Title
Non-classical MOSFETs : design, modeling, and characterization

Author
Yuan, Yu

Publication Date
2012

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Non-Classical MOSFETs: Design, Modeling, and Characterization

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in

Electrical Engineering (Applied Physics)

by

Yu Yuan

Committee in charge:

Professor Yuan Taur, Chair
Professor Peter Asbeck, Co-Chair
Professor Prabhakar Bandaru
Professor Chung-Kuan Cheng
Professor Yuhwa Lo

2012

The dissertation of Yu Yuan is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Co-Chair

Chair

University of California, San Diego

2012

DEDICATION

This dissertation is dedicated to my deceased father-in-law Hongquan Yu, who passed away on March 25, 2011.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Many people have provided help during my graduate studies at UCSD. First and foremost, I am greatly indebted to my advisor, Professor Yuan Taur, for his advice and guidance throughout the journey of my studies and research. I have been inspired not only by his deep insight, but also the way he thinks. Without his patience, support and encouragement, all the works in this dissertation would not have been possible. I would like to express gratitude to my co-advisor, Professor Peter Asbeck, who patiently guided me into III-V device research. His vast knowledge benefited me throughout the joint SRC program on Non-classical CMOS. His perseverance towards scientific research set a great example for me.

I would like to thank my committee members, Professors Prabhakar Bandaru, Chung-Kuan Cheng, and Yu-Hwa Lo, for taking their time to serve as committee members, to review my dissertation and give me valuable comments. I would like to acknowledge Professor Paul McIntyre at Stanford University, for the data they provided. A large part of this dissertation is based on experimental data. Many thanks to my collaborators in Prof. Taur's group, in particular, Jooyoung Song, and Han-Ping Chen, as well as Lingquan Wang in Prof. Asbeck's group. I appreciate my friends who made my life more joyful than it would have been without them. I am not able to cover all of them due to the limited pages.

Last, but not least, I would like to express gratitude to my family. Thank my mother in-law, who has been taking care of me during the completion of the

dissertation. Thank my parents and my sister, who are always there when I need support and encouragement. Special thanks go to my husband, for his continual love and all-around support, which have gone beyond what words can describe.

The text of Chapter II, in part, is a reprint of the material as it appears in "A distributed bulk-oxide trap model for Al_2O_3-InGaAs MOS devices" by Yu Yuan, Bo Yu, Jaesoo Ahn, Paul C. McIntyre, Peter M. Asbeck, Mark J. W. Rodwell and Yuan Taur, submitted to IEEE Transactions on Electron Devices. The dissertation author was the primary investigator and author of this paper.

The text of Chapter IV, in part, is a reprint of the material as it appears in "An analytic model for threshold voltage shift due to quantum confinement in surrounding gate MOSFETs with anisotropic effective mass" by Yu Yuan, Bo Yu, Jooyoung Song, and Yuan Taur, Solid State Electronics, 2009. The dissertation author was the primary investigator and author of this paper.

The text of Chapter IV, in part, is a reprint of the material as it appears in "A two-dimensional analytical solution for short-channel effects in nanowire MOSFETs" by Bo Yu, Yu Yuan, Jooyoung Song, and Yuan Taur, IEEE Transaction on Electron Devices, 2009. The dissertation author was a co-author of this paper.

The text of Chapter IV, in part, is a reprint of the material as it appears in "Scaling of nanowire transistors" by Bo Yu, Lingquan Wang, Yu Yuan, Peter Asbeck, and Yuan Taur, IEEE Transaction on Electron Devices, 2008. The dissertation author was a co-author of this paper.

VITA

2004	B.S., University of Science and Technology of China
2009	M.S., University of California, San Diego
2012	Ph.D., University of California, San Diego

PUBLICATIONS

Y. Yuan, B. Yu, J. Ahn, P. C. McIntyre, P. M. Asbeck, M. J. W. Rodwell, and Y. Taur, "A distributed bulk trap model for Al_2O_3-InGaAs MOS devices," submitted to *IEEE Trans. Electron Devices*.

Y. Yuan, L. Wang, B. Yu, B. Shin, J. Ahn, P. C. McIntyre, P. M. Asbeck, M. J. W. Rodwell, and Y. Taur, "A distributed model for border traps in Al_2O_3-InGaAs MOS devices," *IEEE Electron Device Letters*, vol. 32, pp. 485-487, Apr. 2011.

Y. Yuan, L. Wang, B. Yu, B. Shin, J. Ahn, P. C. McIntyre, P. M. Asbeck, M. J. W. Rodwell, and Y. Taur, "A distributed model for border traps in Al_2O_3-$In_{0.53}Ga_{0.47}As$ MOS devices," in *SRC Techcon*, 2011.

J. Ahn, M. Gunji, I. Geppert, Y. Yuan, Y. Taur, M. Eizenberg, and P. C. McIntyre, "TiO_2/Al_2O_3 bilayer dielectrics as gate oxide for $In_{0.53}Ga_{0.47}As$ metal oxide semiconductor devices," in *MRS Spring Meeting*, 2011.

J. Song, Y. Yuan, B. Yu, W. Xiong and Y. Taur, "Compact modeling of experimental n- and p-channel FinFETs," *IEEE Trans. Electron Devices*, vol. 57, pp. 1369-1374, Jun. 2010.

Y. Yuan, B. Yu, J. Song, and Y. Taur, "An analytic model for threshold voltage shift due to quantum confinement in surrounding gate MOSFETs with anisotropic effective mass," *Solid State Electron.*, vol. 53, pp. 140-144, 2009.

B. Yu, Y. Yuan, J. Song, and Y. Taur, "A two-dimensional analytical solution for short-channel effects in nanowire MOSFETs," *IEEE Trans. Electron Devices*, vol. 56, pp. 2357-2362, Oct. 2009.

J. Song, B. Yu, Y. Yuan, and Y. Taur, "A review on compact modeling of multiple-gate MOSFETs," *IEEE Trans. Circuits and System I: Regular Papers*, vol. 56, pp. 1858-1869, 2009.

U. Singisetti, M. A. Wistey, G. J. Burek, A. K. Baraskar, B. J. Thibeault, A. C. Gossard, M. J. W. Rodwell, B. Shin, E. J. Kim, P. C. McIntyre, B. Yu, Y. Yuan, D. Wang, Y. Taur, P. Asbeck, and Y.-J. Lee, "$In_{0.53}Ga_{0.47}As$ channel MOSFETs with self-aligned InAs source/drain formed by MEE regrowth," *IEEE Electron Device Letters*, vol. 30, pp. 1128-1130, Nov. 2009.

E. J. Kim, E. Chagarov, J. Cagnon, Y. Yuan, A. C. Kummel, P. M. Asbeck, S. Stemmer, K. C. Saraswat, and P. C. McIntyre, "Atomically abrupt and unpinned $Al_2O_3/In_{0.53}Ga_{0.47}As$ interfaces," *J. Appl. Phys.*, vol. 106, 124508, 2009.

B. Yu, L. Wang, Y. Yuan, P. M. Asbeck, and Y. Taur, "Scaling of nanowire transistors," *IEEE Trans. Electron Devices*, vol. 55, pp. 2846-2858, Nov. 2008.

B. Yu, J. Song, Y. Yuan, W.-Y. Lu, and Y. Taur, "A unified analytic drain-current model for multiple-gate MOSFETs," *IEEE Trans. Electron Devices*, vol. 55, pp. 2157-2163, Aug. 2008.

M. J. W. Rodwell, M. Wistey, U. Singisetti, G. Burek, A. Gossard, S. Stemmer, R. Engel-Herbert, Y. Hwang, Y. Zheng, C. Van de Walle, P. Asbeck, Y. Taur, A. Kummel, B. Yu, D. Wang, Y. Yuan, C. Palmstrom, E. Arkun, P. Simmonds, P. McIntyre, J. Harris, M. V. Fischetti, and C. Sachs, "Technology development & design for 22 nm InGaAs/InP-channel MOSFETs," in *IPRM*, 2008, pp. 1-6.

M. Rodwell, M. Wistey, U. Singisetti, G. Burek, A. Gossard, S. Stemmer, R. Engel-Herbert, Y. Hwang, Y. Zheng, C. Van de Walle , P. Asbeck, Y. Taur, A. Kummel, B. Yu, D. Wang, Y. Yuan, C. Palmstrom, E. Arkun, P. Simmonds, P. McIntyre, J. Harris, M. V. Fischetti, C. Sachs, "Device architecture and processing for III-V MOS technology," in *Workshop on Germanium and III-V MOS Technology, European Solid-State Device Research Conference (ESSDERC)*, Edinburgh, UK, 2008.

FIELDS OF STUDY

Major Field: Electrical Engineering

Studies in Applied Physics

Professors Yuan Taur and Peter Asbeck

ABSTRACT OF THE DISSERTATION

Non-Classical MOSFETs: Design, Modeling, and Characterization

by

Yu Yuan

Doctor of Philosophy in Electrical Engineering (Applied Physics)

University of California, San Diego, 2012

Professor Yuan Taur, Chair

Professor Peter Asbeck, Co-Chair

Low power and high density requires scaling of MOSFETs in VLSI. As the Si based bulk MOSFETs scale down to the limit imposed by gate oxide tunneling induced gate leakage, short channel effects (SCEs) induced loss of control on electrostatic integrity, high body doping induced high V_t variation, and band-to-band tunneling induced high substrate leakage, etc., two categories of novel MOSFETs are being intensively investigated: Si multiple gate MOSFETs and high mobility III-V material based MOSFETs. Among all types of Si multiple gate MOSFETs, nanowire MOSFET is drawing quite a few attentions for its superior

electrostatic control through all-around gate structure. High mobility III-V MOSFETs are considered as a principal candidate to achieve high speed without too aggressive scaling, which can keep good control of electrostatic integrity. This dissertation is primarily devoted to modeling and characterization of challenges and features which are becoming pronounced in aggressively scaled MOSFETs and high mobility material based MOSFETs. High-κ dielectric on III-V MOS capacitors are intensively characterized and modeled with the focus on defects at insulator-semiconductor interface as well as inside the oxide, which are grand challenges for III-V MOSFETs. A distributed bulk-oxide trap model is developed to account for the commonly observed frequency dispersion of small signal capacitance-voltage and conductance-voltage data in accumulation and near flat band region. The observed C-V humps in depletion to strong inversion are modeled by interface states model. For III-V MOSFETs design, SCEs and raised source/drain issues are studied using TCAD simulation. Fabricated III-V MOSFETs are characterized and mobility is extracted through experimental current voltage data and multiple frequency gate to channel capacitance measurement and data. For multiple gate Si MOSFETs, this dissertation focuses on nanowire MOSFETs. SCEs based on generalized scale length theory are discussed and compact models are proposed and validated by TCAD simulation. Quantum confinement effects on V_t shift in nanowire MOSFETs with anisotropic effective mass are modeled. Scaling limit is projected for extremely scaled nanowire MOSFETs based on V_t shift sensitivity and scale length theory. Finally, inversion layer capacitance beyond the conven-

tional bulk Si-based MOSFETs is investigated for III-V MOSFETs as well as two typical 3-D transistors, namely symmetric double-gate MOSFETs and nanowire MOSFETs.

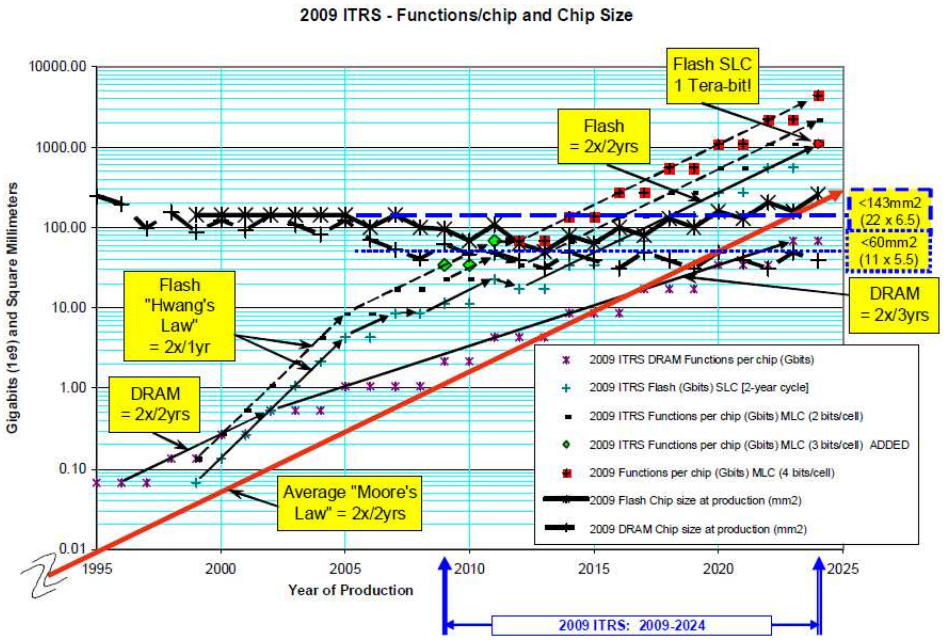
Chapter I

Introduction

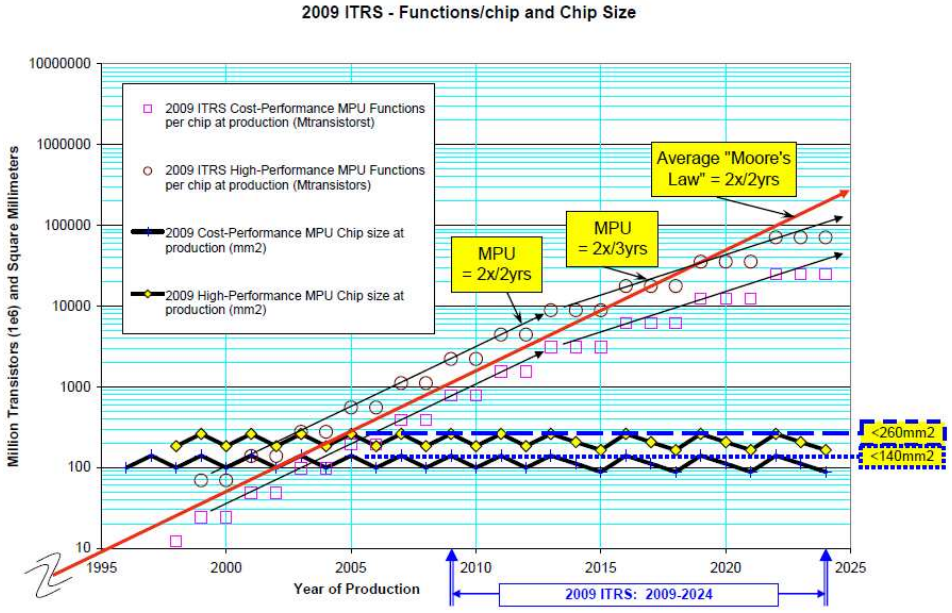
I.1 A Review of CMOS Technology in the Nano Era

The number of transistors on a chip roughly doubles every two years [1]. This is the famous "Moore's Law" which is followed by the semiconductor industry for many decades. Fig. I.1 shows the technology trends for both memory and microprocessor unit (MPU) products summarized by International Technology Roadmap for Semiconductors (ITRS) in 2009 [2], in comparison with "Moore's Law". It is predicted that, after 2013, the "Moore's Law" rate of on-chip transistors for MPU slows to 2× every three years, to match the slower 3-year technology cycle. How could the semiconductor industry maintain the "Moore's Law" for decades? But why will the pace slow down after 2013?

The primary driving source for the "Moore's Law" is the continued scaling down of CMOS transistors to ever smaller physical dimensions. As shown in Fig.



(a)



(b)

Figure I.1: 2009 ITRS product technology trends: (a) memory product functions/chip and industry average "Moore's Law" and chip size trends; (b) MPU product functions/chip and industry average "Moore's Law" and chip size trends. Adapted from [2].

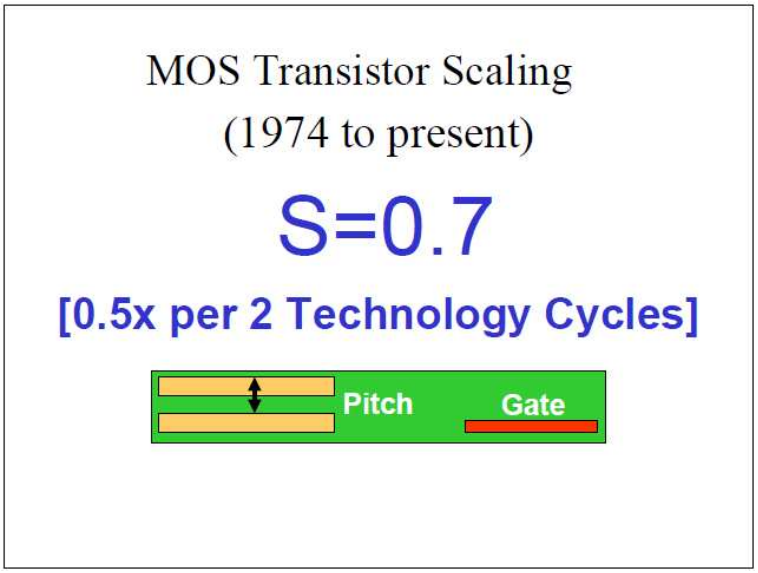


Figure I.2: MOS transistor scaling — 1974 to present: 0.5 shrink factor per two technology cycle. Adapted from [2].

I.2, the half pitch or gate length has always been reduced by half per 2 technology cycle since year 1974. In other words, the CMOS scaling factor between two adjacent technology nodes is about 0.7. Following this historical trend, the CMOS technology entered nano era in early 2000s, and Intel has been holding the leadership of CMOS scaling during this period. Fig. I.3 demonstrates the roadmap of Intel since 90 nm technology node. In may of 2011, Intel announced and demonstrated the first 22 nm microprocessor, codenamed "Ivy Bridge", that will be the first high-volume chip to use 3-D transistors. With this major breakthrough and historic innovation in microchips, Intel continued to successfully introduce leading edge process and products on a 2 year cadence [3]. It is indicated in Fig. I.3 that every step in the nanotechnology era is not straightforward. Besides 3-D transistor innovation, other breakthroughs such as strained silicon and high-k metal gate

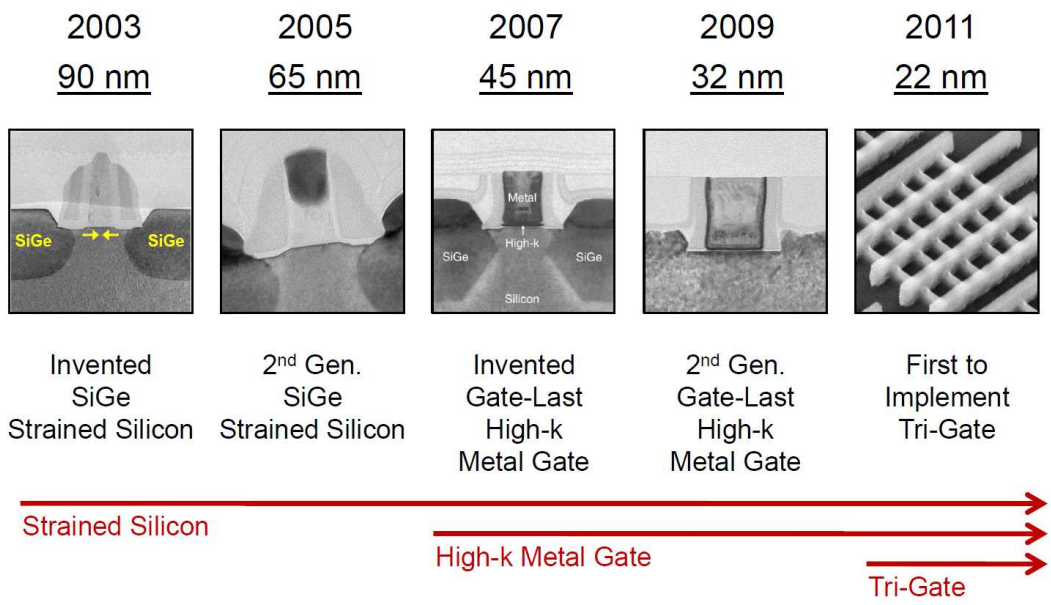


Figure I.3: Intel's roadmap of CMOS scaling since 90 nm technology node. Adapted from [3].

are the miracles to make it happen, i.e., follow the "Moore's Law". People believe that, using the 3-D tri-gate transistors, Intel would be able to follow another round of on-time 2 year cycle for the 16 nm technology node. However, as we are getting more and more close to the physical limits, it becomes more and more difficult to move forward to the next node. It is why people expect slower 3-year technology cycle after the 16 nm node in 2013. Before diving into the potential innovations for future technology, we will review the existing breakthroughs for the CMOS scaling in the nanotechnology era, namely, strained silicon [4]-[12], high-k metal gate [13]-[22], and 3-D transistor [23]-[41].

I.1.A Strained Silicon

It is well known that, in a short-channel MOSFET, the saturation of drain current occurs at a much lower voltage due to velocity saturation. Besides, the CMOS technology does not follow the constant-field scaling, and the electric field has been increasing over the generations. As a result, the effective mobility drops accordingly, due to the surface roughness scattering at high vertical effective field. Therefore, the saturation current deviates more and more from $1/L$ dependence as the channel length L shrinks. It became a headache for CMOS technology in nano era, because scaling down of dimensions would have much less impact on the device performance. Strain engineering has been introduced to address this headache in 90 nm technology node and beyond by many semiconductor manufacturers, especially microprocessor manufacturers [4]-[12]. The purpose of strained silicon is to achieve significant mobility enhancement to compensate the limitation imposed by velocity saturation and mobility degradation at high field. The most challenging part, which results in huge complexity, is that NMOS and PMOS respond differently to different types of strain (compressive or tensile). There are two categories of stress that can lead to strained silicon, namely, biaxial substrate stress or uniaxial-process-induced stress. Both of them have been paid enough attention, and studied extensively. The comparison of biaxial substrate strain and uniaxial process-induced strain by [6] reveals that uniaxial strain is superior than biaxial strain in sub-100 nm CMOS technology in the following aspects. First, uni-

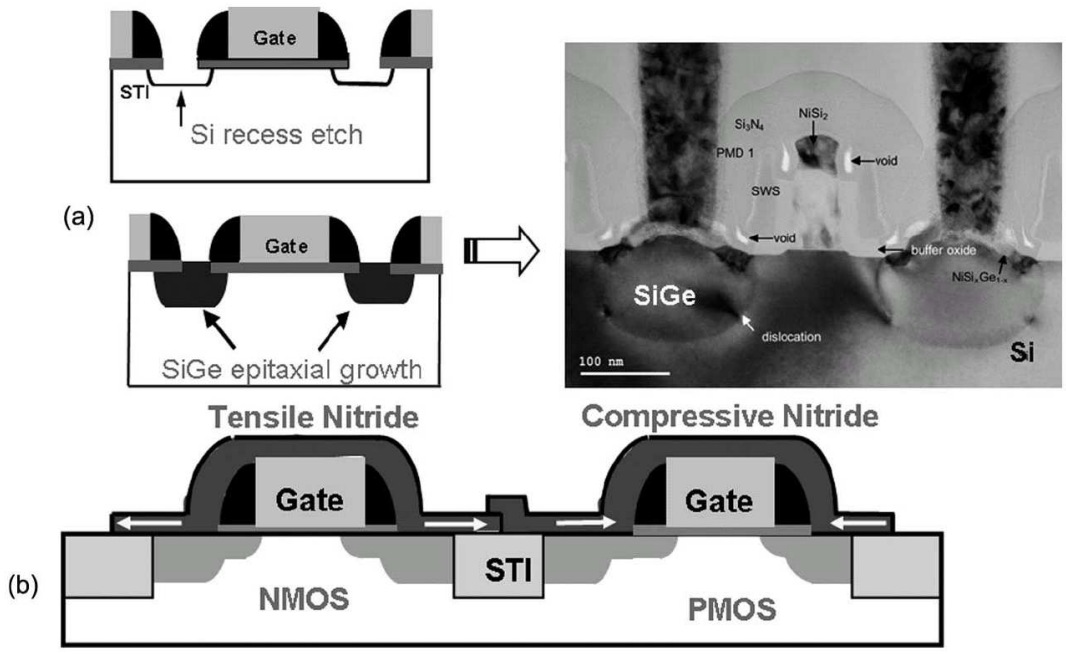


Figure I.4: (a) Strained-Si p-channel MOSFET process flow for the representative stacked gate transistor and TEM cross-sectional view. (b) dual stress liner process architecture with tensile and compressive silicon nitride capping layers. Adapted from [5].

axial strain can provide more significant enhancement of hole mobility than biaxial strain at low strain and high vertical electric field due to differences in the warping of the valence band under strain [6]. Second, uniaxially strained-Si MOSFETs have better performance in nanoscale short channel devices, since the mobility enhancement of electrons and holes are mostly due to reduced conductivity effective mass. Finally, uniaxially strained n-channel bears the threshold voltage shift five times smaller than biaxial strain. Large threshold voltage shift can weaken the performance improvements introduced by mobility enhancement through channel doping adjustment. Furthermore, because different types of uniaxial stress (compressive or tensile) can be realized for p-type and n-type MOSFETs respectively

on the same wafer, uniaxial strain has been chosen by the semiconductor industry for nanoscale CMOS technology. Using embedded SiGe for source/drain is one of the most effective and widely implemented approaches to create large compressive uniaxial stress for p-MOSFETs. This technique consists of steps shown in Fig. I.4(a). Source/drain is first etched to create a silicon recess, and then SiGe is epitaxially grown in the recessed source/drain. Similarly, embedded SiC can be applied to n-MOSFETs for tensile stress [7], but it has not been commercialized yet. Besides embedded SiGe, dual stress liners are also being widely adopted [8]-[11]. The compressive and tensile SiN capping layers are deposited over p-MOSFETs and n-MOSFETs, respectively, as illustrated in Fig. I.4(b). Another popular one is stress memorization technique (SMT), which introduces strain into the silicon channel via a stress memorization of the poly-Si gate [12]. Therefore, it is not valid for high-k metal gate technology. With these novel techniques, both electron and hole mobility can be significantly improved, especially hole mobility, which expects an enhancement of over 200% [5].

I.1.B High-k Metal Gate

Intel pioneered in high-k dielectrics and metal gate, and adopted them since 45 nm technology node [3]. The other semiconductor companies are catching up in 32/28 nm products. It is well known that the short-channel effects (SCEs) are the key challenges with CMOS scaling [43]. To suppress the SCEs, the device has to maintain the electrostatic integrity, i.e., the charge in the channel should be

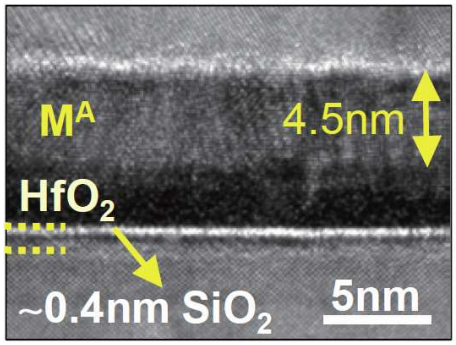


Figure I.5: HRTEM image for the high-k metal gate MOSFET. Adapted from [42].

primarily controlled by the gate but not the drain. For the classical bulk MOSFETs and partially-depleted SOI (PDSOI) MOSFETs, the two major solutions for control of SCEs are reducing the gate oxide thickness and the maximum depletion width, respectively. Both of them are facing the physical limits. We focus on reducing the gate oxide thickness first, and leave the other to the next part. As the gate oxide thickness decreases with CMOS scaling, it is already below 3 nm in the early nanoscale CMOS technology, which is so thin that the oxide layer only consists of several atomic layers. In this case, the gate leakage current becomes a major concern for circuit design, and it is the biggest trouble for further scaling down of oxide thickness. Insulator with higher relative dielectric constant than 3.9 (SiO_2), so-called high-k dielectric, is no doubt the best way to go beyond this limit imposed by gate tunneling leakage. The bad quality of interface between high-k dielectric and silicon is the most prominent difficulty that prevents high-k dielectric to appear earlier in CMOS technology. Actually, having a native oxide with good interface quality is one of the primary reasons that silicon was eventually

chosen for CMOS technology. Although researchers have been able to improve the interface quality for high-k/Si, it is still significantly worse than that of SiO_2/Si. Therefore, a compromised methodology is to have SiO_2 interfacial layer and deposit high-k material on top of interfacial layer, as shown in Fig I.5. In such a way, one can take advantage of low interface states density and large band offset of SiO_2/Si by trading off somewhat in terms of effective oxide thickness (EOT), which is defined as the oxide thickness normalized by dielectric constant with respective to SiO_2. As for the high-k material, it is not that the higher dielectric constant the better. In most of the cases, the material with higher dielectric constant has smaller band offset to silicon so that tunneling is not necessarily lower given the same EOT. Besides, there are many other considerations such as defect density, high-k/SiO_2 interface quality, reliability, etc. According to the most recent reports in literature, HfO_2 seems to be preferred by most of semiconductor manufacturers [13, 14, 16, 18, 20, 21, 42].

The most obvious motivation for metal gate is low gate resistance. Actually, metal is used as the gate material in the very early years, but the semiconductor industry moved away from metal to poly-Si due to a few fabrication considerations, e.g., high temperature tolerance, work function suitability, self-aligned gate technique. With the huge improvements in process techniques, people are able to switch back to metal gate and thus take advantage of low resistivity of metal. Besides low resistivity, metal gate is also able to get rid of poly depletion effect, which has bad impact on gate capacitance, gate resistance, control of SCEs, etc.

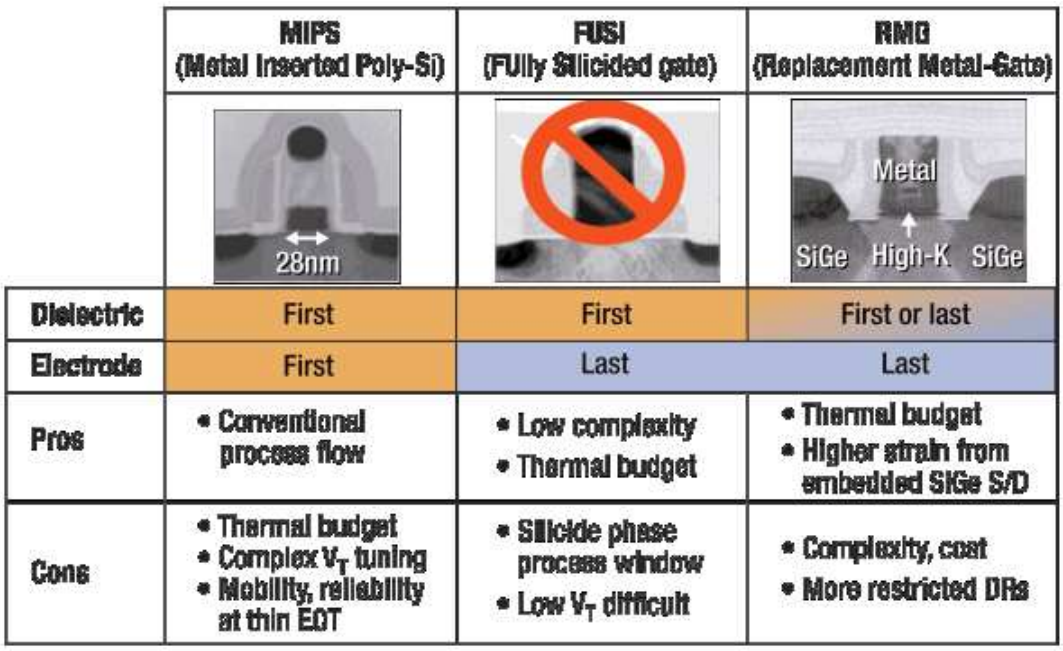


Figure I.6: Pros and cons of different high-k metal gate integration options. Adapted from [44].

One of the key challenges for metal gate is the work function engineering, because n- and p-MOSFETs need different gate work functions to achieve right threshold voltages. This challenge together with some other considerations have led to an intensive debate on whether to use gate-first or gate-last technology [21, 42, 44]. The pros and cons of different high-k metal gate integration options are summarized in Fig. I.6. The fully silicided (FUSI) gate option has already been abandoned, and only gate-first, also commonly called metal inserted Poly-Si (MIPS), or gate-last, also commonly called replacement metal gate (RMG), are actively developed today. MIPS is consistent with the conventional process flow, so it has much better compatibility with the previous circuit design based on SiON Poly technology. On the contrary, RMG has much more restricted design rules due to process complex-

ity, and no doubt this situation is not preferred by circuit designers. However, RMG broadens the range of gate material options for effective work function tuning by allowing low thermal budget after metals deposition, decoupling it from the junction activation [42]. Due to this dilemma, it is quite likely that we will see companies adopting different strategies for the integration of high-k metal gate, depending on their products portfolio [44].

I.1.C 3-D Transistors

For bulk and PDSOI MOSFETs, besides reducing EOT, another effective way to control SCEs is to reduce maximum depletion width, which requires the increasing of channel doping. However, high channel doping causes a lot of negative effects that reduce device performance, increase device leakage, and even enlarge process variability. Some of the typical negative effects include the followings. First of all, high doping would significantly reduce carrier mobility due to strong impurity scattering. Second, dopant fluctuation effect is more severe for higher doping condition, and thus the local variation, i.e., device mismatch, is a bigger concern, especially for the SRAM and analog circuit design. Third, high channel doping results in large band-to-band tunneling, so the gate-induced-drain-leakage (GIDL) current becomes very large. Besides the GIDL current, the diode leakage between drain and body may also increase. Therefore, channel doping level has imposed a limit on the scaling of conventional bulk and PDSOI MOSFETs.

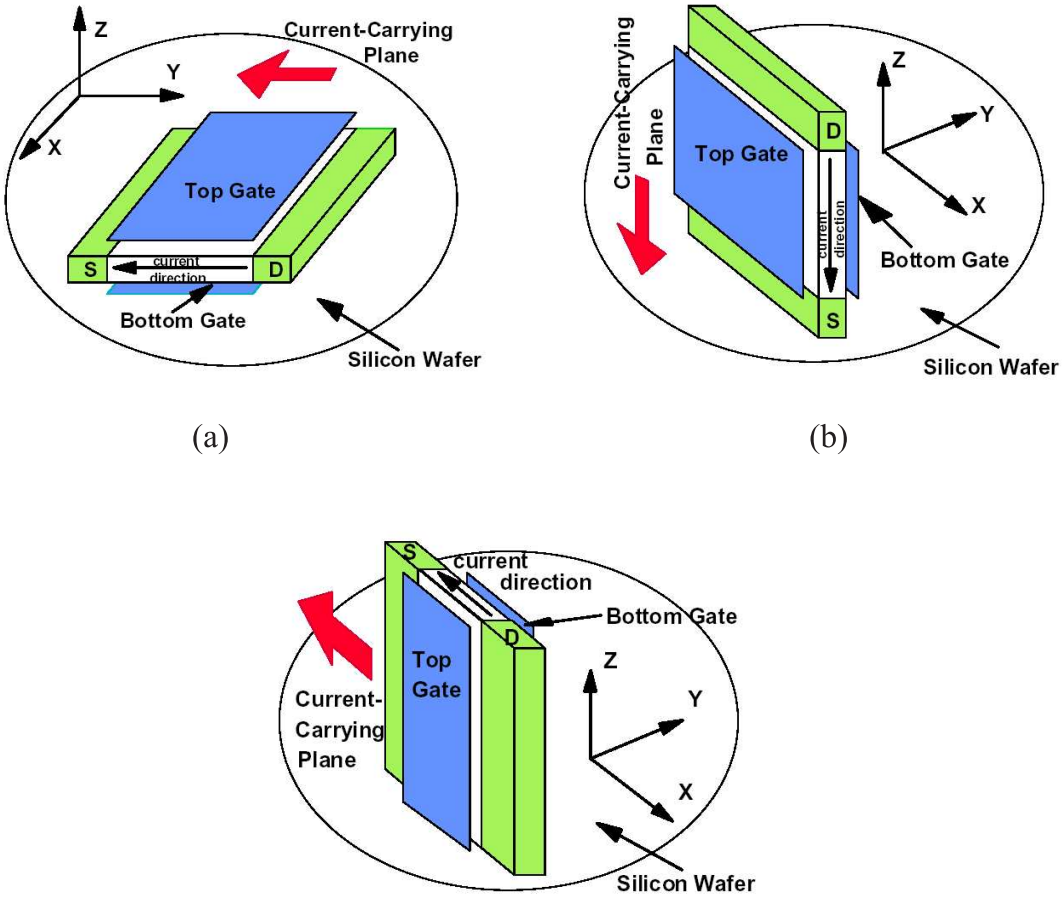


Figure I.7: Different topologies of DG MOSFETs: (a) Planar type; (b) Vertical type; (c) Fin type. Adapted from [46].

Fortunately, many types of multiple-gate (MuG) MOSFETs have been proposed to scale down CMOS technology more aggressively, because of better control SCEs and ideal 60 mV/Decade subthreshold slope [45]. The basic idea of MuG MOSFETs is to use silicon film dimension to control SCEs instead of doping level by taking advantage of very thin silicon film. Among all the MuG MOSFETs, double-gate (DG) MOSFETs received earliest attentions and have been extensively studied since 1980s [23]-[41]. Three possible orientations of DG MOSFET

are illustrated in Fig. I.7. The one in Fig. I.7(a) is the planar type, which is not preferred due to the huge challenge of aligning the top gate and the bottom gate. All the other MuG MOSFETs are called 3-D transistor because of the non-planar nature in device orientation, including the two types of DG showing in Fig. I.7(b) and (c). It has been many years that people believe the Fin type in Fig. I.7(c) is the right choice to extend CMOS technology to 22 nm and beyond, so FinFETs almost becomes an equivalent term to DG MOSFETs. However, in most of the FinFET devices, the gate covers 3 of the 4 sides, therefore, FinFETs can be either DG MOSFETs or tri-gate (TG) MOSFETs depending on the thickness of top gate oxide and aspect ratio of height over width.

As well known, Intel already announced and demonstrated the first 22 nm microprocessor, codenamed "Ivy Bridge", that will be the first high-volume chip to use 3-D TG transistors. The other semiconductor companies are also pushing very hard for TG or FinFET in general to appear in 22/20 nm or 16/14 nm node product [24]-[29]. But one of the key debates is on whether it should be based on SOI or bulk substrate [30]. Intel adopted the bulk substrate [3], and TSMC also leans towards bulk substrate [24]. However, many other companies have been working on SOI-based FinFETs for a while [25]-[29]. Fig. I.8 shows the schematics and TEM pictures for both SOI and bulk FinFETs. Introducing SOI substrate leads to lower junction capacitance, but both alternatives exhibits similar intrinsic device performance [30]. But anyway, the key challenge for both SOI and bulk FinFETs is the threshold voltage engineering. Unlike the bulk or PDSOI MOSFETs, where

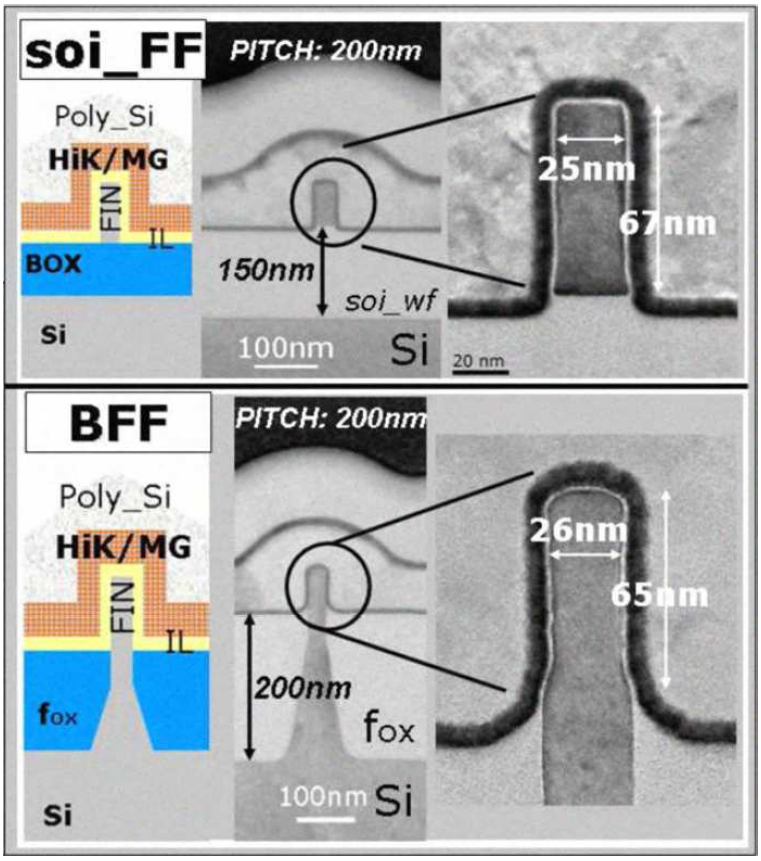


Figure I.8: Schematics and TEM pictures for both SOI and bulk FinFETs. Adapted from [30].

threshold voltage can adjusted with doping profile, threshold voltage engineering for FinFETs relies more on the work function of metal gate material.

I.2 The Future of CMOS Technology

What is the future of CMOS technology after the revolutionary FinFET or TG architecture? It is believed that high-k metal gate will be adopted in all the future CMOS technologies no matter what breakthrough has been achieved. But people would like to step further in the other two directions, namely, increas-

ing mobility and saturation velocity of the channel materials and developing 3-D transistors with better electrostatic integrity. For the first direction, we are almost getting close to the limit of silicon with strain engineering. It is reported that the strain engineering is even applied to the 3-D transistors successfully [24]. The feasible choice to further improve mobility and saturation velocity is to replace silicon with other high mobility materials such as III-V materials or germanium. For the second direction, gate-all-around transistors with either rectangular or circular cross section can provide better control of SCEs than FinFETs.

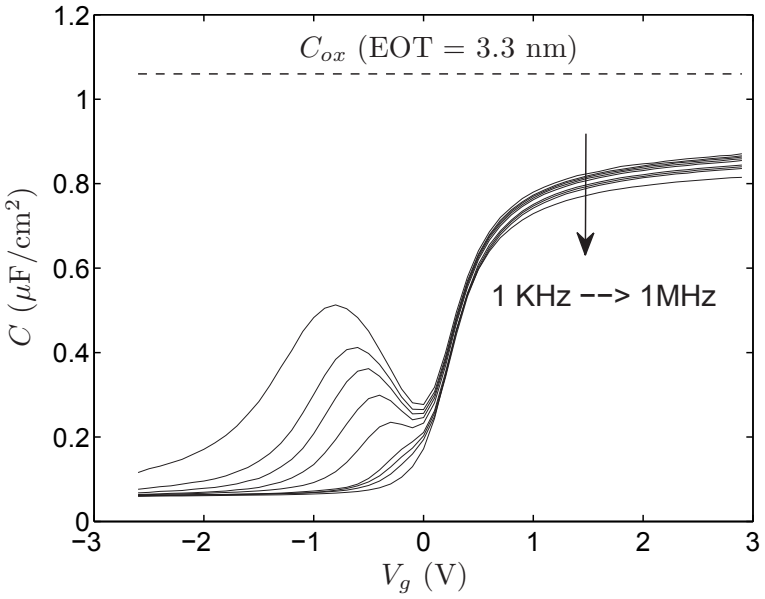
I.2.A III-V/Germanium MOSFETs

III-V materials and germanium have been used in transistors for many decades, such as SiGe BJTs, heterojunction bipolar transistors (HBTs), and high electron mobility transistors (HEMTs). Even the first MOSFET device is based germanium rather than silicon. Having a native oxide with good interface quality is one of the primary reasons that silicon was eventually chosen for CMOS technology. As the strain engineering has pushed silicon to the very limit, III-V materials and germanium are suggested as the future channel materials in MOSFETs, because they have much larger mobilities than silicon or even strained silicon. However, it is not likely to replace silicon with one single material. InGaAs and InP are the ideal candidates for n-type MOSFETs due to their large electron mobilities, but their hole mobilities are comparable to or even lower than silicon. Fortunately, we have GaSb and germanium as good choices for p-type MOSFETs. They have

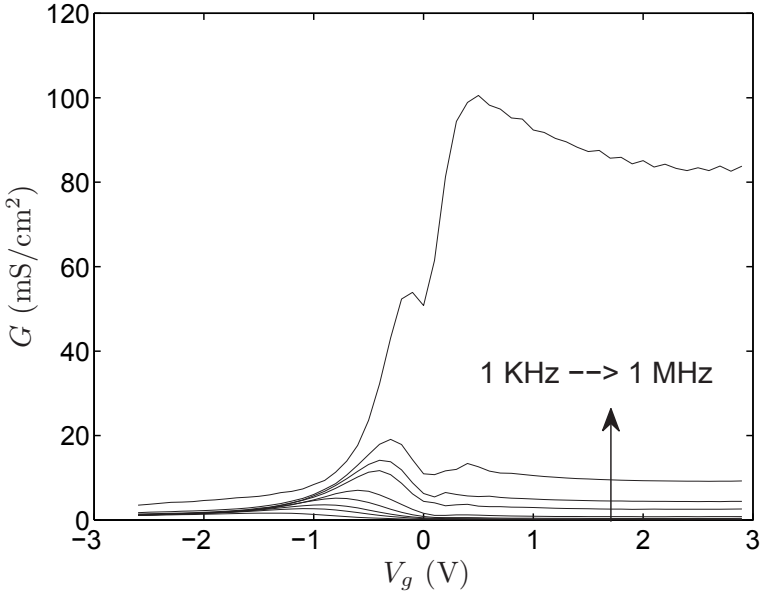
mobility advantage on holes instead of electrons.

III-V/gemanium MOSFETs are quite different from silicon MOSFETs in many different ways. The high mobility is not achieved for free. It is always associated with the small effective mass of electron/hole. Smaller effective mass would introduce much stronger quantum-mechanical effects, therefore the transistor characteristics are more sensitive to the process variations. Besides, small effective mass corresponds to small density-of-states (DOS). In silicon MOSFETs, the inversion layer capacitance is usually dominated by centroid of inversion charge distribution, because of the large DOS of silicon. On the contrary, the semiconductor capacitance of III-V/gemanium MOSFETs is dominated by the DOS. This is one of the key differences between III-V/gemanium and silicon MOSFETs that has to be taken into account in device design and modeling. From the capacitance point of view, this is a positive factor because the total gate capacitance is decreased. But at the same time, the total inversion charge would be reduced, which is not good in terms of total drain current. Another particular benefit of III-V material is that, by tailoring composition, heterostructures offer additional degrees of freedom to optimize the device.

III-V/gemanium MOSFETs have both advantages and disadvantages compared with silicon MOSFETs, but many people think the advantages significantly exceed disadvantages, and put hope on III-V/gemanium MOSFETs for future revolutionary CMOS technology. However, to make it happen, the key success factor



(a)



(b)

Figure I.9: Experimental (a) C-V and (b) G-V data of Al_2O_3/n-InGaAs MOS. Data are provided Professor Paul McIntyre, Stanford University.

is to achieve high quality gate dielectric and semiconductor interface. Different high-k dielectrics have been considered for III-V/gemanium MOSFETs, including

Al_2O_3, HfO_2, ZrO_2, and so on [47]-[63]. At this moment, the trade-off between high permittivity and large barrier is less important, because none of the dielectric materials are showing competitive interface quality close to that of Si/SiO_2 and thus the focus is to improve the interface quality. The large frequency dispersion is commonly observed in the capacitance-voltage (C-V) and conductance-voltage (G-V) data, as exemplified in Fig. I.9. It is widely believed that the future of III-V/gemanium MOSFETs heavily depends on what level of interface quality can be eventually achieved.

I.2.B Nanowire MOSFETs

Transistors with gate all around are usually called nanowire MOSFETs, which can have circular or rectangular cross section, or even some irregular shape. Nanowire MOSFETs can provide the best controllability of SCEs, and thus are being intensively studied [64]-[77]. The rectangular or irregular shaped ones suffer a lot from the corner effect, especially those with significant doping [78, 79]. Therefore, nanowire MOSFETs with circular cross section have generated much more research interests recently. In this dissertation, we will put emphasis on the ones with circular cross section. Fig. I.10 shows both the 3-D schematic view and the TEM images of the twin silicon nanowire MOSFETs fabricated by Samsung.

Similar to the III-V MOSFET case, the best controllability of SCEs is not achieved for free. The semiconductor wire can be considered as 2-D confinement

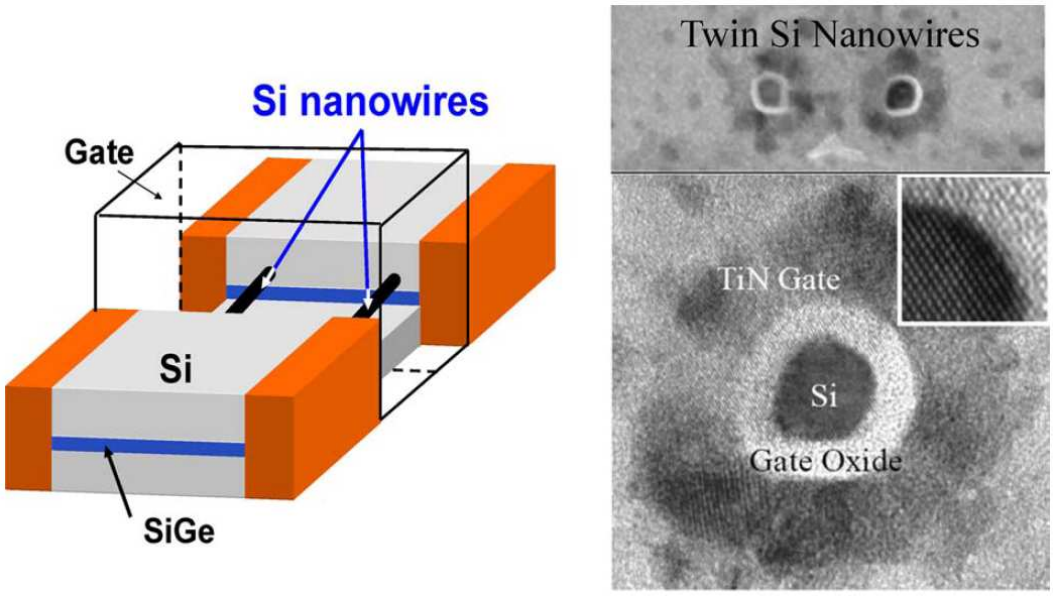


Figure I.10: (a) Three-dimensional schematic view and (b) the TEM images of the twin silicon nanowire MOSFETs fabricated by Samsung. Adapted from [77].

system for carriers, so the quantum mechanical effects are much more severe in nanowire MOSFETs compared with FinFET transistors. A direct output is that threshold voltage shift due to quantum effects would be much more sensitive on the diameter of nanowire, therefore the process variation on diameter could be the key factor that determines the scaling limit. Another disadvantage of 2-D confinement system is that it corresponds to 1-D DOS, which is relatively small due to the $1/sqrt(E - E_0)$ dependence.

I.3 Organization of the Dissertation

The theme of this dissertation is design, modeling and characterization of non-classical MOSFETs. In Chapter I, we review major CMOS technology

in the nanometer era: strained silicon, high-κ metal gate, and 3-D transistors, in particular, FinFET transistor. Then advantages and disadvantages of two primary future CMOS solutions are discussed, namely III-V/germaniam MOSFETs and nanowire MOSFETs. This dissertation includes four chapters in body part.

Chapter II focuses on analysis of high-κ dielectric on InGaAs MOS capacitors. We address the major challenge in high-κ dielectric on III-V MOS: defects or traps inside the insulator as well as at the dielectric-semiconductor interface. A distributed bulk-oxide trap model is developed to account for the commonly observed frequency dispersion of small signal capacitance-voltage and conductance-voltage data in accumulation and near flat band region. The observed C-V humps in depletion to strong inversion are modeled by interface states model.

In Chapter III, design and characterization of III-V MOSFETs are presented. Starting from our 15 nm baseline design, the strategies for further scaling to 11 nm are studied. Potential issues in raised source/drain process are evaluated. The third section of Chapter III deals with III-V MOSFETs characterization and mobility extraction with non-ideal dielectric-semiconductor interface.

In Chapter IV, compact model and scaling limit of nanowire MOSFETs are discussed. First section of this chapter is devoted to compact modeling of SCEs. Quantum confinement effects on V_t shift in nanowire MOSFETs with anisotropic effective mass are modeled as well. Scaling limit is projected for extremely scaled nanowire MOSFETs based on V_t shift sensitivity and scale length theory.

In Chapter V, by invoking both Poisson's and Schrödinger's equations, as well as one subband approximation, analytical results for inversion layer capacitance in symmetric DG and nanowire capacitors are developed. In the derived results, the physical meanings of DOS capacitance and quantum well capacitance are manifest.

The last chapter summarizes the whole dissertation.

Chapter II

High-κ on InGaAs MOS

II.1 Preliminary Analysis on Experimental C-V Characteristics of InGaAs MOS

In this section, experimental frequency dependent capacitance-voltage (C-V) characteristics of InGaAs MOS will be analyzed qualitatively. Before digging into III-V MOS C-V, ideal C-V of Si MOS devices will be reviewed.

Small-signal capacitances measurement at various frequencies is commonly used to characterize metal-oxide-semiconductor devices. The total MOS capacitance per unit area is defined as

$$C = \frac{d(-Q_s)}{dV_g} \tag{II.1}$$

where Q_s is the charge per unit area in Si substrate. With defects charge ignored and poly silicon depletion width small enough, the total MOS capacitance

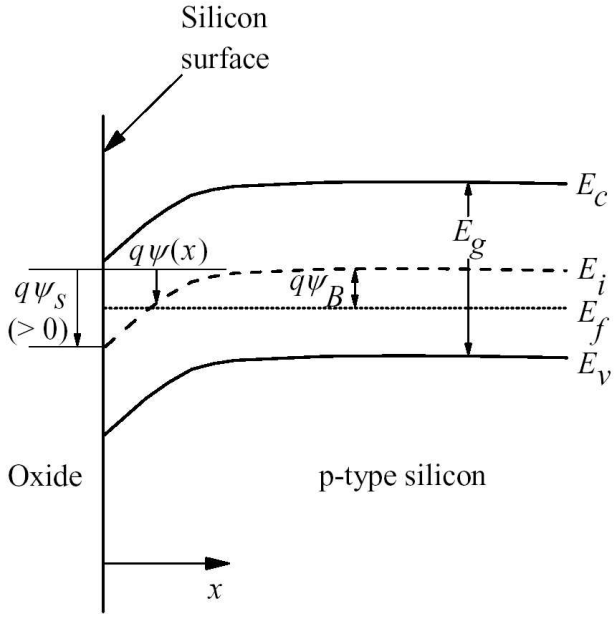


Figure II.1: Energy-band diagram near the silicon surface of a p-type MOS device. The band bending ψ is defined as positive when the bands bend downward with respect to the bulk. Adapted from [43].

in general consists of two components connected in series. One is $C_{ox} = \epsilon_{ox}/t_{ox}$, associated with dielectric constant and thickness of the oxide. The other is semiconductor capacitance C_s dependent on MOS capacitor bias condition. Therefore, C can be express as

$$C = \left(\frac{1}{C_{ox}} + \frac{1}{C_s} \right)^{-1} \tag{II.2}$$

C_{ox} only depends on gate oxide parameters and is constant for an existing MOS capacitor. On the other hand, semiconductor capacitance C_s is a strong function of surface potential ψ_s which is defined in Fig. II.1. C_s is defined as

$$C_s = \frac{d(-Q_s)}{d\psi_s} \tag{II.3}$$

Since the total semiconductor charge Q_s consists of both contributions from the

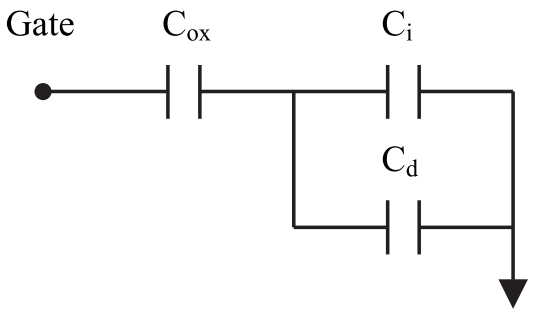


Figure II.2: Equivalent circuit model of an MOS capacitor at low frequency. Semiconductor capacitance is divided into two parts: C_i and C_d.

charge associated with majority carriers Q_d and the charge associated with minority carriers Q_i, C_s can be divided into two parts according to Q_s components:

$$C_s = C_i + C_d \tag{II.4}$$

where C_i and C_d are defined as

$$\begin{aligned} C_i &= \frac{d\left(-Q_i\right)}{d\psi_s} \\ C_d &= \frac{d\left(-Q_d\right)}{d\psi_s} \end{aligned} \tag{II.5}$$

According to the equations (II.2) and (II.4), an equivalent circuit for MOS capacitance C at low frequency is shown in Fig. II.2. In Fig. II.3, three curves at different frequencies overlap on accumulation side because of instantaneous majority response to both gate bias sweep and small-signal modulation. Q_i and C_i are both negligible. On inversion side, the MOS capacitor is such biased that inversion charge is much more than depletion charge at quasi-static condition. When frequency is low enough so that generation of minority carriers is able to follow ac voltage, inversion charge capacitance C_i is also much larger than ma-

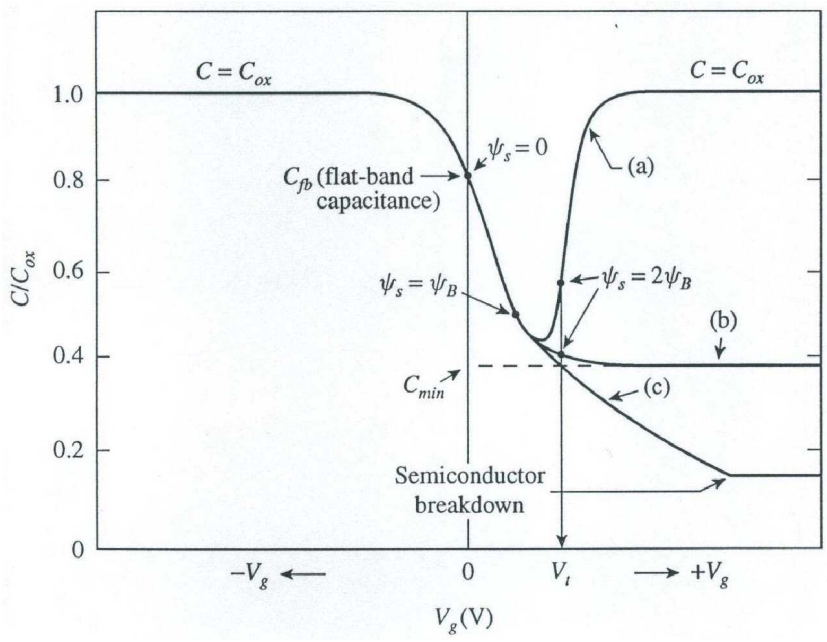


Figure II.3: MOS capacitance-voltage curves: (a) low frequency, (b) high frequency, (c) deep depletion. Adapted from [80].

jority component C_d. The total capacitance will go back up to oxide capacitance C_{ox} as surface potential increases and C_i is even much larger than C_{ox}, as shown in curve (a). Actually, in this case, response of minority carriers screen the ac electric field change from penetrating into space charge region, resulting negligible majority carrier response to ac signal. When frequency is high enough so that generation and recombination of minority carriers are not able to follow ac signal, C_i is negligible and total capacitance will be a series combination of C_{ox} and C_d as shown in (b). For the frequency somewhere in between these two limits, C_s would be frequency dependent on inversion side. Therefore total capacitance also show frequency dependent characteristics in between curves (a) and (b). Actually, it is possible to measure C_i and C_d separately in a split C-V measurement on MOSFET

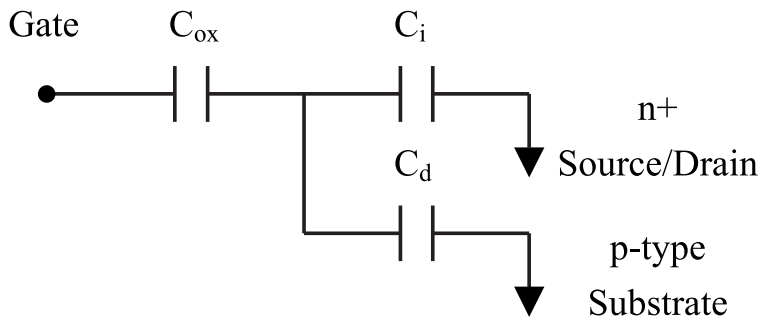


Figure II.4: Equivalent circuit model for split C-V setup. C_i and C_d are connected to different terminals.

[81, 82]. The equivalent circuit model for split C-V setup (Fig. II.4) is slightly different from that of an MOS capacitor. In the split C-V measurement, the n+ Source/Drain part of the total MOS capacitance is C_{inv}

$$C_{inv} = \frac{d\left(-Q_i\right)}{dV_g} = \frac{C_{ox}C_i}{C_{ox} + C_i + C_d} \tag{II.6}$$

and the p-type Substrate part of the total MOS capacitance is C_D

$$C_D = \frac{d\left(-Q_d\right)}{dV_g} = \frac{C_{ox}C_d}{C_{ox} + C_i + C_d} \tag{II.7}$$

The importance of C_{inv} is manifest. Being the slope of DC Q_i–V_g curves, C_{inv} can be integrated to regenerate Q_i–V_g curves. Such a technique is used, for example, in channel mobility measurements where the inversion charge density must be determined accurately.

An example of split C-V measurement results is demonstrated in Fig. II.5. It can be observed that before inversion layer forms only C_D contributes. But after inversion layer forms, C_{inv} becomes dominant quickly, and C_D drops significantly due to the strong screening by inversion layer. Therefore, it is a good

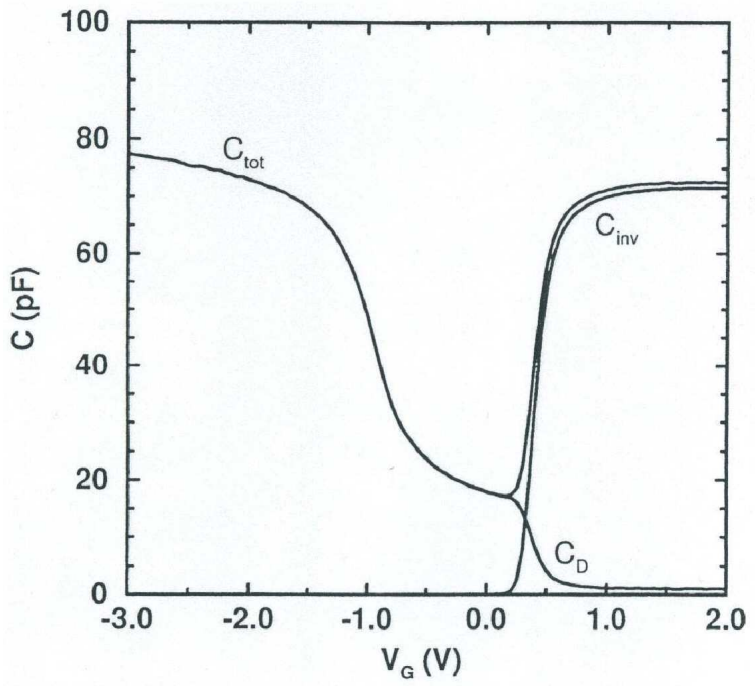


Figure II.5: An example of split C-V measurement results. C_d dominates in the accumulation region, and C_i dominates in the inversion region.

approximation in the strong inversion region that

$$C_{inv} = \frac{C_{ox}C_i}{C_{ox} + C_i} \tag{II.8}$$

i.e., C_{inv} equals C_{ox} and C_i connected in series.

It's worth of note that the above discussion on inversion and accumulation region is only suitable for small C_{ox} and semiconductor that has large conduction band and valence band density of states (DOS), such as Si, where C_s is much larger than C_{ox} in inversion and accumulation. As device dimensions scale down, dielectric becomes thinner and C_{ox} becomes larger, the classic theory is no longer valid. Furthermore, III-V substrate materials such as InGaAs has lower conduction band DOS compared with Si, resulting in a different comparison relationship between

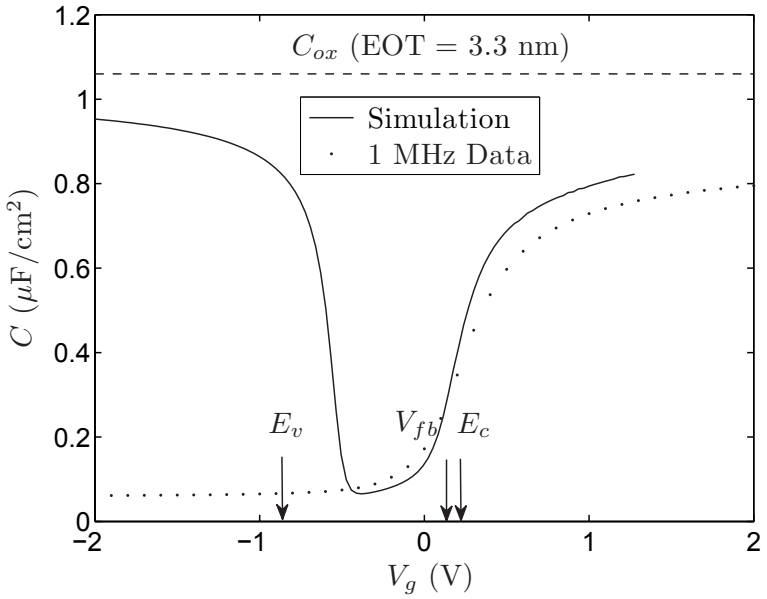


Figure II.6: Simulated quasi-static Al_2O_3/n-$In_{0.53}Ga_{0.47}As$ MOS C-V (solid line) compared with experimental C-V data (dots) at 1 MHz. $C_{ox} = 1.06\mu F/cm^2$

C_{ox} and C_s.

Fig. II.6 shows simulated ideal quasi-static C-V in solid line in comparison with experimental high frequency C-V at 1 MHz in dots for Al_2O_3/n-$In_{0.53}Ga_{0.47}As$ MOS with n-type doping density of 2×10^{16} cm^{-3}. C_{ox} is indicated in the figure by a dashed line. Conduction band edge E_c, Valence band edge E_v and flat-band condition is also labeled in lateral axis V_g. There are several different features of InGaAs MOS ideal quasi-static C-V compared with Si ideal quasi-static C-V with thick oxide: (a) Total capacitance in inversion deviates from C_{ox}; (b) Total capacitance in accumulation deviates from C_{ox}; (c) Non-symmetry between inversion and accumulation to some extent. The simulated quasi-static C-V includes quantum-mechanical effects and non-parabolic effects. Because of quantum effects,

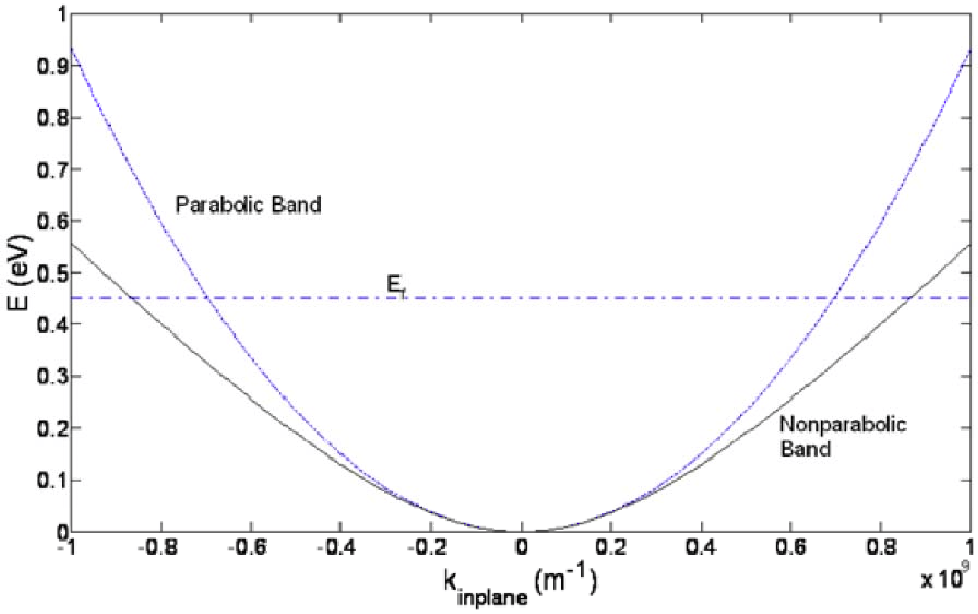


Figure II.7: A comparison on E-k dispersions between parabolic band description and non-parabolic band description. Adapted from [83].

inversion charge is pushed away from oxide-semiconductor interface and inversion charge capacitance C_i is no longer much larger than C_{ox}, and therefore simulated C-V in inversion (negative V_g) deviates from one of serial components C_{ox} [82]. On the other side, because of small electron effective mass and therefore small DOS in conduction band, majority charge capacitance is also not much larger than C_{ox} [82]. So simulated C-V in accumulation (positive V_g) would be considerably lower than dielectric capacitance C_{ox}. However, besides the small DOS degradation to accumulation capacitance, non-parabolic effects actually compensate this low DOS degradation [83]. Fig. II.7 schematically compares E-k dispersions between parabolic band description and non-parabolic band description. As a result of smaller effective mass of InGaAs, and hence smaller DOS, electrons will populate

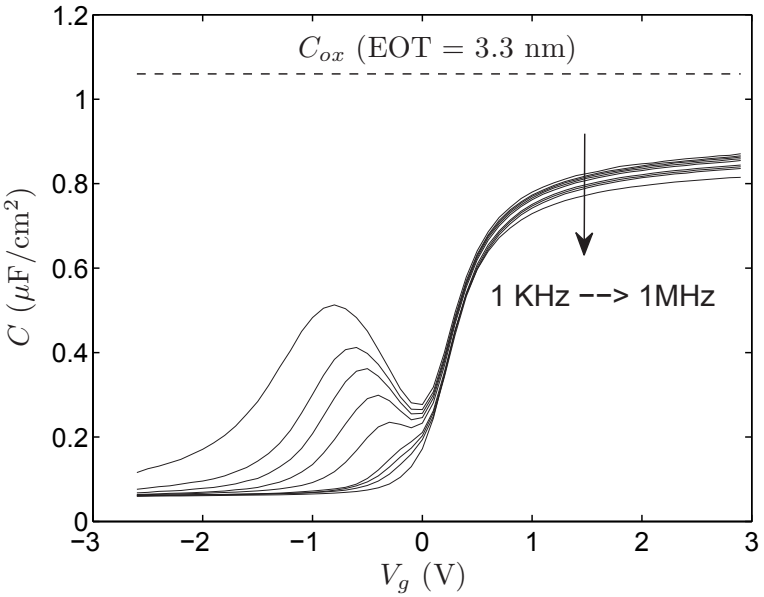


Figure II.8: Experimental C-V of Al_2O_3/n-$In_{0.53}Ga_{0.47}As$ MOS C-V. The measurement frequencies are 1 KHz, 2 KHz, 3 KHz, 5 KHz, 10 KHz, 30 KHz, 50 KHz, 100 KHz, 1 MHz.

higher energy states at a given gate bias and C_{ox}. Effective mass from parabolic approximation at bottom of E-k relation cannot adequately describe the electronic structure at higher energies. Non-parabolic feature should be taken into account for electrons at higher energy, where actual DOS is higher than calculated from parabolic approximation. Therefore small DOS is counteracted by non-parabolic feature to some extent. Thus the final C-V with all these effects taken into account shows non-symmetry of some degree.

With the ideal III-V quasi-static C-V established, we will take a look at experimental multiple frequency C-V for Pt/Al_2O_3/n-$In_{0.53}Ga_{0.47}As$ MOS device, as shown in Fig. II.8 (Data are from McIntyre's group, Stanford University). The device is fabricated under similar processing procedures as in Ref. [84]. The

semiconductor layer structure is 1 μm 2×10^{16} cm^{-3} doped n-$In_{0.53}Ga_{0.47}As$ on 100 nm 5×10^{18} cm^{-3} doped n-$In_{0.53}Ga_{0.47}As$ on n^+ InP substrate. The Al_2O_3 film is prepared by 50 cycles of atomic layer deposition (ALD) with trimethyl aluminum precursor and water vapor oxidant. The sample is then annealed in forming gas for 30 minutes at 400°C.

To characterize the MOS device, the first step is to extract oxide capacitance C_{ox}. We cannot take low frequency capacitance on inversion side as C_{ox} because of the following reasons: (a) On inversion side, 1 KHz is not low enough for true inversion to take place. (b) Even in quasi-static C-V measurement, where true inversion can happen, 1-D quantum confinement on inversion charge will make semiconductor capacitance less than infinity compared with oxide capacitance and total capacitance somewhat smaller than C_{ox}. (c) Considering the non-ideality of measured device sample, defects or traps response will make the measured capacitance distorted and related with C_{ox} in a complicated way. Total capacitance on accumulation side cannot be taken as C_{ox} either for the following reasons: (a) Small conduction band DOS leads to small semiconductor capacitance and thus makes total capacitance lower than C_{ox}. (b) Total capacitance could be either higher or lower than C_{ox} due to traps in oxide. To extract oxide capacitance accurately, we need to compare experimental data with ideal quasi-static simulation. C-V data at highest measurement frequency 1 MHz are least affected by trap response and thus most close to ideal MOS case. When simulated total capacitance matches with 1 MHz data, C_{ox} is obtained from simulation.

In Fig. II.8, a few features can be observed: (a) Frequency dispersion from large positive gate bias through zero gate bias. (b) Frequency dispersion and humps in negative gate bias. It's worth to mention that these features are not unique for Al_2O_3/n-$In_{0.53}Ga_{0.47}As$ MOS capacitor. Actually they are quite common in high-κ - III-V semiconductor MOS devices. So it's of great importance that we understand these features physically and come up with appropriate models to account for the observations and calibrate for essential device parameters, such as defects density.

Conventionally interface states are considered as the cause for frequency dispersion in multi-frequency C-V. Interface states are defects located at the dielectric - semiconductor interface, which have one or more energy levels within the semiconductor bandgap. These defects can exchange charge with the semiconductor. Specifically, they can interact with the semiconductor conduction band by capturing or emitting electrons and with the valence band by capturing or emitting holes. These capture and emission processes are governed by Shockley-Read-Hall (SRH) mechanism. Capture or emission occurs when interface states change occupancy. When a small AC voltage applies to the gate of an MOS capacitor, the band edges move toward or away from the Fermi level alternately. Carriers are captured or emitted, changing occupancy of interface trap levels in a small energy interval a few kT/q wide centered about the Fermi level. This capture and emission of majority carriers causes an energy loss observed at all frequencies except the very lowest (to which interface traps immediately respond) and the very

highest (to which no interface trap response occurs) [85]. Thus a capacitance in series with a conductance can be utilized to model the process of traps' changing occupancy under small signal AC voltage drive. Another type of defects, called bulk-oxide traps, are located inside the dielectric and with energy level distributed throughout wide dielectric bandgap. Bulk-oxide traps can also exchange charge with semiconductor bands through a different mechanism - quantum tunneling. The two types of defects both contribute to the observed multi-frequency C-V and conductance-voltage (G-V) as well. Before characterizing and modeling experimental frequency dependent C-V and G-V, it's necessary to review the conductance method - the most accurate and commonly used methodology for interface states characterization in Si MOS.

II.2 Review of Conductance Method

By assuming an interface trap to be an SRH center located at the oxide-semiconductor interface, the equivalent circuit for an MOS capacitor with interface traps at a single energy E within the bandgap can be derived as shown in Fig. II.9 according to [85, 86]. C_{ox} is the oxide capacitance. C_{inv} is the inversion capacitance. C_{dep} is the depletion capacitance. C_T is the trap capacitance given by

$$C_T = \frac{q^2 N_T}{kT} f_0(1 - f_0) \tag{II.9}$$

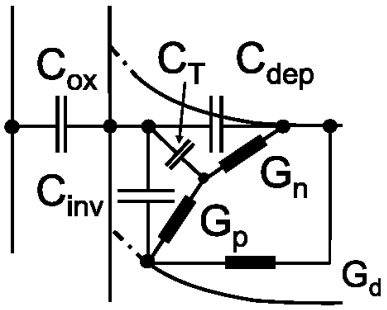


Figure II.9: General equivalent circuit for an n-type MOS capacitor with single interface trap energy level. Adapted from [86].

where N_T is interface states density in unit of cm^{-2}. f_0 is the Fermi function of traps in equilibrium and is given by

$$f_0(E) = \frac{1}{1 + \exp\left[(E - E_f)/kT\right]} \tag{II.10}$$

From (II.9), we can see that C_T is non-negligible only when its energy level E is close to Fermi level E_f within several kT. G_n and G_p are conductances associated with charge exchange between trap states and conduction band and valence band, respectively. They are given by

$$G_n = \frac{q^2 N_T c_n n_s}{kT}(1 - f_0) \tag{II.11}$$

and

$$G_p = \frac{q^2 N_T c_p p_s}{kT} f_0 \tag{II.12}$$

where c_n and c_p are trap capture probability coefficients for electrons and holes respectively. n_s and p_s are the electron and hole densities per unit volume at the interface. G_d models the diffusion-induced inversion response. The equivalent circuit in Fig. II.9 involves traps interaction with both conduction band and valence

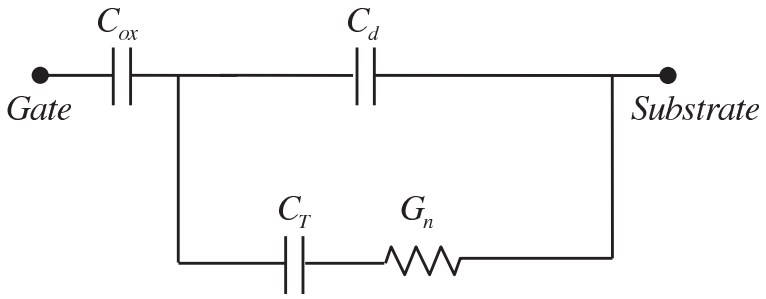


Figure II.10: Equivalent circuit for an n-type MOS capacitor with single interface trap energy level biased in depletion.

band. When the number of holes dominate, G_p is much larger than G_n, interface states exchange traps mainly with conduction band and vice versa. In accumulation and depletion, this general equivalent circuit can be reduced to Fig. II.10, which only includes one trap branch associated with majority carrier energy band. The total admittance Y_{tot} can be easily derived from Fig. II.10,

$$Y_{tot} = j\omega C_d + j\omega C_T G_n (G_n + j\omega C_T)^{-1} \tag{II.13}$$

If we define equivalent parallel capacitance C_p and equivalent parallel conductance G_p as shown in Fig. II.11, we can obtain C_p and G_p/ω, which both contain interface traps information:

$$C_p = C_T \left[1 + (\omega\tau)^2\right]^{-1} + C_d \tag{II.14}$$

$$\frac{G_p}{\omega} = C_T \omega\tau \left[1 + (\omega\tau)^2\right]^{-1} \tag{II.15}$$

It's not straight forward to extract interface states properties from C_p since one also needs to know C_d.

In real devices, interface states are distributed in closely spaced energy

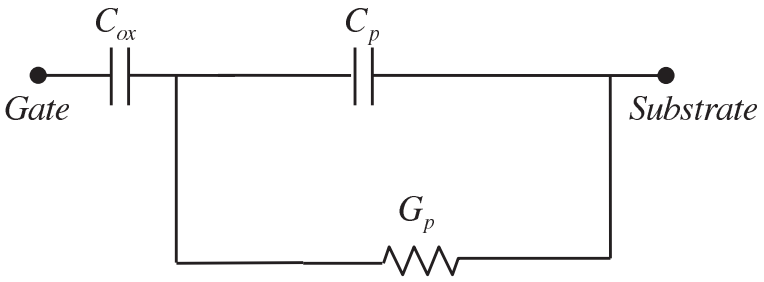


Figure II.11: Equivalent circuit defining equivalent parallel capacitance C_p and equivalent parallel conductance G_p.

levels throughout semiconductor bandgap. The corresponding equivalent circuit for devices biased in depletion is shown in Fig. II.12. If we assume continuous energy distribution of interface states, N_T in (II.9) becomes interface states in unit of $\text{cm}^{-2}\text{Joule}^{-1}$, and we denote this energy distributed density as D_{it} to distinguish from single energy level interface states density. By further assume a constant D_{it} over energy, we can derive total equivalent parallel capacitance and conductance

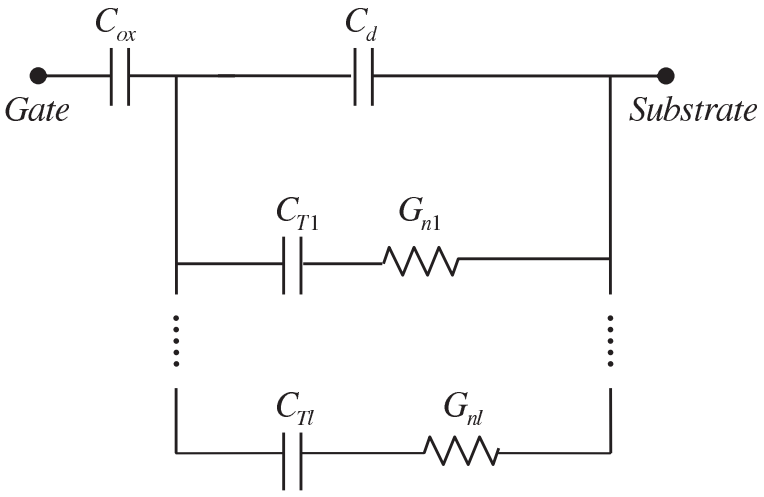


Figure II.12: Equivalent circuit for an n-type MOS capacitor biased in depletion with interface traps distributed in energy.

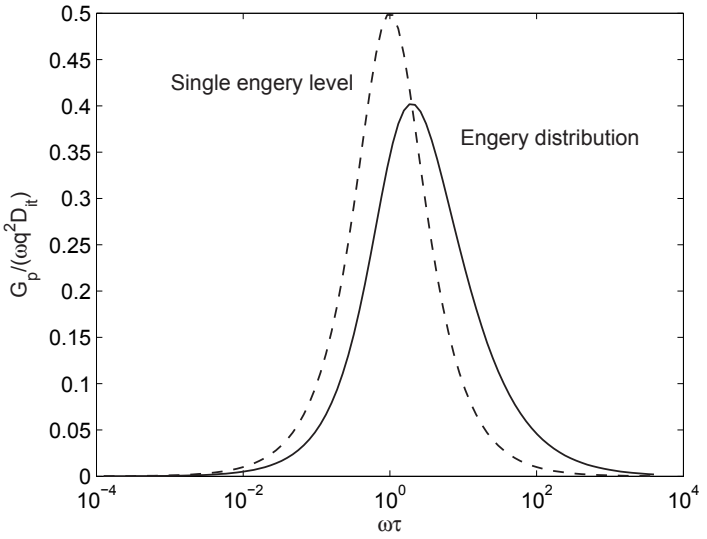


Figure II.13: G_p/ω vs. $\omega\tau$ for single energy level traps (dashed line) and energy distributed traps (solid line). Adapted from [85].

for traps distributed in energy by integrating (II.14) and (II.15) over energy with N_T replaced with D_{it} [85]. The results are

$$C_p = q^2 D_{it} (\omega\tau)^{-1} \arctan(\omega\tau) + C_d \tag{II.16}$$

$$\frac{G_p}{\omega} = q^2 D_{it} (2\omega\tau)^{-1} \ln\left[1 + (\omega\tau)^2\right] \tag{II.17}$$

Fig. II.13 shows G_p/ω vs. $\omega\tau$ for single energy level traps (dashed line) and energy distributed traps (solid line). The peak of the curve for energy distributed case is smaller than single energy case. In addition, energy distributed case has larger spread-out in frequency and a shift to larger frequency compared to single energy case. From G_p/ω peak, one can extract D_{it} information as follows,

$$D_{it} = \frac{2.5}{q^2} \left(\frac{G_p}{\omega}\right)_{peak} \tag{II.18}$$

Till now we have reviewed the basic theory of conductance method. In next section, we will review another D_{it} extraction method based on C-V stretch-out and apply it to the data in Fig. II.8.

II.3 D_{it} Extraction from High Frequency C-V Stretchout

In this section, we will review interface states density extraction method based on high frequency experimental C-V and simulated ideal high frequency C-V. The method is applied to characterize Al_2O_3/n-$In_{0.53}Ga_{0.47}As$.

In high frequency C-V measurement, traps cannot respond to small signal modulation due to their small time constant, so the measured total capacitance is the same as ideal MOS capacitance without traps. As discussed in Section II.1, the total capacitance is serial combination of C_{ox} and C_s, as shown in Fig. II.14. One the other side, the gate bias sweep in C-V measurement is at very low rate. In this case, traps can fully respond to gate bias change during sweep without delay. The total capacitance is C_{ox} in series with the parallel combination of C_s and C_{it}, as shown in Fig. II.15. We can extract D_{it} from comparison between measured

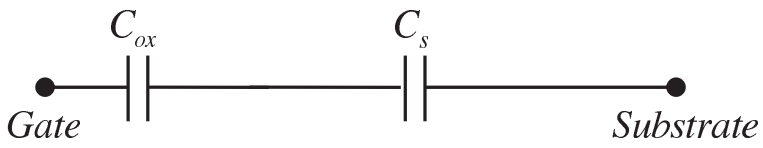


Figure II.14: Equivalent circuit for HF capacitance measurement, where no traps respond to small signal modulation, as well as for ideal simulation without traps present.

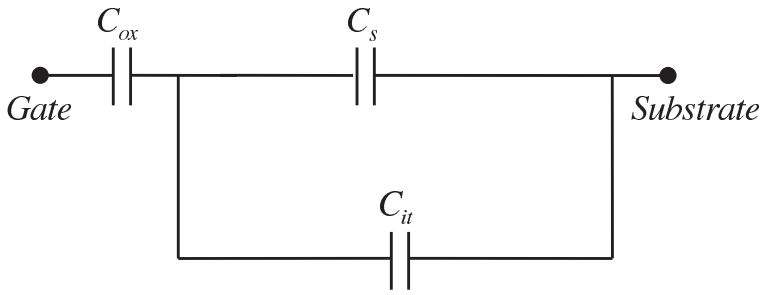


Figure II.15: Equivalent circuit for DC gate bias sweep in capacitance measurement.

HF C-V and simulated ideal C-V. For the simulated ideal C-V, both small signal capacitance and DC sweep capacitance can be represented by Fig. II.14. A given capacitance corresponds to the same surface potential but different gate bias in measured HF C-V and simulated ideal C-V due to traps charge/discharge in gate bias sweep in measurement. So trap density information is embedded in the slope degradation of measured HF C-V. From Fig. II.14, we have

$$\left(\frac{d\psi_s}{dV_g}\right)_{sim} = \frac{C_{ox}}{C_{ox}+C_s} \tag{II.19}$$

From Fig. II.15, we have

$$\left(\frac{d\psi_s}{dV_g}\right)_{data} = \frac{C_{ox}}{C_{ox}+C_s+C_{it}} \tag{II.20}$$

ψ_s is surface potential. $dC/d\psi_s$ in both simulation and HF experimental data are the same. So we can multiply $dC/d\psi_s$ to both sides of (II.19) and (II.20), we have

$$\left(\frac{dC}{dV_g}\right)_{sim} = \frac{C_{ox}}{C_{ox}+C_s}\left(\frac{dC}{d\psi_s}\right) \tag{II.21}$$

$$\left(\frac{dC}{dV_g}\right)_{data} = \frac{C_{ox}}{C_{ox}+C_s+C_{it}}\left(\frac{dC}{d\psi_s}\right) \tag{II.22}$$

D_{it} can be derived from the ratio of (II.21) and (II.22):

$$D_{it} = \frac{1}{q^2}\left[\frac{(dC/dV_g)_{sim}}{(dC/dV_g)_{data}} - 1\right](C_s + C_{ox}) \tag{II.23}$$

At certain capacitance, we can extract D_{it} from slopes of HF experimental C-V and simulated ideal C-V. Since certain capacitance corresponds to certain surface potential and Fermi level position, we can further mapping capacitance to Fermi level positions in the band diagram. Because traps with energy level close to Fermi level dominate in charging/discharging process, D_{it} vs. Fermi level positions is equivalent to D_{it} vs. trap energy. Thus we can extract D_{it} energy profile through this method. It is worth to mention that this method is only suitable for depletion and moderation accumulation, but not applicable to inversion and strong accumulation. In inversion and strong accumulation, HF C-V is rather flat, no information can be obtained from almost zero C-V slope. The methodology is used to extract D_{it} for Al_2O_3/n-$In_{0.53}Ga_{0.47}As$ MOS devices.

Both data of Al_2O_3/n-$In_{0.53}Ga_{0.47}As$ MOS with and without forming gas anneal (FGA) are investigated to quantitively evaluate the effects of FGA to interface defects. Experimental multiple frequency C-V of the FGA Al_2O_3/n-$In_{0.53}Ga_{0.47}As$ MOS is shown in Fig. II.8. Experimental multiple frequency C-V of the Al_2O_3/n-$In_{0.53}Ga_{0.47}As$ MOS without FGA is shown in Fig. II.16. C_{ox} is extracted to be 1.06 $\mu F/cm^2$ by comparing HF experimental C-V with simulated ideal C-V in terms of accumulation capacitance shown in Fig. II.6. Also labeled in Fig. II.6 are V_g values for E_c, E_v, and V_{fb}, where the Fermi level crosses those energies at

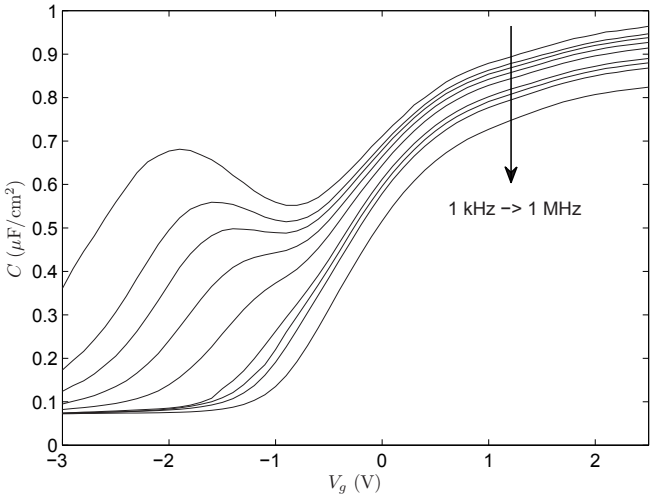


Figure II.16: Experimental C-V of Al_2O_3/n-$In_{0.53}Ga_{0.47}As$ MOS without FGA. The measurement frequencies are 1 kHz, 2 kHz, 3 kHz, 5 kHz, 10 kHz, 30 kHz, 50 kHz, 100 kHz, 1 MHz. Data are from McIntyre's group, Stanford University.

the surface. Most of the steep transition region in the 1 MHz C-V curve is actually near flatband or in moderate accumulation, not depletion. C-V curves are shifted horizontally to account for fixed charge influence on flatband voltage. The C-V slope degradation of data can be observed by the comparison, which is due to traps' charging/discharging as Fermi level moves when sweep gate bias. Fig. II.17 shows the extracted D_{it} energy profile for both FGA device and no FGA device. It can be observed that FGA reduces D_{it} approximately by half in depletion and moderate accumulation. D_{it} increases toward higher energy above conduction band edge. Conventional interface states are only distributed in semiconductor bandgap. Therefore, the extracted D_{it} in semiconductor conduction band could be another type of traps. Considering the observed frequency dispersion in multi-frequency C-V in strong accumulation, traps distributed above conduction band

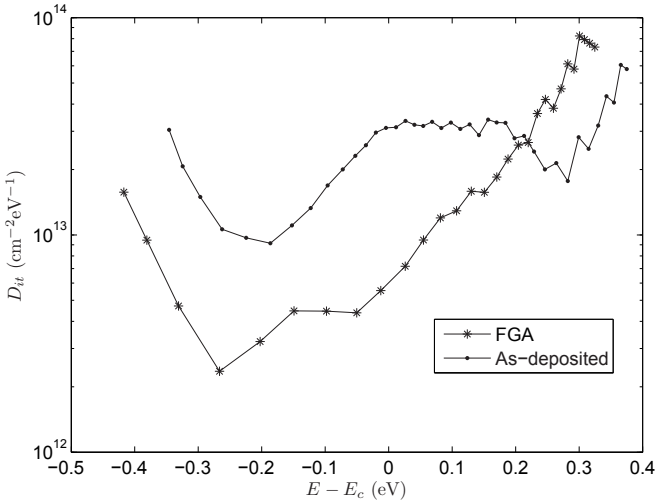
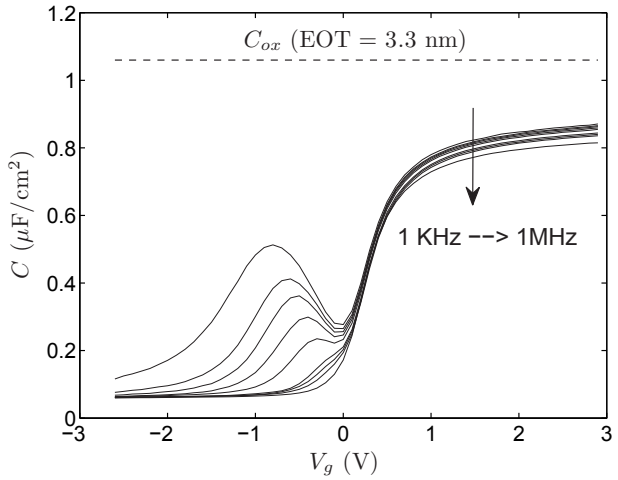


Figure II.17: The extracted D_{it} profile from HF C-V slope degradation in Al_2O_3/n-$In_{0.53}Ga_{0.47}As$ with FGA (star) and without FGA (dot).

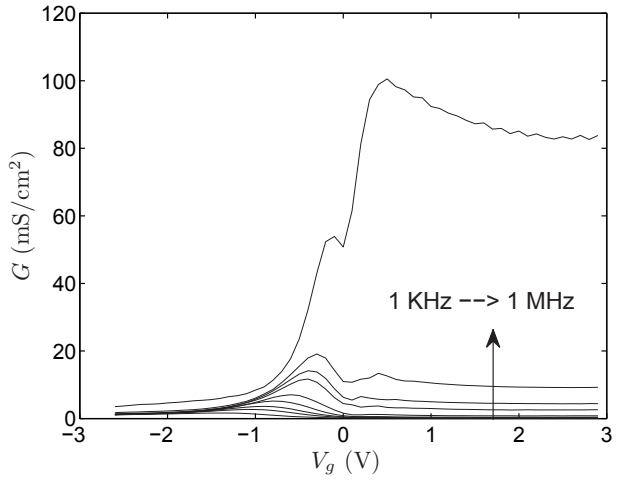
edge must include slow traps with time constant comparable with midgap interface states to produce frequency dispersion between 1 kHz to 1 MHz. In next section, we will investigate the "slow" traps and develop a circuit model to account for the effects on small signal admittance.

II.4 A Distributed Bulk-Oxide Trap Model for the Al_2O_3-InGaAs MOS Devices

Recently, III-V compound semiconductor MOSFETs have been intensely investigated to replace silicon CMOS for high-performance digital applications. In many reports in the literature [87, 88, 89, 90, 91, 92], frequency dispersion is commonly observed in the capacitance-voltage and conductance-voltage data of



(a)



(b)

Figure II.18: Experimental (a) C-V and (b) G-V data of Al_2O_3/n-InGaAs MOS at 1 KHz, 2 KHz, 3 KHz, 5 KHz, 10 KHz, 30 KHz, 50 KHz, 100 KHz, and 1 MHz.

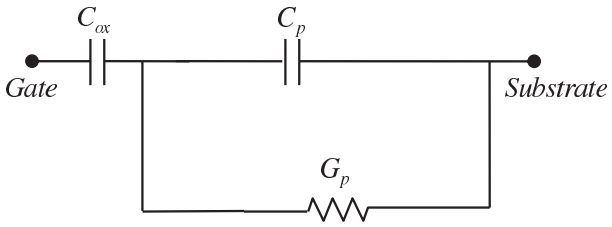
high-κ/III-V MOS devices. Examples are shown in Fig. II.18 for ALD Al_2O_3 on n-type $In_{0.53}Ga_{0.47}As$ MOS capacitor. The dispersion in the strong accumulation region cannot be explained by the conventional interface states whose time constant in such bias regions is far shorter than the period of typical measurement frequencies, 1 KHz to 1 MHz [85, 93]. On the other hand, trap states inside the

gate insulator, called bulk-oxide traps or border traps, do have long time constants as they interact with the conduction band electrons via tunneling [94]. In addition, when the conventional conductance method [85] for interface states is applied to the high-to-low transition region (maximum slope) of the C-V data, the frequency dispersion of conductance does not follow the well known peak behavior. And the C-V stretch-out with respect to the ideal curve indicates an interface state density far exceeding that extracted from the dispersion in that region. Such a discrepancy can be resolved by a bulk-oxide trap model in which the low-frequency component causing C-V stretch-out is larger than the high-frequency component responsible for dispersion.

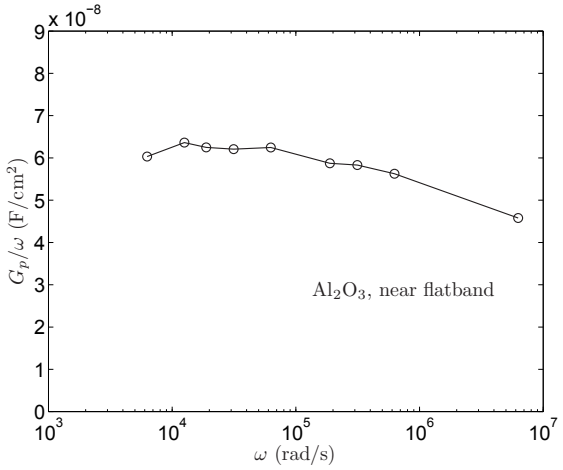
In previous publications, bulk-oxide traps are modeled by a lumped R-C circuit [95, 96], which does not reflect the distributed nature of border trap capacitance over the depth of the gate insulator. We developed a distributed bulk-oxide trap model to explain the dispersion in both strong accumulation and the flatband region of Al_2O_3/n-$In_{0.53}Ga_{0.47}As$ MOS data [97, 98]. The model is also applied to account for the C-V stretch-out in the device.

For the C-V in Fig. II.18, also shown in Fig. II.8, we have extracted its C_{ox} and trap density through stretch-out in Section II.3. In this section, we will focus on the modeling of frequency dispersion in strong accumulation and the flatband region, as well as the stretch-out.

It has been reported that parasitic resistance in series with the MOS capac-



(a)



(b)

Figure II.19: (a) Equivalent circuit for C_p and G_p. (b) Experimental G_p/ω vs. $log(\omega)$ at $V_g = 0.3$ V.

itor could make the apparent C_{tot} lower at high frequencies, e.g., 1 MHz. We rule this out as the cause of the dispersion in the measured C-V because the dispersion persists to as low frequencies as a few KHz, where the series resistance plays no role. Moreover, the estimated spreading resistance in the substrate for the 100 μm dot size is less than a few ohms — several orders of magnitude smaller than the capacitive reactance at 1 MHz. It cannot be explained by the conventional interface states either, since their time constant in accumulation is in nano-seconds, far shorter than the period of the highest measurement frequency, 1 MHz. Near

the flatband or in moderate accumulation, the parallel conductance in the semiconductor, G_p, defined in Fig. II.19, is calculated from the measured $C_{tot}(\omega)$ and $G_{tot}(\omega)$ for a given bias point. Instead of exhibiting a peak behavior as predicted in the standard interface-state model, the G_p/ω vs. $\log(\omega)$ plot in Fig. II.19 is rather featureless, indicative of wide distribution of trap time constants more in line with the bulk-oxide trap model.

The C-V humps in the negative V_g bias region corresponding to depletion and weak inversion do exhibit a peak G_p/ω behavior and are attributed to a localized density of interface states. Analysis of these features will be covered in next section.

II.4.A Distributed Bulk-oxide Trap Model

In MOS devices, traps in the bulk gate dielectric film can exchange charge with mobile carriers in the semiconductor bands through tunneling. Fig. II.20 shows schematically the tunneling process between bulk-oxide traps and conduction band in an n-type MOS device biased in accumulation. The time constant associated with charge exchange between bulk-oxide traps and semiconductor is governed by the tunneling mechanism which gives an exponential dependence on the trap distance x from the interface [94, 95]:

$$\tau(x) = f_0 \tau_0 e^{2\kappa x}. \tag{II.24}$$

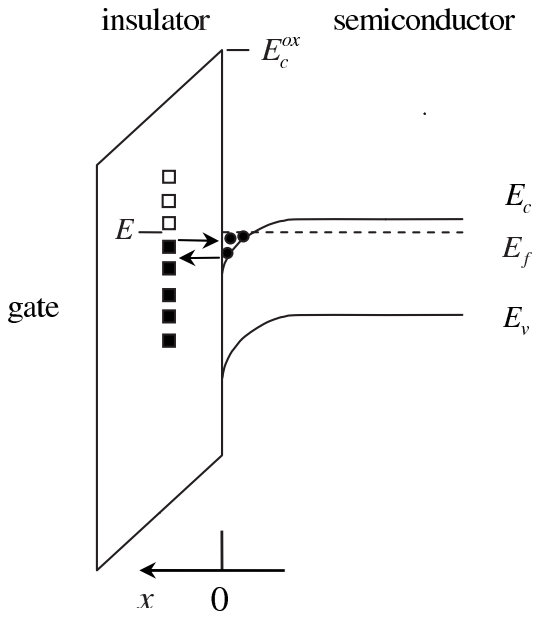


Figure II.20: Schematic diagram of tunneling between bulk-oxide traps in gate insulator and conduction band of semiconductor.

Here τ_0 is the time constant of interface traps inversely proportional to the carrier density at the semiconductor surface, f_0 is the Fermi-Dirac function that a trap at energy E is occupied by an electron, and κ is the attenuation coefficient for an electron wave function of energy E decaying under an energy barrier $E_C^{ox} > E$:

$$\kappa = \sqrt{2m^*(E_C^{ox} - E)}/\hbar. \tag{II.25}$$

m^* is electron effective mass in the dielectric film and E_C^{ox} is the energy of the top of the dielectric barrier as indicated in Fig. II.20.

For a given gate DC bias, bulk-oxide traps at a certain depth x and energy E change occupancy in response to a small signal AC modulation. Bulk-oxide traps at energy close to $E = E_f$ are most responsible for the small-signal capacitance. The effects of bulk-oxide traps at a specific depth and energy on the

small-signal MOS admittance can be modeled by a serial combination of capacitance and conductance. The bulk-oxide traps within an incremental depth Δx at x and an incremental energy ΔE at E are represented by an incremental capacitance $\Delta C_{bt}(E,x)$ and an incremental conductance $\Delta G_{bt}(E,x)$ connected in series. If the density per volume per energy of bulk-oxide traps is N_{bt} in units of $\mathrm{cm}^{-3}\mathrm{Joule}^{-1}$, then [85, 94]

$$\Delta C_{bt}(E,x) = \frac{f_0(1-f_0)q^2 N_{bt}}{kT}\Delta E \Delta x. \tag{II.26}$$

$\Delta G_{bt}(E,x)$ and $\Delta C_{bt}(E,x)$ are related by the time constant $\tau(x)$:

$$\Delta C_{bt}(E,x)/\Delta G_{bt}(E,x) = \tau(x) = f_0\tau_0 e^{2\kappa x}. \tag{II.27}$$

To integrate for a continuous energy distribution of bulk-oxide traps, the serial connection of $\Delta C_{bt}(E,x)$ and $\Delta G_{bt}(E,x)$ at a given x must first be converted to a parallel combination of incremental admittances. Because the factor $f_0(1-f_0)$ is sharply peaked at $E = E_f$, κ of (II.25) is set to be a constant with $E = E_f$ in the integration. The total incremental admittance at x is then:

$$\begin{aligned} \Delta Y_{bt}(x) &= \int_E \frac{1}{\frac{1}{j\omega\Delta C_{bt}(E,x)} + \frac{1}{\Delta G_{bt}(E,x)}} \\ &= \frac{q^2 N_{bt}\ln(1+j\omega\tau_0 e^{2\kappa x})}{\tau_0 e^{2\kappa x}}\Delta x \end{aligned} \tag{II.28}$$

For a continuous distribution of bulk traps throughout the oxide thickness, the equivalent circuit of the MOS device is of a distributed form shown in Fig. II.21, where the oxide capacitance is broken into an infinite number of serial segments with branches of $\Delta Y_{bt}(x)$ connected at different depths. Here, ϵ_{ox} is the permittivity of the insulator and C_s is the semiconductor capacitance.

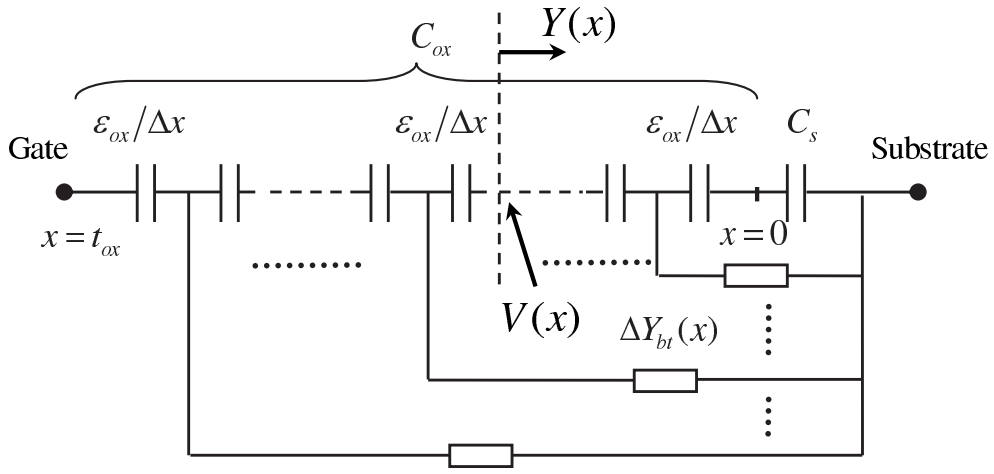


Figure II.21: Equivalent circuit for bulk-oxide traps distributed over the depth of the insulator. The semiconductor capacitance is represented by C_s.

If we define $Y(x)$ to be the equivalent admittance at a point x looking into the semiconductor in Fig. II.21, the recursive nature of the distributed circuit gives the admittance of the next point, $x + \Delta x$, as

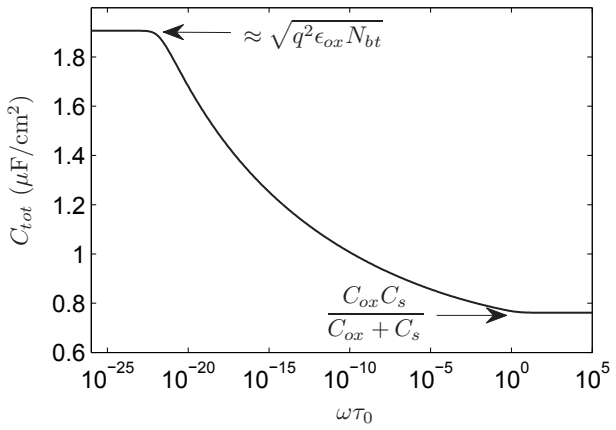
$$Y(x + \Delta x) = \Delta Y_{bt}(x) + \frac{1}{\frac{\Delta x}{j\omega\epsilon_{ox}} + \frac{1}{Y(x)}}. \tag{II.29}$$

Substituting (II.28) for $\Delta Y_{bt}(x)$, the first-order terms in Δx then yield a differential equation for Y(x):

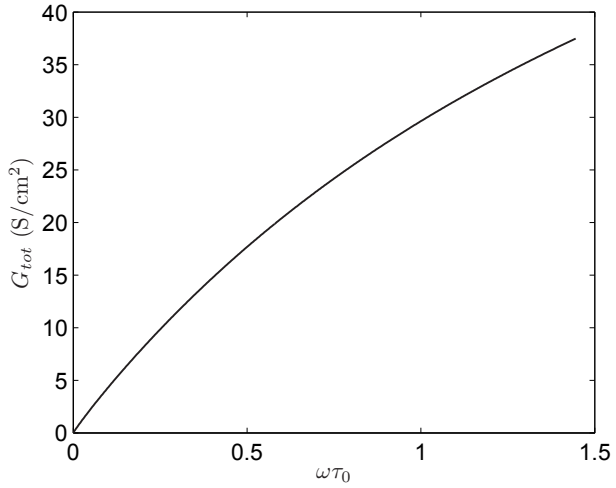
$$\frac{dY}{dx} = -\frac{Y^2}{j\omega\epsilon_{ox}} + \frac{q^2 N_{bt} \ln(1 + j\omega\tau_0 e^{2\kappa x})}{\tau_0 e^{2\kappa x}}. \tag{II.30}$$

The boundary condition is $Y(x = 0) = j\omega C_s$. This differential equation is the correct one to use over the one we derived earlier without energy integration [97]. Although the previously derived equation had a different last term, it only made minor differences in most numerical results.

In general, (II.30) needs to be solved numerically to obtain the total admit-



(a)



(b)

Figure II.22: An example of numerical solution to Eq. II.30: (a) Real and (b) imaginary parts of $Y(x = t_{ox})$ vs. $\omega\tau_0$ with $N_{bt} = 4.2 \times 10^{19}$ cm^{-3}eV^{-1} and $\tau_0 = 2.3 \times 10^{-10}$ s.

tance seen by the gate,

$$Y(x = t_{ox}) \equiv G_{tot} + j\omega C_{tot}. \tag{II.31}$$

A typical example of the solutions C_{tot} vs. $\ln\omega$ and G_{tot} vs. ω is given in Fig. II.22.

In the high frequency limit, $\omega\tau_0 \geqslant 1$, none of the bulk-oxide traps respond to the

AC signal and C_{tot} equals C_{ox} in series with C_s as expected. For the measurement frequencies of 1 KHz to 1 MHz, $1.4 \times 10^{-6} < \omega\tau_0 < 1.4 \times 10^{-3}$, C_{tot} varies linearly with $\ln(1/\omega)$, and G_{tot} linearly with ω, i.e., $G_{tot}/\omega \approx constant$. Both are consistent with the data trends in Fig. II.18. The constant G_{tot}/ω reflects the fact that, for a given gate bias, response of bulk-oxide traps spans a wide spectrum of frequencies due to their depth distribution – a clear distinction from conventional interface traps [85].

For $\omega = 0$ or DC, Fig. II.21 becomes a purely capacitive circuit and (II.30) is reduced to a real equation for the capacitance $C(x)$:

$$\frac{dC}{dx} = -\frac{C^2}{\epsilon_{ox}} + q^2 N_{bt}. \tag{II.32}$$

The boundary condition is $C(x = 0) = C_s$. For uniform N_{bt}, (II.32) can be solved analytically to yield

$$C\,(x) = C_0 \frac{(C_s + C_0)\exp\left(2qx\sqrt{N_{bt}/\epsilon_{ox}}\right) + (C_s - C_0)}{(C_s + C_0)\exp\left(2qx\sqrt{N_{bt}/\epsilon_{ox}}\right) - (C_s - C_0)}. \tag{II.33}$$

Here, $C_0 = q\sqrt{\epsilon_{ox}N_{bt}}$. If $2qt_{ox}\sqrt{N_{bt}/\epsilon_{ox}} \gg 1$, then $C(x = t_{ox}) \approx \sqrt{q^2\epsilon_{ox}N_{bt}}$ (left plateau in Fig. II.22(a)), insensitive to C_s. This is, of course, only a matter of theoretical interest as in practice, it would take much longer than the age of the universe to charge up all the bulk traps in the oxide!

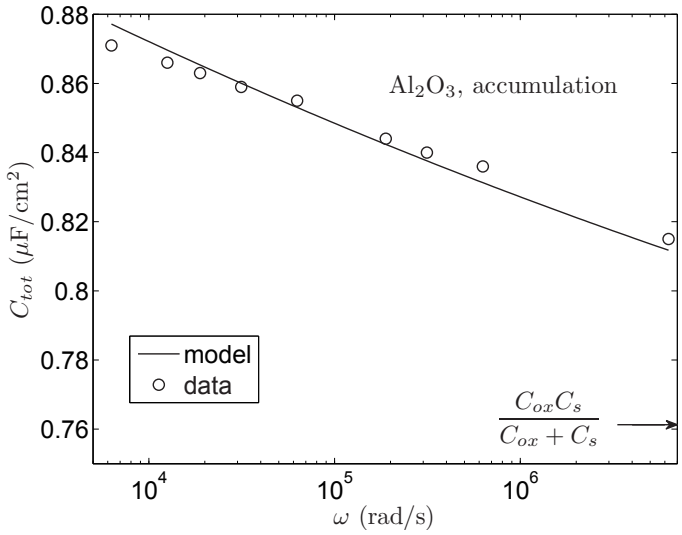
Of particular interest is the case in accumulation where C_s is very high. From equation (II.33), $C_{tot}(\mathrm{DC}) \approx C_0 \coth(C_0/C_{ox})$, always larger than C_{ox}. This is in contrast with the interface state or lumped-circuit border trap models which

do not produce dispersion when shorted out by a large semiconductor capacitance. Dispersion in accumulation is therefore a good indicator of distributed bulk oxide traps.

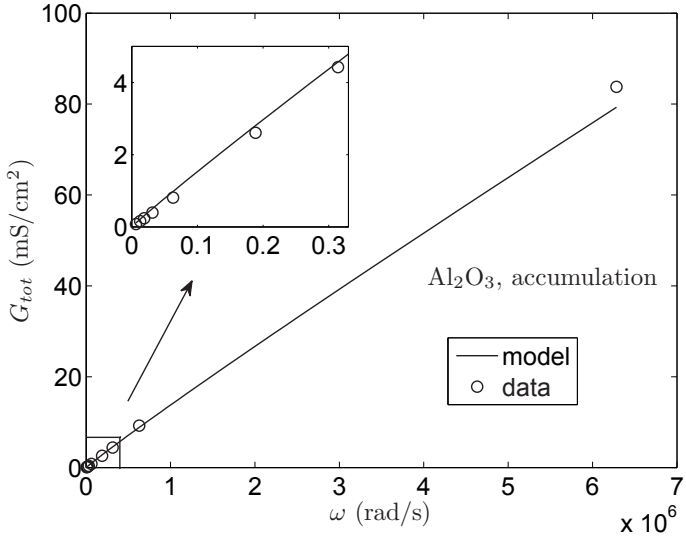
II.4.B Correlation of Model with Multi-Frequency C-V and G-V Data in Strong Accumulation and Near Flatband

The experimental capacitance and conductance versus frequency data in strong accumulation in Fig. II.18 (Al_2O_3/n-InGaAs at $V_g = 2.9$ V) are compared to model calculations in Fig. II.23. For model parameters, the semiconductor capacitance C_s is chosen such that the serial combination of C_{ox} and C_s gives a C_{tot} slightly below the measured 1 MHz capacitance at $V_g = 2.9$ V. κ is calculated from (II.25) with $m^* = 0.5m_0$ and $E_C^{ox} - E = 1.99$ eV. Both the slopes of C_{tot} versus $\ln(1/\omega)$ and G_{tot} versus ω are sensitive to the bulk-oxide trap density N_{bt}. By choosing a single fitting parameter, uniform $N_{bt} = 4.2 \times 10^{19}$ cm^{-3}eV^{-1}, good agreement is achieved between the model and the measured C_{tot}, G_{tot} data from 1 KHz to 1 MHz in Figs. II.23(a) and (b). The parameter τ_0 has little effect on C_{tot}, G_{tot} slopes until $\omega\tau_0 \rightarrow 1$ (Fig. II.22) which is much beyond 1 MHz.

Fig. II.24 shows the model correlation with data near the flatband voltage at $V_g = 0.3$ V, i.e., in the region of steep C-V transition in Fig. II.18. A lower bulk-oxide trap density of $N_{bt} = 2.2 \times 10^{19}$ cm^{-3}eV^{-1} is found to fit both the capacitance and conductance versus frequency data. Note that the zero-depth trap time constant τ_0 here is much longer than that in strong accumulation (Fig.

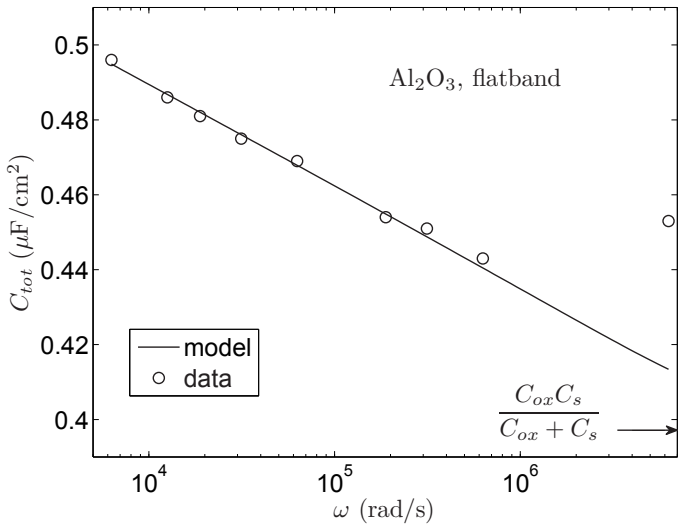


(a)

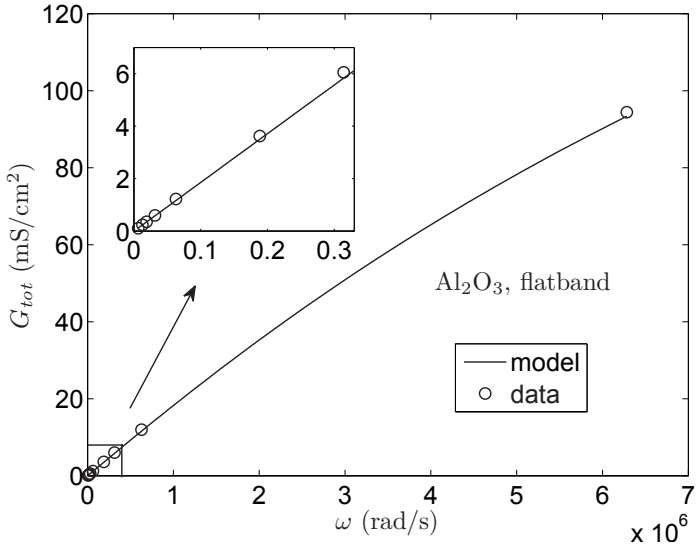


(b)

Figure II.23: Al_2O_3 MOS experimental (a) $C_{tot}(\omega)$ and (b) $G_{tot}(\omega)$ dispersion data (open circles) at $V_g = 2.9$ V in Fig. II.18 compared to those calculated from the distributed bulk-oxide trap model (solid lines). A single bulk-oxide trap density, $N_{bt} = 4.2 \times 10^{19}$ cm^{-3}eV^{-1}, is assumed in both $C_{tot}(\omega)$ and $G_{tot}(\omega)$ calculations. The rest of the model parameters are $C_{ox} = 1.06$ μF/cm^2, $t_{ox} = 5$ nm, $C_s = 2.7$ μF/cm^2, $\kappa = 5.1$ nm^{-1}, and $\tau_0 = 2.3 \times 10^{-10}$ s.

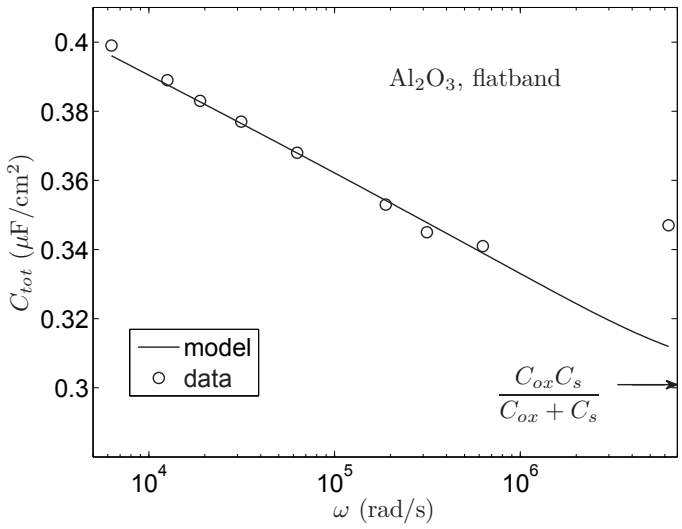


(a)

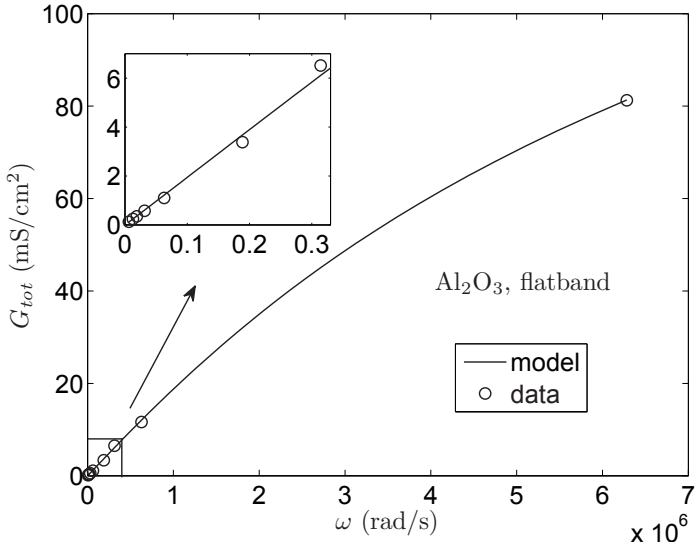


(b)

Figure II.24: Al_2O_3 MOS experimental (a) $C_{tot}(\omega)$ and (b) $G_{tot}(\omega)$ dispersion data (open circles) at $V_g = 0.3$ V in Fig. II.18 compared to those calculated from the distributed bulk-oxide trap model (solid lines). A single bulk-oxide trap density, $N_{bt} = 2.2 \times 10^{19}$ cm^{-3}eV^{-1}, is assumed in both $C_{tot}(\omega)$ and $G_{tot}(\omega)$ calculations. The rest of the model parameters are $C_{ox} = 1.06$ μF/cm^2, $t_{ox} = 5$ nm, $C_s = 0.635$ μF/cm^2, $\kappa = 5.47$ nm^{-1}, and $\tau_0 = 1.35 \times 10^{-7}$ s.

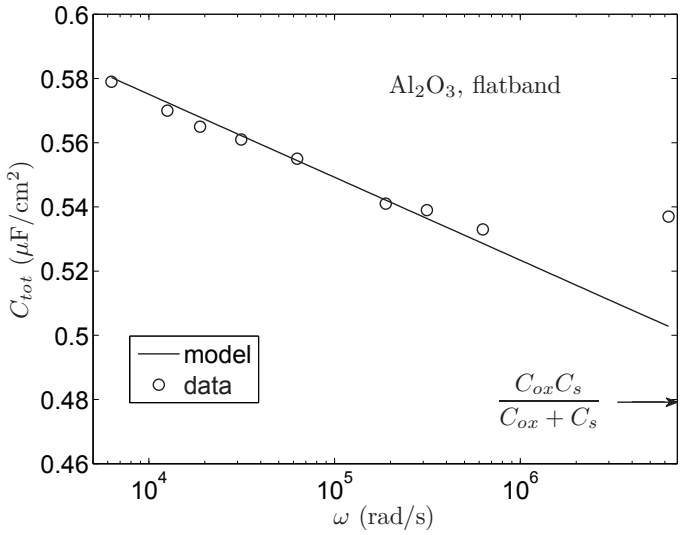


(a)

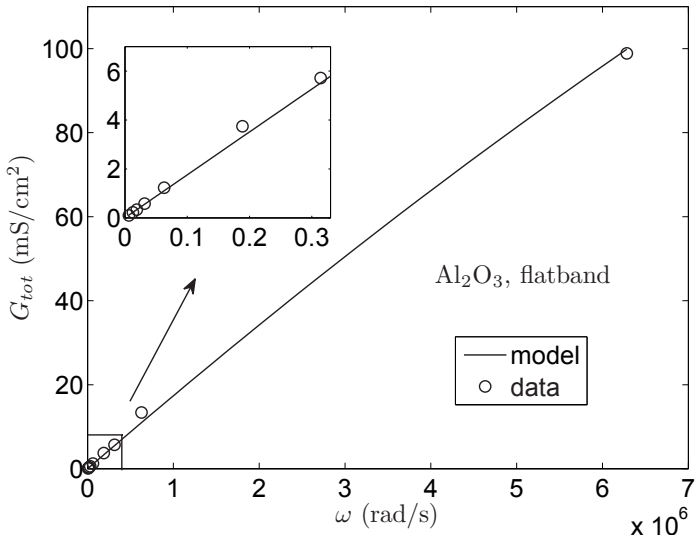


(b)

Figure II.25: Al_2O_3 MOS experimental (a) $C_{tot}(\omega)$ and (b) $G_{tot}(\omega)$ dispersion data (open circles) at $V_g = 0.2$ V in Fig. II.18 compared to those calculated from the distributed bulk-oxide trap model (solid lines). A single bulk-oxide trap density, $N_{bt} = 1.81 \times 10^{19}$ cm^{-3}eV^{-1}, is assumed in both $C_{tot}(\omega)$ and $G_{tot}(\omega)$ calculations. The rest of the model parameters are $C_{ox} = 1.06$ μF/cm^2, $t_{ox} = 5$ nm, $C_s = 0.42$ μF/cm^2, $\kappa = 5.5$ nm^{-1}, and $\tau_0 = 2.5 \times 10^{-7}$ s.



(a)



(b)

Figure II.26: Al_2O_3 MOS experimental (a) $C_{tot}(\omega)$ and (b) $G_{tot}(\omega)$ dispersion data (open circles) at $V_g = 0.4$ V in Fig. II.18 compared to those calculated from the distributed bulk-oxide trap model (solid lines). A single bulk-oxide trap density, $N_{bt} = 2.6 \times 10^{19}$ cm^{-3}eV^{-1}, is assumed in both $C_{tot}(\omega)$ and $G_{tot}(\omega)$ calculations. The rest of the model parameters are $C_{ox} = 1.06$ μF/cm^2, $t_{ox} = 5$ nm, $C_s = 0.875$ μF/cm^2, $\kappa = 5.43$ nm^{-1}, and $\tau_0 = 6 \times 10^{-8}$ s.

II.23) such that nonlinearity starts to show up in G_{tot} at 1 MHz. This is consistent with the lower surface electron density near the flatband. Fig. II.25 and Fig. II.26 show the model correlation with data at $V_g = 0.2$ V and $V_g = 0.4$ V, respectively.

II.4.C Correlation of Model with C-V Stretchout Near Flatband

It is seen in Fig. II.6 that in addition to giving rise to C and G dispersions, oxide traps also cause a degradation of the dC_{tot}/dV_g slope near the flatband and moderate accumulation compared to the ideal simulated curve. Such a "stretchout" of the 1 MHz C-V is commonly employed to extract the interface state density (Terman method) based on the relation $\Delta V_g/\Delta\Psi_s = 1 + (C_s + C_{it})/C_{ox}$, where Ψ_s is the surface potential and C_{it} is the capacitance due to interface states. In Section II.3, this method is applied to the observed C-V slope degradation in Fig. II.6 and yields an interface-state density of 5.4×10^{12} cm^{-2}eV^{-1} at $V_g = 0.3$ V. This level of interface-state density would result in a much higher dispersion than the experimental data. For example, it leads to a G_p/ω peak value of 0.35 μF/cm^2 - about 5.5 times the data in Fig. II.19(b).

The discrepancy between low frequency stretchout and high frequency dispersion is readily resolved with the bulk-oxide trap model which predicts much richer low frequency component than high frequency (Fig. II.22(a)). Physically, the slow sweep rate of V_g allows charging and discharging of more traps by tunneling deeper into the oxide. An expression for the frequency dependent $\Delta V_g/\Delta\Psi_s$ can

be derived from the distributed bulk-oxide trap model. We define the differential potential at a point x in the oxide in Fig. II.21 as $V(x)$, then

$$V(x+\Delta x)-V(x)=\frac{Y(x)V(x)}{j\omega\left(\epsilon_{ox}/\Delta x\right)}. \tag{II.34}$$

where $Y(x)$ is the admittance at x solved by (II.30). Therefore,

$$\frac{dV}{dx}=\frac{Y(x)V(x)}{j\omega\epsilon_{ox}}. \tag{II.35}$$

and

$$\frac{\Delta V_g}{\Delta\Psi_s}=\frac{V(x=t_{ox})}{V\left(x=0\right)}=\exp\left[\frac{1}{j\omega\epsilon_{ox}}\int_0^{t_{ox}}Y(x)dx\right]. \tag{II.36}$$

For DC, $Y(x)=j\omega C\left(x\right)$, where $C\left(x\right)$ is given by (II.33) for uniform N_{bt}. (II.36) gives

$$\left(\frac{\Delta V_g}{\Delta\Psi_s}\right)_{trap}=\frac{\left(C_s+C_0\right)\exp\left(2C_0/C_{ox}\right)-\left(C_s-C_0\right)}{2C_0}\exp\left(-\frac{C_0}{C_{ox}}\right). \tag{II.37}$$

where $C_0=q\sqrt{\epsilon_{ox}N_{bt}}$. Note that this DC expression assumes bulk oxide traps all the way to the metal gate are engaged. In reality, the sweep rate of V_g is finite and we expect an experimental $\Delta V_g/\Delta\Psi_s$ value lower than the DC value of (II.37).

For the Al_2O_3 MOS biased at $V_g=0.3$ V, $\Delta V_g/\Delta\Psi_s$ is computed from (II.36) using the uniform N_{bt} and other parameters extracted from high frequency C and G dispersions in Fig. II.24. Fig. II.27 plots the real part of $\Delta V_g/\Delta\Psi_s$ with respect to $\left(\Delta V_g/\Delta\Psi_s\right)_{notrap}=\left(C_{ox}+C_s\right)/C_{ox}$ as a function of frequency (solid curve). The imaginary part is negligible. The experimental stretchout of V_g can be estimated by comparing the dC_{tot}/dV_g slope of the 1 MHz C-V data to that of the ideal simulated C-V in Fig. II.6. However, since both N_{bt} and τ_0

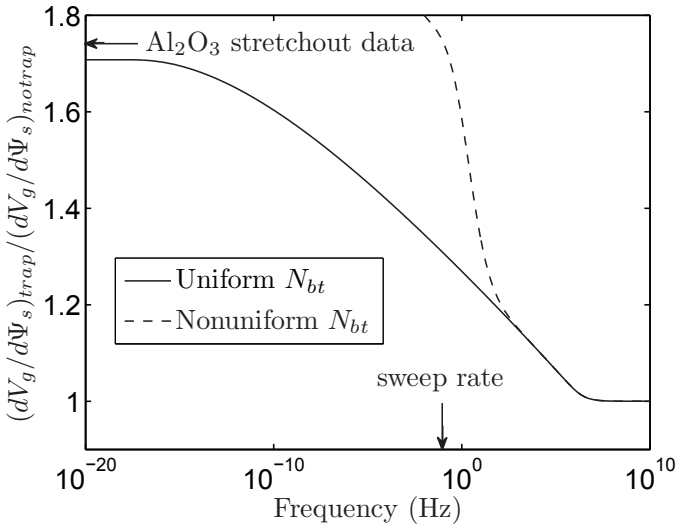


Figure II.27: Real part of $(\Delta V_g/\Delta\Psi_s)_{trap}/(\Delta V_g/\Delta\Psi_s)_{notrap}$ vs. frequency at $V_g = 0.3$ V for Al_2O_3 MOS with uniform N_{bt} and other parameters extracted from high frequency C and G dispersions in Fig. II.24 (solid line). The dashed line shows the ratio calculated with a nonuniform N_{bt} that produces the observed stretchout at a frequency in the range of the C-V sweep rate.

are bias dependent, the contribution of bulk oxide traps to the 1 MHz C_{tot} is also bias dependent and has the effect of steepening the slope. A more accurate way is to determine the slope using the extracted $C_{tot,\omega\tau_0\gg1} = C_sC_{ox}/(C_s + C_{ox})$ near $V_g = 0.3$ V. This is shown in Fig. II.28, which gives a $dC_{tot,\omega\tau_0\gg1}/dV_g$ slope a factor of 1.74 smaller than that of the simulated C-V with no stretchout. This factor is compared to the computed ratio in Fig. II.27 since $dC_{tot,\omega\tau_0\gg1}/dV_g$ is inversely proportional to $\Delta V_g/\Delta\Psi_s$ for the same $C_{tot,\omega\tau_0\gg1}$ and Ψ_s between the two curves. While the calculated DC ratio of 1.71 is comparable to the measured inverse-slope ratio, the calculated ratio of 1.31 at ~ 0.1 Hz corresponding to the sweep rate is lower. The model-to-data match can be fine tuned by allowing nonuniform N_{bt}, with higher trap densities toward the gate. The dashed curve in Fig. II.27 shows

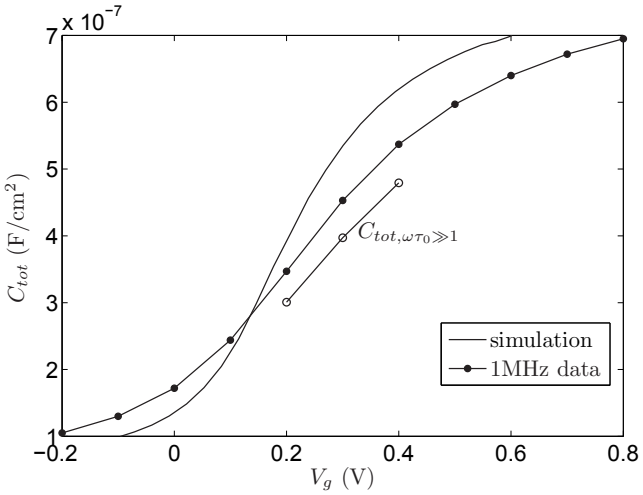


Figure II.28: Total capacitance of ideal simulation (solid line) and 1 MHz data (solid dots), as well as $C_{tot,\omega\tau_0\gg1}$ from model fitting (open circles) vs. V_g near the flatband bias condition. The extracted N_{bt} at $V_g = 0.2$ V and $V_g = 0.4$ V are 1.81×10^{19} cm^{-3}eV^{-1} and 2.6×10^{19} cm^{-3}eV^{-1}, respectively. The extracted τ_0 at $V_g = 0.2$ V and $V_g = 0.4$ V are 2.5×10^{-7} s and 6×10^{-8}, respectively.

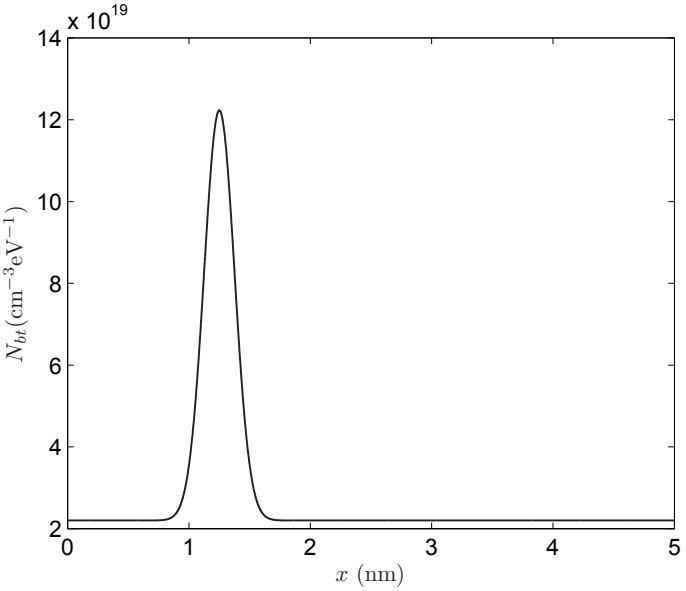


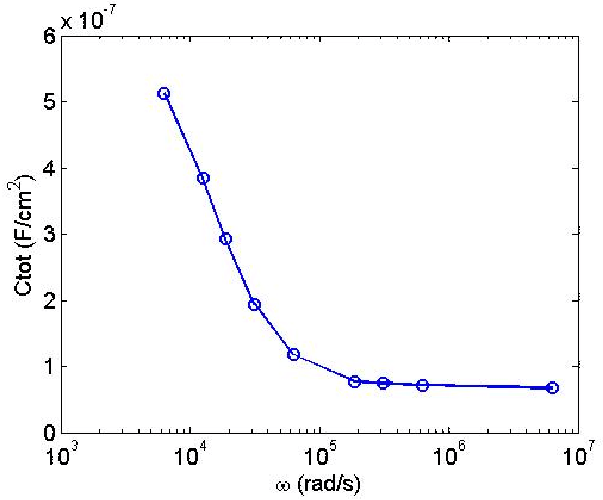
Figure II.29: Non-uniform N_{bt} vs. x used in stretch-out calculation of dashed curve in Fig. II.27.

the model calculation with a nonuniform N_{bt} shown in Fig. II.29 that reproduces the experimental $(\Delta V_g/\Delta\Psi_s)_{trap}$ at 0.1 Hz without affecting the dispersions at higher frequencies.

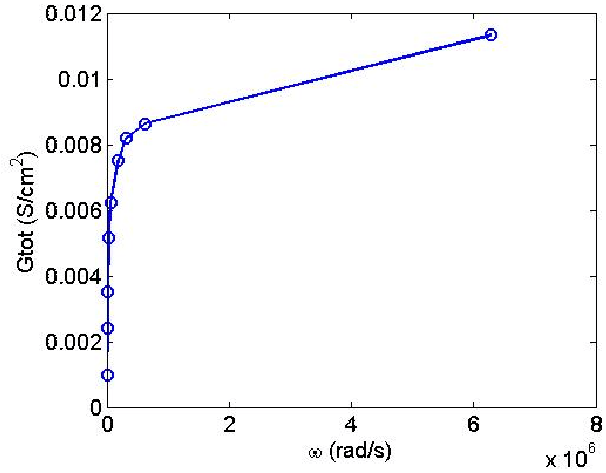
II.5 Modeling of Al_2O_3/n-$In_{0.53}Ga_{0.47}As$ C-V and G-V in Depletion, Weak Inversion and Strong Inversion

In previous section, we have discussed the distributed bulk-oxide trap model, which can explain frequency dispersion in C-V and G-V in strong and moderate accumulation. In this section, we will investigate frequency dispersion on the left side of Fig. II.8, especially the observed humps between $V_g = -1$ V and $V_g = 0$ V. Does the dispersion follow the bulk-oxide trap model? We will take a look at the trend of the dispersion around the humps. Fig. II.30 shows (a) capacitance and (b) conductance frequency dispersion data at $V_g = -0.8$ V of Fig. II.8. The capacitance does not follow $\propto \log(\omega)$ trend and conductance does not follow $\propto \omega$ trend. So the dispersion at humps could not be explained by bulk-oxide trap model.

Fig. II.31 compares experimental Al_2O_3/n-$In_{0.53}Ga_{0.47}As$ multi-frequency C-V in Fig. II.8 with quasi-static C-V from Sentaurus simulation. The humps in 1 kHz C-V occur in depletion and weak inversion where the inversion-layer capacitance is still low. So the humps should be due to interface states response. The quasi-static C-V shown in Fig. II.31 does not take interface states induced stretch-out into account. In real cases, the interface states charging and discharging



(a)



(b)

Figure II.30: (a) capacitance and (b) conductance frequency dispersion data at $V_g = -0.8$ V of Fig. II.8.

during gate bias sweep leads to stretch-out of C-V so that strong inversion takes place at more negative gate voltage than the shown ideal curve. We will adopt the full interface state model from Nicollian and Brews [85]. The full equivalent circuit of MOS with interface states is shown in Fig. II.32. C_D is the depletion

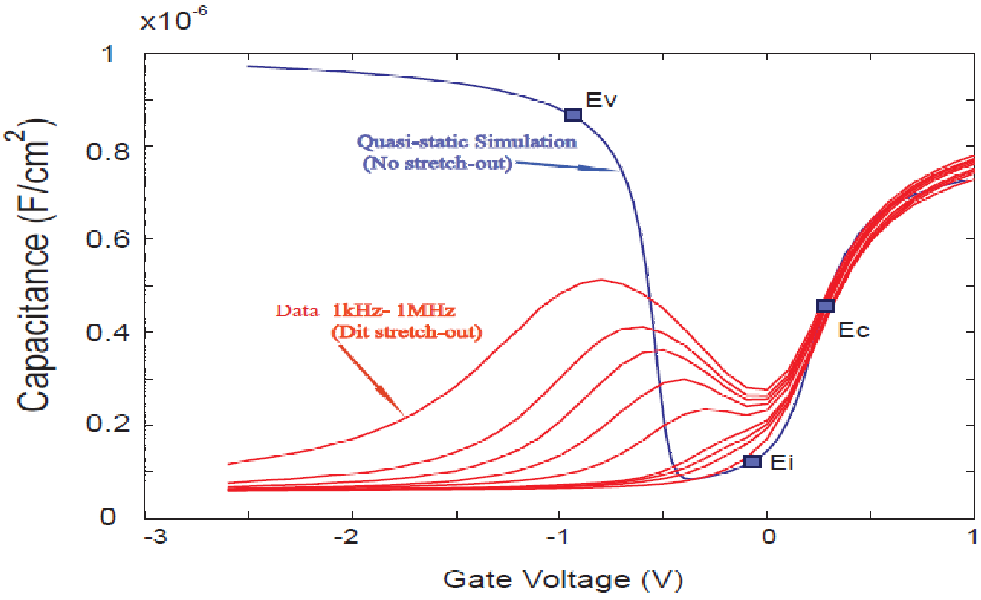


Figure II.31: Experimental Al_2O_3/n-$In_{0.53}Ga_{0.47}As$ multi-frequency C-V in Fig. II.8 is compared with quasi-static C-V from simulation with $C_{ox} = 1.06\mu F/cm^2$. The locations of conduction band edge E_c, valence band edge E_v and intrinsic Fermi level E_i are labeled.

charge capacitance; C_I is the inversion charge capacitance. C_{Tn}, C_{Tp} and G_{gr} are proportional to interface states density D_{it}, and are defined as

$$C_{Tn} = qD_{it}\tau_n^{-1}\int_0^1 (1-f)\left[j\omega f(1-f) + f\tau_p^{-1} + (1-f)\tau_n^{-1}\right]^{-1} df \qquad \text{(II.38)}$$

$$C_{Tp} = qD_{it}\tau_p^{-1}\int_0^1 f\left[j\omega f(1-f) + f\tau_p^{-1} + (1-f)\tau_n^{-1}\right]^{-1} df \qquad \text{(II.39)}$$

$$G_{gr} = -qD_{it}\tau_n^{-1}\tau_p^{-1}\int_0^1 \left[j\omega f(1-f) + f\tau_p^{-1} + (1-f)\tau_n^{-1}\right]^{-1} df \qquad \text{(II.40)}$$

where

$$\tau_n = (\sigma_n v_{th}^n n_s)^{-1} \qquad \text{(II.41)}$$

$$\tau_p = (\sigma_p v_{th}^p p_s)^{-1} \qquad \text{(II.42)}$$

τ_n is the time constant for a trap to capture an electron from conduction band, and τ_p is for trap to capture a hole from valence band. As V_g goes negative,

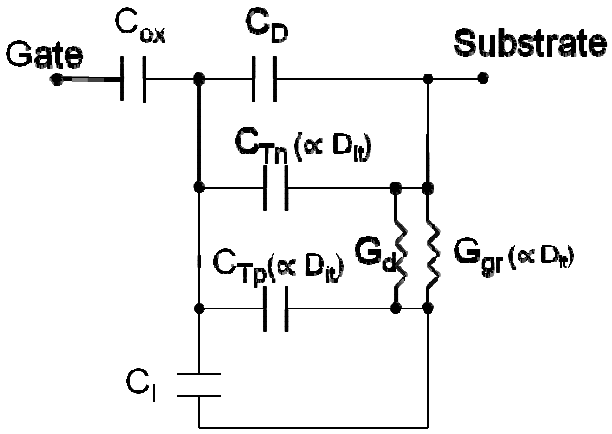


Figure II.32: Full equivalent circuit of MOS with interface states. Adapted from [85].

electron density at interface n_s decreases and hole density at interface p_s increases. As a result, τ_n increases and τ_p decreases. At different regions such as depletion, weak inversion and strong inversion, the full equivalent circuit can be simplified

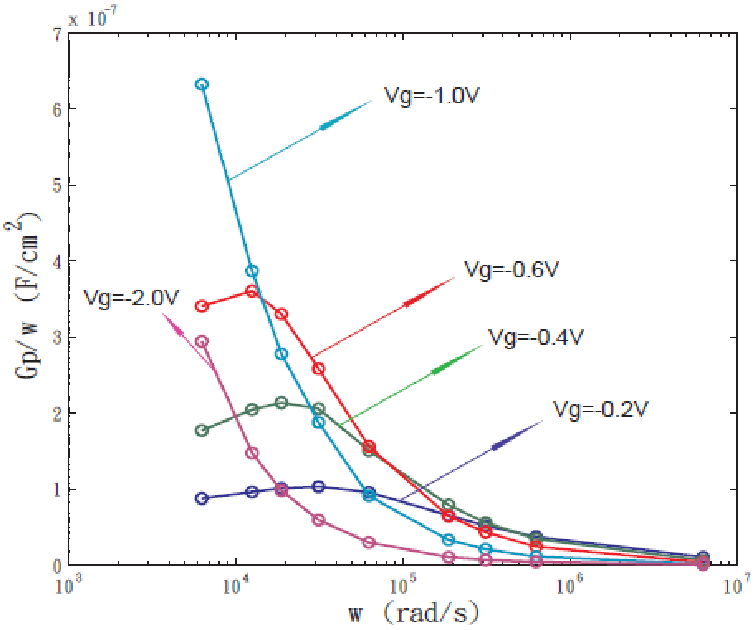
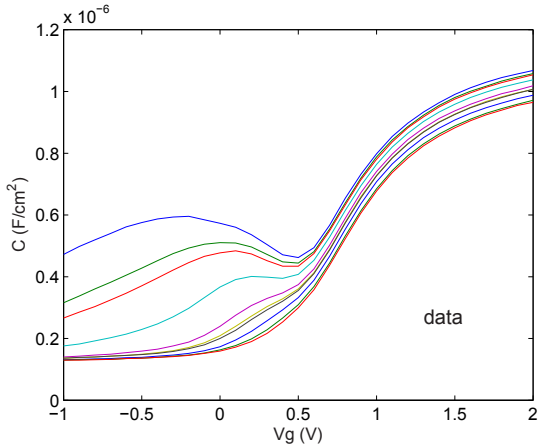
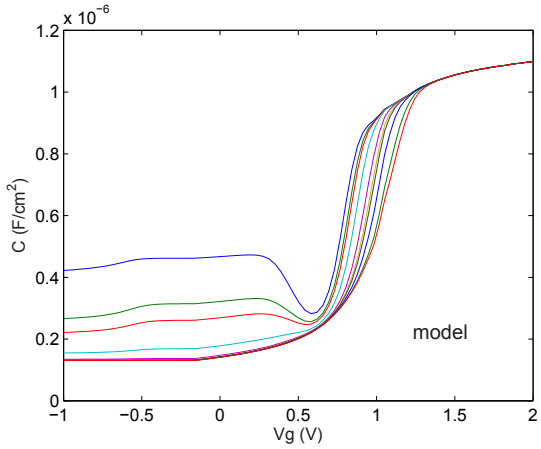


Figure II.33: Experimental G_p/ω of Al_2O_3/n-$In_{0.53}Ga_{0.47}As$ MOS in depletion region.



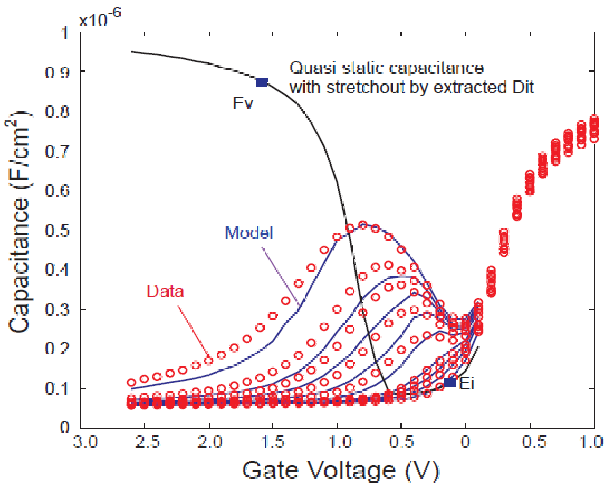
(a)



(b)

Figure II.34: (a) C-V data and corresponding (b) C-V fitting using $D_{it} = 1 \times 10^{13}$ $\mathrm{cm}^{-2}\mathrm{eV}^{-1}$ in interface states model.

according to the comparison between τ_n and τ_p. In depletion, $\tau_n < \tau_p$, C_{Tp} is negligible and there is no C_I. The equivalent circuit reduces to C_{ox} in series with a parallel combination of C_p and G_p, which are defined in (II.16) and (II.17). The G_p/ω exhibits a peak at $\omega\tau_n = 1.98$ as shown in solid curve in Fig. II.13, which is the theoretical basis of the well-known "conductance method".



(a)

Figure II.35: C-V fitting for the data of Fig. II.8. By courtesy of Han-Ping Chen.

Fig. II.33 shows the experimental G_p/ω in depletion region. The curves exhibit consistent trend with the model: as V_g goes negative, τ_n increases and the peak appears at lower frequency. D_{it} increase toward lower frequency can be observed from the peak magnitude increases toward lower frequency. Previously, a uniform D_{it} of 1×10^{13} cm^{-2}eV^{-1} was used for fitting data from a different sample. The results are shown in Fig. II.34. We can see that the fitting is not satisfactory. Interface states distribution should be non-uniform or localized. Another student Han-Ping Chen continues the work of modeling with non-uniform D_{it}, which leads to much better fitting of multi-frequency C-V in Fig. II.35.

II.6 Summary

In this chapter, we discuss the characteristics of high-κ III-V MOS capacitors, focusing on D_{it} and bulk-oxide traps extraction and modeling. The small signal admittance of an Al_2O_3/n-InGaAs MOS sample at multiple frequencies are analyzed. The ideal C-V is simulated to extract C_{ox}. Total trap density is extracted from high frequency stretch-out. However, it is much larger than the number predicted from conductance method. A distributed bulk-oxide trap model based on tunneling between the semiconductor surface and trap states in the gate insulator is developed. It differs fundamentally from the conventional interface state model in that there is a wide frequency spectrum of bulk-oxide trap response at a given gate bias. It is more physical than previously published lumped circuit models in the literature. The model is validated with the Pt/Al_2O_3/n-$In_{0.53}Ga_{0.47}As$ dispersion data in strong accumulation and near the flatband. Unlike surface states which are in units of areal density, bulk-oxide traps are characterized by a volume density, extracted from fitting of the capacitance and conductance data. It is further shown that the commonly employed method of extracting the interface state density from C-V stretchout could yield unphysical numbers inconsistent with the high-frequency dispersion data. On the other hand, the bulk-oxide trap model, in particular, with nonuniform trap density in the oxide film, can explain C-V stretchout independent of the dispersion at higher frequencies. Non-uniform D_{it} yields better fitting of the C-V humps on inversion side in the framework of full

conductance method.

The text of Chapter II, in part, is a reprint of the material as it appears in "A distributed bulk-oxide trap model for Al_2O_3-InGaAs MOS devices" by Yu Yuan, Bo Yu, Jaesoo Ahn, Paul C. McIntyre, Peter M. Asbeck, Mark J. W. Rodwell and Yuan Taur, submitted to IEEE Transactions on Electron Devices. The dissertation author was the primary investigator and author of this paper.

Chapter III

Design and Characterization of III-V MOSFETs

III.1 Further Scaling of 15 nm Baseline Design to 11 nm - SCEs Consideration

We will review our baseline sub-22 nm MOSFETs design [82] before discussing further scaling options. The schematic diagram of the baseline device design is shown in Fig. III.1. This is a thin-BOX-SOI-like structure. There are several design considerations to be captured briefly. First, a conductor not too far from the channel layer is needed to confine the lateral source-drain field for better short channel effects control. So the InGaAs under wide bandgap InAlAs barrier should be highly doped below channel region. Second, a self-aligned shallow Ti implant can be made after gate patterning to damage the p+ regions under

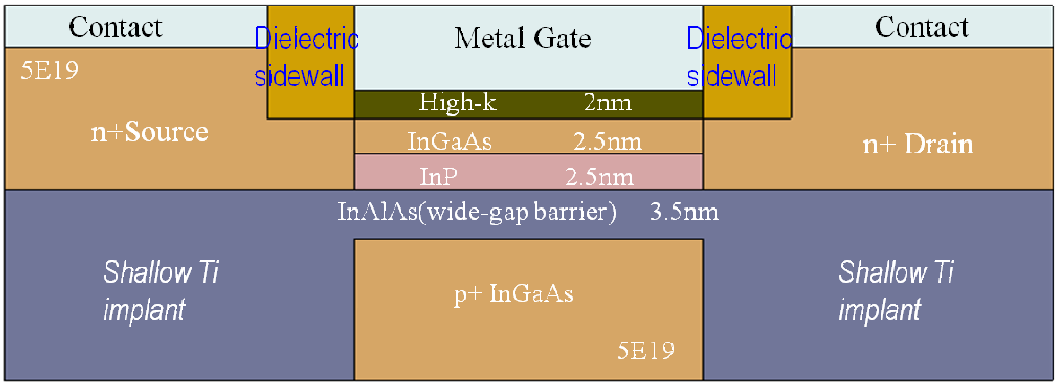


Figure III.1: Schematic diagram of the baseline device design of thin-BOX-SOI-like III-V MOSFETs for sub-22 nm scaling [82].

the source/drain to avoid excessive parasitic capacitance and resistance. Finally, raised source/drain can help to alleviate source starvation effects by increasing the chance of electrons' moving direction change through scattering [83]. The baseline design can be scaled to 15 nm in terms of drain induced barrier lowering (DIBL) and high drain Vt roll-off sensitivity on change length, as shown in Fig. III.2. Further scaling to 11 nm gate length from baseline design will be investigated. Two difference strategies will be considered: vertical scaling and lateral scaling. We set the following criteria for electrostatic integrity: (a) DIBL $<$ 100 mV/V; (b) High drain V_t roll-off $<$ 50 mV/10%L_g. In vertical scaling, we will consider reducing the channel thickness while keeping the same high-κ EOT. The original total channel thickness is 5 nm. Thinner thickness of 3 nm or 4 nm InGaAs will be used in TCAD simulation to calculate DIBL and V_t roll-off at multiple gate length. The results are shown in Fig. III.3 and Fig. III.4 for V_t roll-off and DIBL, respectively. When channel thickness decreases to 3 nm, V_t roll-off sensitivity falls

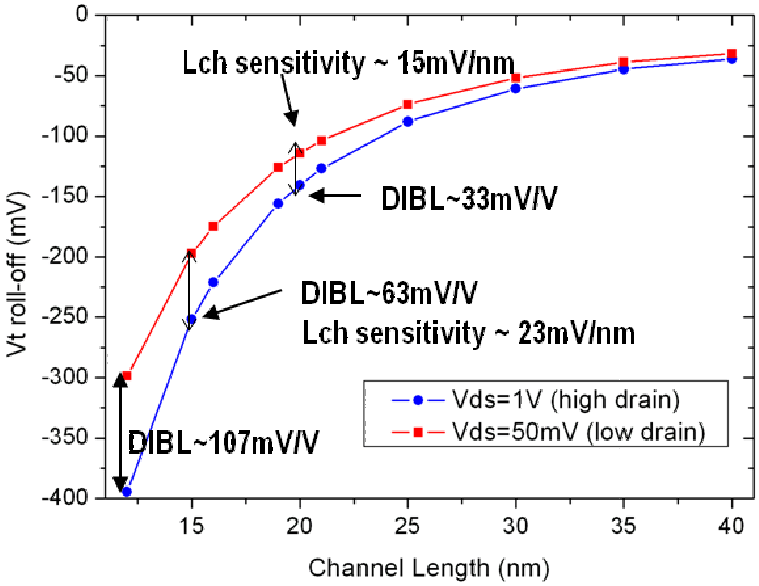


Figure III.2: V_t roll-off vs. channel length for baseline design. DIBL and V_t roll-off sensitivity are also labeled [82].

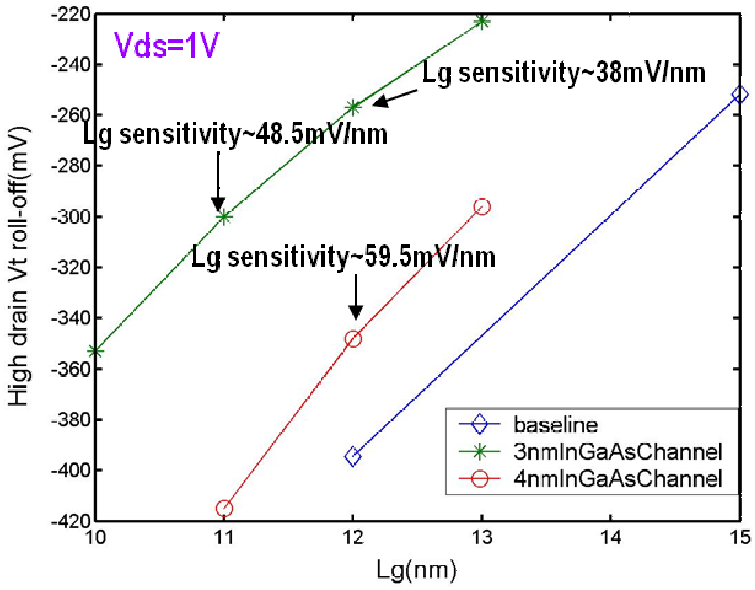


Figure III.3: High drain V_t roll-off vs. channel length for baseline design (diamond), and reduced channel thickness design of 3 nm (star) and 4 nm (circle). V_t roll-off sensitivity on channel length sensitivity is labeled.

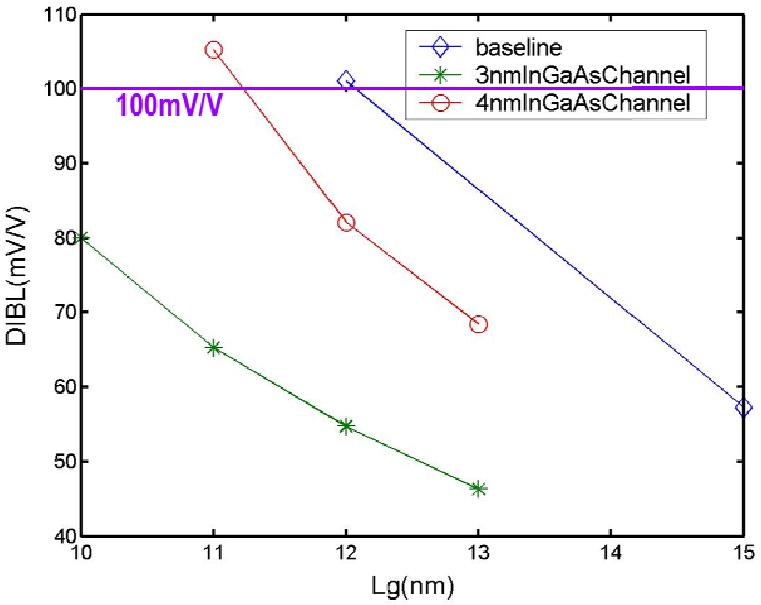


Figure III.4: DIBL vs. channel length for baseline design (diamond), and reduced channel thickness design of 3 nm (star) and 4 nm (circle). The criterion of 100 mV/V is labeled. The simulation is conducted at $V_{ds} = 1$ V.

below 50 mV/10%L_g and DIBL drops below 100 mV/V at $L_g = 11$ nm. So the device can be scaled to 11 nm with 3 nm channel thickness but not with 4 nm

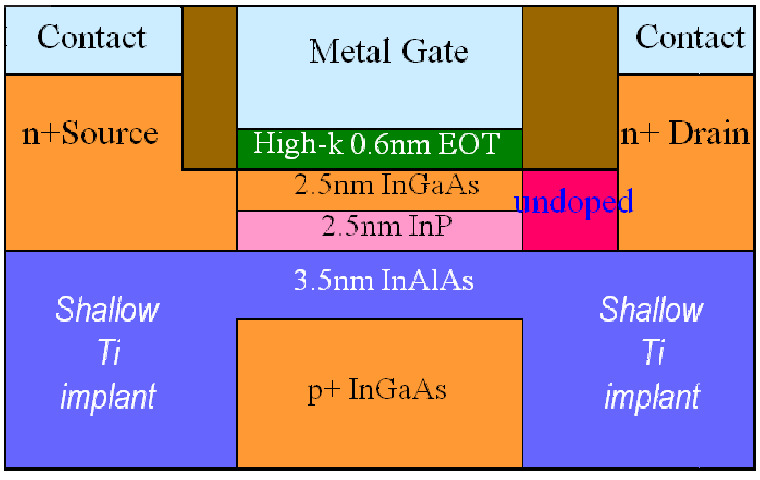


Figure III.5: Schematic diagram of the modified baseline device design with undoped underlap region on the drain side (in pink color).

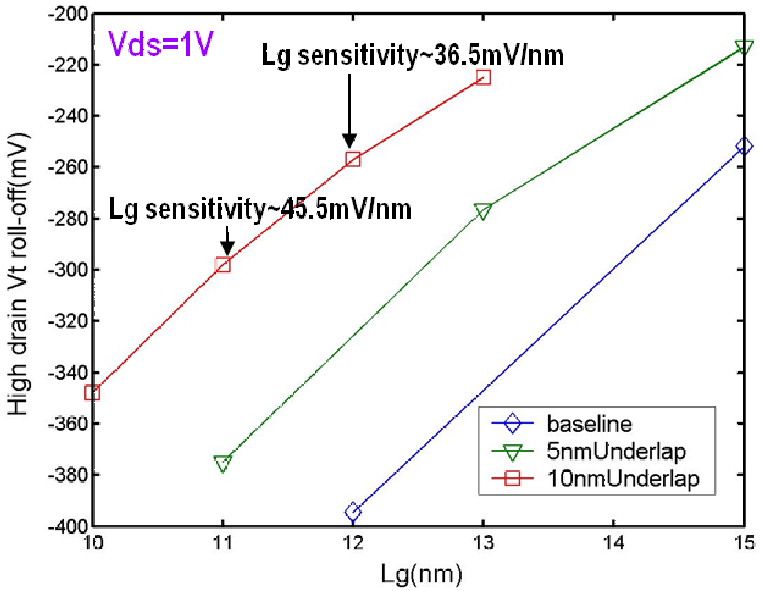


Figure III.6: High drain V_t roll-off vs. channel length for baseline design (diamond), adding 5 nm underlap (triangle) and adding 10 nm underlap (square) design. V_t roll-off sensitivity on channel length sensitivity is labeled.

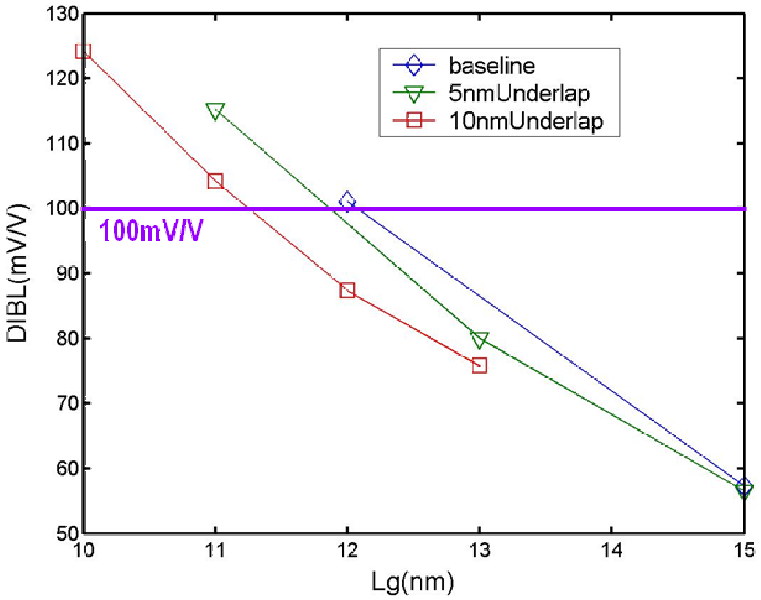


Figure III.7: DIBL vs. channel length for baseline design (diamond), adding 5 nm underlap (triangle) and adding 10 nm underlap (square) design. The criterion of 100 mV/V is labeled. The simulation is conducted at $V_{ds} = 1$ V.

channel thickness. Compared to 4 nm channel thickness, DIBL improves significantly for 3 nm channel thickness. In lateral scaling, we will consider asymmetric device with undoped underlap region on the drain side, as shown in Fig. III.5. Adding undoped underlap region plays a similar role in improving SCE as increasing channel length. The effects on high drain V_t roll-off and DIBL are illustrated in Fig. III.6 and Fig. III.7, respectively. The device can barely be scaled to 11 nm channel length with 10 nm underlap region. Compared with DIBL improvement by vertical scaling, lateral underlap is less effective in reducing DIBL.

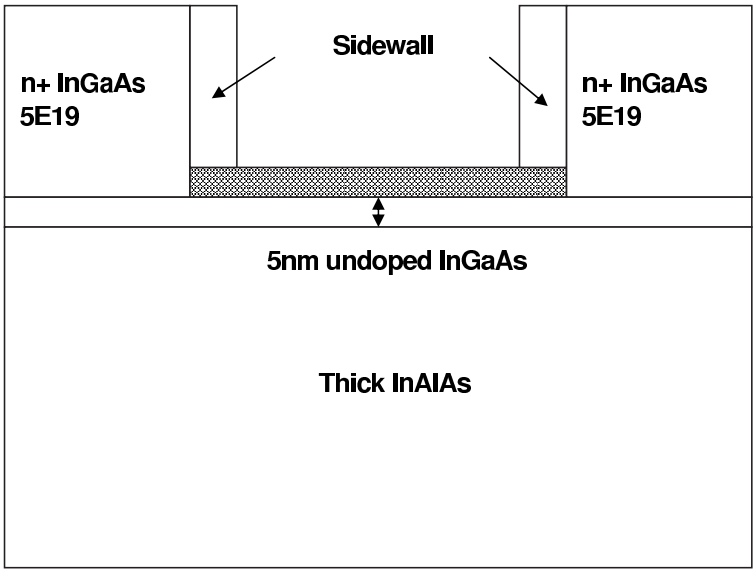


Figure III.8: Schematic diagram of the raised source/drain structure, optimized to avoid "source starvation" [83].

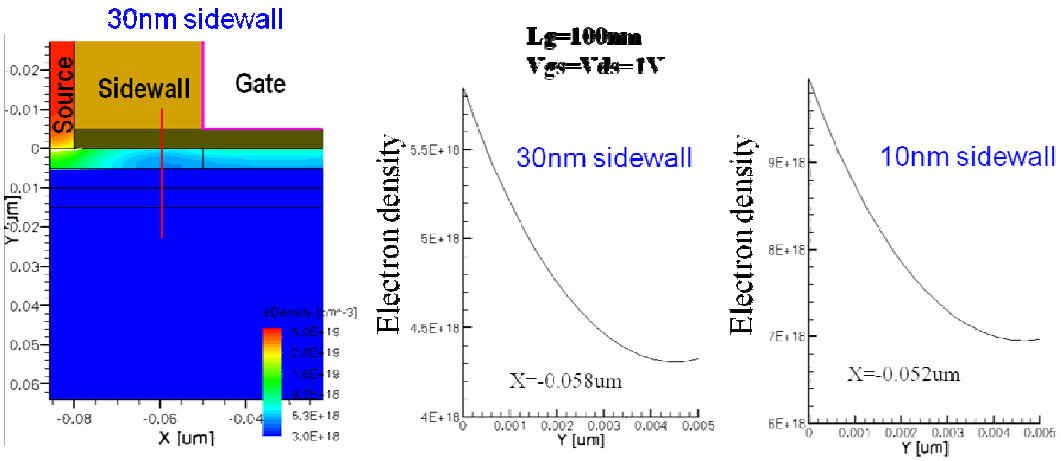


Figure III.9: Simulated electron density vs. y-axis under the sidewall on the source side for 10 nm sidewall and 30 nm sidewall with 100 nm gate length at $V_{gs} = 1$ V and $V_{ds} = 1$ V.

III.2 Raised Source/Drain Process Issues

Two source/drain (S/D) patterns have been considered in our III-V MOSFET design. One is recessed S/D structure and the other is raised S/D structure. We will focus on raised S/D structure shown in Fig. III.8. On both sides of undoped InGaAs channel, there are extensions of undoped InGaAs. These undoped InGaAs extensions might introduce extra series resistance to the MOSFET. We are interested in electron density in the undoped InGaAs extension on the source side because of its effect on the device current. The region under the sidewall is more difficult to obtain electrons compared to the region under n+ source. Electrons in the underlap regions come primarily from three sources: spillover from n+ raised source; fringing capacitance induced electrons; spillover from gated channel. Electron density under the sidewall is very sensitive to sidewall thickness. Elec-

tron density from Sentaurus simulation is plotted in Fig. III.9 for 10 nm and 30 nm sidewall. Electron density in the underlap region is more than 4×10^{18} cm^{-3} even for 30 nm thick sidewall, which is acceptable. Relatively low electron density under the sidewall causes extra series resistance to the MOSFETs. This can be observed from the simulated band diagram shown in Fig. III.10. The source is on the left side. The kink of the electron Fermi level on the source side indicates the I-R drop caused by the side wall. The series resistance can be estimated through the I-R drop and the MOSFET current at $V_{gs} = 1$ V and $V_{ds} = 1$ V. The source series resistance is less than 15 Ohm·μm for 10 nm wide side wall. Even for 30 nm side wall, the source series resistance is only about 75 Ohm·μm, which will not significantly affect the device performance. I_{ds} vs. V_{gs} is plotted in Fig. III.11 for different sidewall thickness. Transconductance degrades as the sidewall thickness increases because of the series resistance effects. Threshold voltage increases as the sidewall thickness increases due to larger source to drain separation and hence better short channel effects control. 30 nm sidewall results in 25% degradation in "on" current.

III.3 III-V MOSFETs I-V, C-V, and Mobility Extraction

To characterize III-V MOSFETs with high-κ gate insulator, interface traps effects on device performance cannot be ignored due to the their large number. Therefore, we will discuss the effects of interface traps on device performance:

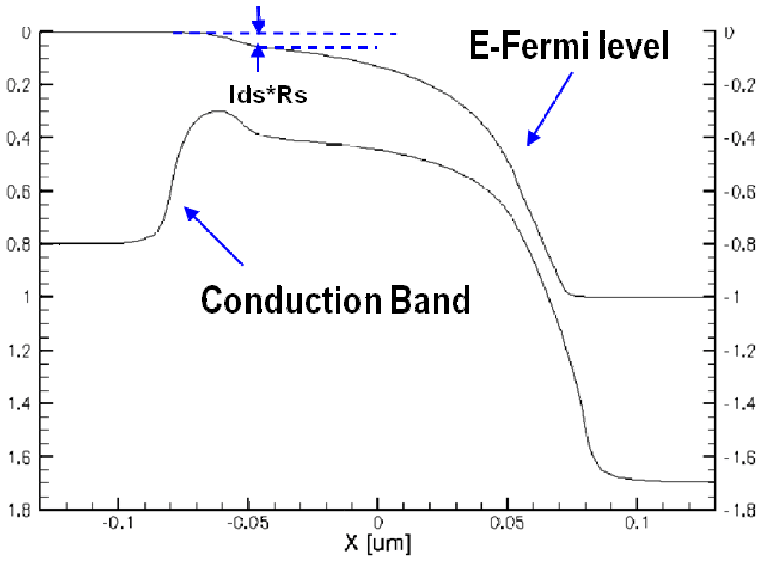


Figure III.10: The simulated band diagram for the MOSFETs structure in Fig. III.8, with 100 nm gate length at $V_{gs} = 1$ V and $V_{ds} = 1$ V.

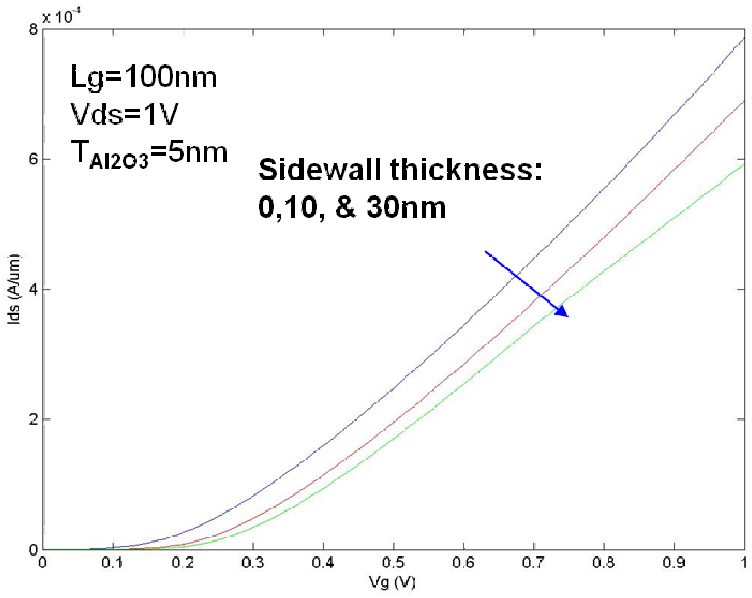
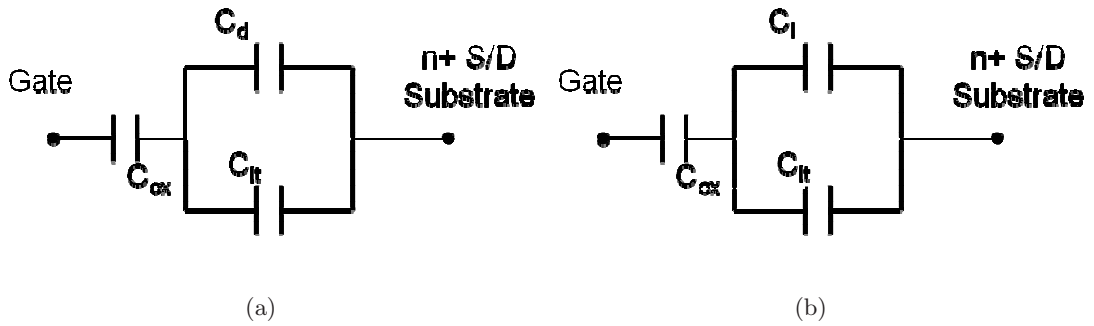


Figure III.11: The simulated I_{ds} vs. V_{gs} for the MOSFETs structure in Fig. III.8 of 100 nm gate length and various sidewall thickness. $V_{ds} = 1$ V.

subthreshold degradation in the subthreshold region; inversion charge capacitance degradation above threshold. Fig. III.12 shows the equivalent circuits for MOS-



(a) (b)

Figure III.12: The equivalent circuits for MOSFETs with interface traps in (a) subthreshold region and (b) above threshold region.

FETs with interface traps in (a) subthreshold region and (b) above threshold region for DC condition. Interface traps' occupancy change follows DC gate bias sweep without delay, so it can be modeled by a pure capacitance, in parallel with depletion capacitance in subthreshold region and with inversion capacitance in above threshold condition. These two equivalent circuits will be utilized in the characterization of InGaAs MOSFETs with Al_2O_3 gate insulator.

In Fig. III.13, I_{ds} at $V_{gs} = 0.05$, 1, and 2.5 V are plotted. Poor subthreshold slope of 720 mV/dec is observed due to large D_{it}. According to Fig. III.12 (a), the subthreshold slope is given by

$$S.S. = \frac{C_{ox} + C_d + C_{it}}{C_{ox}} \times 60mV/dec \tag{III.1}$$

C_d is negligible compared to C_{it} due to large D_{it} and therefore large C_{it}. D_{it} is estimated to be $\sim 1 \times 10^{14}$ $cm^{-2}eV^{-1}$ from (III.1).

To extract effective mobility of the device, mobile charge density needs to be characterized from device measurement. In mature Si technology, a single fre-

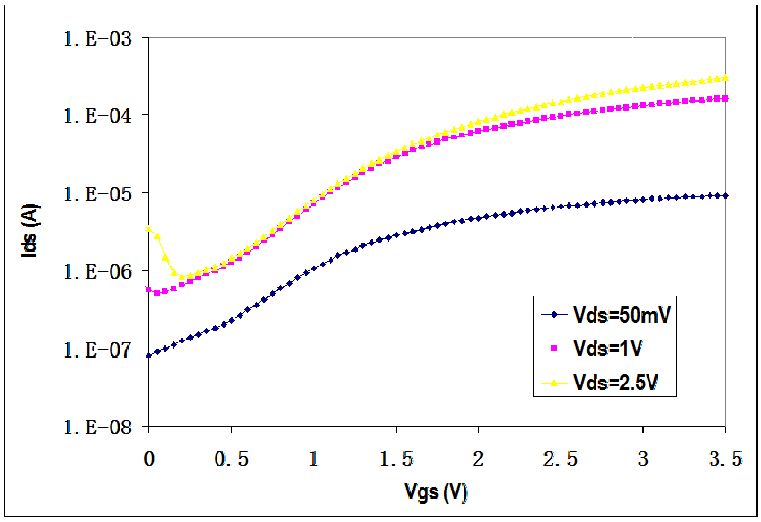


Figure III.13: I_{ds} vs. V_{gs} for three terminal $In_{0.53}Ga_{0.47}As$ channel MOSFETs with 5 nm Al_2O_3 gate insulator and semi-insulating substrate. The device width is 7 μm; gate length is 10 μm. From Prof. Mark Rodwell, University of California, Santa Barbara.

quency split C-V measurement can provide mobile charge information. Since the MOSFET to be characterized is three terminal device, gate to source/drain capacitance can be measured to extract mobile charge density. Another factor to be considered is trap response to C-V measurement. In high-κ III-V MOSFETs, much more traps are present at insulator-semiconductor interface and inside gate insulator than Si MOSFETs. Single frequency C-V is not sufficient to characterized mobile charge density when trap response plays a role. To separate trap response from inversion charge response, gate to channel capacitance is measured at multiple frequencies especially high frequencies. The two probes connect to gate pad and source/drain pad. The C-V data are plotted in Fig. III.14. The multi-frequency gate to channel capacitance exhibits strong frequency dependence. This is consistent with frequency dependent interface traps response. At lower frequen-

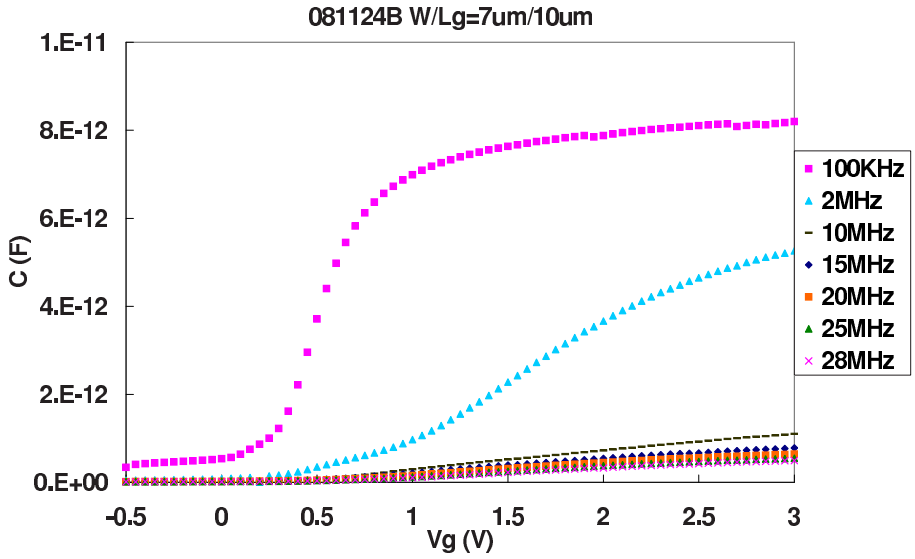


Figure III.14: Multi-frequency gate to channel capacitance for three terminal $In_{0.53}Ga_{0.47}As$ channel MOSFETs with 5 nm Al_2O_3 gate insulator and semi-insulating substrate. The device width is 7 μm; gate length is 10 μm. The sample is provided by Prof. Mark Rodwell, University of California, Santa Barbara.

cies, more trap response are present. So low frequency C-V should mainly corresponds to trap response instead of true inversion charge response. The inversion charge response should be instant and frequency independent in the presence of source/drain electron reservoir. So the C-V data at highest measurement frequency are least contributed by interface traps response and will be used to estimate inversion charge density. When there is no interface traps response, the equivalent circuit in Fig. III.12 (b) reduces to a simple series combination of C_{ox} and C_i, where C_i is inversion charge sheet density differentiation with respect to surface potential. So we can obtain C_i through high frequency gate to channel capacitance with known C_{ox}. On the other hand, I_{ds} is measured in DC condition, when the total equivalent circuit is as shown in Fig. III.12 (b). To calculate the inversion

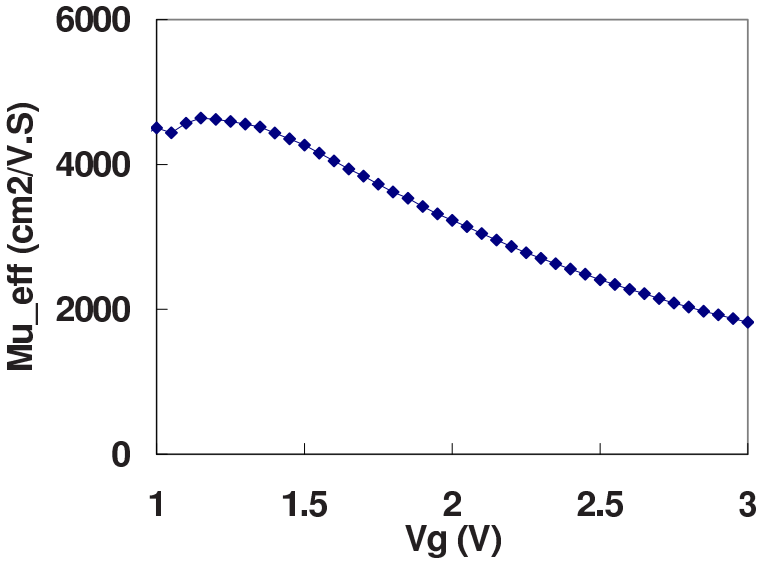


Figure III.15: Estimated electron effective mobility for the three terminal $In_{0.53}Ga_{0.47}As$ channel MOSFETs with 5 nm Al_2O_3 gate insulator and semi-insulating substrate. The device width is 7 μm; gate length is 10 μm.

charge density per unit area vs. gate bias, we need to know C_{inv} which is defined by inversion charge sheet density differentiation with respect to gate voltage V_g. With interface traps present, C_{inv} is given by

$$C_{inv} = \frac{C_{ox}C_i}{C_{ox} + C_i + C_{it}} \tag{III.2}$$

C_{inv} could be degraded significantly by large C_{it}. With known C_{ox}, C_i and C_{it}, we can obtain C_{inv} vs. V_g. Inversion charge sheet density Q_i is obtained by integrating C_{inv} over V_g:

$$Q_i(V_g) = \int_{-\infty}^{V_g} C_{inv}(V_g^{'})dV_g^{'} \tag{III.3}$$

Then electron effective mobility can be estimated through low drain bias I_{ds} vs. V_g data:

$$\mu_{eff} \approx \frac{L_g I_{ds}}{W Q_i V_{ds}} \tag{III.4}$$

where L_g and W are MOSFET gate length and width, respectively; V_{ds} is the drain to source bias. The effective mobility is estimated from V_{ds} = 50 mV I_{ds} vs. V_g data for the L_g = 10 μm and W = 7 μm device, whose I_{ds} vs. V_g and gate to channel capacitance are shown in Fig. III.13 and Fig. III.14, respectively. The result is plotted in Fig. III.15. The estimated effective electron mobility is in the range of 2000 to 5000 $cm^2/V{\cdot}S$. Assuming no interface traps response at highest measurement frequency may overestimate inversion charge density and hence underestimate the mobility.

III.4 Summary

In device design part, the baseline sub-22 nm MOSFETs design can be scaled to 11 nm gate length by shrink the channel thickness to 3 nm and can be barely scaled to 11 nm gate length by adding 10 nm undoped underlap region at the drain side; the series resistance induced by increasing the sidewall thickness to 30 nm in a L_g = 100 nm MOSFET is acceptable. In device characterization part, the mobile charge density in high-κ InGaAs MOSFET is characterized by frequency dependent measurement of gate to channel capacitance and electron effective mobility is extracted to be in the range of 2000 to 5000 $cm^2/V{\cdot}S$.

Chapter IV

Nanowire MOSFETs: Compact Model and Scaling Limit

A continuous analytic drain current model has been developed for long-channel nanowire transistors recently [99]. This model is developed by following the approach for DG MOSFETs in [100] of Professor Taur. It is based on the exact solution of the Poisson's equation, and is derived directly from Pao-Sah integral without charge sheet approximation. Therefore, the model is able to cover all the operation regions (linear, saturation, subthreshold) as well as the transition between them without fitting parameters. On top of the drain current model, the analytic charge and capacitance model has also been obtained to constitute a complete set of compact model for long-channel nanowire transistors, i.e., enable AC and transient CAD simulation [101]. But the focus of this chapter is not on the long-channel nanowire transistors. On the contrary, we will put emphasis on the

compact model of short-channel nanowire transistors, especially those close to the scaling limit, where both short channel effects (SCEs) and quantum effects have to be included.

IV.1 Analytic Model for SCEs in Nanowire MOSFETs

Fig. IV.1 shows the schematic diagram of an undoped (or lightly doped) nanowire MOSFET. In the subthreshold region, both the fixed and mobile charge densities are negligible, therefore Poisson's equation takes the following form in both the silicon and insulator regions:

$$\frac{1}{\rho}\frac{\partial}{\partial\rho}\rho\left(\epsilon\left(\rho\right)\frac{\partial\psi}{\partial\rho}\right)+\frac{\partial}{\partial y}\left(\epsilon\left(\rho\right)\frac{\partial\psi}{\partial y}\right)=0 \tag{IV.1}$$

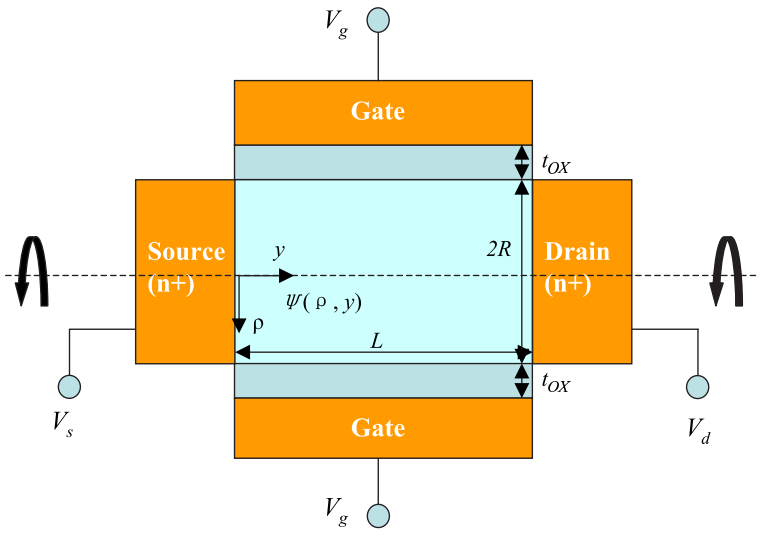


Figure IV.1: Schematic diagram of a nanowire MOSFET (cut through symmetry axis).

where the electrostatic potential $\psi(\rho, y)$ is defined as the intrinsic potential at a point (ρ, y) with respect to the Fermi level in the n+ source, and $\epsilon(\rho)$ is the dielectric constant of silicon or the insulator. If we assume the source and drain junctions are abrupt, the boundary conditions are as follows:

$$\begin{aligned} Gate &: \psi(R_i, y) = V_g - \Delta\phi, 0 \leqslant y \leqslant L \\ Source &: \psi(\rho, 0) = V_s + V_{bi}, 0 \leqslant \rho \leqslant R \\ Drain &: \psi(\rho, L) = V_d + V_{bi}, 0 \leqslant \rho \leqslant R \end{aligned} \tag{IV.2}$$

where $R_i = R + t_{ox}$, $\Delta\phi$ is the work function of gate electrode with respect to the intrinsic silicon, and V_{bi} is the build-in voltage roughly given by $V_{bi} = E_g/2q$. It can be noticed that we have no explicit boundary conditions in the insulator gap regions between the source/drain and gate. Numerical simulation results show that the potential in this region is approximately a logarithm function of ρ. The potential in the insulator gap region can then be expressed by

$$\begin{aligned} \psi(\rho, 0) &= (V_s + V_{bi}) + \frac{V_g - \Delta\phi - V_s - V_{bi}}{\ln(R_i/R)} \ln\frac{\rho}{R}, R < \rho \leqslant R_i \\ \psi(\rho, L) &= (V_d + V_{bi}) + \frac{V_g - \Delta\phi - V_d - V_{bi}}{\ln(R_i/R)} \ln\frac{\rho}{R}, R < \rho \leqslant R_i \end{aligned} \tag{IV.3}$$

By applying the superposition principle, the electrostatic potential in the nanowire MOSFET can be written as

$$\psi(\rho, y) = v(\rho) + u_L(\rho, y) + u_R(\rho, y) \tag{IV.4}$$

where $v(\rho)$ is the 1-D solution of

$$\frac{1}{\rho}\frac{\partial}{\partial\rho}\rho\left(\epsilon(\rho)\frac{\partial v(\rho)}{\partial\rho}\right) = 0 \tag{IV.5}$$

and satisfies gate boundary condition. u_L and u_R are solutions to Poisson's equation and satisfy the source and drain boundary conditions, respectively. For example, u_L is zero on the gate and drain boundaries, but $v + u_L$ satisfies the source boundary condition. Similarly, u_R is zero on the gate and source boundaries, but $v + u_R$ satisfies the drain boundary condition [43, 102].

Owing to the volume inversion effect in the subthreshold region, we can choose $v(\rho) = constant = V_g - \Delta\phi$. Meanwhile, u_L and u_R can be written as $u_L(\rho, y) = \Sigma_{n=1}^{\infty} b_n B_n(\rho, y)$ and $u_R(\rho, y) = \Sigma_{n=1}^{\infty} c_n C_n(\rho, y)$, respectively, where

$$B_n(\rho, y) = \begin{cases} J_0(k_n\rho)\frac{\sinh[k_n(L-y)]}{\sinh(k_nL)}, 0 \leqslant \rho \leqslant R \\ J_0(k_nR)\frac{J_0(k_n\rho)Y_0(k_nR_i)-J_0(k_nR_i)Y_0(k_n\rho)}{J_0(k_nR)Y_0(k_nR_i)-J_0(k_nR_i)Y_0(k_nR)}\frac{\sinh[k_n(L-y)]}{\sinh(k_nL)}, R < \rho \leqslant R_i \end{cases} \tag{IV.6}$$

and

$$C_n(\rho, y) = \begin{cases} J_0(k_n\rho)\frac{\sinh(k_ny)}{\sinh(k_nL)}, 0 \leqslant \rho \leqslant R \\ J_0(k_nR)\frac{J_0(k_n\rho)Y_0(k_nR_i)-J_0(k_nR_i)Y_0(k_n\rho)}{J_0(k_nR)Y_0(k_nR_i)-J_0(k_nR_i)Y_0(k_nR)}\frac{\sinh(k_ny)}{\sinh(k_nL)}, R < \rho \leqslant R_i \end{cases} \tag{IV.7}$$

Here $J_0(r)$ denotes the 0-th order Bessel function of the first kind (Bessel function), and $Y_0(r)$ denotes the 0-th order Bessel function of the second kind (Neumann function) [103].

The implicit boundary conditions at the silicon/insulator interface also require $\epsilon(\rho)(\partial u_{L,R}/\partial\rho)$ to be continuous at $\rho = R$, which lead to:

$$\epsilon_{si}\frac{J_1(k_nR)}{J_0(k_nR)} = \epsilon_{ox}\frac{J_1(k_nR)Y_0(k_nR_i) - J_0(k_nR_i)Y_1(k_nR)}{J_0(k_nR)Y_0(k_nR_i) - J_0(k_nR_i)Y_0(k_nR)} \tag{IV.8}$$

where $J_1(r)$ and $Y_1(r)$ are the 1-st order Bessel function and Neumann function, respectively. This eigenvalue equation is equivalent to that in [104]. By solving

this implicit equation, we can obtain the eigenvalues of k_n, which form a sequence $k_1 < k_2 < k_3 < \cdots$ with $k_n/k_1 \sim n$. The generalized scale length λ is defined by $\lambda = \pi/k_1$.

The eigenfunctions $B_n(\rho, 0)$ $(=C_n(\rho, L))$ are not orthogonal to each other. In order to evaluate the coefficients b_n and c_n, we need to find the orthogonality relationship in this case. It can be approved that the orthogonality relationship takes the form

$$\int_0^{R_i} B_n(\rho, 0) B_m(\rho, 0) \epsilon(\rho) \rho d\rho = 0, n \neq m \tag{IV.9}$$

The key point to prove equation (IV.9) is to analytically carry out the integrals in the form of $\int_{t_1}^{t_2} t f_1(at) f_2(bt) dt$ $(a \neq b)$, where the functions f_1 and f_2 can be either Bessel function J_0 or Neumann function Y_0. Here, we will take $\int_{t_1}^{t_2} t J_0(at) Y_0(bt) dt$ as an example to show how to calculate this type of integral. According to the definitions, one has

$$\frac{1}{t}\frac{d}{dt}\left[t\frac{dJ_0(at)}{dt}\right] + a^2 J_0(at) = 0 \tag{IV.10}$$

$$\frac{1}{t}\frac{d}{dt}\left[t\frac{dY_0(bt)}{dt}\right] + b^2 Y_0(bt) = 0 \tag{IV.11}$$

(IV.10)$\times (tY_0(bt))$ $-$(IV.11)$\times (tJ_0(at))$ leads to

$$\left(a^2 - b^2\right) t J_0(at) Y_0(bt) = -Y_0(bt)\frac{d}{dt}\left[t\frac{dJ_0(at)}{dt}\right] + J_0(at)\frac{d}{dt}\left[t\frac{dY_0(bt)}{dt}\right] \tag{IV.12}$$

Integrating the above equation, we finally obtain

$$\left(a^2 - b^2\right)\int_{t_1}^{t_2} t J_0(at) Y_0(bt) dt = \left[atJ_1(at) Y_0(bt) - btJ_0(at) Y_1(bt)\right]\Big|_{t=t_1}^{t=t_2} \tag{IV.13}$$

By multiplying $u_{L,R}$ with $B_n(\rho,0)\,\epsilon(\rho)\,\rho$ and integrating, we will obtain the coefficients b_n and c_n. b_n can be expressed as

$$b_n = \frac{\int_0^{R_i} [\psi(\rho,0) - v(\rho)]\, B_n(\rho,0)\,\epsilon(\rho)\,\rho d\rho}{\int_0^{R_i} B_n^2(\rho,0)\,\epsilon(\rho)\,\rho d\rho} \tag{IV.14}$$

and c_n can be expressed as

$$c_n = \frac{\int_0^{R_i} [\psi(\rho,L) - v(\rho)]\, C_n(\rho,L)\,\epsilon(\rho)\,\rho d\rho}{\int_0^{R_i} C_n^2(\rho,L)\,\epsilon(\rho)\,\rho d\rho} \tag{IV.15}$$

The above tedious integrals can be carried out by using the basic characteristics of Bessel and Neumann functions [103]. To derive explicit expressions for b_n and c_n, we have to analytically calculate following types of integral: $\int_{t_1}^{t_2} t f_1(at)\,dt$, $\int_{t_1}^{t_2} t\ln t f_1(at)\,dt$, and $\int_{t_1}^{t_2} t f_1(at)\,f_2(at)\,dt$. Here, we will list the result of an example for each type, and the others can be obtained by replacing $J_{0,1}$ with $Y_{0,1}$, or reversely:

$$\int_{t_1}^{t_2} tJ_0(at)\,dt = \left[\frac{tJ_1(at)}{a}\right]\Bigg|_{t=t_1}^{t=t_2} \tag{IV.16}$$

$$\int_{t_1}^{t_2} t\ln tJ_0(at)\,dt = \left[\frac{t\ln tJ_1(at)}{a} + \frac{J_0(at)}{a^2}\right]\Bigg|_{t=t_1}^{t=t_2} \tag{IV.17}$$

$$\int_{t_1}^{t_2} tJ_0(at)\,Y_0(at)\,dt = \left[\frac{t^2J_1(at)\,Y_1(at)}{2} + \frac{t^2J_0(at)\,Y_0(at)}{2}\right]\Bigg|_{t=t_1}^{t=t_2} \tag{IV.18}$$

According to these intermediate results, we can obtain the final expressions of b_n and c_n as

$$b_n = E(V_s + V_{bi} + \Delta\phi - V_g) \tag{IV.19}$$

and

$$c_n = E(V_d + V_{bi} + \Delta\phi - V_g) \tag{IV.20}$$

where

$$E = \frac{\pi^2}{2} \frac{1}{\ln\left(R_i/R\right)} \frac{\epsilon_i}{\epsilon_{si}} \frac{J_0\left(k_n R\right)}{A \cdot B + \frac{\pi^2}{4} k_n^2 R^2 \cdot C} \tag{IV.21}$$

with

$$\begin{aligned} A &= \frac{J_0\left(k_n R\right)}{J_0\left(k_n R\right) Y_0\left(k_n R_i\right) - J_0\left(k_n R_i\right) Y_0\left(k_n R\right)} \\ B &= \frac{J_1\left(k_n R\right)}{J_1\left(k_n R\right) Y_0\left(k_n R_i\right) - J_0\left(k_n R_i\right) Y_1\left(k_n R\right)} \\ C &= \left(1 - \frac{\epsilon_i}{\epsilon_{si}}\right) J_0^2\left(k_n R\right) + \left(1 - \frac{\epsilon_{si}}{\epsilon_i}\right) J_1^2\left(k_n R\right) \end{aligned} \tag{IV.22}$$

Note that b_n and c_n are bias dependent. The V_d dependence in c_n is responsible for drain induced barrier lowering (DIBL), and the V_g dependence in both b_n and c_n is responsible for subthreshold slope degradation, which will be discussed later.

For $L > 1.3\lambda$, the $u_{L,R}$ series decay rapidly since $k_n/k_1 \sim n$. Therefore only the lowest order terms in the $u_{L,R}$ series need to be kept. Then the expression for the full 2-D potential in the subthreshold region is $\psi\left(\rho, y\right) = V_g - \Delta\phi + b_1 B_1\left(\rho, y\right) + c_1 C_1\left(\rho, y\right)$, which in the silicon region is explicitly given by

$$\psi\left(\rho, y\right) = V_g - \Delta\phi + \frac{b_1 \sinh\left[k_1\left(L - y\right)\right] + c_1 \sinh\left(k_1 y\right)}{\sinh\left(k_1 L\right)} J_0\left(k_1 \rho\right) \tag{IV.23}$$

Once the analytical potential is obtained, the subthreshold current can be derived based on the current continuity equation

$$I_{ds}\left(y\right) = \mu\left(2\pi R\right) Q_i\left(y\right) \frac{dV\left(y\right)}{dy} \tag{IV.24}$$

where

$$Q_i\left(y\right) = \frac{q n_i}{R} \int_0^R e^{q[\psi(\rho,y) - V(y)]/kT} \rho d\rho \tag{IV.25}$$

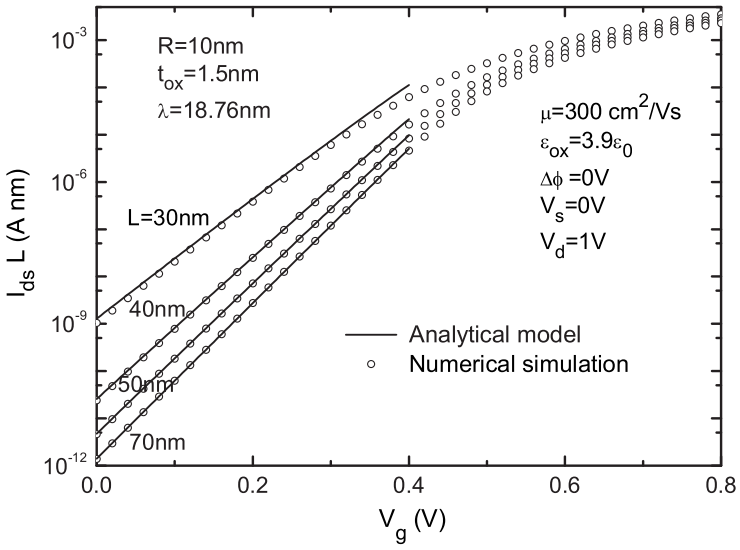


Figure IV.2: $I_{ds}L-V_g$ curves for nanowire MOSFETs obtained from the analytical model (solid lines) in comparison with the 2-D numerical simulation results (open circles).

is the inversion charge per unit gate area. Current continuity requires I_{ds} to be independent of y. Therefore, integration of (IV.24) with respect to y from 0 to L yields

$$\begin{aligned} I_{ds} &= 2\pi\mu q n_i \frac{\int_{V_s}^{V_d} e^{-qV(y)/kT} dV\,(y)}{\int_0^L \frac{dy}{\int_0^R e^{q\psi(\rho,y)/kT}\rho d\rho}} \\ &= 2\pi\mu kT n_i \frac{e^{-qV_s/kT} - e^{-qV_d/kT}}{\int_0^L \frac{dy}{\int_0^R e^{q\psi(\rho,y)/kT}\rho d\rho}} \end{aligned} \tag{IV.26}$$

The subthreshold current can be calculated analytically as a function of V_g, V_s, and V_d according to (IV.26). In Fig. IV.2, we compare the model-predicted $I_{ds}L-V_g$ curves with the numerical simulation results. Because we use cylindrical coordinates instead of Cartesian coordinates in ISE-DESSIS, the numerical simu-

lation is essentially 2-D. We observe excellent agreement in the subthreshold region where the analytical solution is valid.

Threshold voltage roll-off ΔV_t can be obtained from the parallel shift of $I_{ds}L - V_g$ curves (in log-scale) of a short-channel device with respect to the long-channel device, i.e.,

$$I_{ds}L\,(short\ channel) = I_{ds}L\,(long\ channel) \times e^{-\frac{q\Delta V_t}{kT}} \tag{IV.27}$$

The subthreshold drain current expression (IV.26) contains a double integral in the denominator that cannot be carried out analytically. To derive an explicit expression for compact model purposes, we need to simplify the 2-D potential function $\psi(\rho, y)$.

The drain current is predominantly controlled by the point of maximum electron energy barrier (minimum electrostatic potential) in the channel direction located at $(0, y_c)$, where y_c is obtained by solving $\partial\psi\,(\rho, y)\,/\partial y|_{y=y_c} = 0$,

$$y_c = \frac{L}{2} - \frac{1}{2k_1}\ln\frac{c_1 e^{k_1 L/2} - b_1 e^{-k_1 L/2}}{b_1 e^{k_1 L/2} - c_1 e^{-k_1 L/2}} \simeq \frac{L}{2} - \frac{1}{2k_1}\ln\frac{c_1}{b_1} \tag{IV.28}$$

The minimum potential can then be expressed as

$$\psi_{\min} = \psi(0, y_c) = V_g - \Delta\phi + 2\sqrt{b_1 c_1}\,e^{-\frac{k_1 L}{2}} \tag{IV.29}$$

We expand $\psi(\rho, y)$ at $(0, y_c)$ and keep to the second order term,

$$\psi(\rho, y) \simeq \psi_{\min} + k_1^2\sqrt{b_1 c_1}\,e^{-\frac{k_1 L}{2}}\,(y - y_c)^2 - \frac{1}{2}k_1^2\sqrt{b_1 c_1}\,e^{-\frac{k_1 L}{2}}\rho^2 \tag{IV.30}$$

Substituting (IV.30) into the extraction equation (IV.27), one can derive

$$\Delta V_t = \Delta V_{t1} + \Delta V_{t2} + \Delta V_{t3} \tag{IV.31}$$

where ΔV_{t1} corresponds to the minimum potential term given by

$$\Delta V_{t1} = -2\sqrt{b_1 c_1} e^{-\frac{k_1 L}{2}} \tag{IV.32}$$

and ΔV_{t2} and ΔV_{t3} correspond to the modifications induced by the second order terms in the channel direction and radial direction, respectively. They are

$$\Delta V_{t2} = \frac{kT}{q} \ln \left\{ \frac{\sqrt{\pi}}{2D_1 L} \left[\operatorname{erf}\left(D_1 \left(L - y_c\right)\right) + \operatorname{erf}\left(D_1 y_c\right) \right] \right\}$$

$$\Delta V_{t3} = \frac{kT}{q} \ln \left[\frac{D_1^2 R^2}{2\left(1 - e^{-D_1^2 R^2/2}\right)} \right] \tag{IV.33}$$

where $\operatorname{erf}(z)$ is the error function defined as

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-u^2} du \tag{IV.34}$$

and

$$D_1 = \left(\frac{q k_1^2 \sqrt{b_1 c_1} e^{-\frac{k_1 L}{2}}}{kT} \right)^{1/2} \tag{IV.35}$$

Fig. IV.3 shows the model-predicted threshold voltage roll-off ΔV_t as a function of channel length, compared with 2-D numerical simulation results. The shift is evaluated at a constant $I_{ds}L$ level corresponding to the current of the long channel device at $V_g - \Delta\phi = 0$ V, deeply into the subthreshold region. Inasmuch as the subthreshold slope changes slightly when $L > 1.3\lambda$, the choice of constant current level is non-critical. Good agreement has been achieved for channel length down to $\sim 1.3\lambda$.

As can be observed from (IV.20) and (IV.31), DIBL effect is incorporated in the threshold voltage roll-off model through the V_d-dependent parameter c_1, and

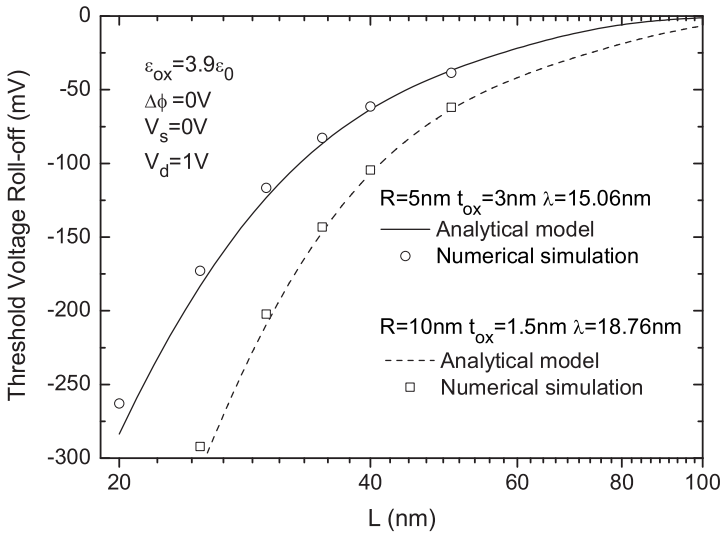


Figure IV.3: Threshold voltage roll-offs for nanowire MOSFETs as functions of channel length L obtained from the analytical model (lines) in comparison with the 2-D numerical simulation results (symbols).

can be estimated by

$$DIBL = \Delta V_t(V_{ds} = 0.05V) - \Delta V_t(V_{ds} = 1V) \tag{IV.36}$$

Fig. IV.4 indicates the validity of our DIBL model by comparison with simulation results. In Fig. IV.5, the model-predicted DIBL effects are compared with the experimental results in [67], where the physical gate lengths of the twin silicon nanowire MOSFETs are reported to be around 30 nm. The underlap regions in the extremely short devices led to an increase of the effective channel length over the physical gate length. An estimated 35 nm effective channel length gives a good match between the analytical model and experimental data.

In order to extract an explicit expression for the subthreshold slope from

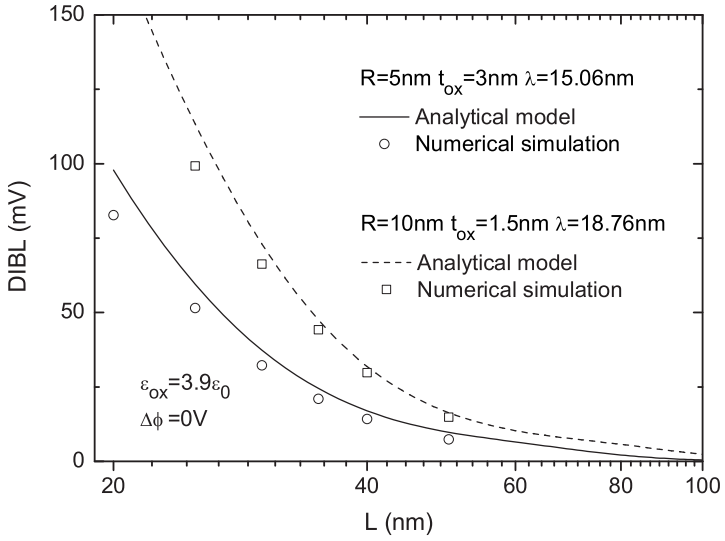


Figure IV.4: DIBL effects for nanowire MOSFETs as functions of channel length L obtained from the analytical model (lines) in comparison with the 2-D numerical simulation results (symbols).

(IV.26), the 2-D potential function is further simplified by $\psi(\rho, y) \simeq \psi_{\min}$. We can approximate the subthreshold current as $I_{ds} \propto e^{q\psi_{\min}/kT}$. Then the inverse subthreshold slope is given by

$$S = \left[\frac{\partial \left(\log_{10} I_{ds}\right)}{\partial V_g}\right]^{-1} \simeq \left(\frac{\partial \psi_{\min}}{\partial V_g}\right)^{-1} \times 60 \; mV/Dec \tag{IV.37}$$

With the definition of $\psi_{\min}$ in (IV.29), one obtains the expression for $\partial\psi_{\min}/\partial V_g$, which is smaller than 1 and thus accounts for the subthreshold slope degradation:

$$\frac{\partial \psi_{\min}}{\partial V_g} = 1 - 2E \frac{\left(V_{bi} + \Delta\phi + \frac{V_d+V_s}{2} - V_g\right) e^{-\frac{k_1 L}{2}}}{\sqrt{(V_s + V_{bi} + \Delta\phi - V_g)(V_d + V_{bi} + \Delta\phi - V_g)}} \tag{IV.38}$$

where E is given by (IV.21) ($n = 1$).

Fig. IV.6 shows that the analytical subthreshold slope for nanowire MOS-

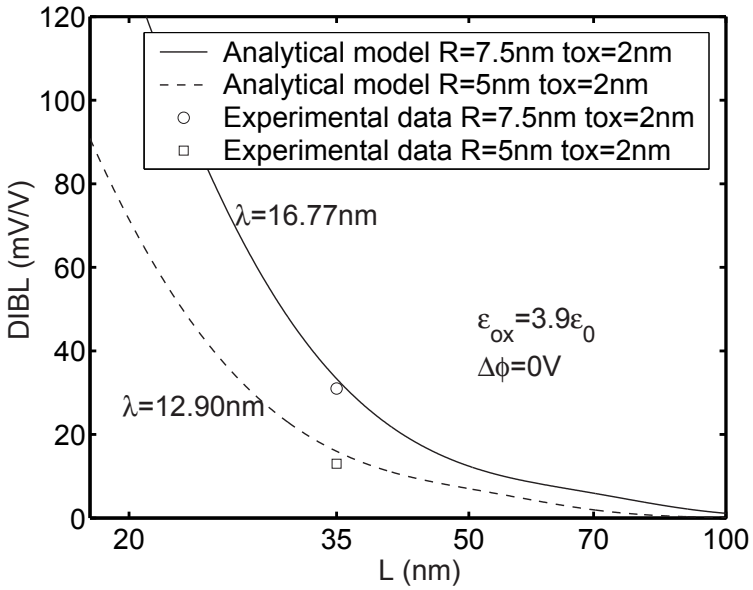


Figure IV.5: DIBL effects for nanowire MOSFETs as functions of channel length L obtained from the analytical model (lines) in comparison with the experimental data in Ref. [67] (symbols).

FETs is in good agreement with 2-D numerical simulation results for $L > 1.3\lambda$. Again, the gate voltage is chosen to satisfy $V_g - \Delta\phi = 0$ V so that the device is biased in the deep subthreshold region. Actually, the subthreshold slope is almost independent of bias condition as long as it is in the subthreshold region.

IV.2 Analytic Model for Threshold Voltage Shift Due to Quantum Effects in Nanowire MOSFETs

In addition to SCEs, quantum-mechanical (QM) effects play more and more important roles in aggressively scaled devices, especially those with nanowire-like Si body (i.e., 2-D quantum confinement). QM effects manifest themselves in two

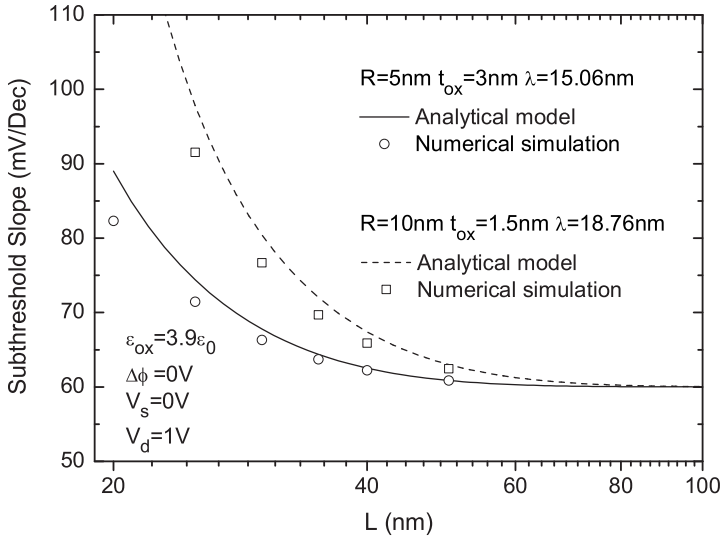


Figure IV.6: Subthreshold slopes for nanowire MOSFETs as functions of channel length L obtained from the analytical model (lines) in comparison with the 2-D numerical simulation results (symbols).

ways [43]. First, the slope of inversion charge versus gate voltage curve is degraded because the inversion charge distribution is pushed further into Si, i.e., away from the oxide interface. Second, the threshold voltage (V_t) is shifted to a higher value due to the higher quantized subband energies. Previously, the QM V_t shift has been analytically modeled for conventional bulk MOSFETs [43], DG MOSFETs [105], and also triple-gate and quadruple-gate MOSFETs [106]. We develop an analytic model for quantum confinement induced V_t shift in undoped or lightly-doped nanowire MOSFETs. To deal with the ellipsoidal constant energy surfaces of the Si conduction band, which are induced by anisotropic effective mass, we use elliptic coordinates, where Mathieu functions are invoked [107].

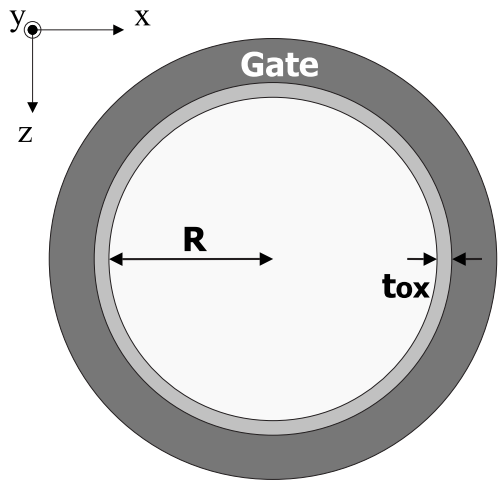


Figure IV.7: Schematic cross section diagram of a nanowire MOSFET. Y is the symmetry axis direction.

Following the general method of Stern and Howard [108] for films, Bescond *et al.* [65] have proposed the effective-mass approach for n-type nanowire MOSFETs with arbitrary orientation very recently. To model the V_t shift caused by quantum confinement in Si nanowire MOSFETs, it is reasonable to make use of Boltzmann statistics as well as parabolic approximation in the weak inversion region. Furthermore, the mobile charge term in the Poisson equation can be neglected, leading to the decoupling of Poisson's and Schrödinger equations. Therefore, in the undoped or lightly-doped body, the potential is almost constant [99], and we can assume the 2-D quantum well to be flat-bottomed with infinitely high potential barriers.

For an arbitrarily oriented undoped or lightly-doped cylindrical nanowire MOSFET, whose schematic cross section diagram is illustrated in Fig. IV.7, the

total energy under weak inversion condition is given by

$$E = E' + \frac{\hbar^2 k_y^2}{2m_d} \tag{IV.39}$$

where E' is the bottom energy of each discrete subband, determined by the eigenvalue equation

$$\left(-\frac{\hbar^2}{2m_x}\frac{\partial^2}{\partial x^2} - \frac{\hbar^2}{2m_z}\frac{\partial^2}{\partial z^2} + V(x,z)\right)\zeta(x,z) = E'\zeta(x,z) \tag{IV.40}$$

with

$$V(x,z) = \begin{cases} 0, & \sqrt{x^2+z^2} < R \\ \infty, & \sqrt{x^2+z^2} \geqslant R \end{cases} \tag{IV.41}$$

Here, m_d is the density-of-state (DOS) effective mass. According to (IV.40), we have chosen a (x,z) coordinate to diagonalize the reduced 2×2 reciprocal effective-mass tensor. m_x and m_z are the corresponding effective masses in the x- and z-directions, respectively. We choose the coordinates such that $m_x \geqslant m_z$.

For the special case of $m_x = m_z$, eigenvalues of E' can be easily obtained with Bessel functions [103]:

$$E' \triangleq E_{mn} = \frac{2\hbar^2 q_{mn}}{m_x R^2} \tag{IV.42}$$

where q_{mn} is determined by the boundary condition

$$J_m(2\sqrt{q_{mn}}) = 0 \quad for\ m = 0,1,2,... \tag{IV.43}$$

Here, $J_m(\rho)$ denotes the Bessel function of the first kind. The subscript n indicates that $2\sqrt{q_{mn}}$ is the n-th zero of $J_m(\rho)$. There is an extra two-fold degeneracy for

$m \neq 0$. [1]

For the general case of $m_x > m_z$, by making a coordinate scaling: $X = \sqrt{m_x/m_0}x$ and $Z = \sqrt{m_z/m_0}z$, where m_0 is the free electron mass, (IV.40) can be transformed to

$$\left(\frac{\partial^2}{\partial X^2} + \frac{\partial^2}{\partial Z^2} + \frac{2m_0E'}{\hbar^2}\right)\zeta(X,Z) = 0 \tag{IV.44}$$

with the boundary condition

$$\zeta(X,Z) = 0, \quad where\ \sqrt{\frac{m_0}{m_x}X^2 + \frac{m_0}{m_z}Z^2} = R. \tag{IV.45}$$

The new boundary is an ellipse instead of a circle. Therefore, to solve the eigenvalue equation (IV.40) generally, we need to use the elliptic coordinates, and deal with Mathieu functions based on separation of variables. The solution in the elliptic coordinates is given by [107]

$$\zeta(\xi,\eta) = \sum_{m=0}^{\infty} A_m C_m(\xi,q)c_m(\eta,q) + \sum_{m=1}^{\infty} B_m S_m(\xi,q)s_m(\eta,q) \tag{IV.46}$$

where the elliptic coordinates (ξ,η) is corresponding to (X,Z) by

$$X = a\cosh\xi\cos\eta \tag{IV.47}$$

$$Z = a\sinh\xi\sin\eta \tag{IV.48}$$

with a defined as the focal distance of the elliptical boundary given by

$$a = R\sqrt{\frac{m_x - m_z}{m_0}}. \tag{IV.49}$$

[1] The two-fold degeneracy is a result of symmetry. Actually, we have another group of characteristic values determined by $J_{-m}\left(2\sqrt{q'_{mn}}\right) = 0$, where $m = 1, 2, ...$ Based on the circular symmetry, it is not surprising that $q'_{mn} = q_{mn}$, which finally leads to the two-fold degeneracy.

Note that we choose $m_x > m_z$. In equation (IV.46), A_m and B_m are coefficients; $c_m(\xi, q)$ and $s_m(\xi, q)$ are the Mathieu functions with order m of cosine and sine types, respectively; $C_m(\xi, q)$ and $S_m(\xi, q)$ are the modified Mathieu functions with order m of cosine and sine types, respectively.[2] Boundary condition yields

$$C_m\left(\xi_b, q_{mn}^{(c)}\right) = 0 \quad for\ m = 0, 1, 2, ... \tag{IV.50}$$

or

$$S_m\left(\xi_b, q_{mn}^{(s)}\right) = 0 \quad for\ m = 1, 2, 3, ... \tag{IV.51}$$

Here, ξ_b is defined by $\xi_b = \cosh^{-1}\sqrt{m_x/\left(m_x - m_z\right)}$. The subscript n indicates that ξ_b is the n-th zero of $C_m(\xi, q_{mn}^{(c)})$ and $S_m(\xi, q_{mn}^{(s)})$. Energy is related with $q_{mn}^{(l)}$ ($l = c$ or s) through the following equation

$$E' \triangleq E_{mn}^{(l)} = \frac{2\hbar^2 q_{mn}^{(l)}}{\left(m_x - m_z\right) R^2} \tag{IV.52}$$

Within the framework of quantum mechanics, the total mobile charge density per unit gate length N^{QM} is given by

$$N^{QM} = N^{1D} \exp\left(\frac{E_f - E_s^{QM}}{k_B T}\right) \tag{IV.53}$$

with

$$N^{1D} = \sum_k N_k^{1D} \tag{IV.54}$$

and

$$N_k^{1D} = \begin{cases} \sum_m \sum_n g\sqrt{\frac{2m_d k_B T}{\pi\hbar^2}} \exp\left(-\frac{E_{mn}^{(c)}}{k_B T}\right) \\ + \sum_m \sum_n g\sqrt{\frac{2m_d k_B T}{\pi\hbar^2}} \exp\left(-\frac{E_{mn}^{(s)}}{k_B T}\right), \ m_x > m_z \\ \sum_m \sum_n gD\sqrt{\frac{2m_d k_B T}{\pi\hbar^2}} \exp\left(-\frac{E_{mn}}{k_B T}\right), \ m_x = m_z \end{cases} \tag{IV.55}$$

[2]The Mathieu and the modified Mathieu functions with order m of cosine (sine) type may be denoted in some references as $ce_m(\xi, q)$ ($se_m(\xi, q)$) and $Ce_m(\xi, q)$ ($Se_m(\xi, q)$), respectively.

Here, E_s^{QM} is the quantum-mechanically calculated conduction band edge. The subscript k denotes valley index. g and m_d are the degeneracy and the DOS effective mass, respectively. $E_{mn}^{(l)}$ and E_{mn} are the subband energies for the cases of $m_x > m_z$ and $m_x = m_z$, which can be solved through Mathieu and Bessel functions, respectively. The coefficient D is responsible for the degeneracy which comes from symmetry: $D = 1$ for $m = 0$; $D = 2$ otherwise. Without QM effects, the classical theory leads to

$$N^{CL} = \pi R^2 N^{3D} \exp\left(\frac{E_f - E_s^{CL}}{k_B T}\right) \tag{IV.56}$$

where E_s^{CL} is the classical counterpart of E_s^{QM}, and N^{3D} is the 3-D effective DOS, i.e., $N^{3D} = N_c$ for the conduction band. Equating N^{QM} in (IV.53) to N^{CL} in (IV.56) yields an analytic model for the shift of the QM potential compared to the classical one:

$$\Delta\psi_s^{QM} = \frac{E_s^{QM} - E_s^{CL}}{-q} = \frac{k_B T}{q} \ln\left(\frac{\pi R^2 N^{3D}}{N^{1D}}\right). \tag{IV.57}$$

Knowing $\Delta\psi_s^{QM}$, one can easily calculate the V_t shift due to the QM effects as [43]

$$\Delta V_t^{QM} = \frac{S}{60} \Delta\psi_s^{QM} \tag{IV.58}$$

where S is the inverse subthreshold slope in the unit of mV/decade. An analytic model for S has been given in the last section. When SCEs are not severe, S is close to the ideal 60 mV/decade.

Based on above derivation, quantum confinement induced threshold voltage shift can be analytically obtained for nanowire MOSFETs with certain radius

and known effective mass. It has been shown that the effective-mass approach is even applicable to very small size Si nanowires (e.g., $R = 1 - 2$ nm), but in these extreme cases the effective mass varies as radius changes [109, 110]. To calculate corresponding effective masses for different valleys is a complicated research topic by itself, and here we will not put emphasis on it. For simplicity, we will employ effective masses for bulk silicon and concentrate on the subject of this section, which emphasizes the compact model for V_t shift induced by quantum confinement. In the following calculations, we focus on n-type nanowire MOSFETs with three typical orientations: $y = (100), (110), (111)$. The corresponding effective masses and valley degeneracies are listed in Table IV.1, and the corresponding characteristic values ($q_{mn}^{(l)}$ and q_{mn}) are listed in Table IV.2, IV.3, and IV.4 in ascending order. The characteristic values can be easily obtained with the built-in Bessel, Mathieu, and related functions of MATHEMATICA. For each case, we have listed enough number of characteristic values to achieve high accuracy.

In Fig. IV.8, we have shown the model-predicted $\Delta\psi_s^{QM}$ versus R for nanowire nMOSFETs. When R is quite small, the lowest energy subband plays a dominant role, and the carrier population in higher energy subband is negligible. Because of larger lowest subband energy, $\Delta\psi_s^{QM}$ for orientation (111) is expected to be larger than those for orientations (100) and (110), as shown in Fig. IV.8. However, it is not a quite significant difference. $\Delta\psi_s^{QM}$ for (100) and (110) are very close to each other for small R due to the same lowest subband energies, and the subtle difference is accounted for by different degeneracies. For larger R,

Table IV.1: Valley notation, valley degeneracy, and confinement (m_x, m_z) and DOS (m_d) effective masses for electrons in nanowire MOSFETs with different orientations [65]. The effective masses are given in units of the free electron mass m_0.

Orientation	(100)		(110)		(111)
Valley, k	1	2	1	2	1
Degeneracy, g	4	2	2	4	6
m_x	0.916	0.190	0.916	0.315	0.403
m_z	0.190	0.190	0.190	0.190	0.190
m_d	0.190	0.916	0.190	0.553	0.432

the carrier population in higher energy subband becomes significant. Generally, the heavier DOS effective mass leads to carriers occupying relatively lower energy subbands, and thus tends to counteract relatively higher subband energies which are always associated with heavier DOS effective mass. Therefore, we can observe in Fig. IV.8 that $\Delta\psi_s^{QM}$ for all the three orientations have similar behaviors as R becomes larger than $\sim$ 3 nm. The importance of carrier population in higher energy subband for $R >\sim$ 3 nm is further shown in Fig. IV.9. The first ten subbands in all valleys are significant for nanowire MOSFETs, whereas for DG MOSFETs, the first three or four subbands in all valleys are enough until $t_{si} \sim$ 20 nm [105]. This result can be understood from the fact that the subband energy E' in DG

Table IV.2: Characteristic values for the cases of (100) and (110) $k = 1$ in Table IV.1.

	(lmn)	$q_{mn}^{(l)}$	(lmn)	$q_{mn}^{(l)}$	(lmn)	$q_{mn}^{(l)}$
	$(c01)$	3.278	$(s41)$	23.812	$(c81)$	45.515
	$(c11)$	5.517	$(c12)$	28.417	$(s22)$	47.050
	$(c21)$	8.569	$(c61)$	29.585	$(c42)$	47.474
(100)	$(s11)$	11.021	$(s51)$	29.595	$(s81)$	51.811
and	$(c31)$	12.482	$(c22)$	34.056	$(s32)$	54.058
(110)	$(s21)$	14.559	$(s61)$	36.180	$(c91)$	54.835
$k = 1$	$(c41)$	17.282	$(c71)$	37.097	$(c52)$	55.328
	$(s31)$	18.810	$(c32)$	40.391	$(s91)$	60.874
	$(c51)$	22.981	$(s12)$	40.713	$(s42)$	61.764
	$(c02)$	23.508	$(s71)$	43.582	$(c03)$	62.535

Table IV.3: Characteristic values for the cases of (110) $k = 2$ and (111) $k = 1$ in Table IV.1.

	(lmn)	$q_{mn}^{(l)}$	(lmn)	$q_{mn}^{(l)}$	(lmn)	$q_{mn}^{(l)}$
	$(c01)$	0.761	$(c31)$	4.797	$(c22)$	8.415
	$(c11)$	1.691	$(s31)$	5.114	$(c51)$	9.365
(110)	$(s11)$	2.170	$(c12)$	6.211	$(s51)$	9.476
$k = 2$	$(c21)$	3.042	$(c41)$	6.919	$(s22)$	9.721
	$(s21)$	3.461	$(s41)$	7.124	$(c03)$	11.083
	$(c02)$	4.433	$(s12)$	7.500	$(c32)$	11.381
	$(c01)$	1.188	$(c02)$	7.341	$(c22)$	12.845
	$(c11)$	2.468	$(s31)$	7.703	$(c51)$	13.212
(111)	$(s11)$	3.552	$(c41)$	9.708	$(s51)$	13.847
$k = 1$	$(c21)$	4.316	$(c12)$	9.836	$(s22)$	15.661
	$(s21)$	5.381	$(s41)$	10.526	$(c32)$	16.405
	$(c31)$	6.734	$(s12)$	12.512	$(c61)$	17.206

Table IV.4: Characteristic values for the case of (100) $k = 2$ in Table IV.1, which corresponds to Bessel function.

	(mn)	q_{mn}	(mn)	q_{mn}	(mn)	q_{mn}
	(01)	1.446	(31)	10.176	(03)	18.723
(100)	(11)	3.671	(12)	12.306	(51)	19.233
$k = 2$	(21)	6.595	(41)	14.394	(32)	23.819
	(02)	7.618	(22)	17.712	(61)	24.681

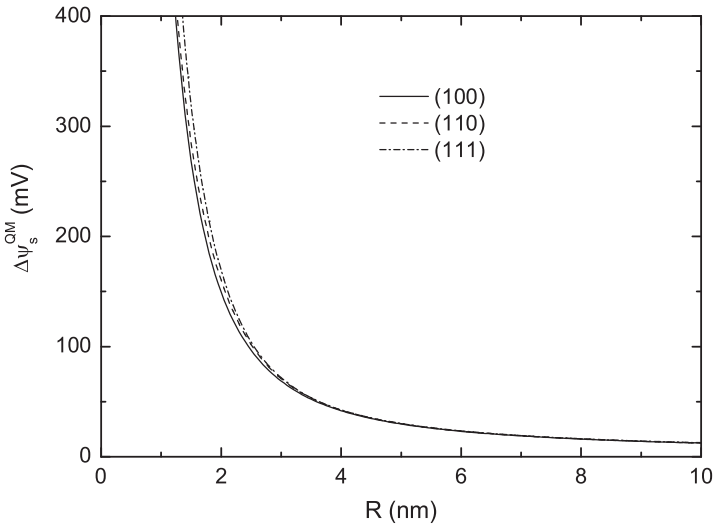


Figure IV.8: Model-predicted $\Delta\psi_s^{QM}$ for nanowire nMOSFETs as functions of R for orientations (100), (110), and (111).

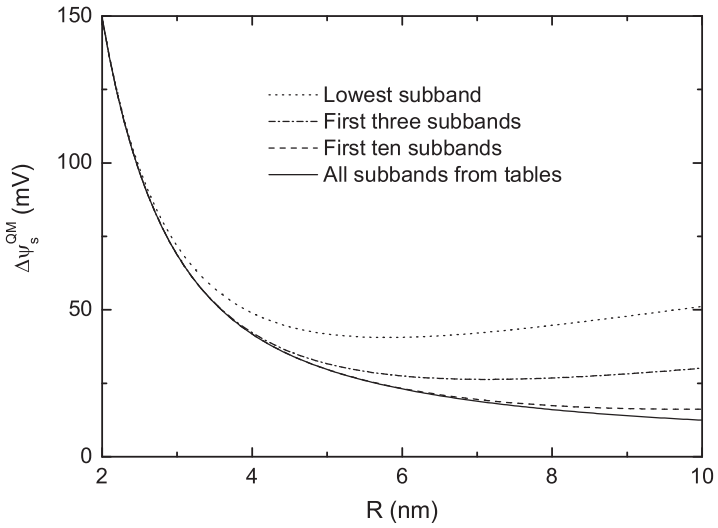


Figure IV.9: $\Delta\psi_s^{QM}$ for nanowire nMOSFETs as functions of R for orientation (100) with different numbers of subbands accounted for: 1) Lowest subband in both valleys; 2) First 3 subbands in both valleys; 3) First 10 subbands in both valleys; 4) All subbands obtained from Table IV.2, IV.3, and IV.4 (corresponding to the so-called model-predicted $\Delta\psi_s^{QM}$).

MOSFETs is proportional to j^2, whereas in nanowire MOSFETs we roughly have $E' \propto j$, where j is the ordinal number for each subband.

Although Mathieu functions provide accurate eigenvalues, to simplify the model, we may substitute reduced isotropic effective mass for anisotropic effective mass. The approximation leads to an isotropic effective mass problem which has solution of Bessel function type. Reduced isotropic effective mass m_r is defined as follows

$$m_r \triangleq \frac{2}{\left(\frac{1}{m_x} + \frac{1}{m_z}\right)} \tag{IV.59}$$

The first 20 subband energies for the valley labeled $k = 1$ of orientation (100)

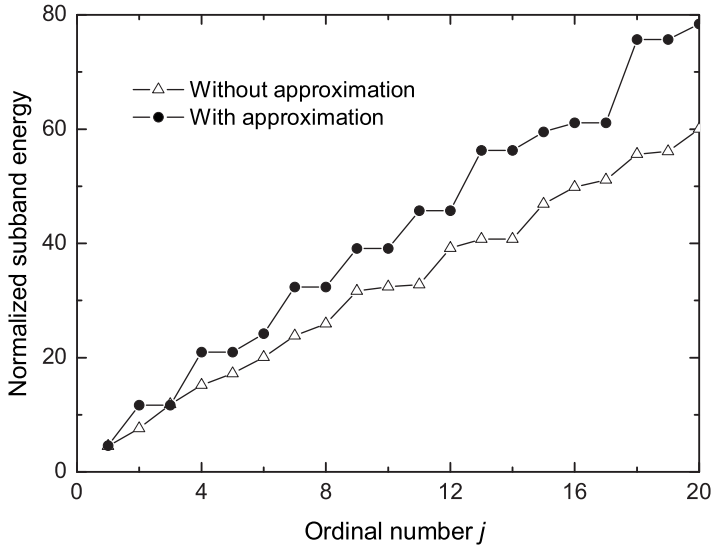


Figure IV.10: Subband energies E' normalized to $2\hbar^2/m_0R^2$ as functions of ordinal number j for the valley labeled $k = 1$ of orientation (100). Degeneracy for $m \neq 0$ is included where the approximation of reduced isotropic effective mass is assumed.

calculated using reduced effective mass are compared with those obtained without approximation in Fig. IV.10, with ordinal number j for each subband as the horizontal ordinate. First of all, Fig. IV.10 validates that subband energy E' in nanowire MOSFETs roughly has linear relationship with ordinal number j. The simplified model yields almost the same lowest subband energy as accurate model. However, the subband energies obtained by the simplified model are obviously larger than accurate values except the first and third ones. Based on the observations in Fig. IV.10, the results of $\Delta\psi_s^{QM}$ versus R in Fig. IV.11 can be expected. For small R, $\Delta\psi_s^{QM}$ is dominated by the lowest subband energy, so the two curves generated with and without the approximation almost overlap. As R increases, the

approximate $\Delta\psi_s^{QM}$ becomes larger than the accurate one due to the overestimation of most subband energies. Actually, we found that the approximation-induced relative error of the lowest subband energy remains quite small over a wide range of the ratio m_x/m_z, as shown in Fig. IV.12. Therefore, with given anisotropic effective mass, V_t shift can be obtained conveniently using the reduced isotropic effective mass for small radius, where only the lowest subband population is important. As mentioned previously, effective mass is dependent on the radius for very small size nanowire MOSFETs. Instead of calculating characteristic value $q_{01}^{(c)}$ of Mathieu function for every different value of m_x/m_z, one can make use of the approximation to obtain V_t shift directly, with characteristic value q_{01} of Bessel function given in Table. IV.4. However, this simplified model with reduced isotropic effective mass cannot completely substitute for the original model due to the limitation of small radius.

This section presents an analytic model for V_t shift due to quantum confinement in nanowire MOSFETs with anisotropic effective mass, which results in the involvement of Mathieu functions. With this compact model, it has been shown that V_t shifts for (100) and (110) oriented nMOSFETs are slightly less serious than that for (111) one when the radius R is small. The model also indicates that several subbands are necessary to obtain accurate V_t shift for some intermediate range of radius. By introducing the reduced isotropic effective mass, a simplified version of the model is proposed and demonstrated to be valid for small radius.

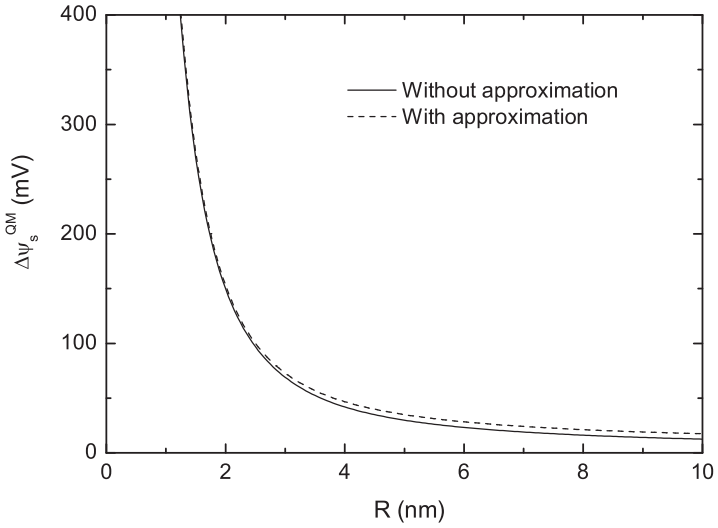


Figure IV.11: $\Delta\psi_s^{QM}$ for nanowire nMOSFETs as functions of R for orientation (100) with effective mass of the valley labled $k = 1$ being: 1) anisotropic effective mass (solid line); 2) reduced isotropic effective mass (dashed line).

IV.3 Scaling Limit of Nanowire MOSFETs

An earlier work (Wang *et al.*) [111] compared the electrostatic integrity of a planar DG and a nanowire silicon MOSFET with respect to their respective quantum confinement effects. They identified a size limit based on the sensitivity of the ground-state energy and therefore of the threshold voltage to the silicon thickness or radius. A first-order scale length parameter was employed without a quantitative criterion on the acceptable short-channel threshold roll-off.

In this section, we discuss the key factors in the scaling of nanowire MOSFETs to 10 nm gate lengths and below. By applying a tolerance criterion for threshold variation with channel length, a general guideline is established between

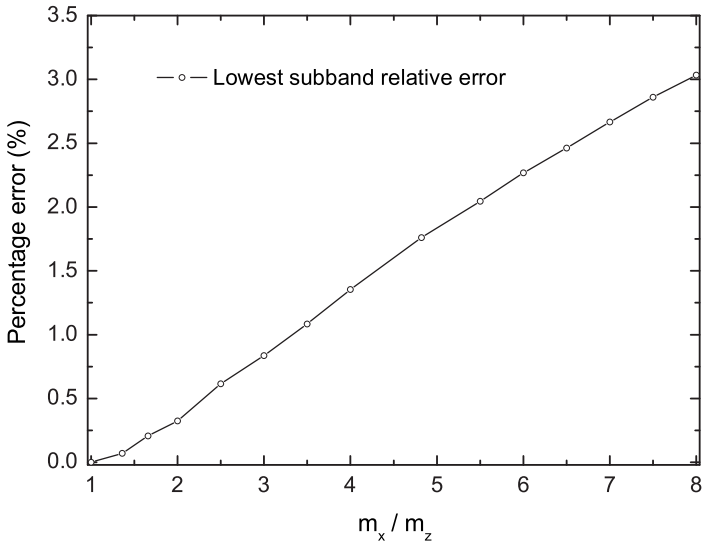


Figure IV.12: Percentage error of the lowest subband energy caused by the approximation of reduced isotropic effective mass, as a function of ratio m_x/m_z.

the gate length and the nanowire size, based on the analytic model of short-channel threshold voltage roll-off for given nanowire dimensions in Section IV.1. In the last section, solving the 2-D Schrodinger's equation in subthreshold for anisotropic effective masses yields discrete energy levels for electrons. The ground-state energy and therefore the threshold voltage is sensitive to the radius of the nanowire below 5-10 nm. Combination with the short-channel scaling guideline allows an estimate of the scaling limits of nanowire MOSFETs as a function of the electron effective mass in the semiconductor.

Fig. IV.13 shows the model-predicted threshold voltage roll-off ΔV_t^{SCE}, according to (IV.31), as a function of normalized channel length L/λ, compared

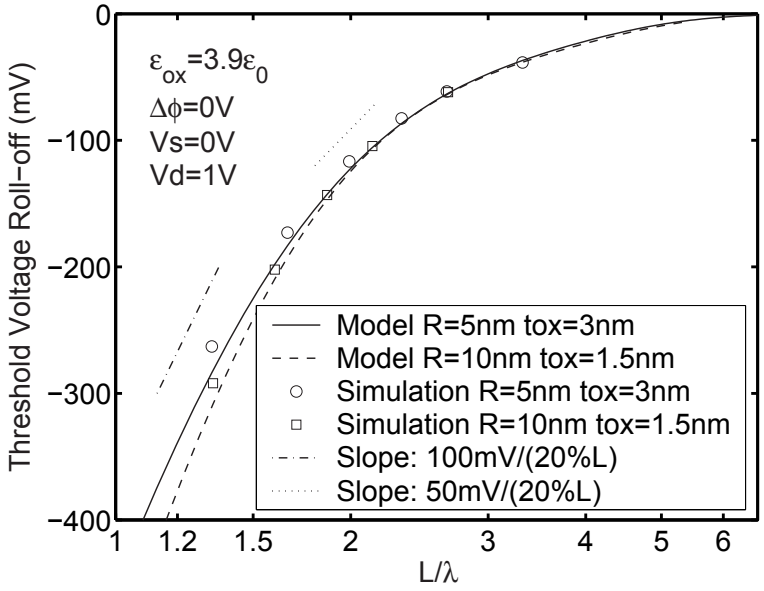


Figure IV.13: Threshold voltage roll-offs for nanowire MOSFETs as functions of normalized channel length L/λ obtained from the analytical model (lines) in comparison with the 2-D numerical simulation results (symbols).

with numerical simulation results. Assuming $\pm 10\%$ variation of channel length L, we obtain a general guideline of $L_{min} \sim 1.3\lambda$ for a tolerance criterion of 100 mV threshold voltage variation, indicated by the slope (chained line) in Fig. IV.13. If we further assume $R = 3t_{ox}$ (SiO_2) for silicon nanowire MOSFETs near the scaling limit, a simple relation between the generalized scale length λ and the nanowire radius R can be extracted from equation (IV.8) as $\lambda = 2.4R$. This leads to a general scaling guideline for nanowire transistors below 10 nm channel lengths: $L_{min} \sim 3R$.

As shown in the last section, $\Delta\psi_s^{QM}$ for all the three orientations of silicon nanowire nMOSFETs are very sensitive to R when $R <\sim 5$ nm. For small R,

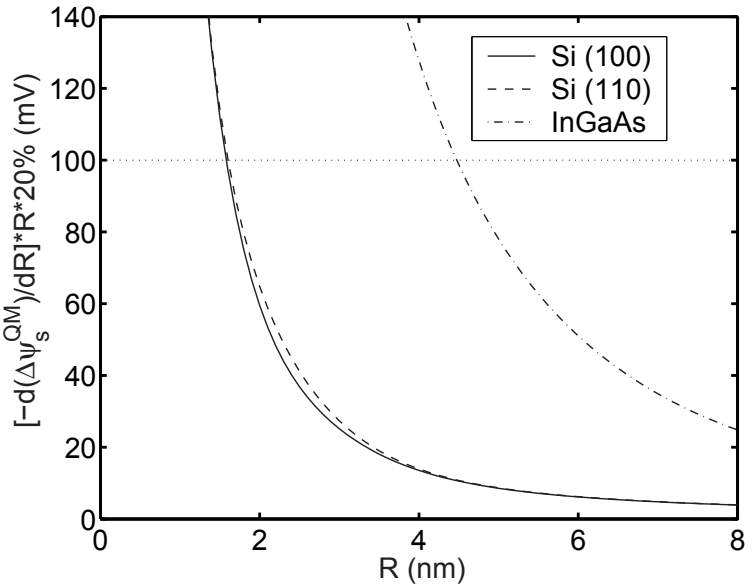


Figure IV.14: Model-predicted total variation of $\Delta\psi_s^{QM}$, $[-d(\Delta\psi_s^{QM})/dR] \times (20\% \cdot R)$, for nanowire nMOSFETs as functions of R, based on an assumption of $\pm 10\%$ variation of R. Nanowires with (100) and (110) orientations for silicon and InGaAs are compared.

$\Delta\psi_s^{QM}$ for orientation (111) is larger than those for orientations (100) and (110), due to the larger lowest subband energy. Therefore, (100) and (110) orientations are better choices than (111) in terms of device scaling limit, and from now on, we only focus on (100) and (110) orientations. The sensitivity can be expressed as the magnitude of variation in $\Delta\psi_s^{QM}$ per variation in R. Because of technological process tolerance on R, this sensitivity is the key factor that will limit the scalability of nanowire MOSFETs. The limit hinges on how precise the nanowire radius can be controlled. If we assume a typical $\pm 10\%$ variation of R statistically, the total variation of $\Delta\psi_s^{QM}$ is

$$\delta\left(\Delta\psi_s^{QM}\right) = -\frac{d\left(\Delta\psi_s^{QM}\right)}{dR} \times (20\% \cdot R) \tag{IV.60}$$

Here, we assume uniform or average variation of R, i.e., local variations are averaged out. In Fig. IV.14, $\delta\Delta\psi_s^{QM}$ is plotted as a function of R. For a criterion of $\delta\Delta\psi_s^{QM} \leqslant 100$ mV, the minimum tolerable R for (100) and (110) silicon nanowire nMOSFETs are both 1.6 nm. Following the above-mentioned general guideline of $L_{min} \sim 3R$ from consideration of SCEs, we obtain the minimum channel length for both (100) and (110) orientations: 4.8 nm. The curve for InGaAs nanowire nMOSFETs is also shown in Fig. IV.14 for comparison. The conduction band of $In_{0.53}Ga_{0.47}As$ has an isotropic effective mass of $0.04m_0$ without degeneracy, and $\Delta\psi_s^{QM}$ for InGaAs nanowire nMOSFETs can also be calculated by equation (IV.57) with only Bessel functions. Due to the small effective mass, the minimum tolerable R for InGaAs nanowire nMOSFETs is 4.5 nm for the same 100 mV criterion. The SCEs of nanowire transistors is independent of effective mass. If we assume a tunneling-limited gate dielectric of $t_{ox} = 1.5$ nm with $\epsilon_{ox} = \epsilon_{InGaAs} = 14\epsilon_0$, the scaling limit for InGaAs nanowire MOSFETs is estimated to be $L_{min} \sim 1.3\pi(R + t_{ox})/\alpha_1 \sim 10$ nm.

For comparison, we have also estimated the scaling limit of DG MOSFETs following the similar criteria. With the 100 mV criterion for $\delta\Delta\psi_s^{QM}$, the minimum tolerable t_{si} for (100) and (110) confinement directions are 1.3 nm and 2.1 nm, respectively [105]. Assuming a scaling limit of $t_{ox} = 0.5$ nm (SiO_2) as with the nanowire before, we obtain the scaling limits for DG MOSFETs in (100) and (110) confinement directions: $L_{min} \sim 1.3\lambda \sim 4.5$ nm and 5.8 nm, respectively.

IV.4 Summary

In this chapter, the compact models for SCEs in nanowire MOSFETs have been presented, including threshold voltage roll-off, DIBL, and subthreshold slope degradation. Besides, quantum confinement effects on V_t shift in nanowire MOSFETs with anisotropic effective mass are also modeled. Based on models for both SCEs and quantum effects, scaling limit is projected for extremely scaled nanowire MOSFETs, in comparison with that of DG MOSFETs.

The text of Chapter IV, in part, is a reprint of the material as it appears in "An analytic model for threshold voltage shift due to quantum confinement in surrounding gate MOSFETs with anisotropic effective mass" by Yu Yuan, Bo Yu, Jooyoung Song, and Yuan Taur, Solid State Electronics, Feb 2009. The dissertation author was the primary investigator and author of this paper.

The text of Chapter IV, in part, is a reprint of the material as it appears in "A two-dimensional analytical solution for short-channel effects in nanowire MOSFETs" by Bo Yu, Yu Yuan, Jooyoung Song, and Yuan Taur, IEEE Transaction on Electron Devices, Oct 2009. The dissertation author was a co-author of this paper.

The text of Chapter IV, in part, is a reprint of the material as it appears in "Scaling of nanowire transistors" by Bo Yu, Lingquan Wang, Yu Yuan, Peter Asbeck, and Yuan Taur, IEEE Transaction on Electron Devices, Nov 2008. The dissertation author was a co-author of this paper.

Chapter V

Physical Understanding of Inversion Layer Capacitance beyond Bulk Silicon MOSFETs

In Chapter II, we put emphasis on the impact of oxide and interface traps on C-V characteristics. While in this chapter, we will focus on the inversion layer capacitance of semiconductor itself and take out the effects of oxide and interface traps in the analysis for the purpose of simplicity. The motivation is to obtain a clear physical understanding of the inversion layer capacitance C_i beyond the conventional bulk or PDSOI silicon-based MOSFETs. The analysis is performed on III-V MOSFETs as well as two typical 3-D transistors, namely symmetric DG MOSFETs and nanowire MOSFETs. The results for III-V MOSFETs are jointly developed by Bo Yu, Dennis Wang, Prof. Peter Asbeck, Prof. Yuan Taur and

myself. The details can be found in [82], and we are not going to repeat it. Instead, we will present our new results of comprehensive equivalent capacitance circuit model for symmetric DG MOSFETs and nanowire MOSFETs.

V.1 Inversion Layer Capacitance in Symmetric DG MOSFETs

In this section, we will extend our comprehensive equivalent capacitance circuit model to symmetric DG MOSFETs, especially DG MOSFETs based on III-V material. Consider an undoped (or lightly doped) symmetric DG MOSFET schematically shown in Fig. V.1. First of all, a DG MOS capacitance can be broken up into an equivalent circuit shown in Fig. V.2. Here, gate insulator capacitance

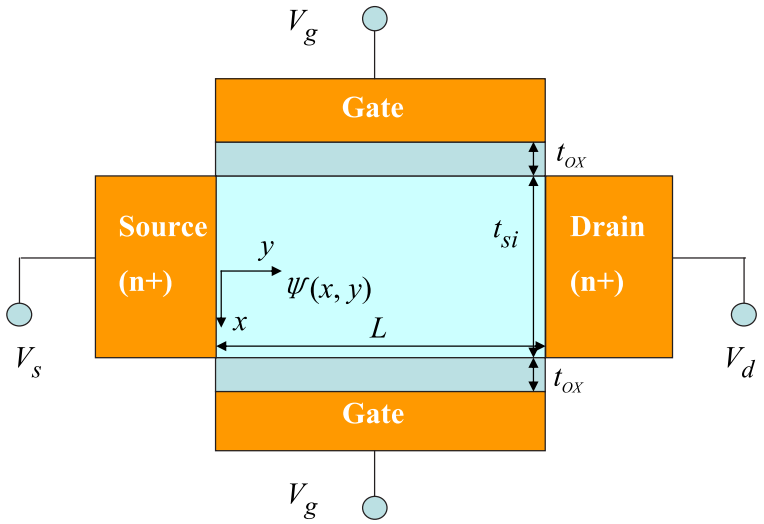


Figure V.1: Schematic diagram of a symmetric DG MOSFET. Symmetry acix is at $x = 0$.

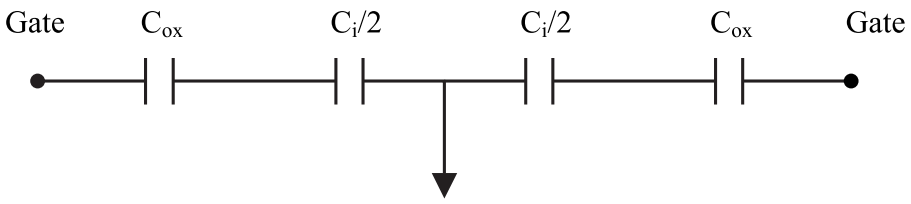


Figure V.2: Equivalent circuit model of a symmetric DG MOS capacitor at low frequency.

C_{ox} is given by $C_{ox} = \epsilon_{ox}/t_{ox}$, and inversion layer capacitance C_i is defined as

$$C_i = \frac{d\,(-Q_i)}{d\psi_s} \tag{V.1}$$

where Q_i is the total inversion charge per unit area, including all the contributions across the semiconductor thin film. The symmetric nature of the equivalent circuit in Fig. V.2 enables us to simplify it by a straightforward folding. The simplified equivalent circuit is illustrated in Fig. V.3.

Similarly, according to the quantum-mechanical theory, one is able to expand Q_i into $\sum_n Q_{i,n}$, where $Q_{i,n}$ is the inversion charge associated with the n-th subband of the quantum well formed by both band offset (structural confinement)

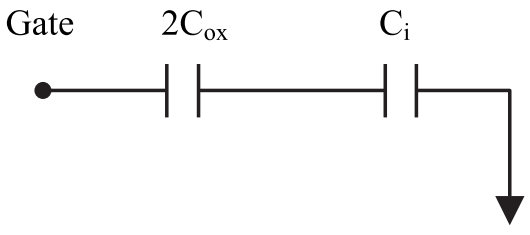


Figure V.3: Simplified equivalent circuit model of a symmetric DG MOS capacitor at low frequency.

and transverse electrical field (electrical confinement). If we define

$$C_{i,n} = \frac{d\left(-Q_{i,n}\right)}{d\psi_s} \tag{V.2}$$

it is straightforward to get

$$C_i = \sum_n C_{i,n} \tag{V.3}$$

Since the surface potential ψ_s is determined by the bending of conduction band edge at the semiconductor/insulator interfaces, it is obvious that

$$C_{i,n} = \frac{d\left(-Q_{i,n}\right)}{d\psi_s} = \frac{d\left(-Q_{i,n}\right)}{d\left(\frac{E_f - E_c(x=\pm t_{si}/2)}{q}\right)} \tag{V.4}$$

where E_f is the Fermi level in the semiconductor, and $E_c(x = \pm t_{si}/2)$ is the surface conduction band edge. We can further decompose $C_{i,n}$ like this:

$$C_{i,n} = \left\{ \left[\frac{d\left(-Q_{i,n}\right)}{d\left(\frac{E_f - E_n}{q}\right)} \right]^{-1} + \left[\frac{d\left(-Q_{i,n}\right)}{d\left(\frac{E_n - E_c(x=\pm t_{si}/2)}{q}\right)} \right]^{-1} \right\}^{-1} \tag{V.5}$$

where E_n is the eigen-energy of the n-th subband. We define

$$C_{DOS,n} = \frac{d\left(-Q_{i,n}\right)}{d\left(\frac{E_f - E_n}{q}\right)} \tag{V.6}$$

since it is the capacitance associated with DOS for the n-th subband. The other term is much more complicated, which contains the information of the solutions of Schrödinger's equation and Poisson's equation. To get its exact value, the two master equations must be solved self-consistently. But anyway it is the capacitance associated with quantum well confinement effect for the n-th subband, so we define

$$C_{qw,n} = \frac{d\left(-Q_{i,n}\right)}{d\left(\frac{E_n - E_c(x=\pm t_{si}/2)}{q}\right)} \tag{V.7}$$

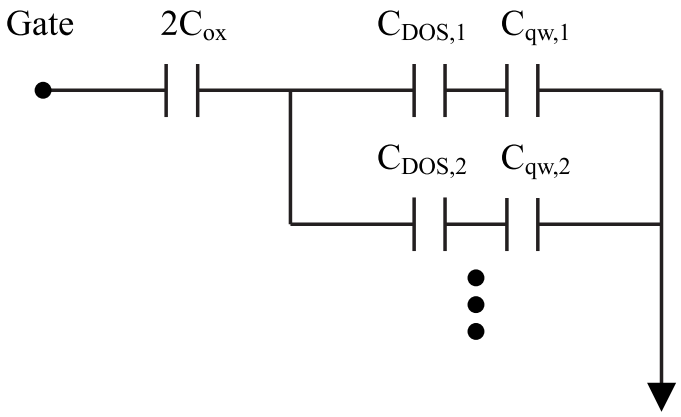


Figure V.4: Comprehensive equivalent circuit model of a symmetric DG MOS capacitor at low frequency.

Finally, equation (V.3) is explicitly written as

$$C_i = \sum_n \frac{1}{\frac{1}{C_{DOS,n}} + \frac{1}{C_{qw,n}}} \tag{V.8}$$

Combining equation (V.8) and Fig. V.3, a more comprehensive equivalent circuit model of a symmetric DG MOS capacitor can be obtained, as illustrated in Fig. V.4.

As we mentioned, $C_{DOS,n}$ is the capacitance associated with DOS for the n-th subband. Here, we will show in details why it is the DOS capacitance. To unveil the physical meaning and derive the analytical expression, we need to make a couple of necessary approximations: 1) No electrons penetrate into the insulator, i.e., the insulator has infinite barrier; 2) the whole device has parabolic dispersion relation. Under these approximations, it is well known based on the quantum-mechanical theory that

$$Q_{i,n} = -\frac{qm^*_{d,n}}{\pi\hbar^2} kT \ln\left[1 + \exp\left(\frac{E_f - E_n}{kT}\right)\right] \tag{V.9}$$

where $m^*_{d,n}$ is the DOS effective mass for the n-th subband. For materials with subband degeneracy (e.g., Si), degeneracy g is incorporated into $m^*_{d,n}$.

Substituting equation (V.9) into equation (V.6) yields

$$C_{DOS,n} = q^2 \frac{m^*_{d,n}}{\pi\hbar^2} \frac{1}{1+\exp\left(\frac{E_n - E_f}{kT}\right)} \tag{V.10}$$

The last term $1/[1+\exp((E_n - E_f)/kT)]$ is in the form of Fermi-Dirac distribution function, which approaches unity when E_f is more than a few kT/q above E_n. We know that $m^*_{d,n}/\pi\hbar^2$ is usually recognized as the 2-D DOS, therefore it is clear now why (V.6) is called the DOS capacitance.

$C_{DOS,n}$ has very clear physical meaning, but not $C_{qw,n}$. The complexity of the self-consistent solutions of Schrödinger's equation and Poisson's equation has been fully lumped into $C_{qw,n}$. The coupling between populations in different subbands is reflected in $C_{qw,n}$, and this coupling prevents us from developing a compact analytical result for $C_{qw,n}$. It is reasonable to make the assumption that only the lowest subband is occupied in ultra scaled DG MOSFETs. Under this approximation, the comprehensive equivalent circuit model of a symmetric DG MOS capacitor shown in Fig. V.4 is reduced to a simplified one, as illustrated in Fig. V.5.

According to the one subband approximation, one can obtain

$$C_{qw,1} = \frac{d\left(-Q_{i,1}\right)/d\psi_s}{d\left(\frac{E_1 - E_c(x=\pm t_{si}/2)}{q}\right)/d\psi_s} = \frac{C_i}{d\left(\frac{E_1 - E_c(x=\pm t_{si}/2)}{q}\right)/d\psi_s} \tag{V.11}$$

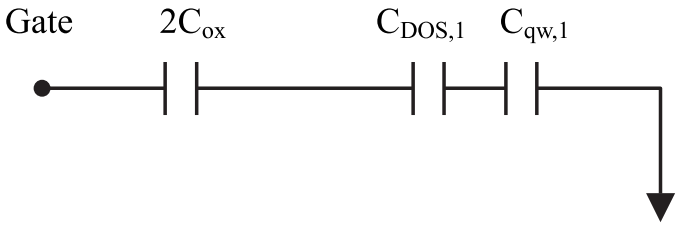


Figure V.5: A simplified equivalent circuit model of a symmetric DG MOS capacitor under the one subband approximation.

Following the quantum-mechanical theory, we get

$$\begin{aligned}
\frac{d\left(\frac{E_1-E_c(x=\pm t_{si}/2)}{q}\right)}{d\psi_s} &= \frac{d\left(\langle\phi_1|\,\frac{H-E_c(x=\pm t_{si}/2)}{q}\,|\phi_1\rangle\right)}{d\psi_s} \\
&= \frac{d\,\langle\phi_1|}{d\psi_s}\left(\frac{H-E_c(x=\pm t_{si}/2)}{q}\right)|\phi_1\rangle \\
&\quad + \langle\phi_1|\,\frac{d\left(\frac{H-E_c(x=\pm t_{si}/2)}{q}\right)}{d\psi_s}\,|\phi_1\rangle \\
&\quad + \langle\phi_1|\left(\frac{H-E_c(x=\pm t_{si}/2)}{q}\right)\frac{d\,|\phi_1\rangle}{d\psi_s} \\
&= \langle\phi_1|\,\frac{d\left(\frac{H-E_c(x=\pm t_{si}/2)}{q}\right)}{d\psi_s}\,|\phi_1\rangle
\end{aligned} \tag{V.12}$$

where H represents the Hamiltonian of the quantum well, and $|\phi_1\rangle$ is the eigen wavefunction corresponding to the lowest subband. It can be approved that

$$\frac{d\,\langle\phi_1|}{d\psi_s}\left(\frac{H-E_c(x=\pm t_{si}/2)}{q}\right)|\phi_1\rangle = \langle\phi_1|\left(\frac{H-E_c(x=\pm t_{si}/2)}{q}\right)\frac{d\,|\phi_1\rangle}{d\psi_s} = 0 \tag{V.13}$$

Since the kinetic operator in Hamiltonian H is independent of ψ_s and the band-offsets between different materials are invariables, it is straightforward to have

$$\langle\phi_1|\,\frac{d\left(\frac{H-E_c(x=\pm t_{si}/2)}{q}\right)}{d\psi_s}\,|\phi_1\rangle = \langle\phi_1|\,\frac{d\,(\psi_s-\psi(x))}{d\psi_s}\,|\phi_1\rangle \tag{V.14}$$

The 1-D Poisson's equation for DG MOS capacitor is given by

$$\frac{d^2\psi}{dx^2} = \frac{q}{\epsilon_s} n(x), -\frac{t_{si}}{2} < x < \frac{t_{si}}{2} \tag{V.15}$$

Integrating it twice, we obtain

$$\psi(x=0) - \psi(x) = \frac{q}{\epsilon_s} \left[\int_0^x x^{'} n(x^{'}) dx^{'} - x \int_0^x n(x^{'}) dx^{'} \right] \tag{V.16}$$

which yields

$$\begin{aligned} \psi_s - \psi(x) &= \frac{q}{\epsilon_s} \left[\int_0^x x^{'} n(x^{'}) dx^{'} - x \int_0^x n(x^{'}) dx^{'} \right] \\ &\quad - \frac{q}{\epsilon_s} \left[\int_0^{\frac{t_{si}}{2}} x^{'} n(x^{'}) dx^{'} - \frac{t_{si}}{2} \int_0^{\frac{t_{si}}{2}} n(x^{'}) dx^{'} \right] \end{aligned} \tag{V.17}$$

Substituting (V.17) into (V.14), we will have four terms, which are calculated separately as follows:

$$\begin{aligned} \langle \phi_1 | \frac{d \int_0^x x^{'} n(x^{'}) dx^{'}}{d\psi_s} | \phi_1 \rangle &= \int_{-\frac{t_{si}}{2}}^{\frac{t_{si}}{2}} |\phi_1(x)|^2 \left(\int_0^x x^{'} \frac{dn(x^{'})}{d\psi_s} dx^{'} \right) dx \\ &= \int_{-\frac{t_{si}}{2}}^{\frac{t_{si}}{2}} \frac{n(x)}{\int_{-\frac{t_{si}}{2}}^{\frac{t_{si}}{2}} n(x) dx} \left(\int_0^x x^{'} \frac{dn(x^{'})}{d\psi_s} dx^{'} \right) dx \\ &= -\int_{-\frac{t_{si}}{2}}^{\frac{t_{si}}{2}} x \frac{dn(x)}{d\psi_s} \left(\int_{-\frac{t_{si}}{2}}^{x} \frac{n(x^{'})}{\int_{-\frac{t_{si}}{2}}^{\frac{t_{si}}{2}} n(x) dx} dx^{'} \right) dx \\ &\quad + \int_0^{\frac{t_{si}}{2}} x \frac{dn(x)}{d\psi_s} dx \end{aligned} \tag{V.18}$$

The second step is based on the one subband assumption and infinite insulator

barrier approximation. The third step is simply integration by parts.

$$\begin{aligned}\langle\phi_1|\frac{d\left[x\int_0^x n(x')dx'\right]}{d\psi_s}|\phi_1\rangle &= \int_{-\frac{t_{si}}{2}}^{\frac{t_{si}}{2}}\frac{xn(x)}{\int_{-\frac{t_{si}}{2}}^{\frac{t_{si}}{2}}n(x)dx}\left(\int_0^x\frac{dn(x')}{d\psi_s}dx'\right)dx \\ &= -\int_{-\frac{t_{si}}{2}}^{\frac{t_{si}}{2}}\frac{dn(x)}{d\psi_s}\left(\int_{-\frac{t_{si}}{2}}^{x}\frac{x'n(x')}{\int_{-\frac{t_{si}}{2}}^{\frac{t_{si}}{2}}n(x)dx}dx'\right)dx\end{aligned} \tag{V.19}$$

Here the second step has utilized the symmetry of $n(x)$ after integration by parts.

$$\langle\phi_1|\frac{d\int_0^{\frac{t_{si}}{2}}x'n(x')dx'}{d\psi_s}|\phi_1\rangle = \int_0^{\frac{t_{si}}{2}}x\frac{dn(x)}{d\psi_s}dx \tag{V.20}$$

$$\langle\phi_1|\frac{d\left[\frac{t_{si}}{2}\int_0^{\frac{t_{si}}{2}}n(x')dx'\right]}{d\psi_s}|\phi_1\rangle = \frac{t_{si}}{2}\int_0^{\frac{t_{si}}{2}}\frac{dn(x)}{d\psi_s}dx \tag{V.21}$$

Combining equations (V.18)-(V.21), one obtain

$$\begin{aligned}\frac{d\left(\frac{E_1-E_c(x=\pm t_{si}/2)}{q}\right)}{d\psi_s} &= \frac{q}{\varepsilon_s}\left(\frac{t_{si}}{2}-x_c-x_{sh}\right)\int_0^{\frac{t_{si}}{2}}\frac{dn(x)}{d\psi_s}dx \\ &= \frac{\frac{t_{si}}{2}-x_c-x_{sh}}{\varepsilon_s}\frac{C_i}{2}\end{aligned} \tag{V.22}$$

where x_c and x_{sh} are defined as follows:

$$x_c = \frac{\int_0^{\frac{t_{si}}{2}}x\cdot n(x)dx}{\int_0^{\frac{t_{si}}{2}}n(x)dx} \tag{V.23}$$

$$x_{sh} = \frac{\int_0^{\frac{t_{si}}{2}}\frac{dn(x)}{d\psi_s}\left(\int_0^x\frac{\left(x-x'\right)n(x')}{\int_0^{\frac{t_{si}}{2}}n(x)dx}dx'\right)dx}{\int_0^{\frac{t_{si}}{2}}\frac{dn(x)}{d\psi_s}dx} \tag{V.24}$$

x_c is the centroid of the inversion charge, but the reference point is the axis of symmetry instead of semiconductor/insulator interface. Therefore, it is physically sound to have $t_{si}/2-x_c$ term in (V.22). x_{sh} is very complicated, but we know that

it is like the square root of a second order central moment. In other words, x_{sh} is related to the spreading of the inversion charge. To show this point manifestly, we can assume Gaussian distribution approximately, i.e., assume

$$\begin{aligned}
\frac{n(x)}{\int_0^{\frac{t_{si}}{2}} n(x)dx} &\simeq \frac{1}{\sqrt{2\pi}\sigma_c}\exp\left[-\frac{(x-x_c)^2}{2\sigma_c^2}\right],\ 0<x<\frac{t_{si}}{2} \\
\frac{\frac{dn(x)}{d\psi_s}}{\int_0^{\frac{t_{si}}{2}} \frac{dn(x)}{d\psi_s}dx} &\simeq \frac{1}{\sqrt{2\pi}\sigma_{av}}\exp\left[-\frac{(x-x_{av})^2}{2\sigma_{av}^2}\right],\ 0<x<\frac{t_{si}}{2}
\end{aligned} \tag{V.25}$$

where x_{av} is the centroid of the differential inversion charge with respect to the axis of symmetry and defined as

$$x_{av} = \frac{\int_0^{\frac{t_{si}}{2}} x\cdot\frac{dn(x)}{d\psi_s}dx}{\int_0^{\frac{t_{si}}{2}} \frac{dn(x)}{d\psi_s}dx} \tag{V.26}$$

Then we can calculated x_{sh} as

$$\begin{aligned}
x_{sh} &= \frac{\int_0^{\frac{t_{si}}{2}} \frac{dn(x)}{d\psi_s}\left(\int_0^x \frac{\left(x-x'\right)n(x')}{\int_0^{\frac{t_{si}}{2}} n(x)dx}dx'\right)dx}{\int_0^{\frac{t_{si}}{2}} \frac{dn(x)}{d\psi_s}dx} \\
&\simeq \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma_{av}}\exp\left[-\frac{(x-x_{av})^2}{2\sigma_{av}^2}\right]\int_{-\infty}^{x}\frac{(x-y)}{\sqrt{2\pi}\sigma_c}\exp\left[-\frac{(y-x_c)^2}{2\sigma_c^2}\right]dydx \\
&= \sqrt{\frac{\sigma_c^2+\sigma_{av}^2}{2\pi}} - \frac{x_c-x_{av}}{2} + O\left(\delta^2\right)
\end{aligned} \tag{V.27}$$

x_c is very close to x_{av}, and σ_c is very close to σ_{av}. So $(x_c - x_{av}), (\sigma_c - \sigma_{av}) \sim O(\delta)$. In the last step of (V.27), we only keep to the first order of δ. Now, it is clear that the primary part of x_{sh} is $\sqrt{(\sigma_c^2+\sigma_{av}^2)/2\pi}$, which reinforces our previous understanding of x_{sh} as a second order moment. The difference between x_c and x_{av} will also make a small part of contribution.

According to (V.27), it is physically meaningful to reformulate (V.22) as

follow:

$$\frac{d\left(\frac{E_1-E_c(x=\pm t_{si}/2)}{q}\right)}{d\psi_s} = \frac{\frac{t_{si}}{2} - x_{cen} - x_{spr}}{\varepsilon_s}\frac{C_i}{2} \tag{V.28}$$

where

$$x_{cen} = \frac{x_c + x_{av}}{2} \tag{V.29}$$

$$\begin{aligned} x_{spr} &= x_{sh} + \frac{x_c - x_{av}}{2} \\ &= \frac{\int_0^{\frac{t_{si}}{2}} \frac{dn(x)}{d\psi_s}\left[\int_0^x \left(x-x'\right)n(x')dx' - \int_x^{\frac{t_{si}}{2}}\left(x-x'\right)n(x')dx'\right]dx}{2\cdot\int_0^{\frac{t_{si}}{2}}\frac{dn(x)}{d\psi_s}dx\cdot\int_0^{\frac{t_{si}}{2}}n(x)dx} \\ &\simeq \sqrt{\frac{\sigma_c^2+\sigma_{av}^2}{2\pi}} \end{aligned} \tag{V.30}$$

It is not difficult to discover that both x_{cen} and x_{spr} will not change if $n(x)$ and $dn(x)/d\psi_s$ are switched.

Substituting equation (V.28) into (V.11) yields

$$C_{qw,1} = \frac{2\epsilon_s}{\frac{t_{si}}{2} - x_{cen} - x_{spr}} \tag{V.31}$$

Because the capacitances are connected in series, it is more convenient to use the concept of effective thickness, which is defined by the reciprocal of capacitance normalized to the dielectric constant ϵ_s. For example, we define

$$C_{DOS,1} = q^2\frac{m_{d,1}^*}{\pi\hbar^2}\frac{1}{1+\exp\left(\frac{E_1-E_f}{kT}\right)} \triangleq \frac{\epsilon_s}{x_{DOS}} \tag{V.32}$$

According to the one subband approximation, we have

$$C_i = \frac{1}{\frac{1}{C_{DOS,1}} + \frac{1}{C_{qw,1}}} \tag{V.33}$$

Combining equations (V.31), (V.32), (V.33), one finally obtain

$$C_i = \frac{\epsilon_s}{x_{DOS} + \frac{t_{si}}{4} - \frac{x_{cen}}{2} - \frac{x_{spr}}{2}} \triangleq \frac{\epsilon_s}{x_i} \tag{V.34}$$

i.e., x_i is the effective thickness associated with the inversion layer capacitance C_i and given by

$$x_i = x_{DOS} + \frac{t_{si}}{4} - \frac{x_{cen}}{2} - \frac{x_{spr}}{2} \tag{V.35}$$

V.2 Inversion Layer Capacitance in Nanowire MOSFETs

In this section, we will extend our comprehensive equivalent capacitance circuit model to nanowire MOSFETs, especially nanowire MOSFETs based on III-V material. Consider an undoped (or lightly doped) nanowire MOSFET schematically shown in Fig. V.6. First of all, a nanowire MOS capacitance can be broken

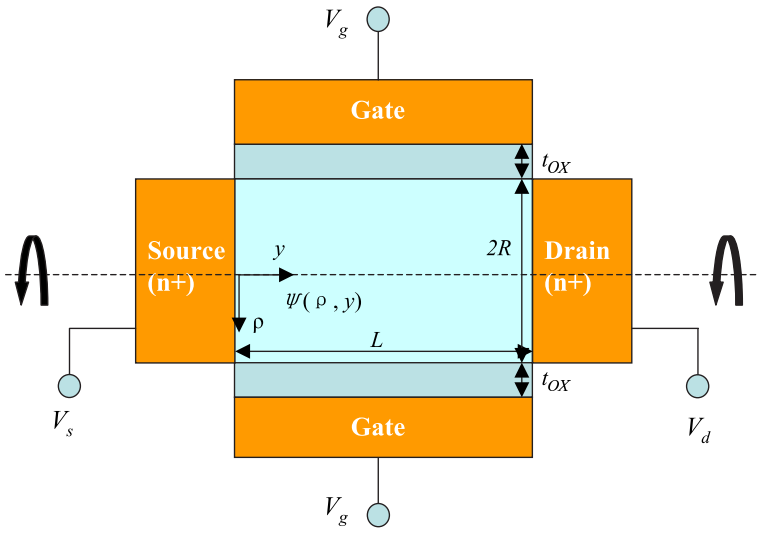


Figure V.6: Schematic diagram of a nanowire MOSFET (cut through symmetry axis).

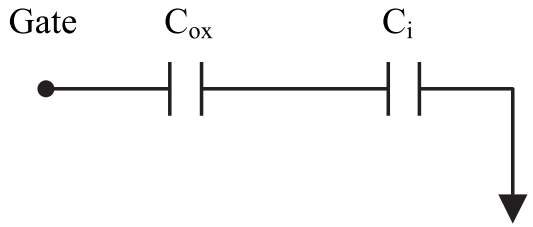


Figure V.7: Equivalent circuit model of a nanowire MOS capacitor at low frequency.

up into an equivalent circuit shown in Fig. V.7. Here, gate insulator capacitance C_{ox} is given by $C_{ox} = \epsilon_{ox}/[R\ln(1 + t_{ox}/R)]$, and inversion layer capacitance C_i is defined as

$$C_i = \frac{d\left(\frac{-Q_m}{2\pi R}\right)}{d\psi_s} \tag{V.36}$$

where Q_m is the total inversion charge per unit length along the channel direction. Here, both C_{ox} and C_i have been normalized by $2\pi R$.

Similarly, according to the quantum-mechanical theory, one is able to expand Q_i into $\sum_n Q_{i,n}$, where $Q_{i,n}$ is the inversion charge associated with the n-th subband of the quantum well formed by both band offset (structural confinement) and transverse electrical field (electrical confinement). Since nanowire is 2-D confinement system, its eigenstates are supposed to have 2 subscript indice. But we can sort and assign the ordinal number n to the eigenstates in energy space. If we define

$$C_{i,n} = \frac{d\left(-Q_{i,n}\right)}{d\psi_s} \tag{V.37}$$

it is straightforward to get

$$C_i = \sum_n C_{i,n} \tag{V.38}$$

Since the surface potential ψ_s is determined by the bending of conduction band edge at the semiconductor/insulator interfaces, it is obvious that

$$C_{i,n} = \frac{d\left(-Q_{i,n}\right)}{d\psi_s} = \frac{d\left(-Q_{i,n}\right)}{d\left(\frac{E_f - E_c(\rho=R)}{q}\right)} \tag{V.39}$$

where E_f is the Fermi level in the semiconductor, and $E_c(\rho = R)$ is the surface conduction band edge. We can further decompose $C_{i,n}$ like this:

$$C_{i,n} = \left\{ \left[\frac{d\left(-Q_{i,n}\right)}{d\left(\frac{E_f - E_n}{q}\right)} \right]^{-1} + \left[\frac{d\left(-Q_{i,n}\right)}{d\left(\frac{E_n - E_c(\rho=R)}{q}\right)} \right]^{-1} \right\}^{-1} \tag{V.40}$$

where E_n is the eigen-energy of the n-th subband. We define

$$C_{DOS,n} = \frac{d\left(-Q_{i,n}\right)}{d\left(\frac{E_f - E_n}{q}\right)} \tag{V.41}$$

since it is the capacitance associated with DOS for the n-th subband. The other term is much more complicated, which contains the information of the solutions of Schrödinger's equation and Poisson's equation. To get its exact value, the two master equations must be solved self-consistently. But anyway it is the capacitance associated with quantum well confinement effect for the n-th subband, so we define

$$C_{qw,n} = \frac{d\left(-Q_{i,n}\right)}{d\left(\frac{E_n - E_c(\rho=R)}{q}\right)} \tag{V.42}$$

Finally, equation (V.38) is explicitly written as

$$C_i = \sum_n \frac{1}{\frac{1}{C_{DOS,n}} + \frac{1}{C_{qw,n}}} \tag{V.43}$$

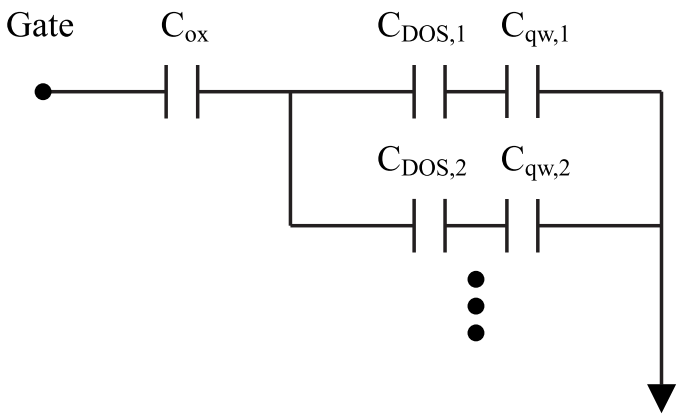


Figure V.8: Comprehensive equivalent circuit model of a nanowire MOS capacitor at low frequency.

Combining equation (V.43) and Fig. V.7, a more comprehensive equivalent circuit model of nanowire MOS capacitor can be obtained, as illustrated in Fig. V.8.

$C_{DOS,n}$ has very clear physical meaning, but not $C_{qw,n}$. The complexity of the self-consistent solutions of Schrödinger's equation and Poisson's equation has been fully lumped into $C_{qw,n}$. The coupling between populations in different subbands is reflected in $C_{qw,n}$, and this coupling prevents us from developing a compact analytical result for $C_{qw,n}$. It is reasonable to make the assumption that only the lowest subband is occupied in ultra scaled nanowire MOSFETs. Under this approximation, the comprehensive equivalent circuit model of a nanowire MOS capacitor shown in Fig. V.8 is reduced to a simplified one, as illustrated in Fig. V.9.

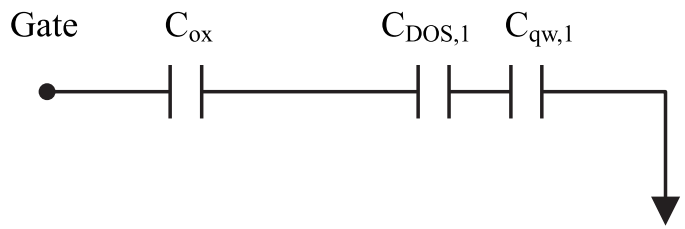


Figure V.9: A simplified equivalent circuit model of a nanowire MOS capacitor under the one subband approximation.

According to the one subband approximation, one can obtain

$$C_{qw,1} = \frac{d\left(-Q_{i,1}\right)/d\psi_s}{d\left(\frac{E_1 - E_c(\rho=R)}{q}\right)/d\psi_s} = \frac{C_i}{d\left(\frac{E_1 - E_c(\rho=R)}{q}\right)/d\psi_s} \tag{V.44}$$

Following the quantum-mechanical theory, we get

$$\begin{aligned}
\frac{d\left(\frac{E_1 - E_c(\rho=R)}{q}\right)}{d\psi_s} &= \frac{d\left(\langle\phi_1| \frac{H - E_c(\rho=R)}{q} |\phi_1\rangle\right)}{d\psi_s} \\
&= \frac{d\langle\phi_1|}{d\psi_s}\left(\frac{H - E_c(\rho=R)}{q}\right)|\phi_1\rangle \\
&\quad + \langle\phi_1| \frac{d\left(\frac{H - E_c(\rho=R)}{q}\right)}{d\psi_s} |\phi_1\rangle \\
&\quad + \langle\phi_1|\left(\frac{H - E_c(\rho=R)}{q}\right)\frac{d|\phi_1\rangle}{d\psi_s} \\
&= \langle\phi_1| \frac{d\left(\frac{H - E_c(\rho=R)}{q}\right)}{d\psi_s} |\phi_1\rangle
\end{aligned} \tag{V.45}$$

where H represents the Hamiltonian of the quantum well, and $|\phi_1\rangle$ is the eigen wavefunction corresponding to the lowest subband. It can be approved that

$$\frac{d\langle\phi_1|}{d\psi_s}\left(\frac{H - E_c(\rho=R)}{q}\right)|\phi_1\rangle = \langle\phi_1|\left(\frac{H - E_c(\rho=R)}{q}\right)\frac{d|\phi_1\rangle}{d\psi_s} = 0 \tag{V.46}$$

Since the kinetic operator in Hamiltonian H is independent of ψ_s and the band-offsets between different materials are invariables, it is straightforward to

have

$$\langle\phi_1|\frac{d\left(\frac{H-E_c(\rho=R)}{q}\right)}{d\psi_s}|\phi_1\rangle = \langle\phi_1|\frac{d\left(\psi_s-\psi(\rho)\right)}{d\psi_s}|\phi_1\rangle \tag{V.47}$$

The 1-D Poisson's equation for nanowire MOS capacitor is given by

$$\frac{d^2\psi}{d^2\rho}+\frac{1}{\rho}\frac{d\psi}{d\rho}=\frac{q}{\epsilon_s}n(\rho), 0<\rho<R \tag{V.48}$$

Integrating it twice, we obtain

$$\psi(\rho=0)-\psi(\rho)=\frac{q}{\epsilon_s}\left[\int_0^{\rho}\ln(\rho')\rho' n(\rho')d\rho' - \ln(\rho)\int_0^{\rho}\rho' n(\rho')d\rho'\right] \tag{V.49}$$

which yields

$$\begin{aligned}\psi_s-\psi(\rho) &= \frac{q}{\epsilon_s}\left[\int_0^{\rho}\ln(\rho')\rho' n(\rho')d\rho' - \ln(\rho)\int_0^{\rho}\rho' n(\rho')d\rho'\right] \\ &\quad -\frac{q}{\epsilon_s}\left[\int_0^{R}\ln(\rho')\rho' n(\rho')d\rho' - \ln(R)\int_0^{R}\rho' n(\rho')d\rho'\right]\end{aligned} \tag{V.50}$$

Substituting (V.50) into (V.47), we will have four terms, which are calculated separately as follows:

$$\begin{aligned}\langle\phi_1|\frac{d\int_0^{\rho}\ln(\rho')\rho' n(\rho')d\rho'}{d\psi_s}|\phi_1\rangle &= \int_0^{R}\frac{n(\rho)\rho}{\int_0^{R}n(\rho)\rho d\rho}\left(\int_0^{\rho}\ln(\rho')\rho'\frac{dn(\rho')}{d\psi_s}d\rho'\right)d\rho \\ &= -\int_0^{R}\ln(\rho)\rho\frac{dn(\rho)}{d\psi_s}\left(\int_0^{\rho}\frac{n(\rho')\rho'}{\int_0^{R}n(\rho)\rho d\rho}d\rho'\right)d\rho \\ &\quad +\int_0^{R}\ln(\rho)\rho\frac{dn(\rho)}{d\psi_s}d\rho\end{aligned} \tag{V.51}$$

The second step is based on the one subband assumption, infinite insulator barrier approximation, and circular symmetry approximation. The third step is simply

integration by parts.

$$\begin{aligned}\langle\phi_1|\frac{d\left[\ln(\rho)\int_0^\rho \rho^{'}n(\rho^{'})d\rho^{'}\right]}{d\psi_s}|\phi_1\rangle &= \int_0^R \frac{\ln(\rho)n(\rho)\rho}{\int_0^R n(\rho)\rho d\rho}\left(\int_0^\rho \rho^{'}\frac{dn(\rho^{'})}{d\psi_s}d\rho^{'}\right)d\rho \\ &= -\int_0^R \rho\frac{dn(\rho)}{d\psi_s}\left(\int_0^\rho \frac{\ln(\rho^{'})n(\rho^{'})\rho^{'}}{\int_0^R n(\rho)\rho d\rho}d\rho^{'}\right)d\rho \\ &\quad +\int_0^R \rho\frac{dn(\rho)}{d\psi_s}d\rho\frac{\int_0^R \ln(\rho)n(\rho)\rho d\rho}{\int_0^R n(\rho)\rho d\rho} \qquad \text{(V.52)}\end{aligned}$$

$$\langle\phi_1|\frac{d\int_0^R \ln(\rho')\rho^{'}n(\rho^{'})d\rho^{'}}{d\psi_s}|\phi_1\rangle = \int_0^R \ln(\rho)\rho\frac{dn(\rho)}{d\psi_s}d\rho \qquad \text{(V.53)}$$

$$\langle\phi_1|\frac{d\left[\ln(R)\int_0^R \rho^{'}n(\rho^{'})d\rho^{'}\right]}{d\psi_s}|\phi_1\rangle = \ln(R)\int_0^R \rho\frac{dn(\rho)}{d\psi_s}d\rho \qquad \text{(V.54)}$$

Combining equations (V.51)-(V.54), one obtain

$$\begin{aligned}\frac{d\left(\frac{E_1-E_c(\rho=R)}{q}\right)}{d\psi_s} &= \frac{q}{\varepsilon_s}\left(Z_c - Z_{sh}\right)\int_0^R \rho\frac{dn(\rho)}{d\psi_s}d\rho \\ &= \frac{Z_c - Z_{sh}}{\varepsilon_s}RC_i \qquad \text{(V.55)}\end{aligned}$$

where Z_c and Z_{sh} are defined as follows:

$$Z_c = \frac{\int_0^R \ln(\frac{R}{\rho})\cdot n(\rho)\rho d\rho}{\int_0^R n(\rho)\rho d\rho} \qquad \text{(V.56)}$$

$$Z_{sh} = \frac{\int_0^R \frac{dn(x)}{d\psi_s}\left(\int_0^\rho \frac{\ln(\frac{\rho}{\rho^{'}})n(\rho^{'})\rho^{'}}{\int_0^R n(\rho)\rho d\rho}d\rho^{'}\right)d\rho}{\int_0^R \rho\frac{dn(\rho)}{d\psi_s}d\rho} \qquad \text{(V.57)}$$

The concept of centroid of the inversion charge in the polar coordinate system is very different from that in Cartesian coordinate. Re^{-Z_c} is the centroid of the inversion charge with the axis of symmetry as the reference point. Z_{sh} is very complicated, but we know that it is like the square root of a second order central moment. In other words, Z_{sh} is related to the spreading of the inversion charge.

Substituting equation (V.55) into (V.44) yields

$$C_{qw,1} = \frac{\epsilon_s}{R\,(Z_c - Z_{sh})} \tag{V.58}$$

Because the capacitances are connected in series, it is more convenient to use the concept of effective thickness, which is defined by the reciprocal of capacitance normalized to the dielectric constant ϵ_s. According to the one subband approximation, we have

$$C_i = \frac{1}{\frac{1}{C_{DOS,1}} + \frac{1}{C_{qw,1}}} \tag{V.59}$$

One finally obtain

$$C_i = \frac{\epsilon_s}{x_{DOS} + R(Z_c - Z_{sh})} \triangleq \frac{\epsilon_s}{x_i} \tag{V.60}$$

i.e., x_i is the effective thickness associated with the inversion layer capacitance C_i and given by

$$x_i = x_{DOS} + R(Z_c - Z_{sh}) \tag{V.61}$$

V.3 Summary

In conclusion, based on both Poisson's equation and Schrödinger's equation, we have developed equivalent capacitance circuit models for both DG and nanowire MOS capacitors by adopting the one subband approximation. The analytical results in terms of effective thicknesses can help understand the physical meanings of inversion layer capacitance in DG and nanowire MOS capacitors, especially those with high mobility materials. Because high mobility is always associated with low

DOS, the major contribution to inversion layer capacitance of high mobility materials is from DOS capacitance rather than the capacitance component due to quantum confinement effect.

Chapter VI

Conclusion

In this dissertation, design, modeling and characterization of non-classical MOSFETs are present, with focus on MOS capacitors and MOSFETs based on III-V materials as well as nanowire MOSFETs.

Unlike Si MOS, III-V MOS devices lack native oxide with good interface quality. Defects or traps at dielectric semiconductor interface and inside the gate insulator are major issue for high-κ III-V MOS devices. Multi-frequency small signal C-V and G-V data are utilized to characterize and model trap effects. Our analysis focuses on InGaAs substrate device. Due to small electron effective mass and hence small conduction band DOS, the Fermi level enters conduction band when N-type MOS capacitors are biased in accumulation. According to conventional interface states theory, interface states have too small time constant to produce the observed frequency dispersion from 1 kHz to 1 MHz in C-V and G-V

when the Fermi level is close to or beyond the conduction band edge. On the other hand, trap states inside the gate insulator, called bulk-oxide traps or border traps, do have long time constants as they interact with the conduction band electrons via tunneling [94]. A distributed bulk-oxide trap model based on tunneling between the semiconductor surface and trap states in the gate insulator is developed. It differs fundamentally from the conventional interface state model in that there is a wide frequency spectrum of bulk-oxide trap response at a given gate bias. It is more physical than previously published lumped circuit models in the literature. The model is validated with the $Pt/Al_2O_3/n\text{-}In_{0.53}Ga_{0.47}As$ dispersion data in strong accumulation and near the flatband. Unlike surface states which are in units of areal density, bulk-oxide traps are characterized by a volume density, extracted from fitting of the capacitance and conductance data. It is further shown that the commonly employed method of extracting the interface state density from C-V stretchout could yield unphysical numbers inconsistent with the high-frequency dispersion data. On the other hand, the bulk-oxide trap model, in particular, with nonuniform trap density in the oxide film, can explain C-V stretchout independent of the dispersion at higher frequencies. On the other hand, the C-V humps on inversion side can be successfully modeled by the conventional interface states theory. In combination with TCAD simulation, interface states density D_{it} vs. energy are extracted. There time constants vs. energy also shows reasonable trend. It is also found that the capacitor enters strong inversion regime at very negative bias although measured capacitance seems dropping down. The true inversion

happened but the response cannot be observed because the frequency is not low enough or the temperature is not high enough.

In III-V MOSFETs device design part, the baseline sub-22 nm MOSFETs design can be scaled to 11 nm gate length by shrink the channel thickness to 3 nm and can be barely scaled to 11 nm gate length by adding 10 nm undoped underlap region at the drain side; the series resistance induced by increasing the sidewall thickness to 30 nm in a $L_g = 100$ nm MOSFET is acceptable. In III-V device characterization, the mobile charge density in high-κ InGaAs MOSFET is characterized by frequency dependent measurement of gate to channel capacitance to eliminate trap response. The electron effective mobility is extracted to be in the range of 2000 to 5000 $cm^2/V{\cdot}S$.

Analytical results of potential and subthreshold current are derived for short-channel nanowire MOSFETs based on generalized scale length theory. 2-D Poisson's equation is solved in both the semiconductor and insulator regions. The compact model for SCEs are derived and validated by numerical simulation. V_t shift due to 2-D quantum confinement in nanowire MOSFETs with anisotropic effective mass is modeled by solving Schrödinger's equation in elliptical coordinates. The scaling limit of nanowire MOSFETs is projected from V_t shift sensitivity and scale length theory.

By invoking both Poisson's and Schrödinger's equations, as well as one subband approximation, analytical results for inversion layer capacitance in symmetric

DG and nanowire capacitors are developed. The derived results can help understand the physical meanings of DOS capacitance and quantum well capacitance, especially in high mobility materials.

Bibliography

[1] G. Moore, "Progress in Digital Integrated Electronics," *IEDM Tech. Dig.*, 11-13, 1975.

[2] *International Technology Roadmap for Semiconductors: 2009 Edition.* [Online] http://www.itrs.net.

[3] M. Bohr and K. Mistry, "Intel's revolutionary 22 nm transistor technology," Intel Newsroom, 2011.

[4] T. Ghani, S. E. Thompson, M. Bohr, *et al.*, "A 90 nm high volume manufacturing logic technology featuring novel 45 nm gate length strained silicon CMOS transistors," *IEDM Tech. Dig.*, pp. 11.6.1-11.6.3, 2003.

[5] S. E. Thompson, G. Sun, Y. S. Choi, and T. Nishida, "Uniaxial-process-induced strained-Si: extending the CMOS roadmap," *IEEE Trans. Electron Devices*, vol. 53, pp. 1010-1020, 2006.

[6] S. E. Thompson, G. Sun, K. Wu, J. Lim, and T. Nishida, "Key differences for process-induced unaixal vs. substrate-induced biaxial stressed Si and Ge channel MOSFETs," *IEDM Tech. Dig.*, pp. 221-224,2004.

[7] B. Yang, *et al.*, "High-performance nMOSFET with in-situ Phosphorus-doped embedded Si:C (ISPD eSi:C) source-drain stressor," *IEDM Tech. Dig.*, pp. 51-54, 2008.

[8] H. S. Yang, *et al.*, "Dual stress liner for high performance sub-45 nm gate length SOI CMOS manufacturing," *IEDM Tech. Dig.*, pp. 1075-1077, 2004.

[9] V. Chan, R. Rengarajan, N. Rovedo, W. Jin, T. Hook, P. Nguyen, *et al.*, "High speed 45 nm gate length CMOSFETs integrated into a 90 nm bulk technology incorporating strain engineering," *IEDM Tech. Dig.*, 2003.

[10] P. R. Chidambaram, B. A. Smith, L. H. Hall, H. Bu, S. Chakravarthi, Y. Kim, *et al.*, "35% drive current improvement from recessed-SiGe drain extensions on 37 nm gate length PMOS," *VLSI Symp. Tech. Dig.*, pp. 48-49, 2004.

[11] S. E. Thompson, *et al.*, "A logic nanotechnology featuring strained silicon," *IEEE Electron Device Lett.*, vol. 25, pp. 191-193, 2004.

[12] C.-H. Chen, T. L. Lee, T. H. Hou, C. L. Chen, C. C. Chen, J. W. Hsu, K. L. Cheng, Y. H. Chiu, H. J. Tao, Y. Jin, C. H. Diaz, S. C. Chen, and M.-S. Liang, "Stress memorization technique (SMT) by selectively strained-nitride capping for sub-65 nm high-performance strained-Si device application," *VLSI Symp. Tech. Dig.*, pp. 56-57, 2004.

[13] M. Chudzik, *et al.*, "High-performance high-κ/metal gates for 45nm CMOS and beyond with gate-first processing," *VLSI Tech. Dig.*, pp. 194-195, 2007.

[14] J. Wang, *et al.*, "Novel channel-stress enhancement technology with eSiGe S/D and recessed channel on damascene gate process," *VLSI Tech. Dig.*, pp. 46-47, 2007.

[15] K. Mistry, *et al.*, "A 45nm logic technology with high-k+ metal gate transistors, strained silicon, 9 Cu interconnect layers, 193nm dry patterning, and 100% Pb-free packaging," *IEDM Tech. Dig.*, pp. 247-250, 2007.

[16] S. Mayuzumi, *et al.*, "Extreme high-performance n- and p-MOSFETs boosted by dual-metal/high-k gate damascene process using top-cut dual stress liners on (100) substrates," *IEDM Tech. Dig.*, pp. 293-296, 2007.

[17] S. Natarajan, *et al.*, "A 32nm logic technology featuring 2nd-generation high-k + metal-gate transistors, enhanced channel strain and $0.171\mu m2$ SRAM cell size in a 291Mb array," *IEDM Tech. Dig.*, pp. 941-943, 2008.

[18] K. Henson, *et al.*, "Gate length scaling and high drive currents enabled for high performance SOI technology using high-κ/metal gate," *IEDM Tech. Dig.*, pp. 645-648, 2008.

[19] J. Huang, *et al.*, "Device and reliability improvement of HfSiON+LaOx/Metal gate stacks for 22nm node application," *IEDM Tech. Dig.*, pp. 45-48, 2008.

[20] T. Ando, *et al.*, "Understanding mobility mechanisms in extremely scaled HfO2 (EOT 0.42 nm) using remote interfacial layer scavenging technique and Vt-tuning dipoles with gate-first process," *IEDM Tech. Dig.*, pp. 423-426, 2009.

[21] L.-Å. Ragnarsson, *et al.*, "Ultra low-EOT (5 Å) gate-first and gate-last high performance CMOS achieved by gate-electrode optimization," *IEDM Tech. Dig.*, pp. 663-666, 2009.

[22] C.-H. Jan, *et al.*, "RF CMOS technology scaling in high-k/metal gate era for RF SoC (system-on-chip) applications," *IEDM Tech. Dig.*, pp. 604-607, 2010.

[23] J. Kavalieros, B. Doyle, S. Datta, G. Dewey,M. Doczy, B. Jin, *et al.*, "Tri-Gate Transistor Architecture with High-k Gate Dielectrics, Metal Gates and Strain Engineering," *VLSI Tech. Dig.*, pp. 50-51, 2006.

[24] C. C. Wu, *et al.*, "High performance 22/20nm FinFET CMOS devices with advanced high-k/metal gate scheme," *IEDM Tech. Dig.*, pp. 600-603, 2010.

[25] V. S. Basker, *et al.*, "A 0.063 μm^2 FinFET SRAM cell demonstration with conventional lithography using a novel integration scheme with aggressively scaled fin and gate pitch," *VLSI Tech. Dig.*, pp. 19-20, 2010.

[26] J. B. Chang, *et al.*, "Scaling of SOI FinFETs down to fin width of 4 nm for the 10nm technology node," *VLSI Tech. Dig.*, pp. 12-13, 2011.

[27] T. Yamashita, *et al.*, "Sub-25nm FinFET with advanced fin formation and short channel effect engineering," *VLSI Tech. Dig.*, pp. 14-15, 2011.

[28] A. Veloso, *et al.*, "Demonstration of scaled 0.099μm^2 FinFET 6T-SRAM cell using Full-Field EUV lithography for (sub-)22nm node single-patterning technology," *IEDM Tech. Dig.*, pp. 301-304, 2009.

[29] N. Horiguchi, *et al.*, "High yield sub-0.1 μm^2 6T-SRAM cells, featuring high-k/metal-gate Finfet devices, double gate patterning, a novel Fin etch strategy, full-field EUV lithography and optimized junction design & layout," *VLSI Tech. Dig.*, pp. 23-24, 2010.

[30] T. Chiarella, *et al.*, "Migrating from planar to FinFET for further CMOS scaling: SOI or bulk?" *Proceeding of ESSCIRC*, pp. 84-87, 2009.

[31] D. Hisamoto, W.-C. Lee, J. Kedzierski, E. Anderson, H. Takeuchi, K. Asano, T.-J. King, J. Bokor, and C. Hu, "A folded-channel MOSFET for deep-subtenth micron era," *IEDM Tech. Dig.*, 1998.

[32] X. Huang, W.-C. Lee, C. Kuo, D. Hisamoto, L. Chang, J. Kedzierski, E. Anderson, H. Takeuchi, Y.-K. Choi, K. Asano, V. Subramanian, T.-J. King, J. Bokor and C. Hu, "Sub-50 nm P-channel FinFET," *IEEE Trans. Electron Devices*, vol. 48, pp. 880-886, 2001.

[33] D. Hisamoto, W.-C. Lee, J. Kedzierski, H. Takeuchi, K. Asano, C. Kuo, E. Anderson, T.-J. King, J. Bokor, and C. Hu, "FinFET-A self-aligned double-gate MOSFET scalable to 20 nm," *IEEE Trans. Electron Devices*, vol. 47, pp. 2320-2325, 2001.

[34] Y.-K. Choi, N. Lindert, P. Xuan, S. Tang, D. Ha, E. Anderson, T.-J. King, J. Bokor, C. Hu, "Sub-20nm CMOS FinFET Technologies," *IEDM Tech. Dig.*, 2001.

[35] B. Yu, L. Chang, S. Ahmed, H. Wang, S. Bell, C.-Y. Yang, C. Tabery, C. Ho, Q. Xiang, T.-J. King, J. Bokor, C. Hu, M.-R. Lin, and D. Kyser, "FinFET scaling to 10 nm gate length," *IEDM Tech. Dig.*, 2002.

[36] H.-S. Wong, K. K. Chan, and Y. Taur, "Self-aligned (top and bottom) double-gate MOSFET with 25 nm thick silicon channel," *IEDM Tech. Dig.*, 1997.

[37] M. Vinet, T. Poiroux, J. Widiez, J. Lolivier, B. Previtali, C. Vizioz, B. Guillaumot, Y. Le Tiec, P. Besson, B. Biasse, F. Allain, M. Casse, D. Lafond, J.-M. Hartmann, Y. Morand, J. Chiaroni, and S. Deleonibus, "Bonded planar double-metal-gate NMOS transistors down to 10 nm," *IEEE Electron Device Lett.*, vol. 26, pp.317-319, 2005.

[38] M. Masahara, T. Matsukawa, K. Ishii, Y. Liu, H. Tanoue, K. Sakamoto, T. Sekigawa, H. Yamauchi, S. Kanemaru, and E. Suzuki, "15-nm-thick Si channel wall vertical double-gate MOSFET," *IEDM Tech. Dig.*, 2002.

[39] Y. Liu, K. Ishii, T. Tsutsumi, M. Masahara, and E. Suzuki, "Ideal rectangular cross-section Si-Fin channel double-gate MOSFETs fabricated using orientation-dependent wet etching," *IEEE Electron Device Lett.*, vol. 24, pp. 484-486, 2003.

[40] L. Chang, Y.-K. Choi, D. Ha, P. Ranade, S. Xiong, J. Bokor, C. Hu and T.-J. King, "Extremely scaled silicon nano-CMOS devices," *Proc. of IEEE*, vol. 91, no. 11, Nov. 2003.

[41] F. Balestra, S. Cristoloveanu, M. Benachir, J. Brini, and T. Elewa, "Double-gate silicon-on-insulator transistor with volume inversion: A new device with greatly enhanced performance," *IEEE Electron Device Lett.*, vol. 8, pp. 410-412, 1987.

[42] A. Veloso, *et al.*, "Gate-last vs. gate-first technology for aggressively scaled EOT logic/RF CMOS," *VLSI Tech. Dig.*, pp. 34-35, 2011.

[43] Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[44] T. Y. Hoffmann, "Integrating high-k/metal gates: gate-first or gate-last?" *Solid State Technology*, Issue 3, 2010.

[45] J.-P. Colinge, "Multiple-gate SOI MOSFETs," *Solid-State Electron.*, vol. 48, pp. 897-905, 2004.

[46] H.-S. Wong, D. J. Frank, P. M. Solomon, D. H. J. Wann, and J. J. Welser, "Nanoscale CMOS," *Proc. IEEE*, vol. 87, p. 537, 1999.

[47] P. D. Ye, G. D. Wilk, J. Kwo, B. Yang, H.-J. L. Gossmann, M. Frei, S. N. G. Chu, J. P. Mannaerts, M. Sergent, M. Hong, K. Ng, and J. Bude, "GaAs MOSFET with oxide gate dielectric grown by atomic layer deposition," *IEEE Electron Device Lett.*, vol. 24, pp. 209-211, 2003.

[48] Y. Xuan, Y. Q. Wu, and P. D. Ye, "High-performance inversion-type enhancement-mode InGaAs MOSFET with maximum drain current exceeding 1 A/mm," *IEEE Electron Device Lett.*, vol. 29, pp. 294-296, 2008.

[49] Y. Xuan, T. Shen, M. Xu, Y. Q. Wu, and P. D. Ye, "High-performance surface channel In-rich $In_{0.75}Ga_{0.25}As$ MOSFETs with ALD high-k as gate dielectric," *IEDM Tech. Dig.*, pp. 371-374, 2008.

[50] Y. Q. Wu, W. K. Wang, O. Koybasi, D. N. Zakharov, E. A. Stach, S. Nakahara, J. C. M. Hwang, and P. D. Ye, "0.8-V supply voltage deep-submicrometer inversion-mode $In_{0.75}Ga_{0.25}As$ MOSFET," *IEEE Electron Device Lett.*, vol. 30, pp. 700-702, 2009.

[51] J. P. de Souza, E. Kiewra, Y. Sun, A. Callegari, D. K. Sadana, G. Shahidi, D. J. Webb, J. Fompeyrine, R. Germann, C. Rossel, and C. Marchiori, "Inversion mode n-channel GaAs field effect transistor with high-k/metal gate," *Appl. Phys. Lett.*, vol. 92, p. 153 508, 2008.

[52] Y. Xuan, Y. Q. Wu, T. Shen, T. Yang, and P. D. Ye, "High performance submicron inversion-type enhancement-mode InGaAs MOSFETs with ALD Al_2O_3, HfO_2 and HfAlO as gate dielectrics," *IEDM Tech. Dig.*, pp. 637-640, 2007.

[53] D. Shahrjerdi, T. Rotter, G. Balakrishnan, D. Huffaker, E. Tutuc, and S. K. Banerjee, "Fabrication of self-aligned enhancement-mode $In_{0.53}Ga_{0.47}As$ MOSFETs with $TaN/HfO_2/AlN$ gate stack," *IEEE Electron Device Lett.*, vol. 29, pp. 557-560, 2008.

[54] I. Ok, H. Kim, M. Zhang, F. Zhu, S. Park, J. Yum, H. Zhao, D. Garcia, P. Majhi, N. Goel, W. Tsai, C. K. Gaspe, M. B. Santos, and J. C. Lee, "Self-aligned n-channel metal-oxide-semiconductor field effect transistor on high-indium-content $In_{0.53}Ga_{0.47}As$ and InP using physical vapor deposition HfO_2 and silicon interface passivation layer," *Appl. Phys. Lett.*, vol. 92, pp. 202 903-202 905, 2008.

[55] H. C. Chin, M. Zhu, X. H. Tung, G. S. Samudra, and Y. C. Yeo, "In situ surface passivation and CMOS-compatible palladium-germanium contacts for surface-channel gallium arsenide MOSFETs," *IEEE Electron Device Lett.*, vol. 29, pp. 553-556, 2008.

[56] J. Q. Lin, S. J. Lee, H. J. Oh, G. Q. Lo, D. L. Kwong, and D. Z. Chi, "Inversion-mode self-aligned $In_{0.53}Ga_{0.47}As$ n-channel metal-oxide-semiconductor field-effect transistor with HfAlO gate dielectric and TaN metal gate," *IEEE Electron Device Lett.*, vol. 29, pp. 977-980, 2008.

[57] S. Koveshnikov, N. Goel, P. Majhi, H. Wen, M. B. Santos, S. Oktyabrsky, V. Tokranov, R. Kambhampati, R. Moore, F. Zhu, J. Lee, and W. Tsai, "$In_{0.53}Ga_{0.47}As$ based metal oxide semiconductor capacitors with atomic layer deposition ZrO_2 gate oxide demonstrating low gate leakage current and equivalent oxide thickness less than 1 nm," *Appl. Phys. Lett.*, vol. 92, pp. 222 904-222 906, 2008.

[58] T. D. Lin, H. C. Chiu, P. Chang, L. T. Tung, C. P. Chen, M. Hong, J. Kwo, W. Tsai, and Y. C. Wang, "High-performance self-aligned inversion-channel $In_{0.53}Ga_{0.47}As$ metal-oxide-semiconductor fieldeffect-transistor with $Al_2O_3/Ga_2O_3(Gd_2O_3)$ as gate dielectrics," *Appl. Phys. Lett.*, vol. 93, pp. 033 516-033 518, 2008.

[59] R. J. W. Hill, D. A. J. Moran, X. Li, H. Zhou, D. Macintyre, S. Thoms, A. Asenov, P. Zurcher, K. Rajagopalan, J. Abrokwah, R. Droopad, M. Passlack, and I. G. Thayne, "Enhancement-mode GaAs MOSFETs with an $In_{0.3}Ga_{0.7}As$ channel, a mobility of over 5000 $cm^2/V{\cdot}s$, and transconductance of over 475 $\mu S/\mu m$," *IEEE Electron Device Lett.*, vol. 28, pp. 1080-1082, 2007.

[60] E. J. Kim, E. Chagarov, J. Cagnon, Y. Yuan, A. C. Kummel, P. M. Asbeck, S. Stemmer, K. C. Saraswat, and P. C. McIntyre, "Atomically abrupt and unpinned $Al_2O_3/In_{0.53}Ga_{0.47}As$ interfaces: experiment and simulation," *J. Appl. Phys.*, vol. 106, p. 124508, 2009.

[61] E. J. Kim, L. Wang, P. M. Asbeck, K. C. Saraswat, and P. C. McIntyre, "Border traps in $Al_2O_3/In_{0.53}Ga_{0.47}As$ (100) gate stacks and their passivation by hydrogen anneals," *Appl. Phys. Lett.*, vol. 96, p. 012906, 2010.

[62] Y. Hwang, R. Engel-Herbert, N. G. Rudawski, and S. Stemmer, "Effect of postdeposition anneals on the Fermi level response of $HfO_2/In_{0.53}Ga_{0.47}As$ gate stacks" *J. Appl. Phys.*, vol. 108, p. 034111, 2010.

[63] G. Brammertz, A. Alian, D. H.-C. Lin, M. Meuris, M. Caymax, and W.-E. Wang, "A combined interface and border trap model for high-mobility substrate Metal-Oxide-Semiconductor devices applied to $In_0.53Ga_0.47As$ and InP capacitors," *IEEE Trans. Electron Device*, vol. 58, pp. 3890-3897, 2011.

[64] Y. Cui, Z. Zhong, D. Wang, W. Wang, and C. M. Lieber, "High performance silicon nanowire field effect transistors," *Nano Lett.*, vol. 3, pp. 149-152, 2003.

[65] M. Bescond, N. Cavassilas, and M. Lannoo, "Effective-mass approach for n-type semiconductor nanowire MOSFETs arbitrarily oriented," *Nanotechnology*, vol. 18, 255201, 2007.

[66] F.-L. Yang, D. H. Lee, H. Y. Chen, C. Y. Chang, S. D. Liu, C. C. Huang, T. X. Chung, H. W. Chen, C. C. Huang, Y. H. Liu, C. C. Wu, C. C. Chen, S. C. Chen, Y. T. Chen, Y. H. Chen, C. J. Chen, B. W. C. P. F. Hsu, J. H. Shieh, H. J. Tao, Y. C. Yeo, Y. Li, J. W. Lee, P. Chne, M. S. Liang, and C. Hu, "5 nm-gate nanowire FinFET," *VLSI Symp. Tech. Dig.*, 2004, pp. 196-197.

[67] S. D. Suk, S.-Y. Lee, S.-M. Kim, E.-J. Yoon, M.-S. Kim, M. Li, C.W. Oh, K. H. Yeo, S. H. Kim, D.-S. Shin, K.-H. Lee, H. S. Park, J. N. Han, C. J. Park, J.-B. Park, D.-W. Kim, D. Park, and B.-I. Ryu, "High performance 5 nm radius twin silicon nanowire MOSFET(TSNWFET): Fabrication on bulk Si wafer, characteristics, and reliability," *IEDM Tech. Dig.*, 2005, pp. 717-720.

[68] K. H. Yeo, S. D. Suk, M. Li, Y.-Y. Yeoh, K. H. Cho, K.-H. Hong, S. K. Yun, M. S. Lee, N. M. Cho, K. H. Lee, D. Y. Hwang, B. K. Park, D.-W. Kim, D. G. Park, and B.-I. Ryu, "Gate-all-around (GAA) twin silicon nanowire MOSFET (TSNWFET) with 15 nm length gate and 4 nm radius nanowires," *IEDM Tech. Dig.*, 2006, pp. 539-542.

[69] S. D. Suk, K. H. Yeo, K. H. Cho, M. Li, Y. Y. Yeoh, S.-Y. Lee, S. M. Kim, E. J. Yoon, M. S. Kim, C. H. Oh, S. H. Kim, D.-W. Kim, and D. G. Park, "High-performance twin silicon nanowire MOSFET (TSNWFET) on bulk Si wafer," *IEEE Trans. Nanotechnol.*, vol. 7, pp. 181-184, 2008.

[70] S. D. Suk, M. Li, Y. Y. Yeoh, K. H. Yeo, K. H. Cho, I. K. Ku, H. Cho, W. Jang, D.-W. Kim, D. Park, and W.-S. Lee, " Investigation of nanowire size dependency on TSNWFET," *IEDM Tech. Dig.*, 2007, pp. 891-894.

[71] K. H. Cho, *et al.*, "Observation of single electron tunneling and ballistic transport in twin silicon nanowire MOSFETs (TSNWFETs) fabricated by top-down CMOS process," *IEDM Tech. Dig.*, 2006.

[72] N. Singh, A. Agarwal, L. K. Bera, T. Y. Liow, R. Yang, S. C. Rustagi, C. H. Tung, R. Kumar, G. Q. Lo, N. Balasubramanian, and D. L. Kwong, "High-performance fully depleted silicon nanowire (diameter $\leqslant 5$ nm) gate-all-around CMOS devices," *IEEE Electron Device Lett.*, vol. 27, no. 5, pp. 383-386, 2006.

[73] N. Singh, F. Y. Lim, W. W. Fang, S. C. Rustagi, L. K. Bera, A. Agarwal, C. H. Tung, K. M. Hoe, S. R. Omampuliyur, D. Tripathi, A. O. Adeyeye, G. Q. Lo, N. Balasubramanian, and D. L. Kwong, "Ultra-narrow silicon nanowire gate-all-around CMOS devices: Impact of diameter, channelorientation and low temperature on device performance," *IEDM Tech. Dig.*, 2006, pp. 548-551.

[74] Y. Tian, R. Huang, Y. Wang, J. Zhuge, R. Wang, J. Liu, X. Zhang, and Y. Wang, "New self-aligned silicon nanowire transistors on bulk substrate fabricated by epi-free compatible CMOS technology: process integration, experimental characterization of carrier transport and low frequency noise," *IEDM Tech. Dig.*, 2007, pp. 895-898.

[75] R. Wang, H. Liu, R. Huang, J. Zhuge, L. Zhang, D.-W. Kim, X. Zhang, D. Park, and Y.Wang, "Experimental investigations on carrier transport in Si nanowire transistors: Ballistic efficiency and apparent mobility," *IEEE Trans. Electron Devices*, vol. 55, pp. 2960-2967, 2008.

[76] T. Yu, R. Wang, R. Huang, J. Chen, J. Zhuge, and Y. Wang, "Investigation of nanowire line-edge roughness in gate-all-around silicon nanowire MOSFETs," *IEEE Trans. Electron Devices*, vol. 57, pp. 2864-2871, 2010.

[77] R. Wang, J. Zhuge, R. Huang, T. Yu, J. Zou, D.-W. Kim, D. Park, and Y.Wang, "Investigation on variability in metal-gate Si nanowire MOSFETs:

analysis of variation sources and experimental characterization," *IEEE Trans. Electron Devices*, vol. 58, pp. 2317-2325, 2011.

[78] J. G. Fossum, J.-W. Yang, and V. P. Trivedi, "Suppression of corner effects in triple-gate MOSFETs ," *IEEE Electron Device Lett.*, vol. 24, pp. 745-747, 2003.

[79] W. Xiong, J. W. Park, and J. P. Colinge, "Corner effect in multiple-gate SOI MOSFETs ," *IEEE Int. SOI Conf.*, pp. 111-113, 2003.

[80] S. M. Sze, *Physics of Semiconductor Devices*. New York: Wiley, 1981.

[81] C. G. Sodini, "Charge accumulation and mobility in thin dielectric MOS transistors," *Solid-State Electron.*, vol. 25, pp. 833-841, 1982.

[82] B. Yu, "Design and modeling of non-classical MOSFETs," Ph.D. dissertation, Univ. California San Diego, CA, 2009.

[83] L. Wang, "Design of scaled electronic devices based on III- V materials," Ph.D. dissertation, Univ. California San Diego, CA, 2009.

[84] E. J. Kim, L. Wang, P. M. Asbeck, K. C. Saraswat, and P. C. McIntyre, "bulk-oxide traps in $Al_2O_3/In_{0.53}Ga_{0.47}As$ (100) gate stacks and their passivation by hydrogen anneals," *Appl. Phys. Lett.*, vol. 96, p. 012906, 2010.

[85] E. H. Nicollian and J. R. Brews, *MOS Physics and Technology*. New York: Wiley, 1982.

[86] K. Martens *et al.*, "On the correct extraction of interface trap density of MOS devices with high-mobility semiconductor substrates," *IEEE Trans. Electron Devices*, vol. 55, pp. 547-556, 2008.

[87] M. Passlack, R. Droopad, and G. Brammertz, "Suitability study of oxide/-gallium arsenide interfaces for MOSFET applications," *IEEE Trans. Electron Devices*, vol. 57, pp. 2944-2956, 2010.

[88] Y. Xuan, Y. Q. Wu, T. Shen, T. Yang, and P. D. Ye, "High performance submicron inversion-type enhancement-mode InGaAs MOSFETs with ALD Al_2O_3, HfO_2 and HfAlO as gate dielectrics," in *IEDM Tech. Dig.*, 2007, pp. 637-640.

[89] N. Goel, P. Majhi, C. O. Chui, W. Tsai, D. Choi, and J. S. Harris, "InGaAs metal-oxide-semiconductor capacitors with HfO_2 gate dielectric grown by atomic-layer deposition," *Appl. Phys. Lett.*, vol. 89, p. 163517, 2006.

[90] E. J. Kim, E. Chagarov, J. Cagnon, Y. Yuan, A. C. Kummel, P. M. Asbeck, S. Stemmer, K. C. Saraswat, and P. C. McIntyre, "Atomically abrupt and unpinned $Al_2O_3/In_{0.53}Ga_{0.47}As$ interfaces: experiment and simulation," *J. Appl. Phys.*, vol. 106, p. 124508, 2009.

[91] G. Brammertz, K. Martens, S. Sioncke, A. Delabie, M. Caymax, M. Meuris, and M. Heyns, "Characteristic trapping lifetime and capacitance-voltage measurements of GaAs metal-oxide-semiconductor structures," *Appl. Phys. Lett.*, vol. 91, p. 133510, 2007.

[92] Y. Hwang, R. Engel-Herbert, N. G. Rudawski, and S. Stemmer, "Effect of postdeposition anneals on the Fermi level response of $HfO_2/In_{0.53}Ga_{0.47}As$ gate stacks" *J. Appl. Phys.*, vol. 108, p. 034111, 2010.

[93] W. Shockley and W. T. Read, Jr., "Statistics of the Recombinations of Holes and Electrons," *Phys. Rev.*, vol. 87, pp. 835-842, 1952.

[94] F. P. Heiman and G. Warfield, "The effects of oxide traps on the MOS capacitance," *IEEE Trans. Electron Devices*, vol. 12, pp. 167-178, 1965.

[95] H. Preier, "Contribution of surface states to MOS impedance," *Appl. Phys. Lett.*, vol. 10, pp. 361-363, 1967.

[96] D. S. L. Mui, J. Reed, D. Biswas, and H. Morkoç, "A new circuit model for tunneling related trapping at insulator-semiconductor interfaces in accumulation," *J. Appl. Phys.*, vol. 72, pp. 553-558, 1992.

[97] Y. Yuan, L. Wang, B. Yu, B. Shin, J. Ahn, P. C. McIntyre, P. M. Asbeck, M. J. W. Rodwell and Y. Taur, "A distributed model for border traps in Al_2O_3-InGaAs MOS devices," *IEEE Electron Device Letters*, vol. 32, pp. 485-487, 2011.

[98] Y. Yuan, B. Yu, J. Ahn, P. C. McIntyre, P. M. Asbeck, M. J. W. Rodwell and Y. Taur, "A distributed bulk-oxide trap model for Al_2O_3-InGaAs MOS devices," Submitted to *IEEE Trans. Electron Devices.*

[99] D. Jimenez, B. Iniguez, J. Sune, L. F. Marsal, J. Pallares, J Roig, and D. Flores, "Continuous analytic I-V model for surrounding-gate MOSFETs," *IEEE Electron Device Lett.*, vol. 25, pp. 571-573, 2004.

[100] Y. Taur, X. Liang, W. Wang, and H. Lu, "A continuous, analytic drain-current model for DG MOSFETs," *IEEE Electron Device Lett.*, vol. 25, pp. 107-109, 2004.

[101] B. Yu, W.-Y. Lu, H. Lu, and Y. Taur, "Analytic charge model for surrounding-gate MOSFETs," *IEEE Trans. Electron Devices*, vol. 54, pp. 492-496, 2007.

[102] D. J. Frank, Y. Taur, and H.-S. P. Wong, "Generalized scale length for two-dimensional effects in MOSFETs," *IEEE Electron Device Lett.*, vol. 19, pp. 385-387, 1998.

[103] Z. X. Wang and D. R. Guo, *Special Functions.* Singapore: World Scientific, 1989.

[104] S.-H. Oh, D. Monroe, and J. M. Hergenrother, "Analytic description of short-channel effects in fully-depleted double-gate and cylindrical, surrounding-gate MOSFETs," *IEEE Electron Device Lett.*, vol. 21, pp. 445-447, 2000.

[105] V. P. Trivdedi and J. G. Fossum, "Quantum-mechanical effects on the threshold voltage of undoped double-gate MOSFETs," *IEEE Electron Device Lett.*, vol. 26, pp. 579-582, 2005.

[106] R. Granzner, F. Schwierz, and V. M. Polyakov, "An analytical model for the threshold voltage shift caused by two-dimensional quantum confinement in undoped multiple-gate MOSFETs," *IEEE Trans. Electron Devices*, vol. 54, pp. 2562-2565, 2007.

[107] N. W. McLachlan, *Theory and Application of Mathieu Functions*. New York: Dover, 1964.

[108] F. Stern and W. E. Howard, " Properties of semiconductor surface inversion layers in the electric quantum limit," *Phys. Rev.*, vol. 163, pp. 816-835, 1967.

[109] J. Wang, A. Rahman, A. Ghosh, G. Klimeck, M. Lundstrom, "On the validity of the parabolic effective-mass approximation for the I-V calculation of silicon nanowire transistors," *IEEE Trans. Electron Devices*, vol. 52, pp. 1589-1595, 2005.

[110] J. P. J. van der Steen, D. Esseni, P. Palestri, L. Selmi, R. J. E. Heuting, "Validity of the parabolic effective mass approximation in silicon and germanium n-MOSFETs with different crystal orientations," *IEEE Trans. Electron Devices*, vol. 54, pp. 1843-1851, 2007.

[111] J. Wang, P. M. Solomon, and M. Lundstrom, "A general approach for the performance assessment of nanoscale silicon FETs," *IEEE Trans. Electron Devices*, vol. 51, pp. 1366-1370, 2004.

[112] S.-Y. Oh, S.-G. Choi, C. G. Sodini, and J. L. Moll, "Analysis of the channel inversion layer capacitance in the very thin-gate IGFET," *IEEE Electron Device Lett.*, vol. 4, pp. 236-239, 1983.

[113] G. Baccarani and M. R. Wordeman, "Transconductance degradation in thin-Oxide MOSFET's," *IEEE Trans. Electron Devices*, vol. 30, pp. 1295-1304, 1983.

[114] A. Hartstein and N. F. Albert, "Determination of the inversion-layer thickness from capacitance measurements of metal-oxide-semiconductor field-effect transistors with ultrathin oxide layers," *Phys. Rev. B*, vol. 38, pp. 1235-1240, 1988.

[115] S. Takagi and A. Toriumi, "Quantitative understanding of inversion-layer capacitance in Si MOSFET's," *IEEE Trans. Electron Devices*, vol. 42, pp. 2125-2130, 1995.