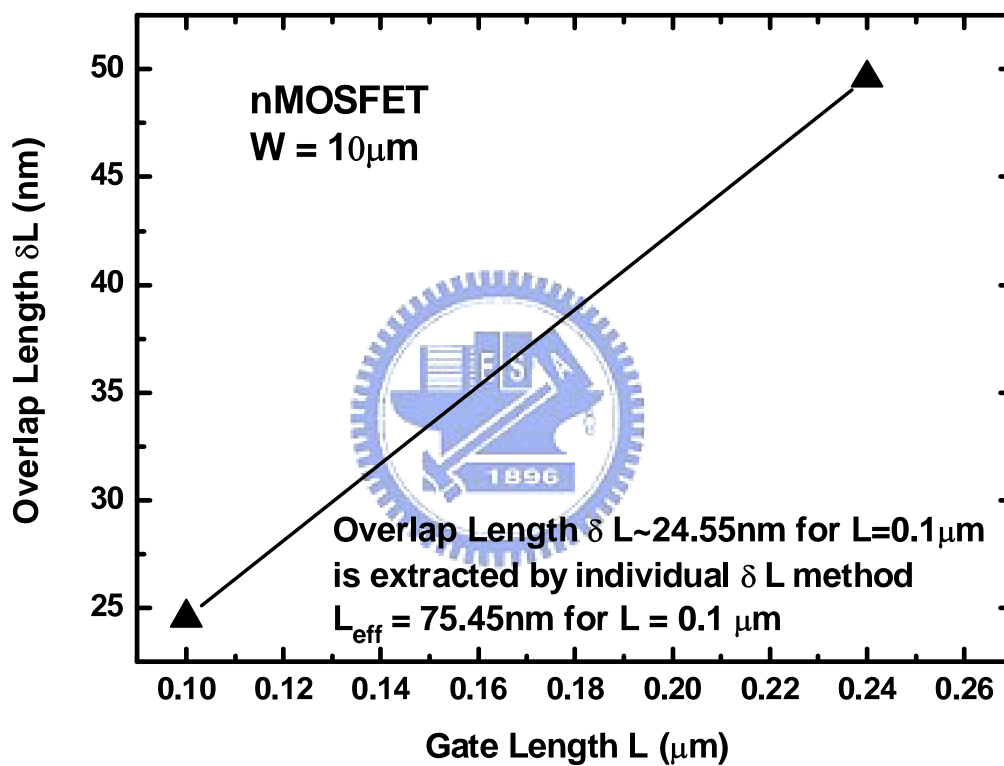


Fig. 18

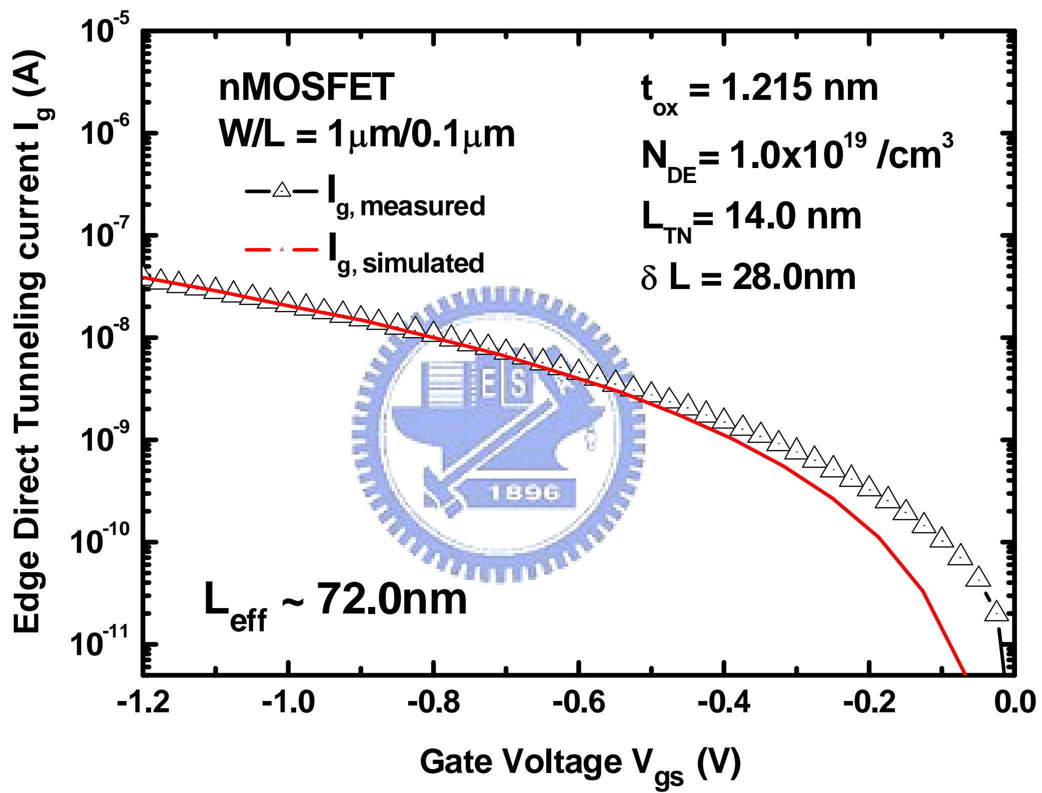


Fig. 19

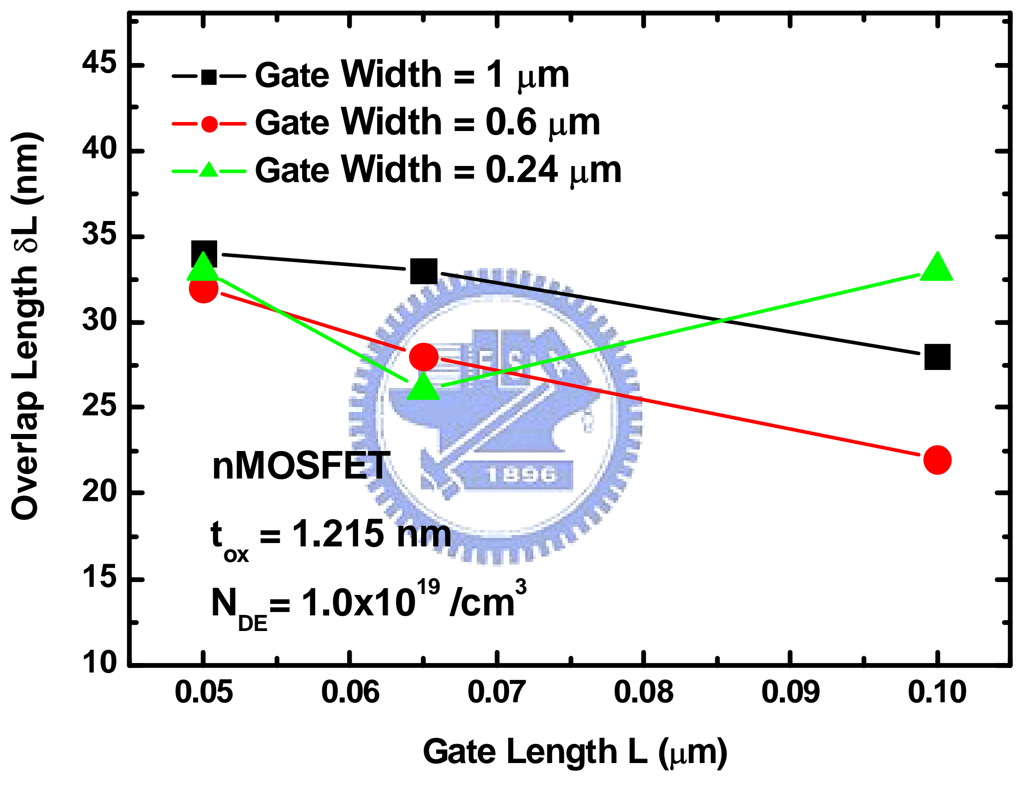


Fig. 20

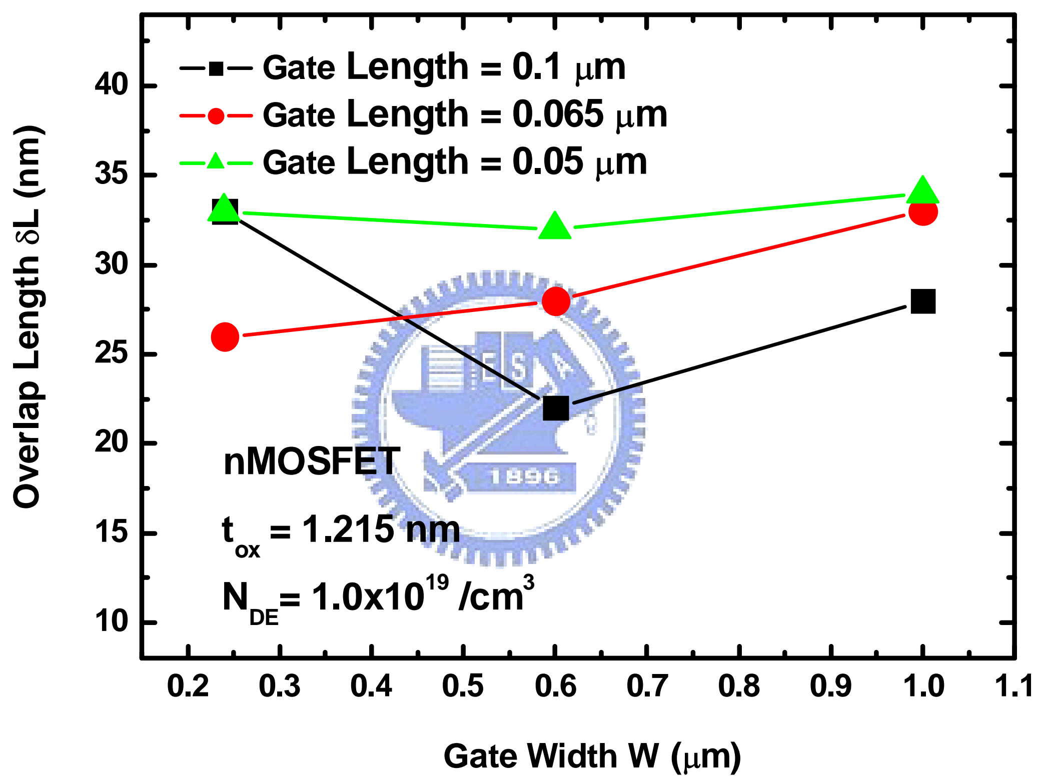


Fig. 21

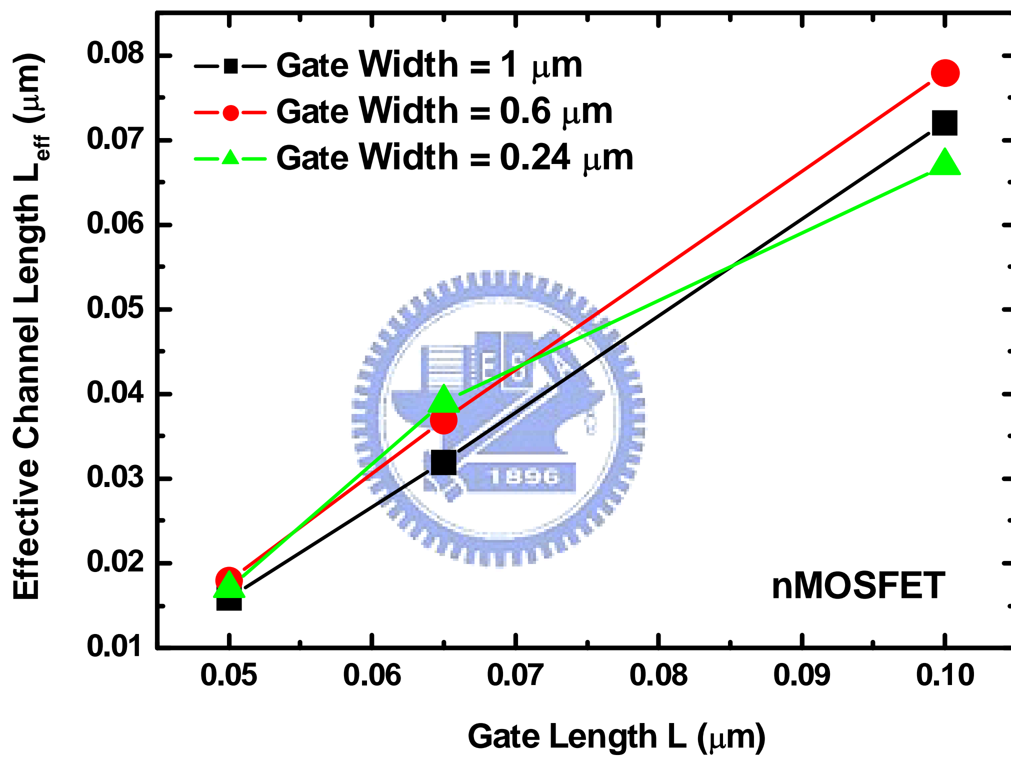


Fig. 22

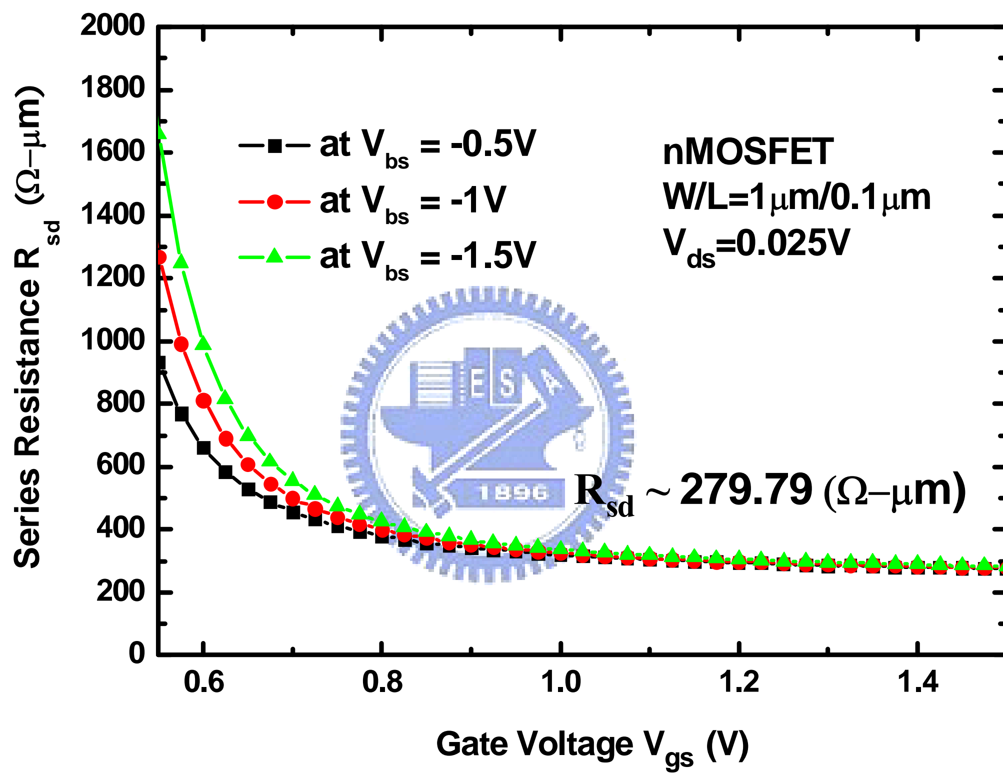


Fig. 23

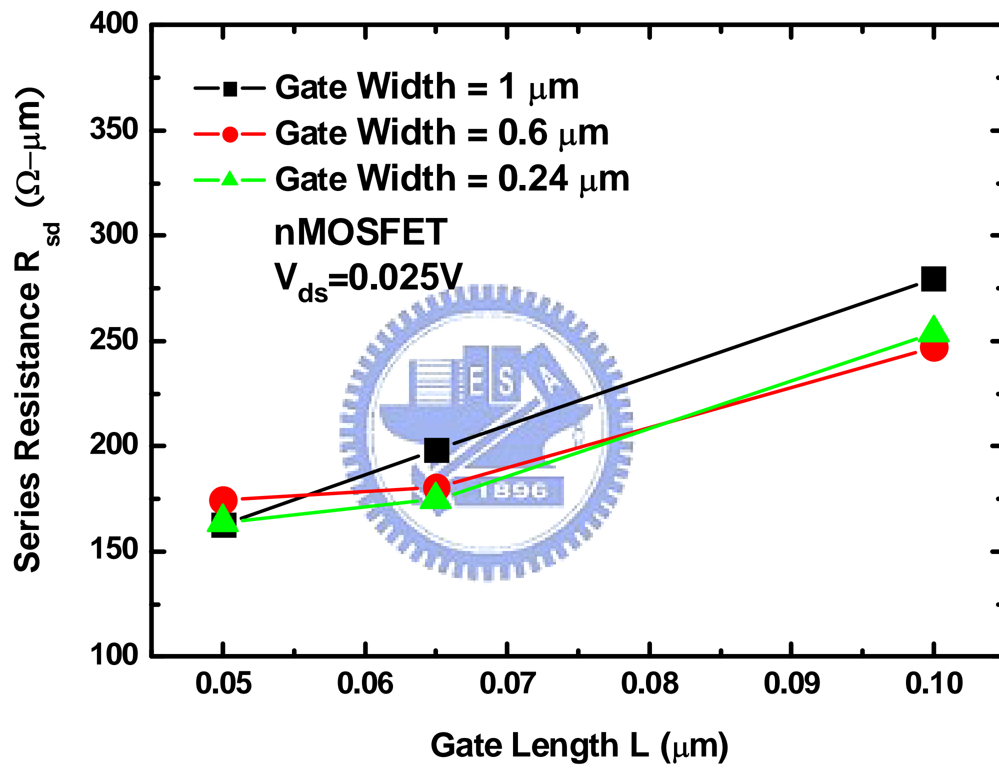


Fig. 24

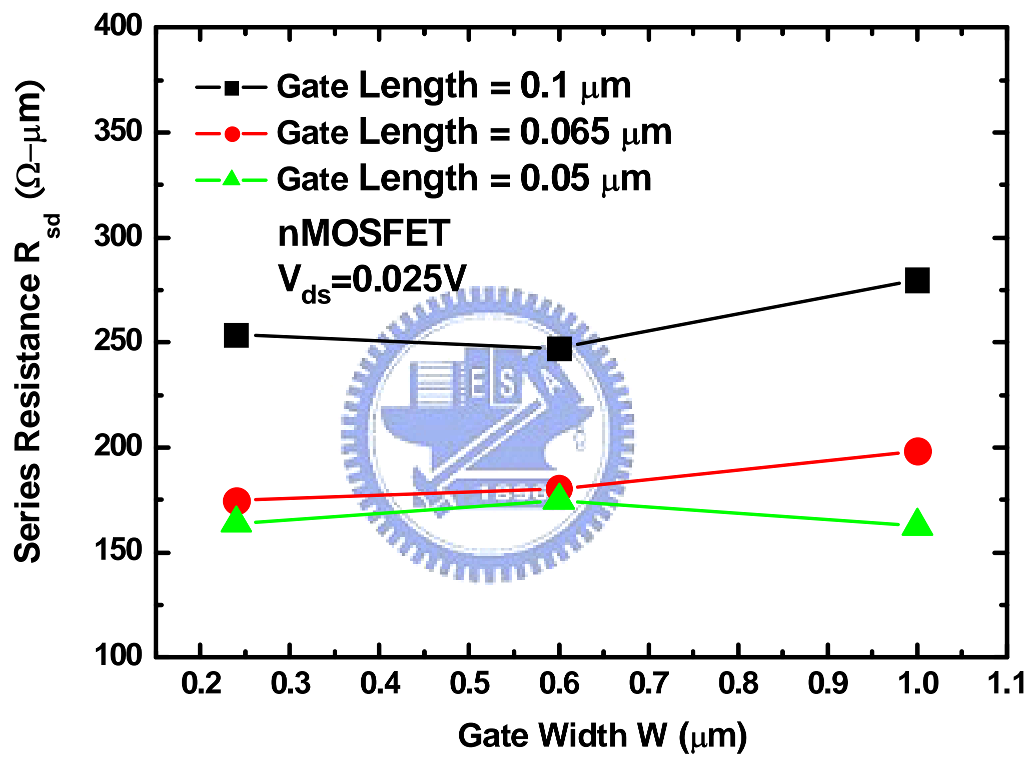


Fig. 25

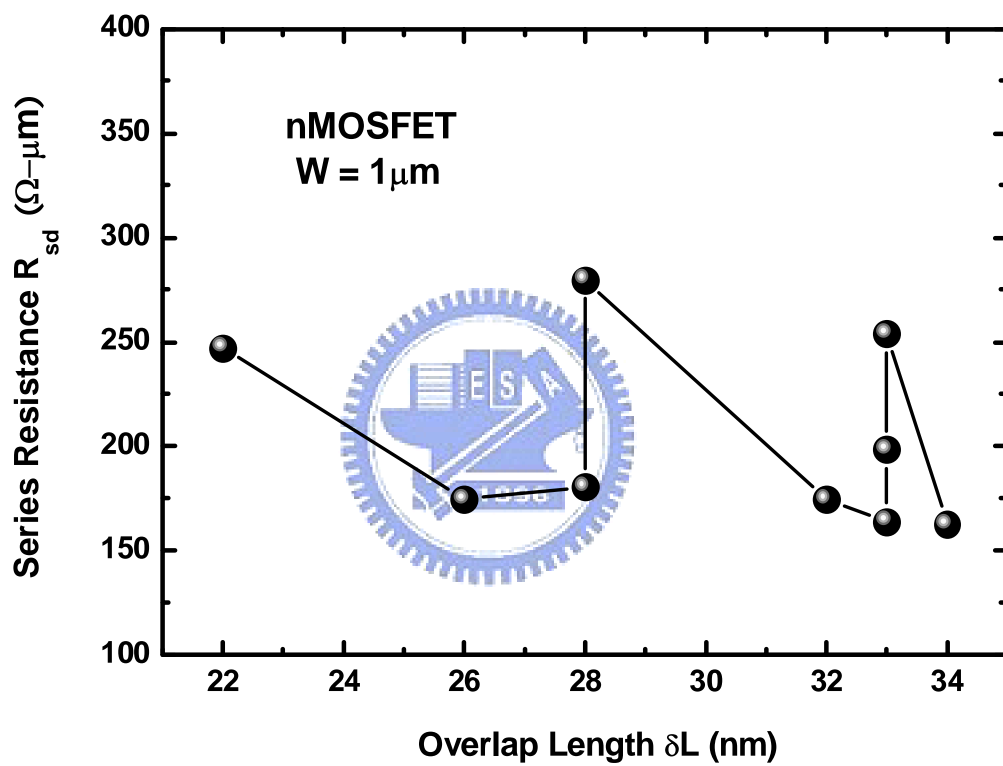


Fig. 26

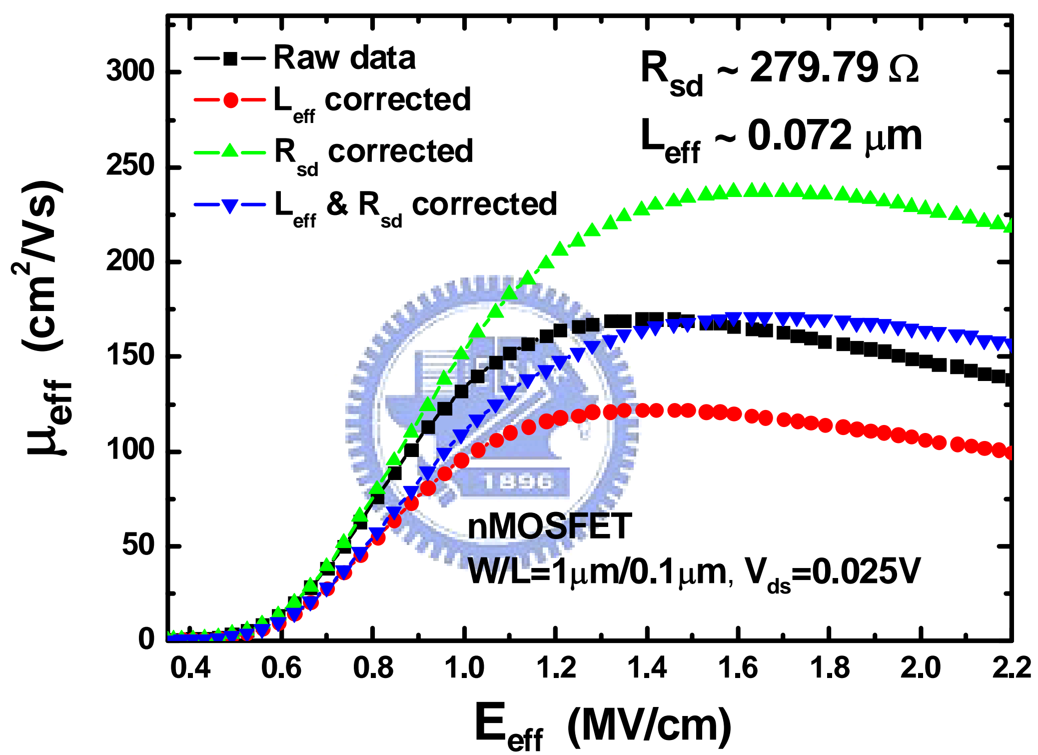


Fig. 27

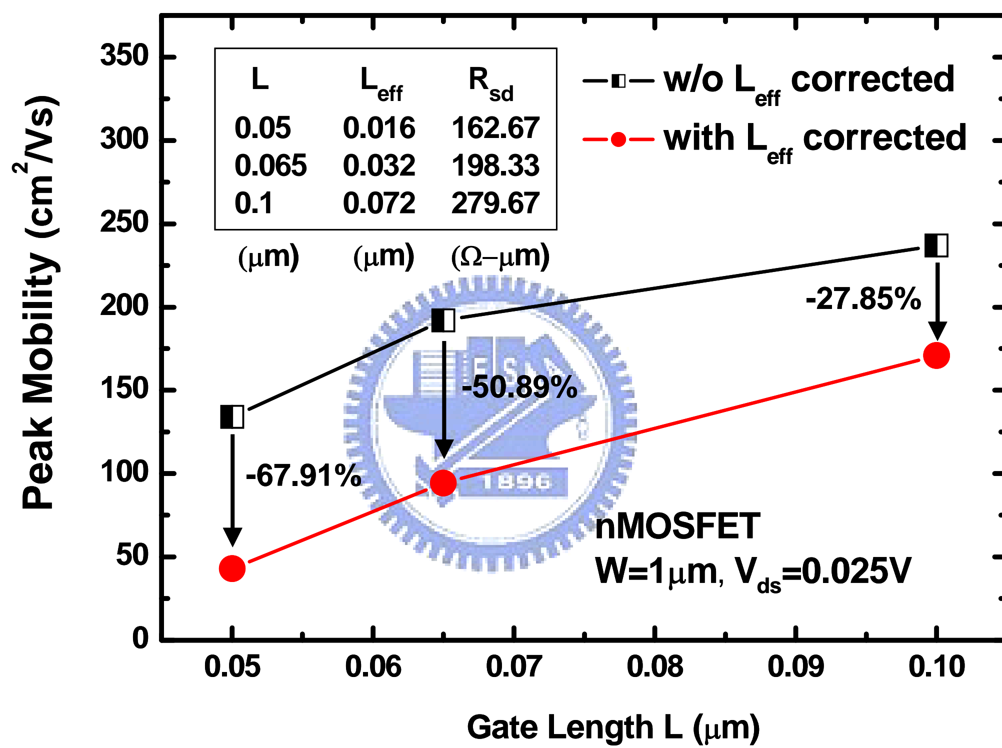


Fig. 28

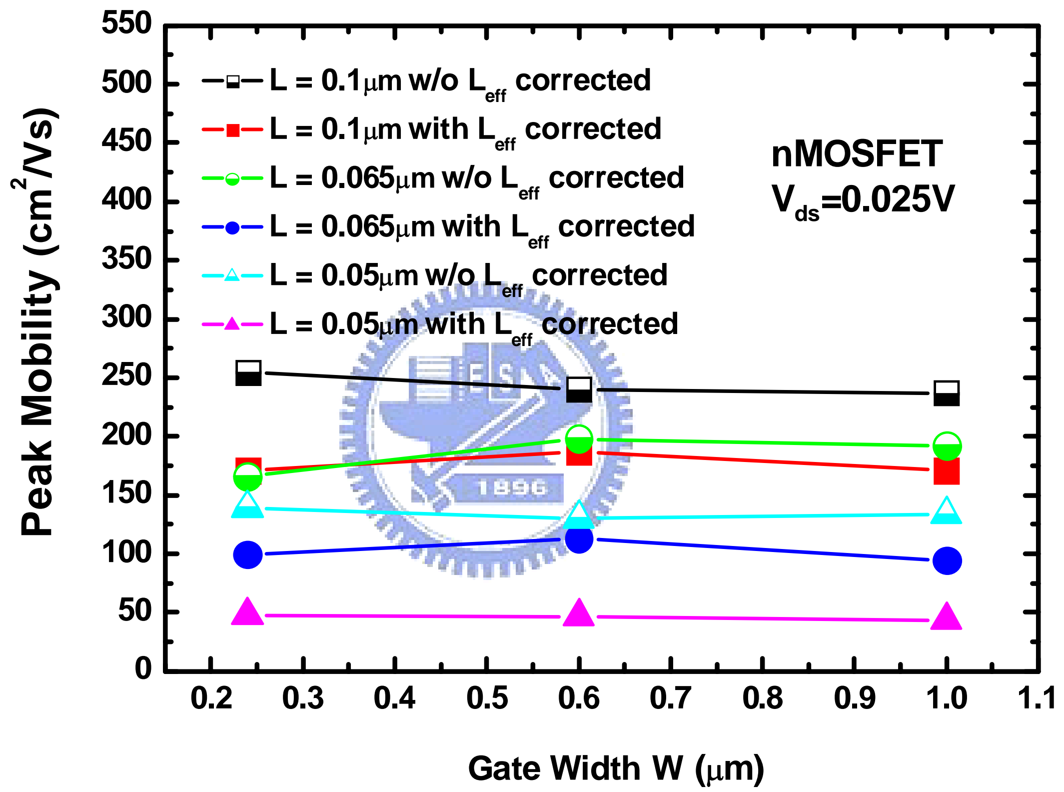


Fig. 29

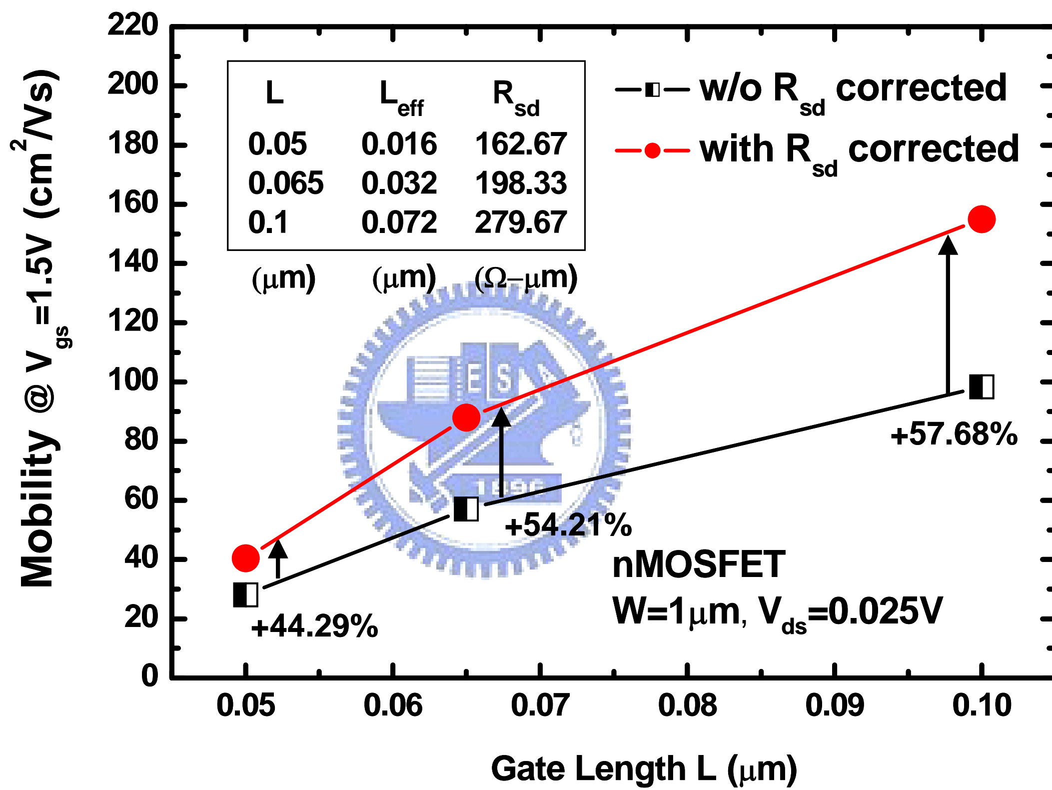


Fig. 30

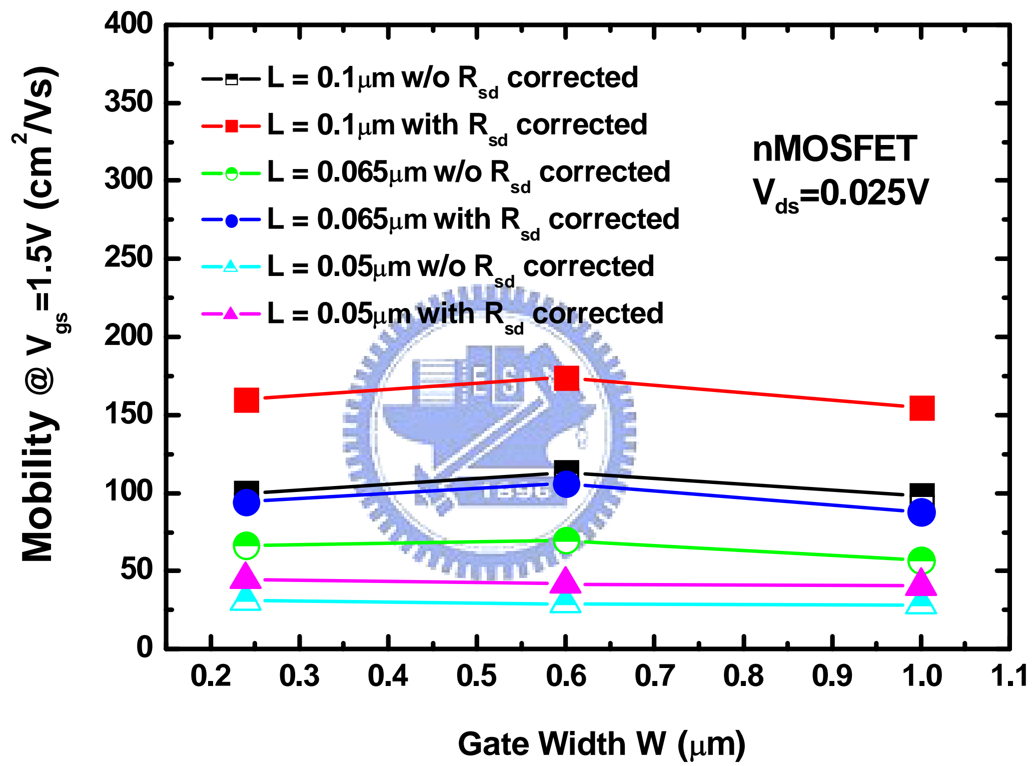


Fig. 31

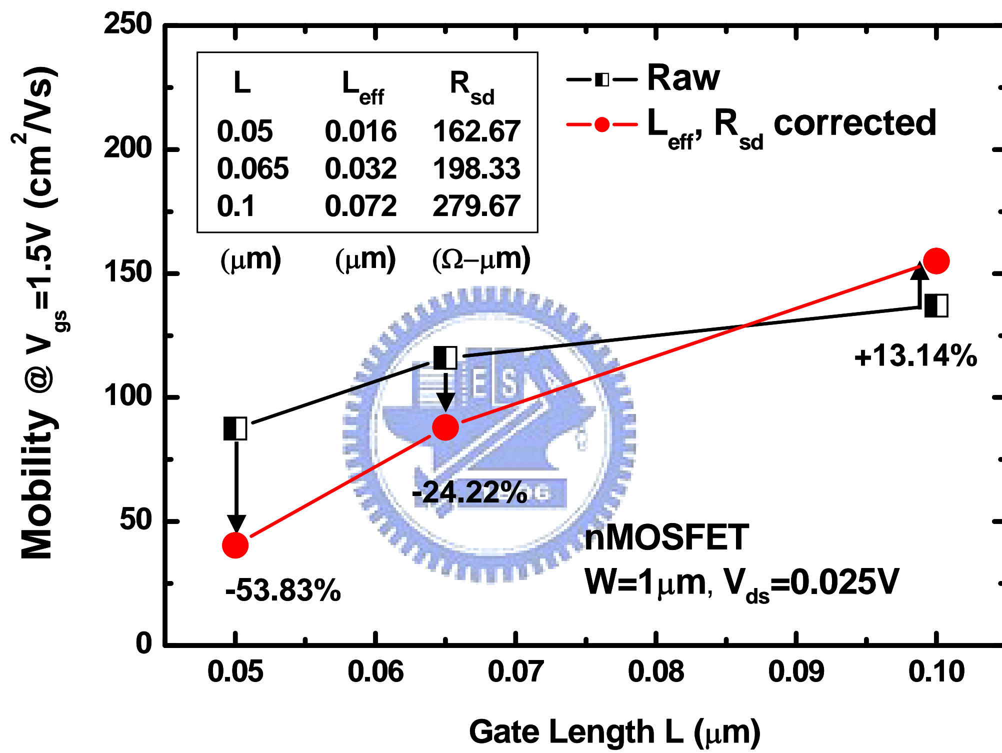


Fig. 32

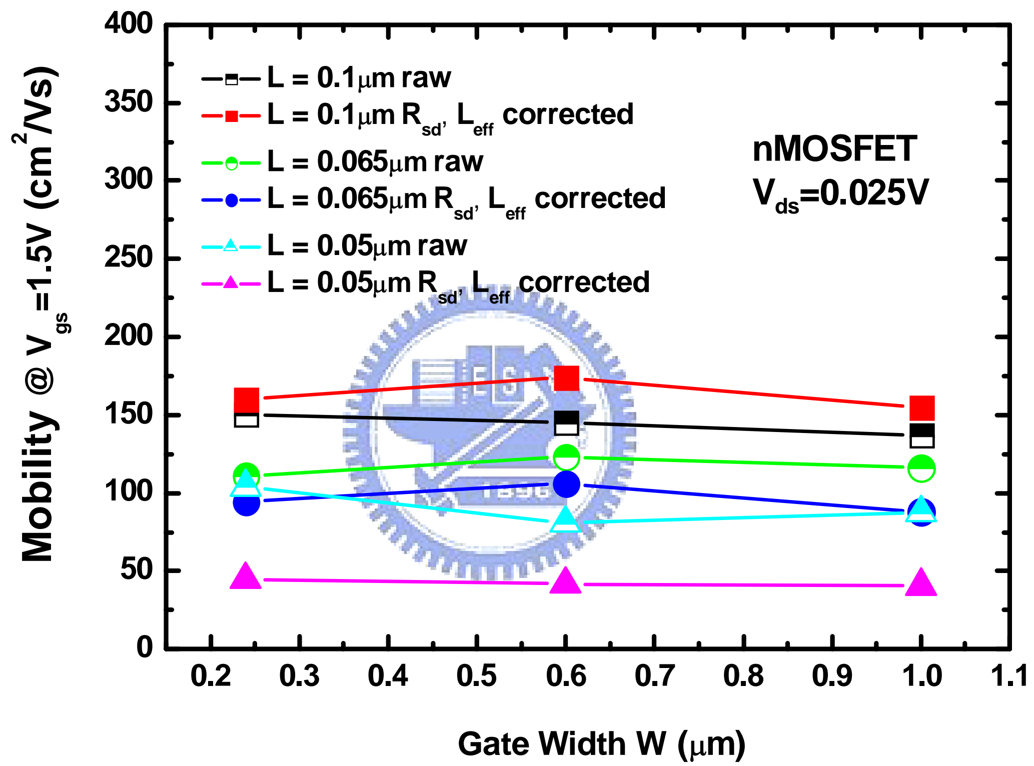


Fig. 33