國立交通大學

電子工程學系電子研究所

碩士論文

金屬閘高介電n型金氧半場效電晶體及鰭式電晶體閘極電子穿隧電流的精確模擬

Accurate Modeling of Gate Electron Tunneling Current in Metal-Gate/High-K nMOSFET and nFinFET

研究生：張洛豪　　Lou-Hao Chang

指導教授：陳明哲　　Prof. Ming-Jer Chen

中華民國 一００ 年 八 月

金屬閘高介電n型金氧半場效電晶體及鰭式電晶體閘極電子穿隧電流的精確模擬

Accurate Modeling of Gate Electron Tunneling Current in Metal-Gate/High-K nMOSFET and nFinFET

研究生：張洛豪　　Lou-Hao Chang

指導教授：陳明哲　　Prof. Ming-Jer Chen

國立交通大學

電子工程學系電子研究所

碩士論文

A Thesis
Submitted to Department of Electronics Engineering &
Institute of Electronics
College of Electrical and Computer Engineering
National Chiao Tung University
in Partial Fulfillment of the Requirements
for the Degree of
Master of Science
in
Electronics Engineering
August 2011
Hsinchu, Taiwan, Republic of China

中華民國 一００ 年 八 月

金屬閘高介電n型金氧半場效電晶體及鰭式電晶體閘極電子穿隧電流的精確模擬

研究生：張洛豪　　指導教授：陳明哲

國立交通大學

電子工程學系電子研究所

摘要

高介電質絕緣層可以抑制閘極漏電流而鰭式金氧半電晶體結構可以改善短通道效應的影響。藉由WKB近似理論建立的閘極穿隧電流已經被發表了。在本篇論文中，將會說明電子通過雙閘極元件中高介電質絕緣層的穿隧模型，而且將此模型應用到n型鰭式金氧半電晶體中。由於元件從單個閘極變成多閘極的結構，電子在各個能帶中擁有的能量公式需要被修正。藉由改變曲線擬合因數，修正後應用於雙層閘集結構的模型對於不同基底厚度依然成立。將滿足修正變數線性方程式中的基底厚度改變，只要知道基底厚度，修正變數就能夠被確定。

除了閘極電容電壓及閘極電流電壓的資料曲線擬合外，量測元件中的相關材料係數可以由進一步的對閘極電流取對數的曲線擬合得到更為正確的結果。由於高介電值層和介面層的介電係數差太大，平緩的過度漸層介於此兩層中加入，能改善對於在高電場中電流的曲線擬合。

由這幾個修正，測量的資料和模擬數據電流可以吻合。這可以更加了解電子在鰭式金氧半電晶體中的穿隧機制。

Accurate Modeling of Gate Electron Tunneling Current in Metal-Gate/High-K nMOSFET and nFinFET

Student: Lou-Hao Chang Advisor: Prof. Ming-Jer Chen

Department of Electronics Engineering and Institute of Electronics

National Chiao Tung University

Abstract

High-K stacks can suppress the gate leakage current while a FinFET structure has benefits of improving the short channel effects. Gate tunneling current model has been established based on WKB approximation. In this thesis, an electron tunneling model through high-K stacks will be constructed for double-gate devices, especially n-type FinFET. The electron subband energy should be modified for the change of the structure from single gate to multiple gates. This model established for double gate structure is also valid for different body thicknesses through different fitting factors used. A linear relation is obtained between body thickness and the subband fitting factor. Once the body thickness is known, the fitting factors can be determined accordingly.

Material parameters of the experimental devices can be accurately determined by a new fitting of dln(Ig)/dVg-Vg in combination with the conventional Cg-Vg and Ig-Vg fittings. A gradual transition layer between high-K layer and interfacial layer can improve the fitting quality at high gate voltages owing to the large difference of permittivity between high-K dielectric and interfacial layer. With these modifications incorporated, good agreements with measured gate tunneling current can be achieved.

The new model can also lead us to a better understanding of the gate tunneling mechanism in FinFET.

Acknowledgement

碩士班兩年時間，在實驗室每周開會時間，陳明哲教授對於學術界的熱情以及學問的淵博，讓我在報告時候，研究瓶頸的時刻，由教授提供的意見及方向，能夠繼續一步步的完成論文的研究，由衷感謝老師這段時間的指導。也謝謝百忙之中抽空參加我口試的口試委員們，對於我的研究上提供意見。

在實驗室中，博士班李建志、許智育、李韋漢與張立鳴學長們對於我課業上、生活上各種問題提供了解決方法，而許智育學長在這段期間，有他的指導與幫助，使我的研究能夠順利進行。

感謝實驗室的研究夥伴們，與光心君、彭霖祥一起互相鼓勵與討論，學弟妹們在課業上以及準備口試上也給予我極大的幫助，跟大家一起度過在實驗室的時間，是這兩年讓我很開心的事。

還有感謝在就學期間，全力支持以及鼓勵我的家人，有了他們的一席鼓勵的話，讓我更有動力去完成及克服遇到的困難。

Contents

Chapter 1

Introduction

1.1 Background

Scaling of device is one of the main targets in IC fabrication technology. From conventional SiO_2 gate oxide MOSFET (metal-oxide-semiconductor field-effect transistor) to high-K material gate dielectrics MOSFET, high-K devices are able to reduce the gate leakage current. With the progressive scaling into 22nm and beyond, short-channel effects (SCE) become a big issue. In order to suppress SCE, increasing channel doping is one of the methods adopted in conventional planar devices. But poor carrier mobility due to high channel doping is the weakness [1]. To overcome the problem, several multiple gate structures have been proposed. Tri-gate device also known as FinFET has been considered as potential candidate for 20 nm gate length. In a FinFET, the gate drapes across the fin, giving the FinFET multiple gates compared to the single gate of the planar transistor. So, the control of the gates on the channel in a FinFET is stronger than in a conventional MOSFET as the gate voltage is applied from several sides and not just from the top. One important feature of multiple gate devices is that they do not need heavy channel doping to suppress short-channel effects.

The MOSFET gate oxide thickness is rapidly approaching the direct tunneling limit that ultimately leads to intolerably large standby power and impractical applications. On the other hand, although high-K stacks exhibit lower gate tunneling current, short-channel effect is more and more serious while pushing planar device into nanometer region. The key feature of FinFET is the strong gate control of the channel region suppressing effectively the short-channel effects. Since the FinFET

requires ultra-thin gate dielectrics, gate tunneling current is an important factor in these devices, because it is related to the limitation of gate oxide scaling. So far, there was no accurate model of FinFET gate tunneling current. Thus, the study in this direction is necessary and important.

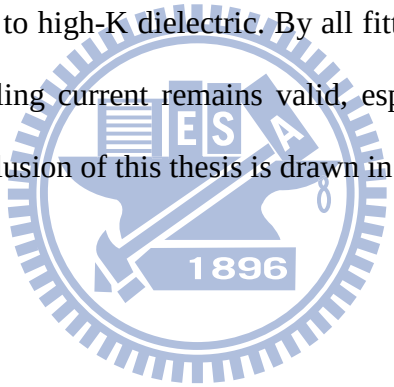
1.2 Arrangement of This Thesis

First of all, building tunneling current model for n-type FinFET is the main purpose of this thesis. The direct tunneling physical model for oxide dielectric is introduced in the first part of Chapter 2. This part includes four key parameters: the inversion layer charge density, the electron impact frequency on interface, the WKB transmission probability, and the reflection correction factor, as will be explained one by one. Because the gate material of experimental FinFET device usually contains high-K dielectric, modifying the model originally for conventional gate oxide SiO_2 is essential. Modified WKB transmission probability and reflection correction factor both are introduced in the second part of Chapter 2. In order to calculate the electron tunneling probability through the high-K stacks, the electron subband energy is necessary to determine. In the third part of Chapter 2, we employ a simple method to calculate subband energy.

Building the gate current model for FinFET is introduced in Chapter 3. It is divided into three parts. First, because of the device structure change from planar to multiple gate, full depletion situation needs to be considered in the double gate model. Depletion charge density calculation is introduced in the first part. Next, the devices of two gates induce confinement to the electrons while the body thickness is scaling down. So the structure confinement makes a significant difference in calculated subband energy between planar and multiple gate devices, which will be introduced in the second part. Lastly, in addition to conventional Ig-Vg and Cg-Vg fittings,

dlnIg/dVg-Vg fitting is important for high-K stacks devices. We can find that it might make an erroneous fitting result if the dlnIg/dVg fitting is not performed.

After the model is constructed, the experimental results about n-type FinFET are shown in Chapter 4. By fitting experimental data, we can extract the process and material parameters of the device. In the beginning, the gate current calculated by analytical model deviates from experimental one. Modified tunneling models are introduced by adding transition layer between the high-K dielectric and interfacial layer. Linear transition layer and parabolic transition layer are the two models adopted in this study. We can also prove that these models are valid for simulating the gate current. The dlnIg/dVg fitting is also done to help determine accurately the gate material parameters related to high-K dielectric. By all fittings and modifications, the analytical model for tunneling current remains valid, especially above threshold of operation. Finally, the conclusion of this thesis is drawn in Chapter 5.



Chapter 2

Physical Model for Planar FET

First, the principle of gate electron direct tunneling in an n^+poly/SiO_2/p-substrate structure will be explained. Then, by substituting high-K for SiO_2 as gate dielectric material, a physical model of electron tunneling across high-K stacks is formed.

2.1 Tunneling Current Model for Oxide Dielectric

The basic band diagram of n^+poly/SiO_2/p-substrate is shown in Fig. 1. Quantum mechanical calculation for the inversion layer in substrate and modified Wentzel-Kramers-Brillouin (WKB) approximation for the transmission probability across SiO2 layer are employed in this model [2],[3]. The direct tunneling electron current model is made up of four key parameters: the inversion layer charge density, the electron impact frequency on interface, the WKB transmission probability, and specially, the reflection correction factor.

$$J_g = qN_{i,j}f_jT_{WKB}T_R \tag{2-1}$$

where

J_g is the electron tunneling current;

q is the elemental charge;

$N_{i,j}$ is the inversion layer charge per unit area with *j*th subband in the *i*th valley;

f_j is the electron impact frequency on SiO_2/Si interface;

T_{WKB} is the WKB transmission probability;

T_R is the reflection correction factor

A. *Inversion Layer Charge ($N_{i,j}$)*

Using the density of states for a two-dimensional electron gas (2DEG) and Fermi-Dirac statics, the inversion layer charge density of each energy subband can be derived as

$$N_{i,j} = \left(\frac{k_B T}{\pi \hbar^2}\right) g_i m_{di} \, ln\left[1 + exp\left(\frac{E_F - E_{ij}}{k_B T}\right)\right] \quad (2\text{-}2)$$

where g_i is the degeneracy of the *i*th valley; m_{di} is the density of states electron effective mass in the *i*th valley and E_{ij} is the energy level of the *j*th subband in the *i*th valley.

B. *Electron Impact Frequency (f_j)*

The electron impact frequency can be described as [4]

$$f_j = \left[2 \int_0^{z_j} \frac{1}{v_{Si\perp}(x)} dx\right]^{-1} \quad (2\text{-}3)$$

where z_j epresents the classical turning point in silicon for electrons in each subband and $v_{Si\perp}$ the interface-normal group velocity component of electron wave packet, which can be expressed as

$$v_{Si\perp}(x) = \sqrt{\frac{2\left(E_{ij} - qV(x)\right)}{m_{z,i}}} \quad (2\text{-}4)$$

$m_{z,i}$ is the longitudinal or transverse effective mass in two or four fold valleys and $V(x)$ is the potential well which can be approximated as a triangle-like electrostatic potential:

$$V(x) = \frac{\varepsilon_{ox} F_{ox} x}{\varepsilon_{Si}} \quad (2\text{-}5)$$

Substituting (2-4) and (2-5) into (2-3), then we can obtain the impact frequency

$$f_j = \left[2 \int_0^{z_j} \frac{1}{v_{Si\perp}(x)} dx\right]^{-1} = \frac{q \varepsilon_{ox} |F_{ox}|}{2 \varepsilon_{Si}} \left(2 m_{z,i} E_{ij}\right)^{-\frac{1}{2}} \quad (2\text{-}6)$$

C. *WKB Transmission Probability (T_{WKB})*

To calculate the tunneling probability across the oxide barrier in Fig. 2, a modified WKB approximation is used. The modified WKB approximation includes T_{WKB} and T_R [3],[4]. The former is the usual WKB tunneling probability, effective for smooth potential barriers; the latter is a correction factor for reflections from potential discontinuities. Both are combined in form:

$$T = T_{WKB} T_R \tag{2-7}$$

$$T_{WKB} = exp\left(-2\int_0^{t_{ox}} \kappa(x)\, dx\right) \tag{2-8}$$

$\kappa(x)$ is the magnitude of the carriers imaginary wave vector within the bandgap between the oxide layer

$$\kappa(x) = \sqrt{\frac{2m_{ox}[E-qV(x)]}{\hbar^2}} \tag{2-9}$$

Then substituting (2-9) into (2-8) yields

$$T_{WKB} = exp\left\{-2\left|\int_0^{t_{ox}} \sqrt{\frac{2m_{ox}[E-qV(x)]}{\hbar^2}}\, dx\right|\right\} = exp\left[\frac{4\sqrt{2m_{ox}}\left(\phi_{an}^{\frac{3}{2}} - \phi_{cath}^{\frac{3}{2}}\right)}{3q\hbar|F_{ox}|}\right] \tag{2-10}$$

ϕ_{cath} and ϕ_{an} represent the magnitude of the electron (tunneling from the *j*th subband in the *i*th valley) energy with reference to oxide conduction band in the channel and in the gate, respectively. They are calculated by

$$q\phi_{cath} = q\phi_{ox} - E_{ij} \tag{2-11}$$

and

$$q\phi_{an} = q\phi_{ox} - E_{ij} - qF_{ox}t_{ox} \tag{2-12}$$

where ϕ_{ox} is the Si-SiO_2 conduction band discontinuity.

The total energy is composed of transverse and longitudinal energies

$$E = \frac{\hbar^2(\kappa_x^2+\kappa_y^2)}{2m_t} + E_j \tag{2-13}$$

In this equation, m_t means the transverse effective mass and E_j is the energy in longitudinal direction.

D. Reflection Correction Factor (T_R)

Reflection factor is related to the type of material the wave penetrates. There are two interfaces in n^+poly-SiO_2-p-substrate structure, Si/SiO_2 and SiO_2/n^+poly. So the reflection correction factor is

$$T_R = T_{R1}T_{R2} = \frac{4\nu_{Si\perp}(E)\times\nu_{ox}(\phi_{cath})}{\nu_{Si\perp}^2(E)+\nu_{ox}^2(\phi_{cath})} \times \frac{4\nu_{Si\perp}(E+q|F_{ox}|t_{ox})\times\nu_{ox}(\phi_{an})}{\nu_{Si\perp}^2(E+q|F_{ox}|t_{ox})+\nu_{ox}^2(\phi_{an})} \quad (2\text{-}14)$$

where T_{R1} and T_{R2} are the reflection factors at the interface of Si/SiO_2 and SiO_2/n^+poly, respectively. $\nu_{Si\perp}(E)$ is the group velocity of electrons incident at the silicon-oxide interface

$$\nu_{Si\perp}(E) = \sqrt{\frac{2E_{ij}}{m_z}} \quad (2\text{-}15)$$

$\nu_{ox}(\phi_{cath})$ is the magnitude of the purely imaginary group velocity of electron at the cathode side of SiO_2:

$$\nu_{ox}(\phi_{cath}) = \frac{1}{\hbar}\frac{d\phi_{cath}}{d\kappa_{ox}} = \frac{\hbar^2\kappa_{ox}^2}{2m_{ox}} \quad (2\text{-}16)$$

On the other hand, $\nu_{Si\perp}(E+q|F_{ox}|t_{ox})$ means the group velocity of electrons leaving the oxide-polygate interface and $\nu_{ox}(\phi_{an})$ is the magnitude of the purely imaginary group velocity of electron at the anode side of SiO_2.

According to electron tunneling current model in (2-1), four important key parameters have been introduced one by one, (2-2), (2-6), (2-10) and (2-14). As electron tunneling is closely related to its subband energy, E_{ij}, it will be introduced in detail latter.

2.2 Modified Tunneling Current Model for High-K Gate Stacks

For highly scaled devices, thinner oxide induces a larger gate leakage current. In order to maintain device performance in the scaling direction, use of high permittivity is one of the solutions. In addition to dielectric layer change, poly gate is replaced by metal gate and thereby the poly depletion can be eliminated. The band diagram of metal gate/high-K/IL/p-substrate is shown in Fig. 3. Similarly, WKB approximation is employed to build the tunneling current model, along with some parameters modified.

A. Modified WKB Transmission Probability for High-K Stacks (T_{WKB})

To calculate the tunneling probability in high-K case, electron not only goes through the interfacial layer (IL) but also the high-K dielectric layer [4]. Consequently, T_{WKB} has two parts in terms of interfacial layer and high-K dielectric:

$$T_{WKB} = exp\left[\left(-2\int_0^{t_{IL}} \kappa(x)\,dx\right) + \left(-2\int_{t_{IL}}^{t_{high-K}} \kappa(x)\,dx\right)\right] \quad (2\text{-}17)$$

The above equation can apply to high-K stacks, but it is just one of the several cases. The band diagram changes with varying gate bias voltage, meaning that lower tunneling barrier height corresponds to higher gate voltage. Three cases are shown in Fig. 4. $\phi_{IL,cath}$ is denoted as the barrier height for tunneling electrons with reference to oxide conduction band at IL/p-substrate interface and $\phi_{IL,an}$ is the barrier height at high-K/IL interface in interfacial layer. $\phi_{high-K,cath}$ and $\phi_{high-K,an}$ are also the barrier heights with the former at high-K/IL interface in high-K dielectric and the latter at metal-gate/high-K interface. They are calculated by

$$q\phi_{IL,cath} = q\phi_{IL} - E_{ij} \quad (2\text{-}18)$$

$$q\phi_{IL,an} = q\phi_{IL} - E_{ij} - qF_{IL}t_{IL} \quad (2\text{-}19)$$

$$q\phi_{high-K,cath} = q\phi_{high-K} - E_{ij} - qF_{IL}t_{IL} \quad (2\text{-}20)$$

$$q\phi_{high-K,an} = q\phi_{high-K} - E_{ij} - qF_{IL}t_{IL} - qF_{high-K}t_{high-K} \quad (2\text{-}21)$$

By Gauss' law we can get $V_{high\text{-}K}$ (potential drop across high-K dielectric) easily,

$$F_{IL} = \frac{V_{IL}}{t_{IL}} \tag{2-22}$$

$$F_{high-K} = \frac{V_{high-K}}{t_{high-K}} = F_{IL}\frac{\varepsilon_{IL}}{\varepsilon_{high-K}} \tag{2-23}$$

The three tunneling cases can be classified by defining the underlying conditions:

Case 1 --- $\phi_{IL,cath} > 0, \phi_{IL,an} > 0, \phi_{high-K,cath} > 0 \; and \; \phi_{high-K,an} > 0$

Case 2 --- $\phi_{IL,cath} > 0, \phi_{IL,an} > 0, \phi_{high-K,cath} > 0 \; and \; \phi_{high-K,an} \leq 0$

Case 3 --- $\phi_{IL,cath} > 0, \phi_{IL,an} > 0, \phi_{high-K,cath} \leq 0 \; and \; \phi_{high-K,an} \leq 0$

Substituting the three cases into (2-17), the three conditions of tunneling probability are

Case 1 ---

$$T_{WKB} = exp\left[\frac{4\sqrt{2m_{IL}}\left(\phi_{IL,an}^{\frac{3}{2}} - \phi_{IL,cath}^{\frac{3}{2}}\right)}{3q\hbar|F_{IL}|}\right] exp\left[\frac{4\sqrt{2m_{high-K}}\left(\phi_{high-K,an}^{\frac{3}{2}} - \phi_{high-K,cath}^{\frac{3}{2}}\right)}{3q\hbar|F_{high-K}|}\right]$$

Case 2 ---

$$T_{WKB} = exp\left[\frac{4\sqrt{2m_{IL}}\left(\phi_{IL,an}^{\frac{3}{2}} - \phi_{IL,cath}^{\frac{3}{2}}\right)}{3q\hbar|F_{IL}|}\right] exp\left[\frac{4\sqrt{2m_{high-K}}\left(0 - \phi_{high-K,cath}^{\frac{3}{2}}\right)}{3q\hbar|F_{high-K}|}\right]$$

Case 3 ---

$$T_{WKB} = exp\left[\frac{4\sqrt{2m_{IL}}\left(\phi_{IL,an}^{\frac{3}{2}} - \phi_{IL,cath}^{\frac{3}{2}}\right)}{3q\hbar|F_{IL}|}\right]$$

(2-24)

B. Modified Reflection Correction Factor for High-K Stacks (T_R)

According to the reflection correction factor for SiO_2, T_R needs to be considered at three interfaces. The correction factor of interface reflection typically approaches

unity for metal/high-K dielectric interface and for the interface of high-K gate stacks, so we only consider the reflection at the Si/IL interface in our model. Therefore, the correcting T_R for high-K stacks is as follows:

$$T_R = \frac{4\nu_{Si\perp}(E) \times \nu_{IL}(\phi_{IL,cath})}{\nu_{Si\perp}^2(E)+\nu_{IL}^2(\phi_{IL,cath})} \quad (2\text{-}25)$$

2.3 Subband Energy Calculation

In this part, the method of calculating the electron subband energy will be introduced. We employ a simplified method to calculate the quantum mechanical effect in the inversion layer of a p-type Si substrate. Four physical values are estimated in this method; they are the charge densities in depletion and inversion, the depletion region band bending, and the inversion layer band bending. Those characteristic parameters of p-type substrate under the inversion conditions are denoted

N_{dep} is the depleted space charge;

N_{inv} is the interfacial inversion charge;

ϕ_s is the semiconductor surface potential or surface band bending;

ϕ_{dep} is the band bending due to the depletion charge.

One assumption needs to be used to while simplifying the subband energy model: the quantum confinement phenomenon of the MOS structure can be treated in a triangular well approximation, so the potential at the semiconductor surface layer follows a linear variation. This approximation has been shown to describe the electron behavior adequately as validated by self-consistent Schrödinger and Poisson equations solving.

The gate voltage induces surface band bending; the inversion charge and depletion

charge each are related to corresponding surface band bending. The total charge per unit area (Q) below the MOS gate is the sum of the inversion electron charge(N_{inv}) and the depletion charge(N_{dep}). The total charge from the law of electrostatics is

$$N_{inv} + N_{dep} = Q(\phi_s) \tag{2-26}$$

First, the surface potential is essential for calculating the total charge.

$$\phi_{dep} = \phi_s - \frac{qN_s\bar{z}_{qm}}{\varepsilon_{Si}\varepsilon_0} - \frac{kT}{q} \tag{2-27}$$

The second term of the right side of (2-27) stems from the influence of inversion charge N_{inv}; the third term $\frac{kT}{q}$ from the gradual transition of the space charge region into the substrate, and the term ϕ_{dep} due to the space charge N_{dep}.

$$N_{dep} = \sqrt{\frac{2\varepsilon_{Si}\varepsilon_0\phi_{dep}N_{sub}}{q}} \tag{2-28}$$

$\bar{z}_{qm}$ in (2-27) is the average equivalent widths of the quantum confined electron gas, as described by

$$\bar{z}_{qm} = \sum_{i,j}\frac{z_{ij}N_{ij}}{N_{inv}} \tag{2-29}$$

z_{ij} is weighted with the corresponding subband occupation factor N_{ij}, thus constituting the mean quantum mechanical channel width $\bar{z}_{qm}$. In this case, the wave functions are given by Airy functions. So the mean subband width z_{ij} is calculated by

$$z_{ij} = \int|\Psi_{ij}(z)|^2 z\,dz = \frac{2E_{ij}\varepsilon_{Si}}{3q\varepsilon_{ox}E_{ox}} \tag{2-30}$$

The energy eigenvalues E_{ij} can be expressed as

$$E_{ij} = \left(\frac{\hbar^2}{2m_i}\right)^{\frac{1}{3}}\left[\frac{3\pi q\varepsilon_{ox}F_{ox}\left(j-\frac{1}{4}\right)}{2\varepsilon_{Si}}\right]^{\frac{2}{3}} \tag{2-31}$$

where m_i is the normal mass for two fold valley or four fold valley, F_{ox} is the oxide electric field and the index j is the subband number. N_{ij}, the remaining parameter in (2-29), can be derived from Fermi-Dirac statistics:

$$N_{ij} = \int_{E_{ij}}^{\infty} D_i(E) f(E) dE = \frac{m_{di}}{\pi\hbar^2} ln\left[1 + exp\left(\frac{E_F - E_{ij}}{kT}\right)\right] \quad (2\text{-}32)$$

where $D_i(E)$ is the density of states of the subband for two dimentional gas and $f(E)$ is the Fermi-Dirac occupation factor.

Chapter 3

Physical Model for FinFET

In this chapter, multiple gate devices are introduced, including gate tunneling current and electron energy in double gate and FinFET devices. Fig. 5 reveals that the width of FinFET is insignificant compared to the whole gate length ($W_{Fin} + 2H_{Fin}$), we can regard FinFET as double gate structure.

There are some changes to the tunneling model because of the different physical structural change from planar to double gate. The equations of depletion charge density and subband energy have to be modified since the Si body part is affected by both the front and back gate.

3.1 Depletion Charge Density Calculation

It is a significant improvement in the double gate structure that the substrate is controlled by two gates (front gate and back gate). Owing to its capability of sensitively influencing the channel, to suppress the short channel effects, the body doping can be decreased. While body thickness is reduced, full depletion in the body happens under the small gate voltage condition. So the task to calculate depletion charge density needs to be considered in two distinct situations:

Case 1 --- $If \sqrt{\frac{2\varepsilon_{Si}\varepsilon_0\phi_{dep}N_{sub}}{q}} < \frac{N_{sub}t_{body}}{2} \Longrightarrow N_{dep} = \sqrt{\frac{2\varepsilon_{Si}\varepsilon_0\phi_{dep}N_{sub}}{q}}$

Case 2 --- $If \sqrt{\frac{2\varepsilon_{Si}\varepsilon_0\phi_{dep}N_{sub}}{q}} > \frac{N_{sub}t_{body}}{2} \Longrightarrow N_{dep} = \frac{N_{sub}t_{body}}{2}$

Case 1 is the condition that the channel works in depletion region. In this case, the depletion charge density is calculated by the aforementioned equation (2-28). Case 2 is the condition that the device works in inversion region. In this case, the charges

attributed by the doping concentration are totally depleted. Therefore, the depletion charge density is equal to half of the body doping charge density. It is not the total charge density because one gate just controls half of the charge in the channel. As long as the body doping is large enough, full depletion is going to happen later due to large gate voltage needed.

3.2 Subband Energy Calculation for FinFET

In double gate devices, gate leakage current and subthreshold leakage both are affected by the quantization of electron energy. Owing to small body thickness, structure confinement needs to be considered in calculating electron energy. Surface band bending changes the electric field in the substrate, so field confinement is another factor [5],[6]. The energy level associated with the jth subband of the ith valley (longitudinal or transverse) is given by

$$E_{ij} = \left(\frac{\hbar^2}{2m_i}\right)^{\frac{1}{3}}\left[\frac{3\pi q\varepsilon_{ox}F_{ox}\left(j-\frac{1}{4}\right)}{2\varepsilon_{Si}}\right]^{\eta} + \frac{j^2(2\pi\hbar)^2}{8m_i t_{body}^2} \quad (4\text{-}1)$$

The first term in (4-1) is field confinement factor and the second is structure confinement one. η is a fitting factor (theoretically ≈ 2/3), m_i is the electron effective mass at *i*th valley and t_{body} is the thickness of body between front gate dielectric and back gate dielectric.

From Fig. 6, it can be observed that, due to the absence of the bulk charge in the double gate device, the surface electric field is negligible below threshold. Therefore, in the double gate device, the electron quantization occurs principally due to the structural confinement below threshold. However, above threshold, an increase in gate voltage increases the electric field (due to higher inversion charge) which also increases the subband energies [7].

The analytical models for E_{ij} closely follow the numerical simulations by Schred

(Fig. 7). Evidence to validate the analytical model is given in Fig. 8, which shows that the gate current density values obtained from the numerical simulation and the analytical model are equal. Fig. 9 shows the subband energy in varying body doping concentration, effective oxide thickness and metal work function. We can find that no matter how the characteristics of the device change, the subband energies from analytical model and numerical simulation are nearly identical to each other, valid in the same fitting factors. Only the altered body thickness can change fitting factors. We obtain the four subband fitting factors of 0.6715, 0.682, 0.67, and 0.68 for 10nm t_{body} (Fig. 10). These fitting factors all approach 2/3 as mentioned in [3],[5],[6].

The subband energy for body thicknesses from 10nm to 50nm all are closely comparable between analytical model and numerical simulation (Fig. 10). Fitting factor versus body thickness is plotted in Fig. 11. The fitting factors symbols for different oxide thicknesess represent the best fitting situations in Fig. 10. We can find that the fitting factor decreases as oxide thickness increases. It can be observed from Fig. 11 that data points are distinguished into two parts, with the first subband below second subband. All of the fitting factors lead to a linear relationship. Now, we have two sets of fitting factors, best fitting factors and linear fitting factors. The former is shown in Fig. 11 by discrete points and the latter by solid linear lines. Then, by using the fitting factors from the linear relation, the resulting subband energy does match the simulated value (Fig. 12). In order to testify the validity of the linear fitting factors, gate current from best factors and linear factors is shown in Fig. 13. The gate current curves in best and linear factors are almost equal as shown in Fig. 13(a) for body thicknesses of 10 and 30 nm. The gate current values for body thickness of 20nm, 40nm and 50nm are not shown explicitly but very small gate current change calculated with linear fitting with respect to the best fitting as shown in Fig. 13(b) versus body thickness. Hence, a compact model for double gate structure is

established by means of the linear equation of fitting factors shown in Fig. 11.

3.3 Parameters in FinFET Gate Tunneling Current Model

Transmission probability calculation is the most important part in this physical model. According to (2-24), $m_{IL}, m_{high-K}, \phi_{IL,an}, \phi_{IL,cath}, \phi_{high-K,an}, \phi_{high-K,cath}$ and F_{IL}, F_{high-K} are essential for resolving this equation. How to find the tunneling effective masses in interfacial layer (m_{IL}) and high-K layer (m_{high-K}) will be explained later. In order to obtain $\phi_{IL,an}, \phi_{IL,cath}, \phi_{high-K,an}$ and $\phi_{high-K,cath}$, ϕ_{IL} and ϕ_{high-K} are the two parameters used in the equations from (2-18) to (2-21). The electric field in interfacial layer (F_{IL}) and high-K dielectric (F_{high-K}) are defined in (2-22) and (2-23).

According to [8],[9], a method of extracting gate material parameters has been proposed, as long as gate current is dominated by direct tunneling or Fowler-Nordheim tunneling (F-N tunneling) from the plot of dlnIg/dVg versus Vg (Fig. 14). The transition of direct tunneling and F-N tunneling across high-K has a peak of dlnIg/dVg, and as a consequence, the position of the peak over can provide a direct estimate of metal work function and high-K electron affinity. From Fig. 14(b) and Fig. 14(c), ϕ_{high-K} determines the peak position; m_{high-K} determines the peak height. So, adjusting ϕ_{high-K} can shift the fitting curve of dlnIg/dVg versus Vg until the position of the peak is the same as the experimental value. Then, adjusting m_{high-K} can change the height of dlnIg/dVg peak until it approaches the experimental value. The parameters related to high-K are determined from the above method. Then, band offset with respect to silicon and the tunneling effective mass in interfacial layer, ϕ_{IL} and m_{IL}, can be extracted by fitting gate current versus gate voltage. Higher ϕ_{IL} decreases the gate current because the electric barrier height can significantly determine the tunneling probability. Larger ϕ_{IL} means that carriers

need more energy to tunnel across the interfacial layer, leading to gate tunneling current drop. Higher m_{IL} also decreases the gate current because the electric effective mass can significantly affect the tunneling probability as shown in equation (2-24). The differential value of the band offset between the anode side and cathode side is negative. It means that the higher the effective mass is, the less likely the tunneling occurs. We can therefore extract the relevant HKMG material parameters by combining conventional Cg-Vg and Ig-Vg curve fittings with the new dlnIg/dVg-Vg curve fitting.

Chapter 4

Metal-Gate/High-*K* FinFET: Experiment and Fitting

The experimental devices were n-type FinFET with metal-gate/high-K/IL system, as schematically depicted in Fig. 3 in terms of energy band diagram in flat-band condition. Because of the small ratio of top gate width to total fin width, we can regard the FinFET structure as double gate structure, as shown in Fig. 5.

The electron tunneling current from inversion layer was measured with source, drain and bulk tied to ground. The measured terminal current is shown in Fig. 15(a). The case of Vd = 0.05V is shown in Fig. 15(b). The following process parameters were obtained by Cg-Vg fitting using the Schrödinger-Poisson equation solver Schred [10], as depicted in Fig. 16. The results are that the metal work function Φ_m is 4.6eV, the EOT is 0.8nm, and the p-type substrate doping concentration N_{sub} is 1×10^{18} cm^{-3}. Then, we took the permittivity of HfO_2 bases high-K (ε_k) as the literature value of 22 ε_0 [11] and to meet EOT (0.8nm), the permittivity of IL (ε_{IL}) is determined to be 6.6 ε_0. Corresponding band offsets of IL (φ_{IL}) to silicon conduction and valence band are therefore 2.44 eV [12].

The gate current was measured with source, drain, and bulk tied to the ground. The measured result has been depicted in Fig. 15(a) versus Vg. In the gate current fitting, the actual electron tunneling current is close to the electron tunneling current from the first and second subband. For the purpose of saving the computation time, we just calculate the first and second subband in this work. Many simplifications are employed in our tunneling simulation, but the results are reasonable as compared with experimental results as depicted in Fig. 17. The tunneling model is constructed by using a simplified treatment of electron subband energy quantization and a modified

WKB approximation of electron transmission probability through the high-K stacks.

From the fitting result in Fig. 17, the simulation by our model is not correct for the gate bias below threshold voltage. The reason for this failure can be attributed to the fact that the gate current is not dominated by direct tunneling in threshold region but our tunneling current model is limited to pure tunneling. Another drawback is that a serious deviation occurs at high gate voltage. This is due to the very large difference of permittivity between high-K dielectric (22 ε_0) and IL (6.6 ε_0). Thus, a gradual transition (intermixing) layer between high-K dielectric and IL needs to be considered in the calculation. The tunneling effective masses of transition layer can be assumed to be linear or parabolic type, as schematically plotted in Fig. 18. The current fitting results are shown in Fig. 19. We find that fitting quality can be improved with the transition layer included, especially for the parabolic one. Although the simulator cannot work in subthreshold region, WKB approximation still show good agreements with the experimental curve even when the gate bias is large enough. The following discussion focuses on the range of the current closely falling the range of gate bias larger than the threshold voltage.

Now, all the model parameters are known, except ϕ_{high-K}, m_{IL} and m_{high-K}. In order to find the parameters related to high-K dielectric, dlnIg/dVg fitting is used again. As mentioned in Chapter 3.3, ϕ_{high-K} and m_{high-K} dominate the position of the peak and the height of the peak in the dlnIg/dVg versus Vg, respectively. ϕ_{high-K} is adjusted first until the position of the dlnIg/dVg peak approaches the experimental data ($\sim$1.5V). Next, m_{high-K} is changed until the height of the dlnIg/dVg peak closely meets the experimental data ($\sim 7V^{-1}$). The fitting result is shown in Fig. 20 and the extracted results are $\phi_{high-K} = 1.07\ eV$ and $m_{high-K} = 0.02m_0$.

Next, we determine the parameters related to interfacial layer (IL) by the Ig-Vg

fitting. The experimental and fitting gate current without transition layer are plotted in Fig. 17 and Fig. 19 with linear and parabolic transition layer. In the Ig-Vg fitting, the gate current does not change by adjusting the value of ϕ_{high-K} or m_{high-K} in case 3 (only direct tunneling through the IL). No matter how the high-K dielectric parameters change, gate tunneling current is not affected at high gate voltage.

Chapter 5

Conclusion

FinFET gate current model has been established by modifying the gate current model for metal-gate/high-K device. Because of the small ratio between top width and the whole width, double gate tunneling current model can be applied to FinFET. A compact model for the structure that is similar to double gate devices has been established.

As the width of FinFET is insignificant compared to the whole gate length, the fitting factor of electron subband energy has regularity in the region of body thickness between 10nm and 50nm. When the body thickness is fixed in double gate structure, the fitting factor does not need to adjust. The model can straightforwardly yield the correct subband energy. For the case of varying body thickness, we can obtain a reasonable fitting factor from the linear relationship. Conventionally, Ig-Vg and Cg-Vg curve fittings are used to extract the parameters. Because of the apparent deviation of gate current fitting at high voltage, transition layer between the high-K dielectric and the interfacial layer is incorporated. Linear and parabolic-type transition layer are adopted in the model. By taking into account the transition layer, the gate tunneling current fitting and dlnIg/dVg fitting can be significantly improved. It is found that parabolic transition layer can produce a better agreement with experiment than linear transition layer.

A peak of dlnIg/dVg indicates a transition of direct tunneling and F-N tunneling across a high-K part, and as a consequence, the position of the dlnIg/dVg peak over Vg can provide a direct estimate of the parameters in high-K. To accurately extract the material and process parameters in the metal-gate high-K dielectrics, we have

systematically constructed a new fitting scheme over the dlnIg/dVg versus Vg curve, along with the combination of Cg-Vg and Ig-Vg fittings. In addition, we have demonstrated that the conventional method without the dlnIg/dVg fitting might lead to erroneous results, Thus, the dlnIg/dVg fitting should be taken into account in the assessment of the metal-gate high-K material parameters in FinFET devices.

References

[1] Amitava DasGupta, “Multiple Gate MOSFETs: The Road to Future,” in *Proceedings of the 2007 International Workshop on the Physics of Semiconductor Devices (IWPSD)*, pp. 96-101, Sep. 2007.

[2] Leonard F. Register, Elyse Rosenbaum, and Kevin Yang, “Analytic model for direct tunneling current in polycrystalline silicon-gate metal-oxide-semi-conductor devices,” *Appl. Phys. Lett.*, vol. 74, no. 3, pp. 457-459, Jan. 1999.

[3] Kuo-Nan Yang, Huan-Tsung Huang, Ming-Chin Chang, Che-Min Chu, Yuh-Shu Chen, Ming-Jer Chen, Yeou-Ming Lin, Mo-Chiun Yu, Simon M. Jang, Douglas C. H. Yu, and M. S. Liang, “A physical model for hole direct tunneling current in p^+ poly-gate PMOSFETs with ultrathin gate oxides,” *IEEE Trans. Electron Devices*, vol. 47, no. 11, pp. 2161-2166, Nov. 2000.

[4] Yijie Zhao and Marvin H. White, “Modeling of direct tunneling current through interfacial oxide and high-*K* gate stacks,” *Solid-State Electronics*, *vol.48*, no. 10-11, pp. 1801-1807, Dec. 2003.

[5] Saibal Mukhopadhyay, Keunwoo Kim, Ching Te Chuang, and Kaushik Roy, “Modeling and Analysis of Leakage Currents in Double-Gate Technologies,” *IEEE Trans. Computer–Aided Design of Integrated Circuits and Systems*, vol. 25, no. 10, pp. 2052-2061, Oct. 2006.

[6] Saibal Mukhopadhyay, Keunwoo Kim, Jae-Joon Kim, Shih-Hsien Lo, Rjiv V. Joshi, Ching-Te Chuang, and Kaushik Roy, “Estimation of gate-to-channel tunneling current in ultra-thin oxide sub-50 nm double gate devices,” *Microelectronics Journal*, vol. 38, no. 8-9, pp. 931-941, Jan. 2006.

[7] Leland Chang, Kevin J. Yang, Yee-Chia Yeo, Igor Polishchuk, Tsu-Jae King, and Chenming Hu, “Direct-Tunneling Gate Leakage Current in Double-Gate and Ultrathin Body MOSFETs,” *IEEE Trans. Electron Devices*, vol. 49, no. 12, pp. 2288-2295, Dec. 2002.

[8] Sufi Zafar, Cyril Cabral, Jr., R. Amos, and A. Callegari, “A method for measuring barrier heights, metal work functions and fixed charge densities in metal/SiO_2/Si capacitors,” *Appl. Phys. Lett.*, vol. 80, no. 25, pp. 4858-4860, Jun. 2002.

[9] Chih-Yu Hsu, Hua-Gang Chang, and Ming-Jer Chen, “A Method of Extracting Metal-Gate High-*k* Material Parameters Featuring Electron Gate Tunneling Current Transition,” *IEEE Trans. Electron Devices*, vol. 58, no. 4, pp. 953-959, Apr. 2011.

[10] Schred. [Online]. Available: http://nanohub.org/resources/schred

[11] Y. T. Hou, M. F. Li, H. Y. Yu, and D. L. Kwong, “Modeling of tunneling currents through HfO_2 and $(HfO_2)_x(Al_2O_3)_{1-x}$ gate stacks,” *IEEE Electron Device Lett.*, vol. 24, no. 2, pp. 96-98, Feb. 2003.

[12] Hongyu Yu, Yong-Tian Hou, Ming-Fu Li, and Dim-Lee Kwong, “Investigation of Hole Tunneling Current Through Ultrathin Oxynitride/Oxide Stack Gate Dielectrics for p-MOSFETs,” *IEEE Trans. Electron Devices*, vol. 49, no. 7, pp. 1158-1164, Jul. 2002.

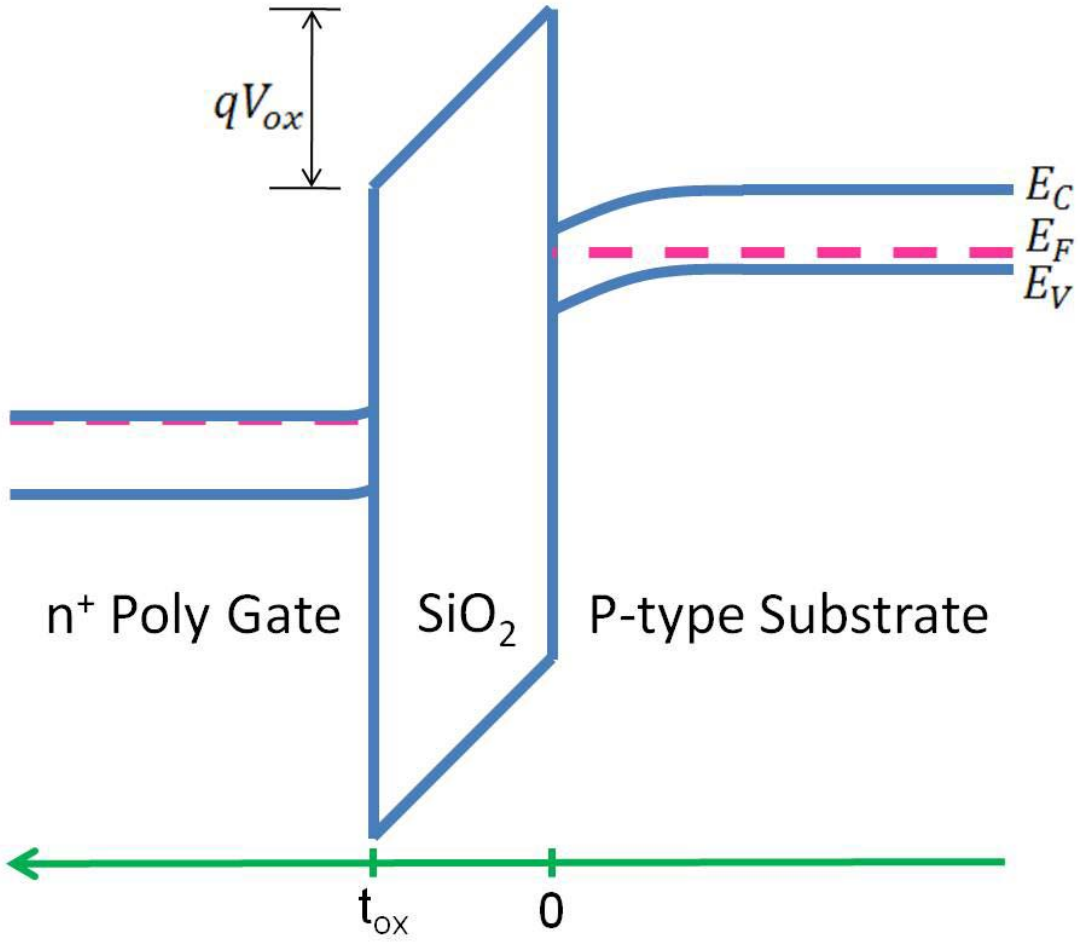


Fig. 1 Schematic of energy band diagram of the n^+ poly-gate/SiO_2/p-Si system.

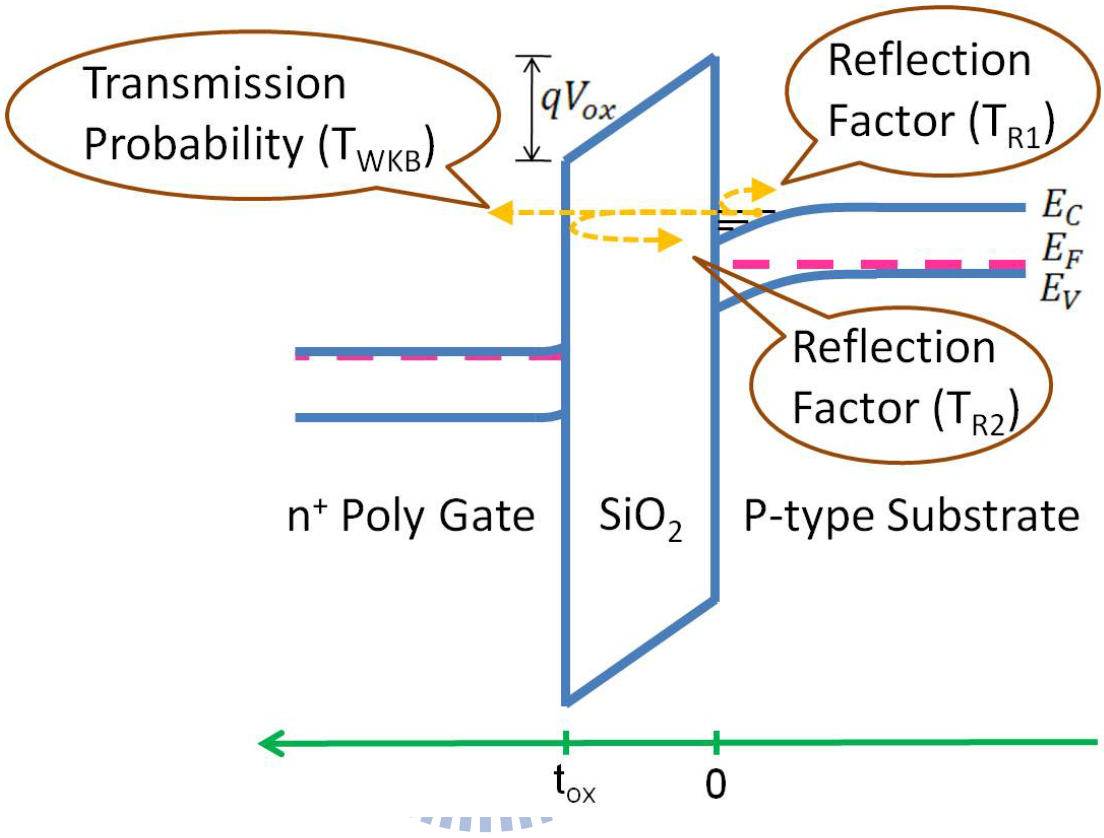


Fig. 2 Tunneling probability illustration.

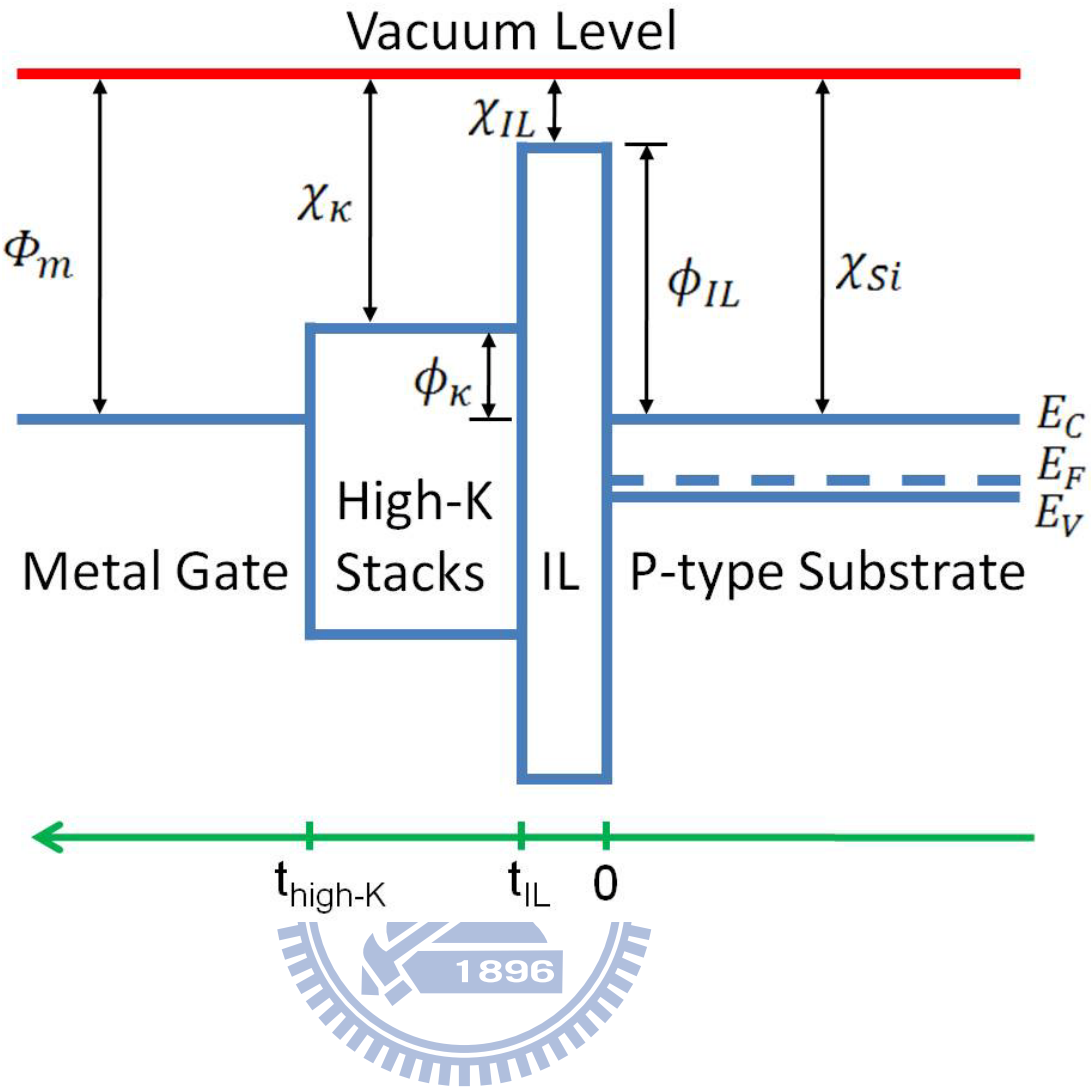


Fig. 3 Schematic of energy band diagram of the metal-gate/high-K/IL/p-Si system.

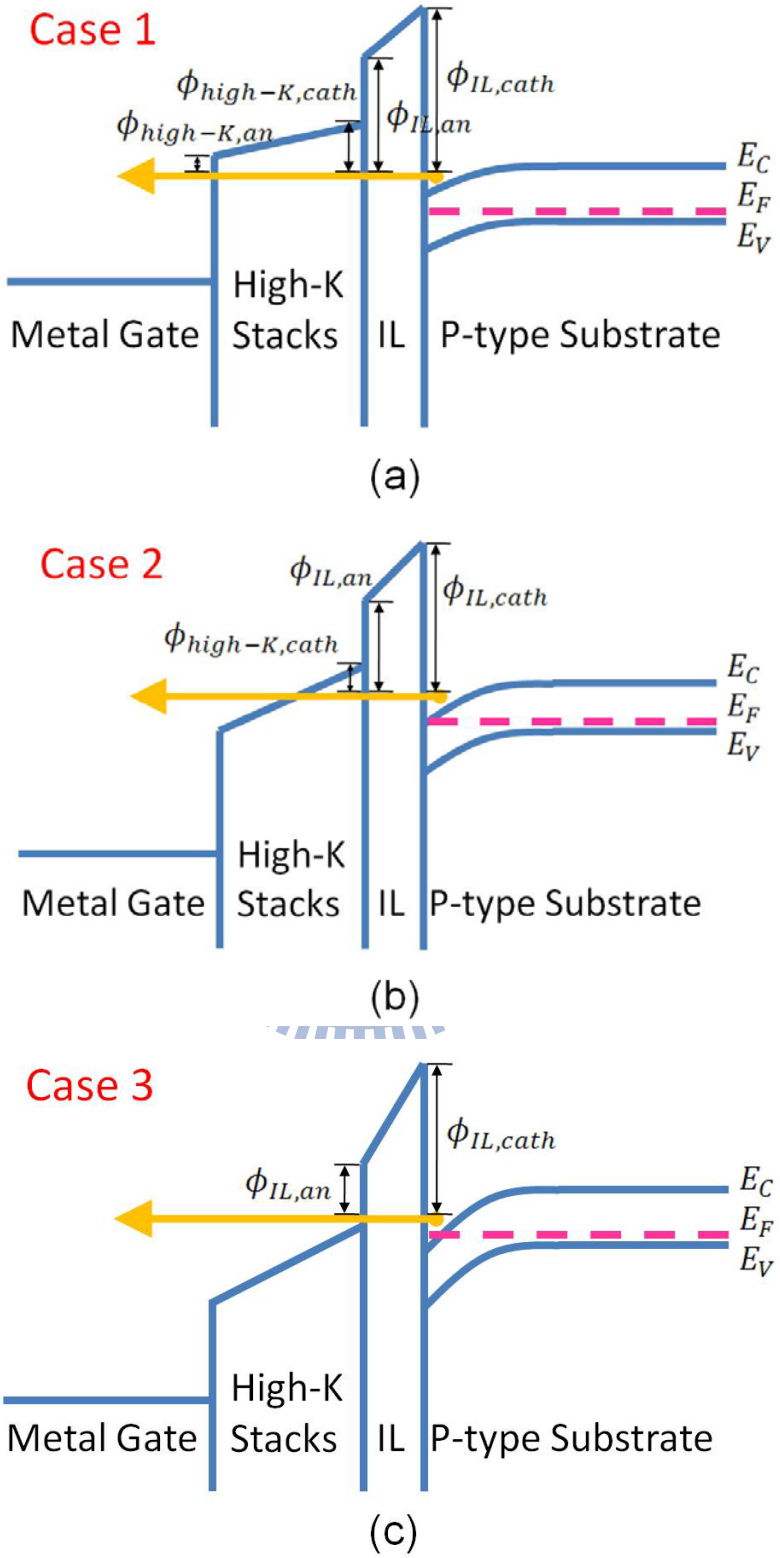


Fig. 4 Schematic description of three cases. (a)Case 1: direct tunneling through the two layers. (b)Case 2: F-N tunneling through the high-K stacks. (c)Case 3: direct tunneling through the IL

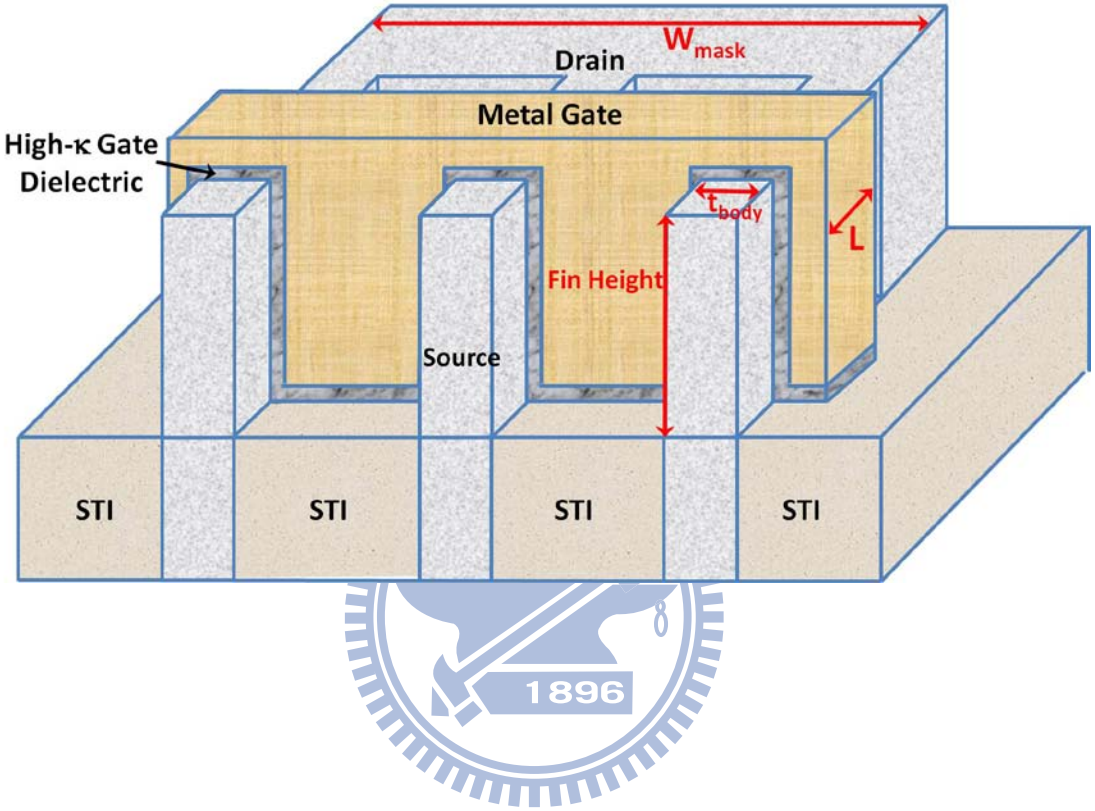
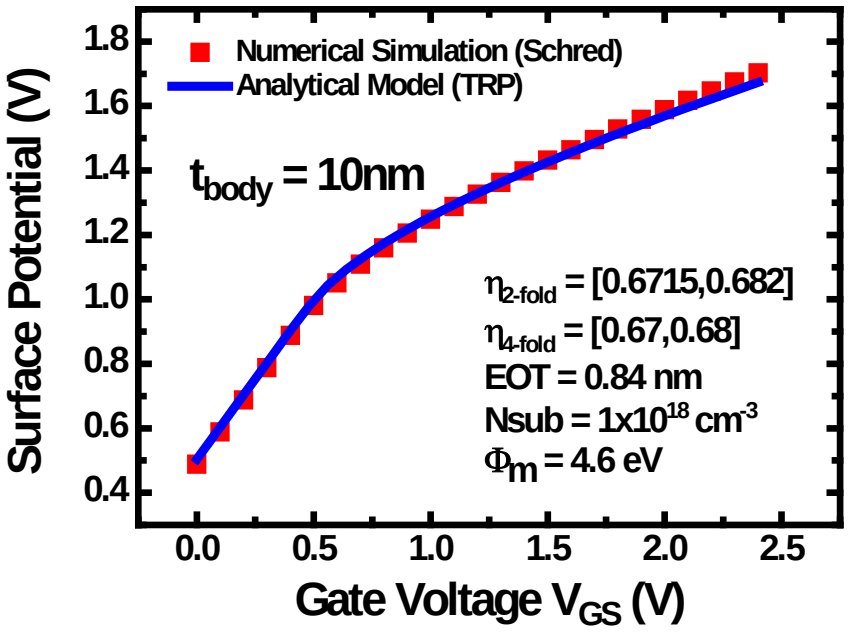
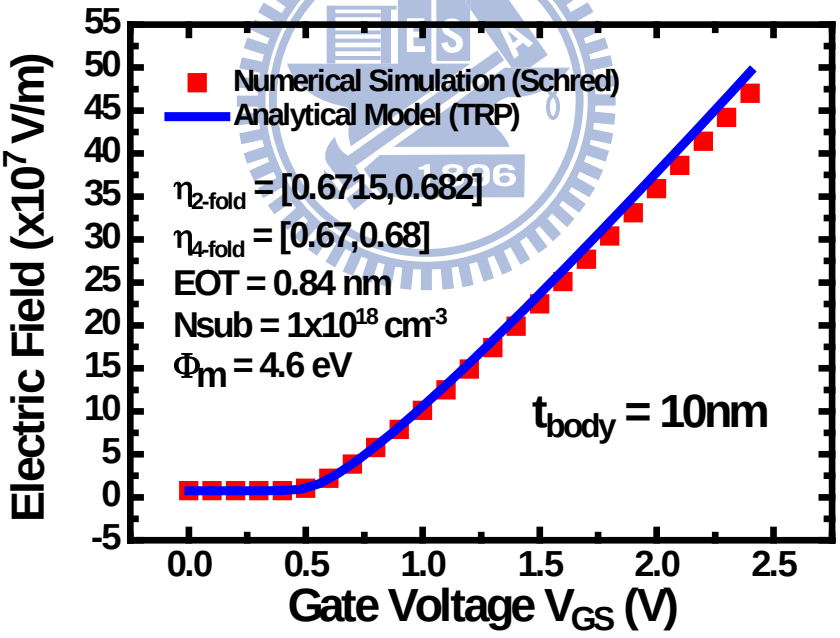


Fig. 5 Schematic of cross-sectional view of FinFET. FinFET with small fin width is like a double gate structure.

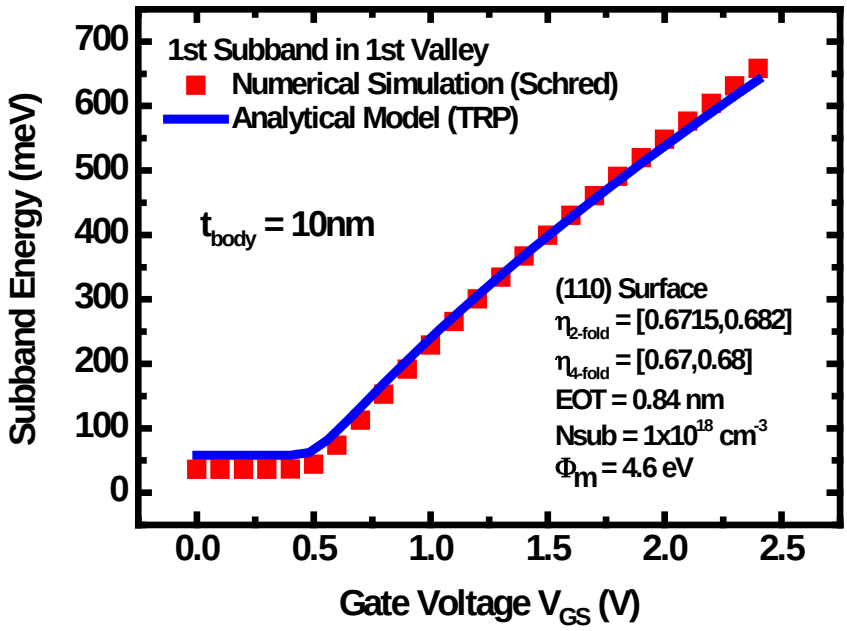


(a)

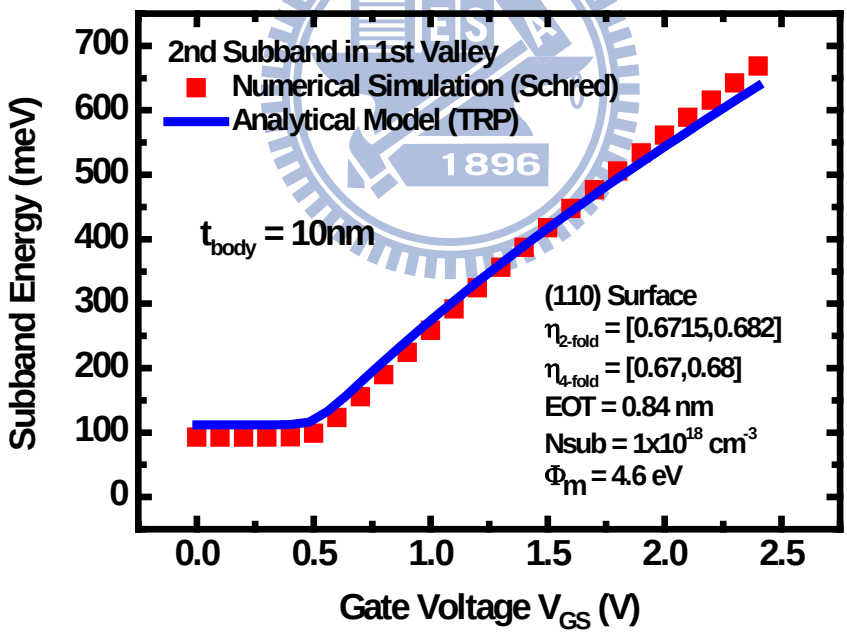


(b)

Fig. 6 Comparison of the self-consistent Schrödinger-Poisson simulation and analytical model for (a)surface potential and (b)electric field.

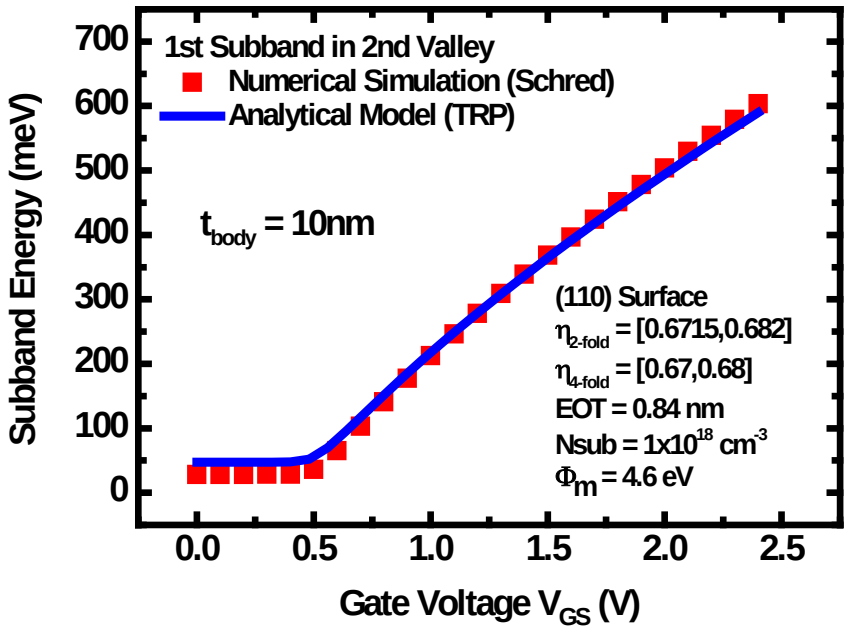


(a)

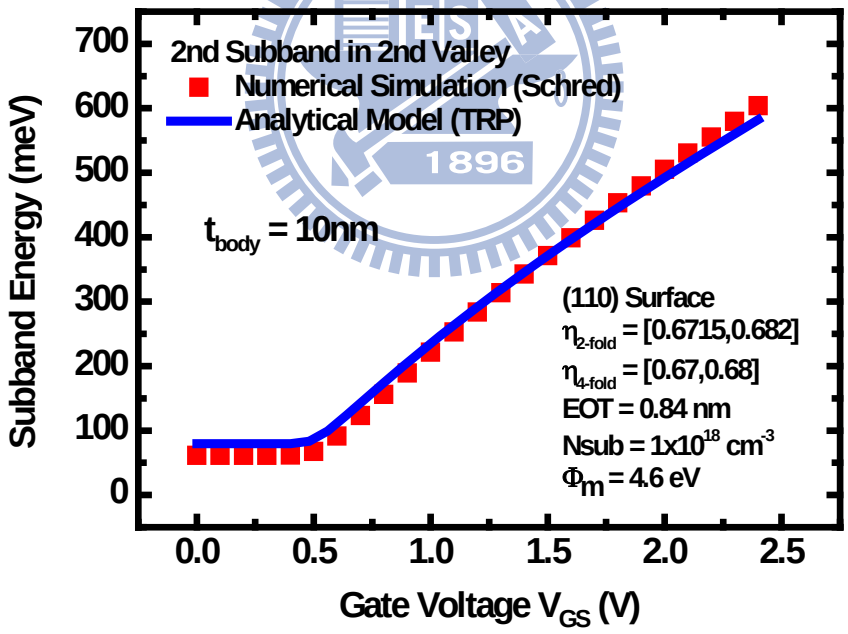


(b)

Fig. 7-1 The analytically calculated (lines) subband energy versus gate voltage and the comparison with those (symbols) of the self-consistent Schrödinger-Poisson simulation. (a)E(1,1); (b)E(2,1); (c)E(1,2); (d)E(2,2).



(c)



(d)

Fig. 7-2 The analytical calculated (lines) subband energy versus gate voltage and the comparison with those (symbols) of the self-consistent Schrödinger-Poisson simulation. (a)E(1,1); (b)E(2,1); (c)E(1,2); (d)E(2,2).

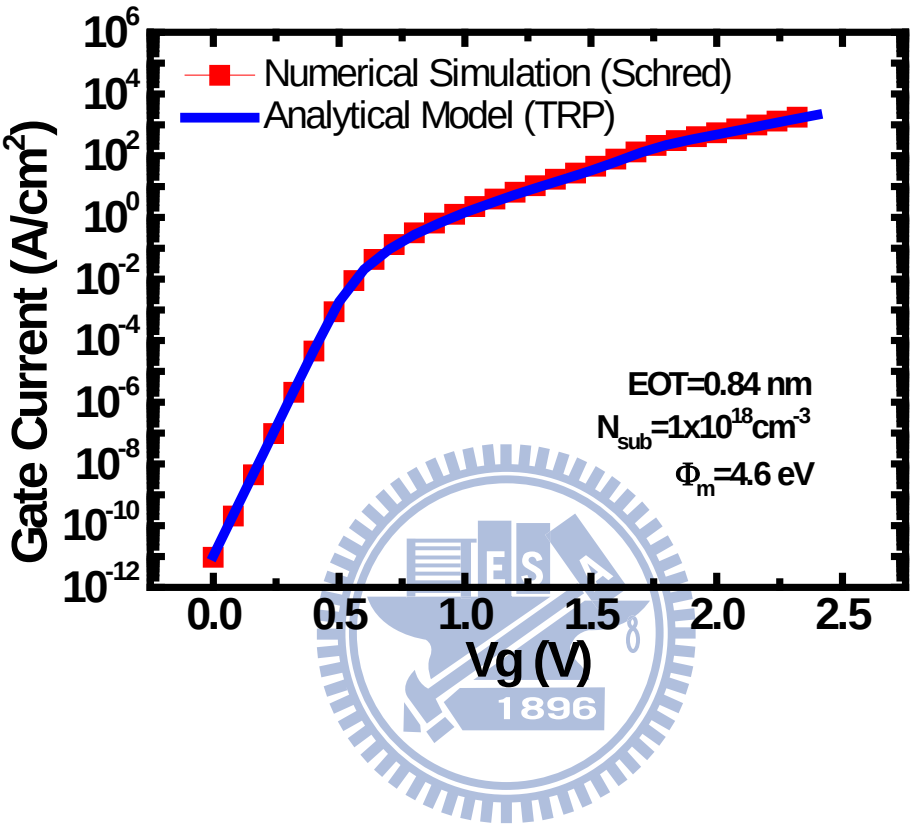


Fig. 8 Comparison of gate tunneling current versus gate voltage from the self-consistent Schrödinger-Poisson (symbols) simulation and analytical model (lines).

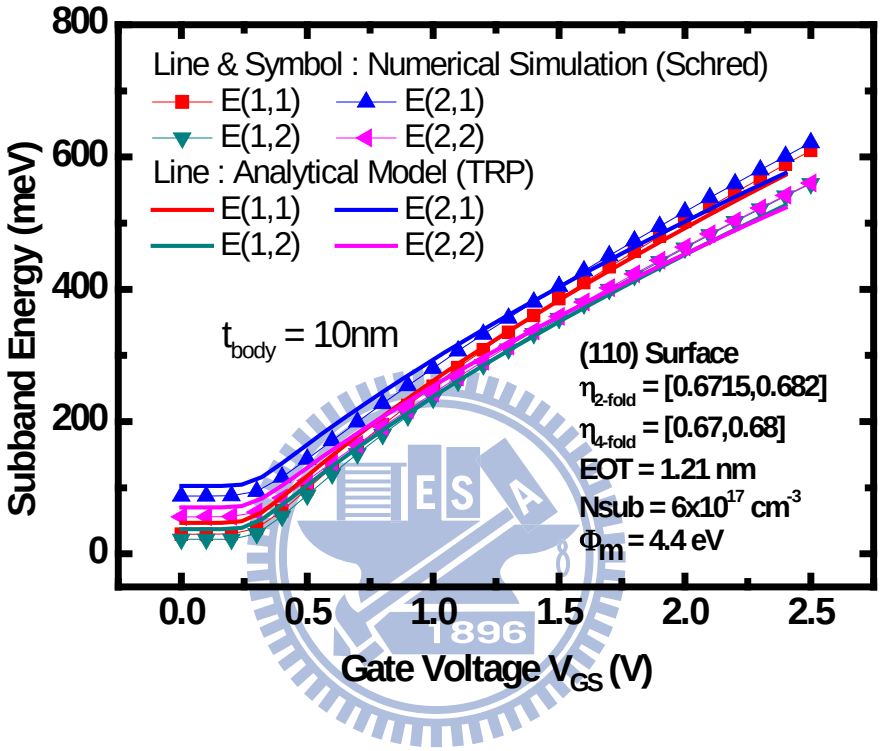
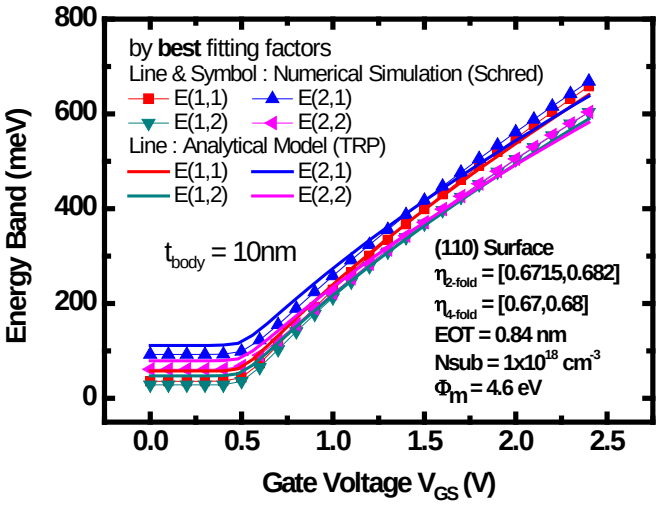
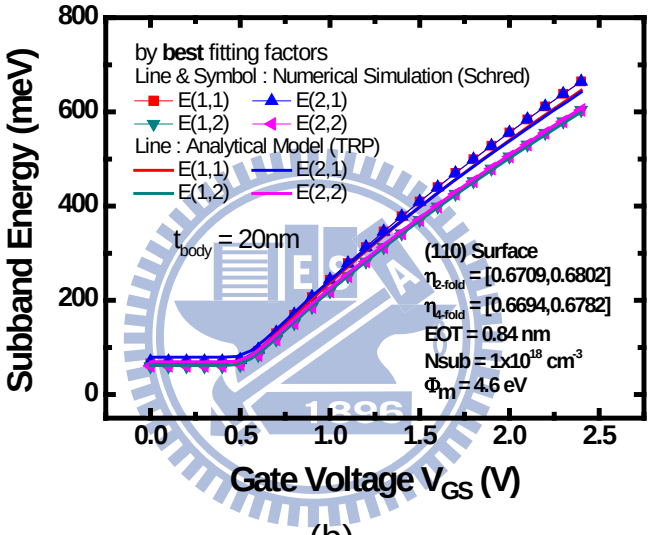


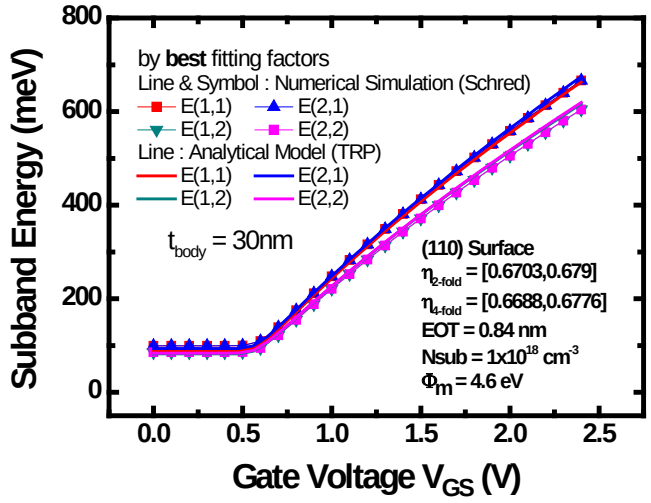
Fig. 9 The analytically calculated (lines) subband energy versus gate voltage and the comparison with those (symbols) of the self-consistent Schrödinger-Poisson simulation for EOT = 1.21nm, Φ_m = 4.4 eV, N_{sub} = 6x10^{17} cm^{-3}, and t_{body} = 10 nm.



(a)

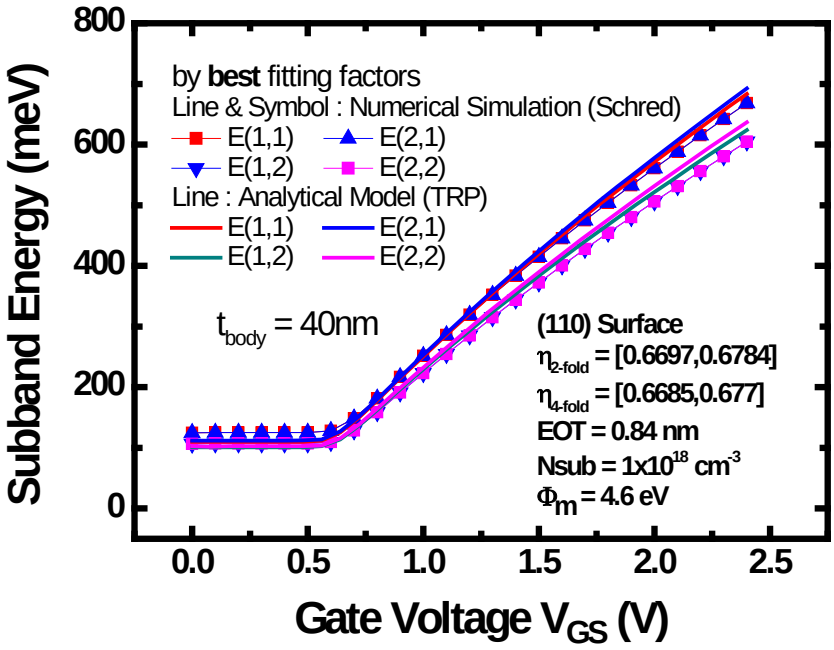


(b)

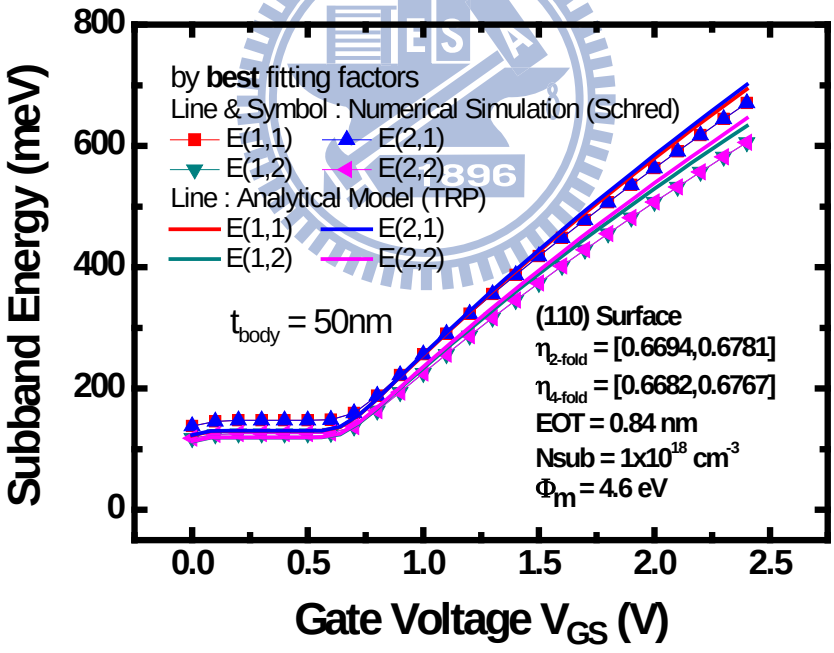


(c)

Fig. 10-1 The subband energy (best fitting) for different body thicknesses. (a)t_{body}=10nm; (b)t_{body}=20nm; (c)t_{body} = 30nm; (d)t_{body} = 40nm; (e)t_{body} = 50nm.



(d)



(e)

Fig. 10-2 The subband energy (best fitting) for different body thicknesses. (a)t_{body}=10nm; (b)t_{body}=20nm; (c)t_{body} = 30nm; (d)t_{body} = 40nm; (e)t_{body} = 50nm.

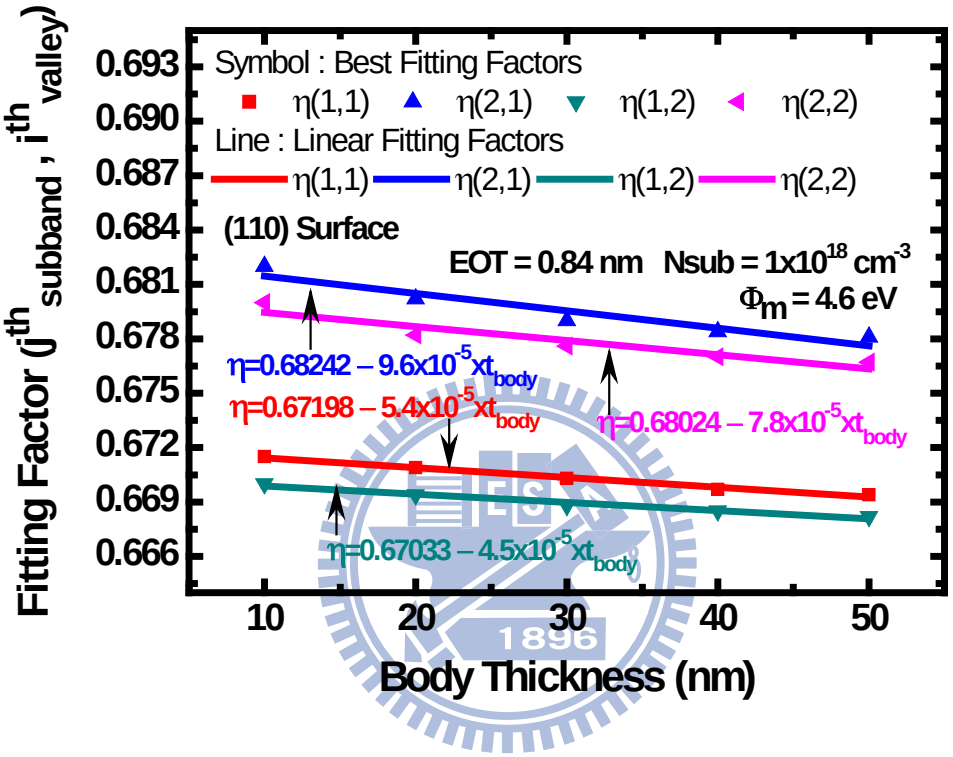
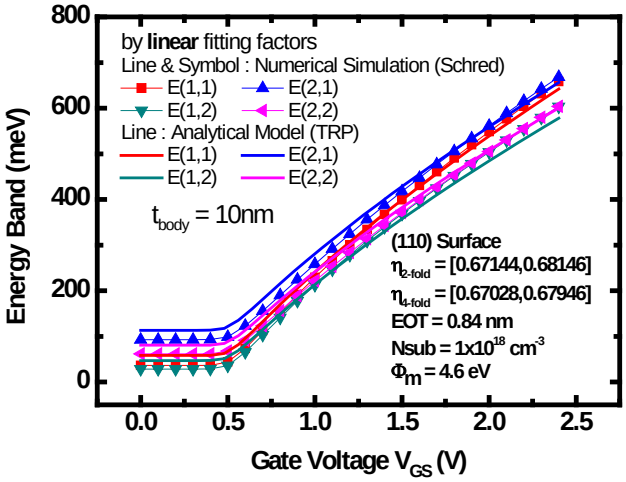
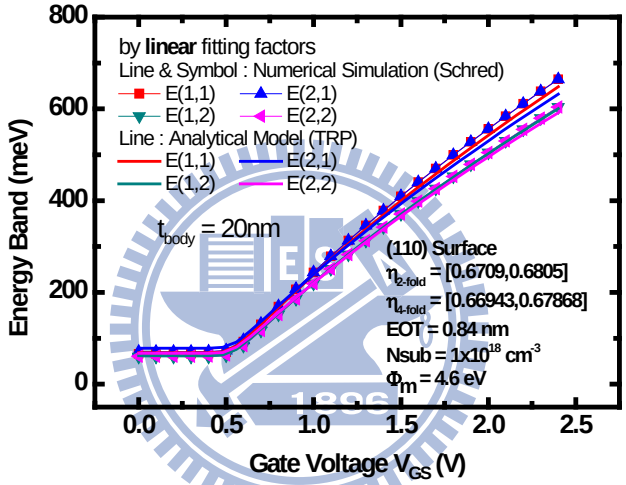


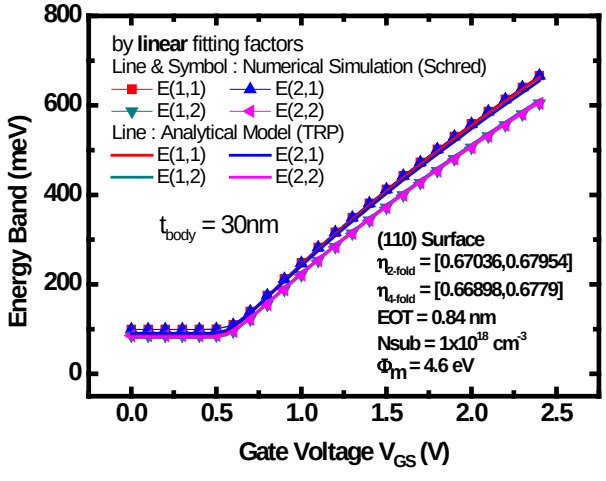
Fig. 11 Subband fitting factors versus body thickness. Symbols are the values for best fitting and lines are for linear fitting.



(a)

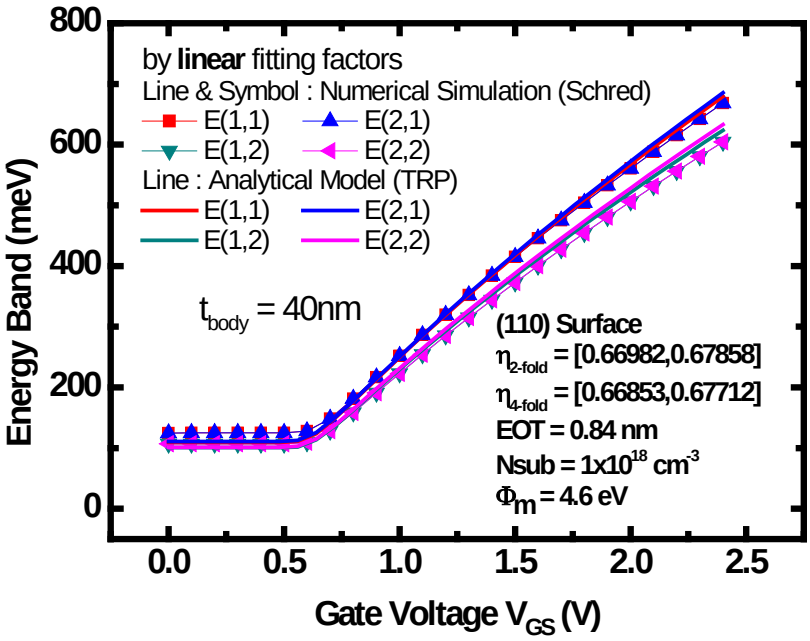


(b)

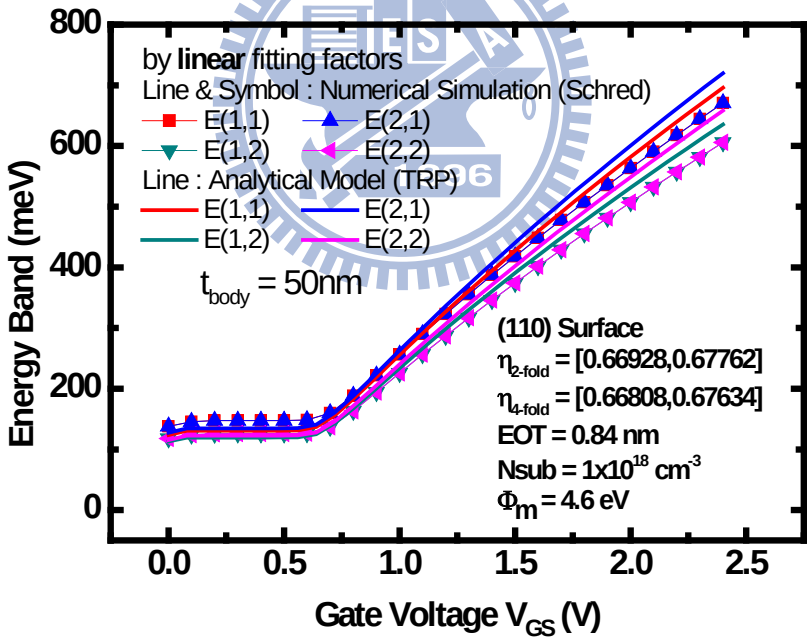


(c)

Fig. 12-1 The case of subband energy (linear relationship) for different body thicknesses. (a)t_{body}=10nm; (b)t_{body}=20nm; (c)t_{body} = 30nm; (d)t_{body} = 40nm; (e)t_{body} = 50nm.



(d)



(e)

Fig.12-2 The case of subband energy (linear relationship) for different body thicknesses. (a)t_{body}=10nm; (b)t_{body}=20nm; (c)t_{body} = 30nm; (d)t_{body} = 40nm; (e)t_{body} = 50nm.

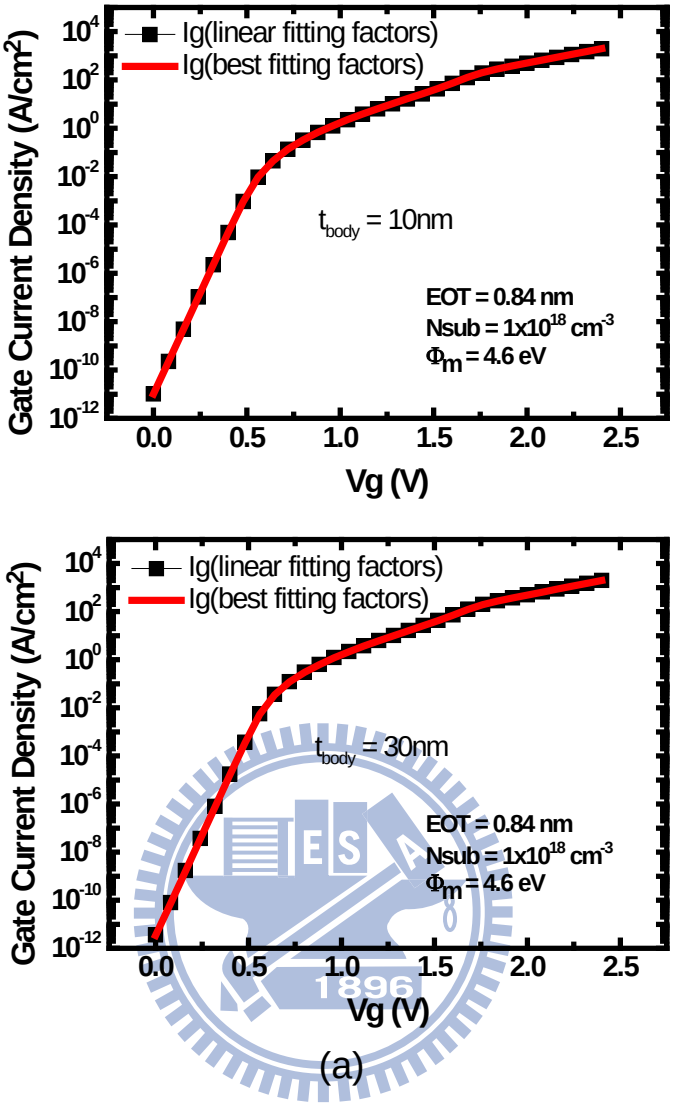
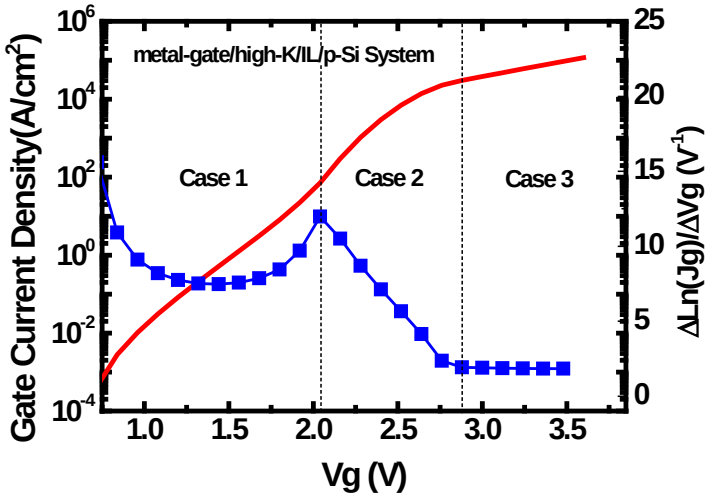
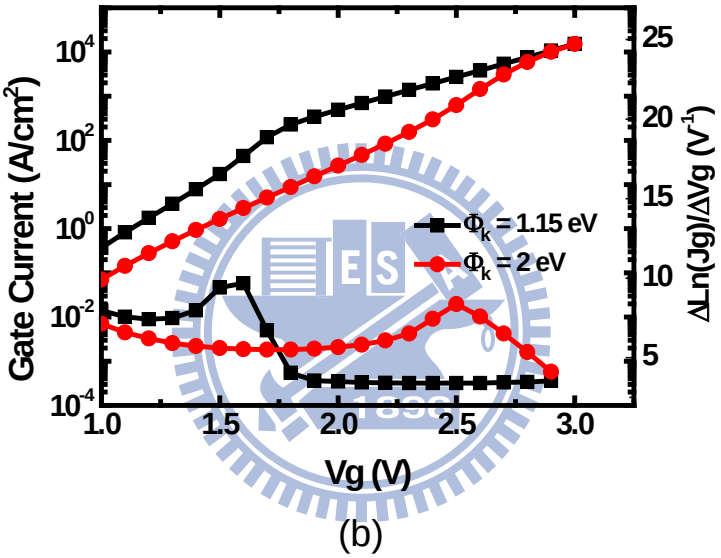


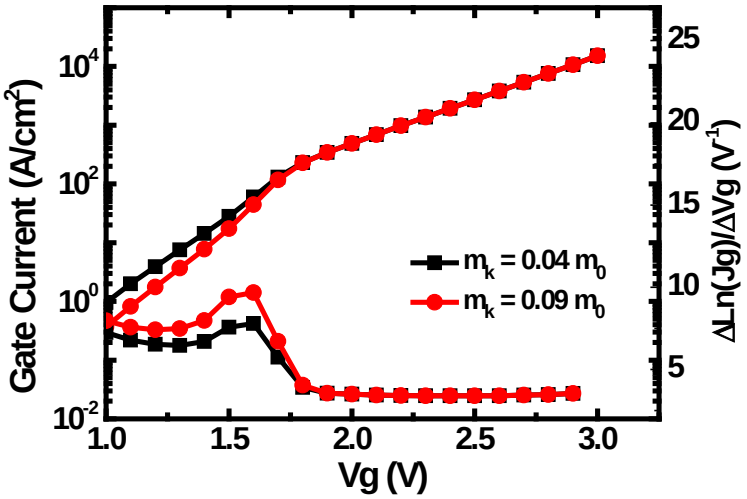
Fig. 13 (a) Comparison of Ig(linear factors) with Ig(best factors) for t_{body}=10nm and t_{body}=30nm. (b) Gate current change by linear and best factors.



(a)



(b)



(c)

Fig. 14 (a) The demonstration of three different tunneling regions. And Ig and dlnIg/dVg versus Vg for (b) varying m_k (c) varying Φ_k.

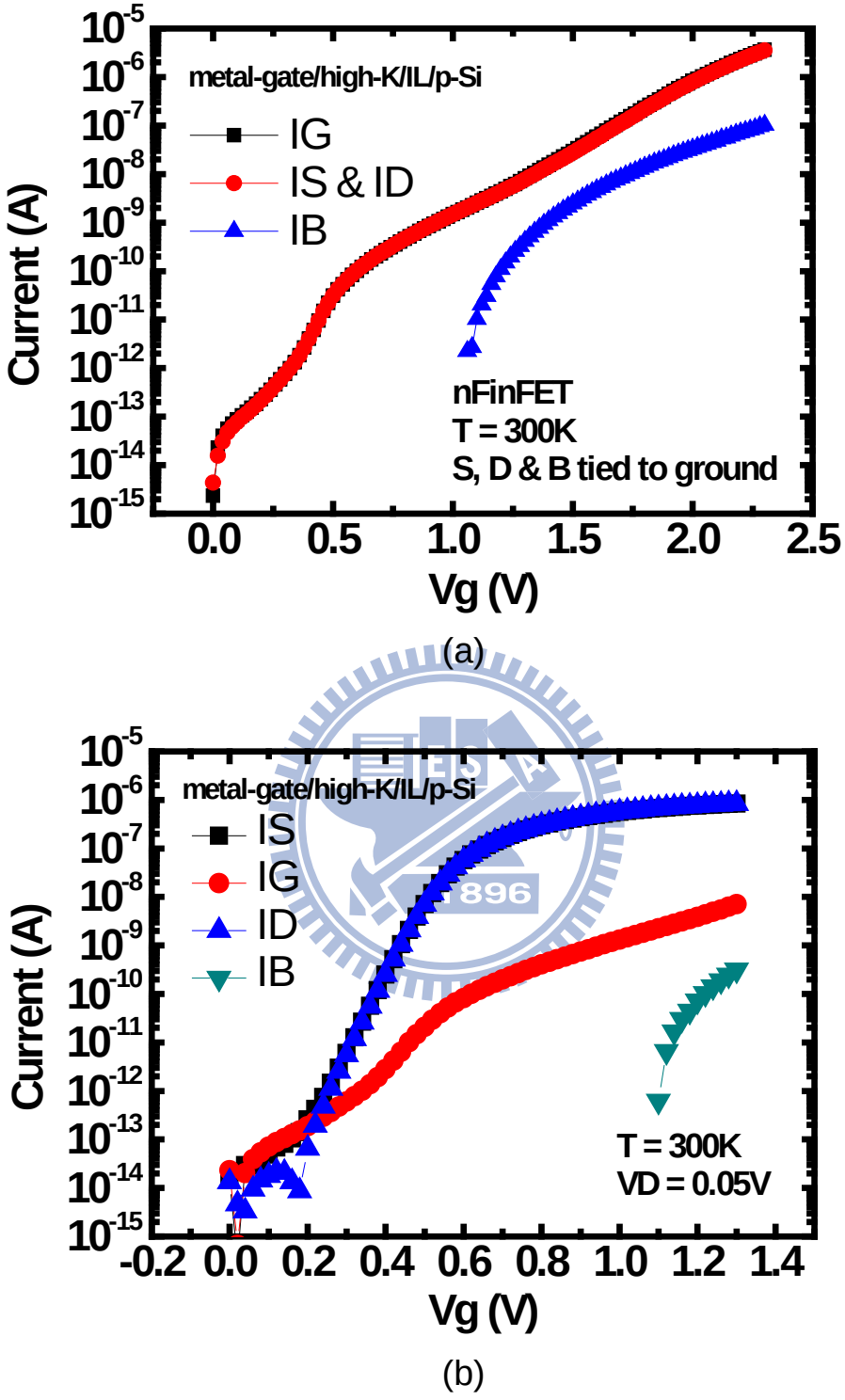


Fig. 15 The measured terminal current versus gate voltage for nFinFET. (a) terminal current with source, drain and bulk tied to ground and (b) terminal current at Vd = 0.05 V.

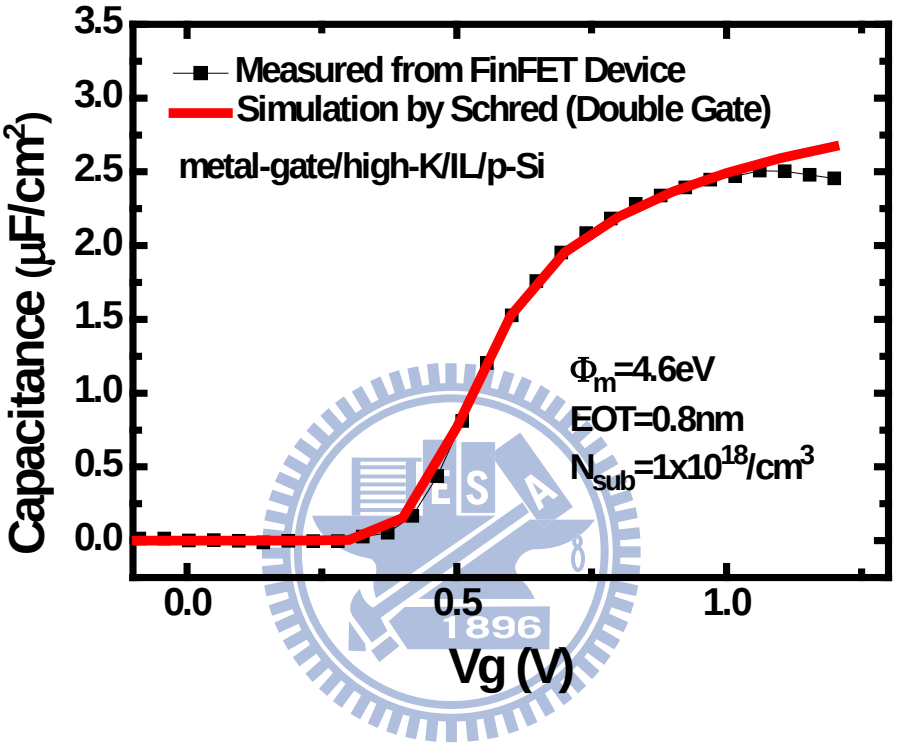


Fig. 16 Experimental (symbol) and simulated (line) C_g versus V_g for n-type FinFET. The extracted process parameters are EOT = 0.8 nm, Φ_m = 4.6 eV and N_{sub} = 1×10^{18} cm^{-3}

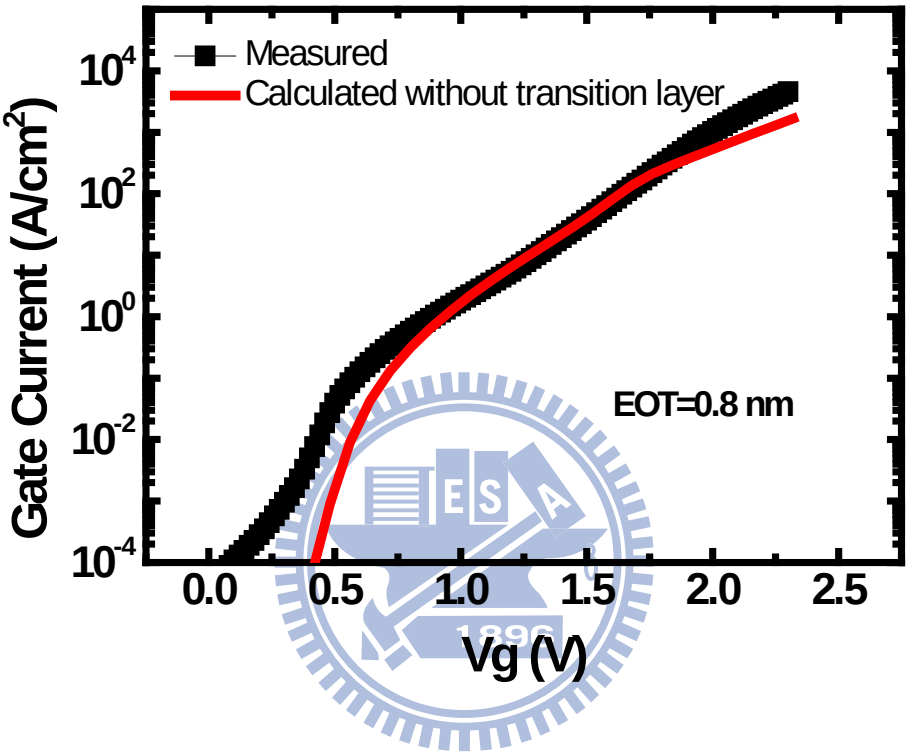


Fig. 17 Comparison of experimental (symbols) electron gate current with calculated (lines) results. Fitting parameters are φ_k = 1.07 eV, m_k = 0.02 m_o, m_{IL} = 1.22 m_o, t_k = 1.2 nm, and t_{IL} = 1 nm.

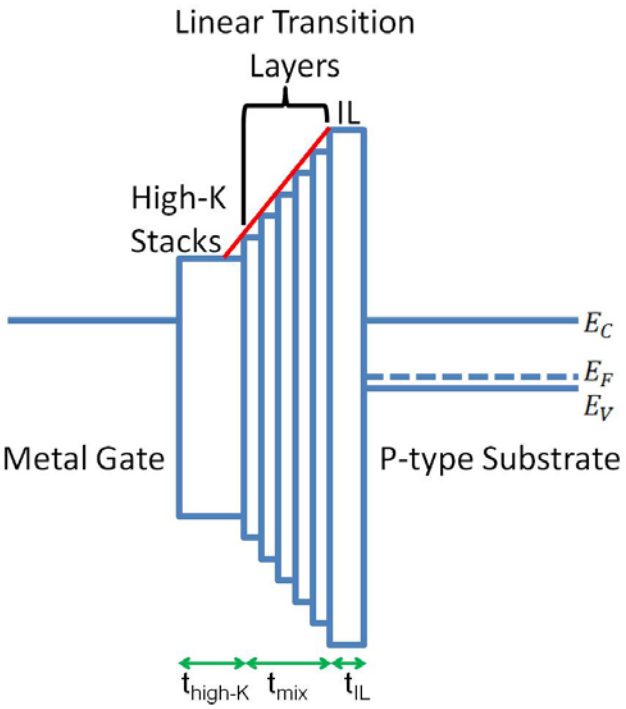


Fig. 18 Schematic of the energy band diagram for (a) a linear gradual transition layer and (b) a parabolic gradual transition layer.

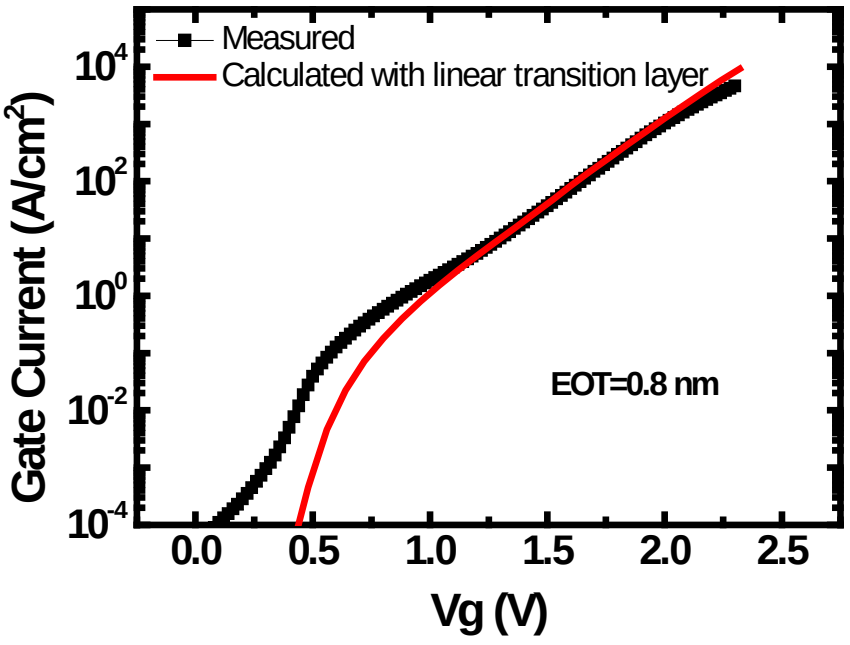


Fig. 19 Comparison of experimental (symbols) electron gate current with calculated (lines) results in the presence of a transition layer. (a) Linear gradual transition layer, φ_k = 1.07 eV, m_k = 0.02 m_o, m_{IL} = 0.8 m_o, t_k = 0.3nm , t_{mix} = 1.31 nm, and t_{IL} = 0.6 nm. (b) Parabolic gradual transition layer, φ_k = 1.07 eV, m_k = 0.02 m_o, m_{IL} = 1.39 m_o, t_k = 0.4 nm , t_{mix} = 1 nm, and t_{IL} = 0.59 nm.

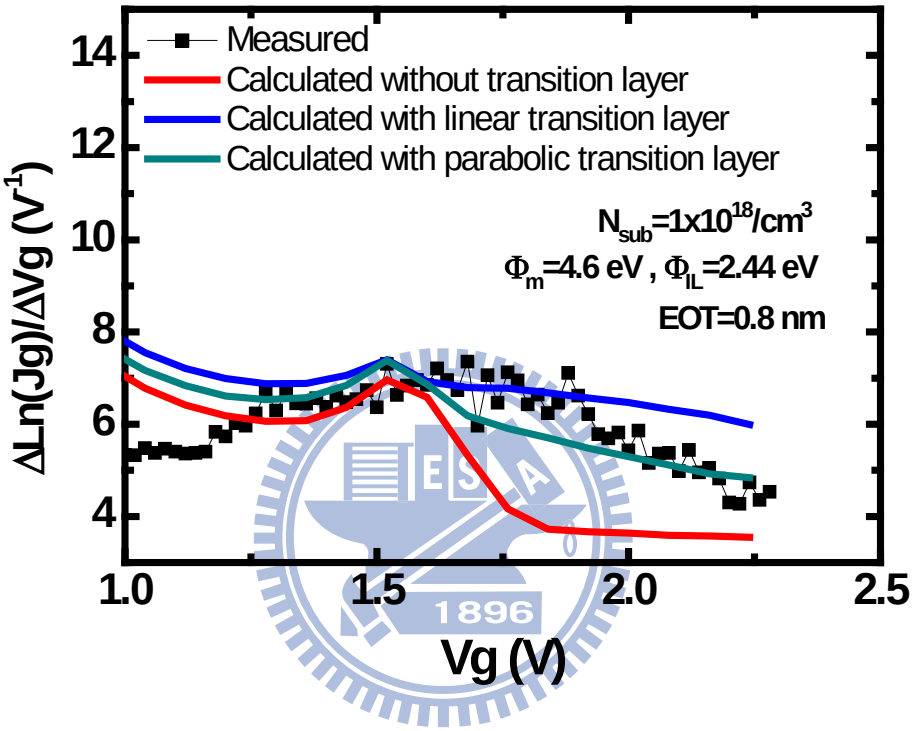


Fig. 20 Comparison of experimental (symbols) $d\ln I_g/dV_g$ versus V_g with calculated (lines) results in the presence of a transition layer. The fitting parameters are: (I)for no transition layer, m_k = 0.02 m_0, m_{IL} = 1.22 m_0, t_k = 1.2 nm, and t_{IL} = 1 nm; (II) for linear transition layer, m_k = 0.02 m_0, m_{IL} = 0.8 m_0, t_k = 0.3 nm, t_{mix} = 1.31 nm, and t_{IL} = 0.6 nm; (III) for linear transition layer, m_k = 0.02 m_0, m_{IL} = 1.39 m_0, t_k = 0.4 nm, t_{mix} = 1 nm, and t_{IL} = 0.59 nm


E S A
1896